

Discrete distributions

Introduction

Here we will be looking at four discrete distributions:

The **uniform distribution**, which assumes that all values have an equal chance of occurring.

The **Bernoulli distribution**, which can model the outcome of a simple event.

The **binomial distribution**, which can model the outcomes of n simple events. This is useful for modelling the number of deaths or retirements amongst a fixed number of policyholders.

The **Poisson distribution**, which is useful for modelling the number of claims a general insurance company receives per unit time.

For each of these distributions we will look at graphs of the probability functions, derive some basic results and find probabilities.

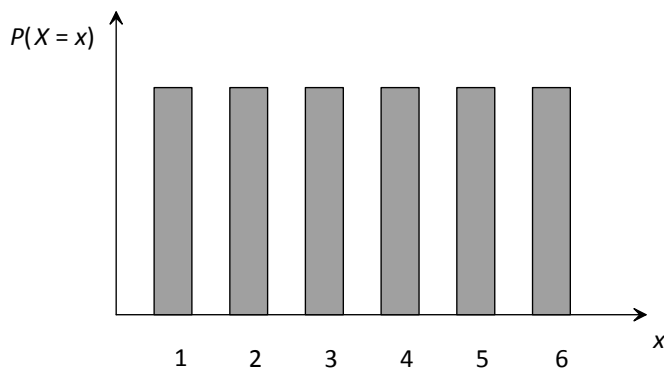
The discrete uniform distribution

General features of the discrete uniform distribution

If we roll a fair die, we have the following probability distribution:

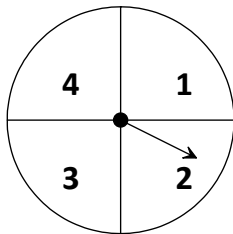
| | | | | | | |
|------------|---------------|---------------|---------------|---------------|---------------|---------------|
| x | 1 | 2 | 3 | 4 | 5 | 6 |
| $P(X = x)$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

The graph of this probability function is:



Each whole number has the *same* probability of occurring. Well in the same way that people who wear uniform look the same, we call this the **discrete uniform distribution**.

Now if we were to spin a fair four-sided spinner, we would obtain the following probability distribution:



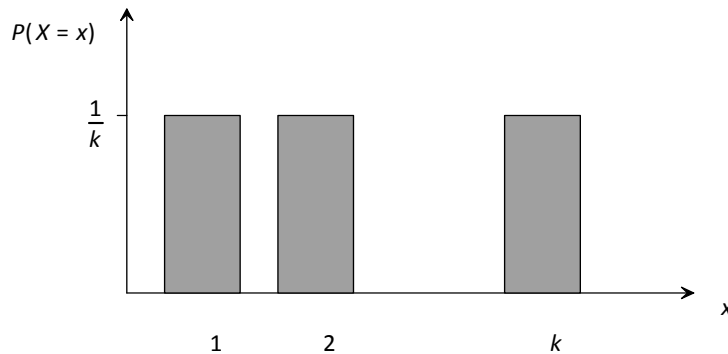
| | | | | |
|------------|---------------|---------------|---------------|---------------|
| x | 1 | 2 | 3 | 4 |
| $P(X = x)$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |

So for 6 discrete values each had probability $\frac{1}{6}$ and for 4 discrete values each had probability $\frac{1}{4}$. So in general, if we had k discrete numbers each would have probability of $\frac{1}{k}$.

The probability function of the discrete uniform distribution

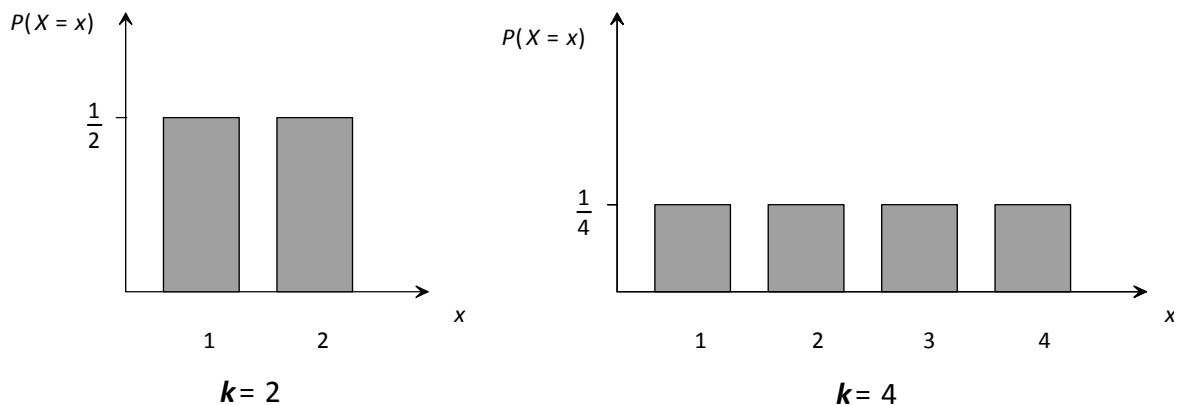
As we have seen, the general case where we have k whole numbers each with the same probability of happening is given by:

$$P(X = x) = \frac{1}{k} \quad x = 1, 2, \dots, k$$



This is a probability function as each probability is non-negative and all the probabilities add up to 1.

To get a better 'feel' for this distribution, we'll look at two discrete uniform distributions with different values for k :



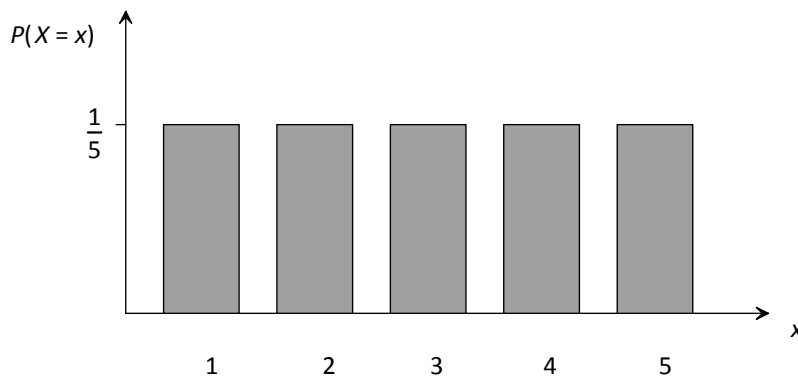
We can see that as k increases, we are sharing the total probability of 1 between more numbers. Hence, the probability of each number occurring is lower.

We doubled the number of values (from 2 to 4) and so the probability is halved (from $\frac{1}{2}$ to $\frac{1}{4}$) as it is shared over twice as many numbers.

Moments of the uniform distribution

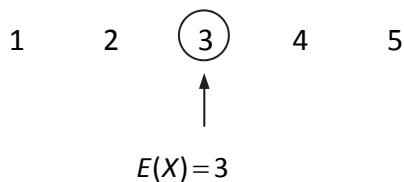
To find the moments, let's look at the discrete uniform distribution with $k = 5$:

$$P(X = x) = \frac{1}{5} \quad x = 1, 2, 3, 4, 5$$



Mean

Since the distribution is symmetrical the mean is in the middle (*ie* it is the median value):



We could also have found the mean by using the formula:

$$E(X) = \sum xP(X = x)$$

Question 1.1

Use this formula to show that the mean is 3 for:

$$P(X = x) = \frac{1}{5} \quad x = 1, 2, 3, 4, 5$$

Variance

There's no easy way to get the variance except to use the formula:

$$\text{var}(X) = E(X^2) - E^2(X)$$

Remembering that:

$$E(X^2) = \sum x^2 P(X = x)$$

Question 1.2

Use these formulae to show that the variance is 2 for:

$$P(X = x) = \frac{1}{5} \quad x = 1, 2, 3, 4, 5$$

Median

Since the graph is symmetrical the median is also in the middle and is the same as the mean. Alternatively, we could count through the probabilities to find which value is half way through the distribution (or better still use the graph of the CDF).

Question 1.3

Count through the probabilities to show that the median is 3 (same as $E(X)$) for:

$$P(X = x) = \frac{1}{5} \quad x = 1, 2, 3, 4, 5$$

Mode

Since $P(X = x)$ is the same for all values in the range there is no mode, as no value occurs with a greater probability than the rest.

In general, the results are:

Discrete uniform distribution

$$P(X = x) = \frac{1}{k} \quad x = 1, 2, \dots, k$$

$$E(X) = \frac{1}{2}(k + 1)$$

$$\text{var}(X) = \frac{1}{12}(k^2 - 1)$$

The proofs can be found in Appendix A.

Probabilities for a discrete uniform distribution

Probabilities can be calculated easily using the probability function.

Question 1.4

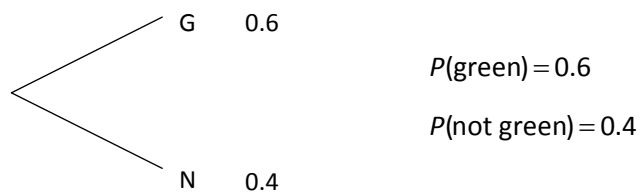
In a game of chance, a large spinner with the numbers 1 – 20 is spun. If the probability of getting any number is the same, calculate the probability of obtaining:

- (i) an even number
- (ii) a number greater than 18.

The Bernoulli distribution

General features of the Bernoulli distribution

Suppose I'm driving down the road and I come to a traffic light. The light is either green (in which case I can go) or it's not green (in which case I'll have to stop). If the probability that the traffic light is green when I reach it is 0.6, we have:



Well this isn't much of a random variable – as we should be assigning probabilities to *numbers*.

What we'll do is use the number 1 for a 'success' (*ie* a green traffic light) and the number 0 for a 'failure' (*ie* a not green traffic light).

$$P(X=1) = 0.6$$

$$P(X=0) = 0.4$$

We could think of these values for X as the number of successes; 1 success (*ie* 1 green traffic light) and 0 successes (*ie* no green traffic lights).

We can write this random variable more concisely using a clever trick:

$$P(X=x) = 0.6^x 0.4^{1-x} \quad x = 0, 1$$

When we have 0 successes (*ie* a failure):

$$P(X=0) = 0.6^0 0.4^1 = 0.4$$

When we have 1 success:

$$P(X=1) = 0.6^1 0.4^0 = 0.6$$

The probability function of the Bernoulli distribution

In general, if we have a probability p of success, we get:

$$P(X = 1) = p$$

$$P(X = 0) = q \quad \text{where } q = 1 - p$$

or we could write this using our clever shortcut:

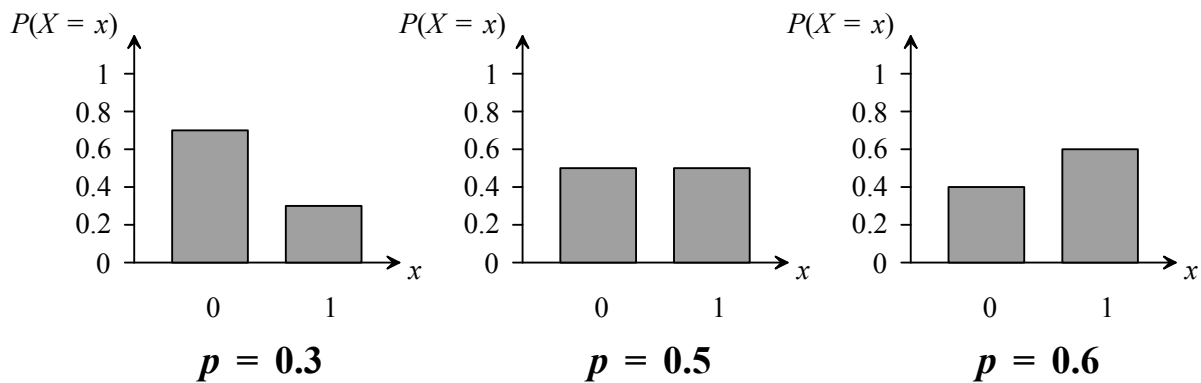
$$P(X = x) = p^x q^{1-x} \quad x = 0, 1$$

This distribution is called the **Bernoulli distribution** (named after its creator Jacob Bernoulli) with parameter p . Since p is a probability we must have $0 \leq p \leq 1$.

The shortcut way of writing 'X has a Bernoulli distribution with parameter p ' is:

$$X \sim \text{Bernoulli}(p)$$

So what does this distribution look like?



Above are three Bernoulli distributions with different values of p . As we can see, the greater the value of p , the greater the success ($X = 1$) bar and the smaller the failure ($X = 0$) bar.

Moments of the Bernoulli distribution

Mean

We can find the mean by using the formula:

$$E(X) = \sum xP(X = x)$$

Question 1.5

Use this formula to show that $E(X) = p$ for the Bernoulli distribution:

$$P(X = 1) = p$$

$$P(X = 0) = q$$

Variance

We can find the variance using the formula:

$$\text{var}(X) = E(X^2) - E^2(X)$$

where:

$$E(X^2) = \sum x^2P(X = x)$$

Question 1.6

Use these formulae to show that $\text{var}(X) = pq$ for the Bernoulli distribution:

$$P(X = 1) = p$$

$$P(X = 0) = q$$

Median

We can get the median by counting through the probabilities to find which of the two values is half way through the distribution.

Mode

The mode is whichever value (0 or 1) has the greatest probability.

In summary:

Bernoulli distribution, $Bernoulli(p)$

$$P(X=1)=p \quad \text{or} \quad P(X=x)=p^x q^{1-x} \quad x=0,1$$

$$P(X=0)=q$$

$$E(X)=p$$

$$\text{var}(X)=pq$$

We will see later that the Bernoulli distribution is a special case of the Binomial distribution. So we only need to use the Binomial formulae.

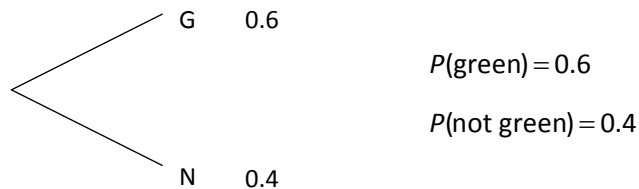
Probabilities of an Bernoulli distribution

There are no probabilities to calculate as $P(X=0)$ and $P(X=1)$ are both given in the definition of the Bernoulli distribution.

The binomial distribution

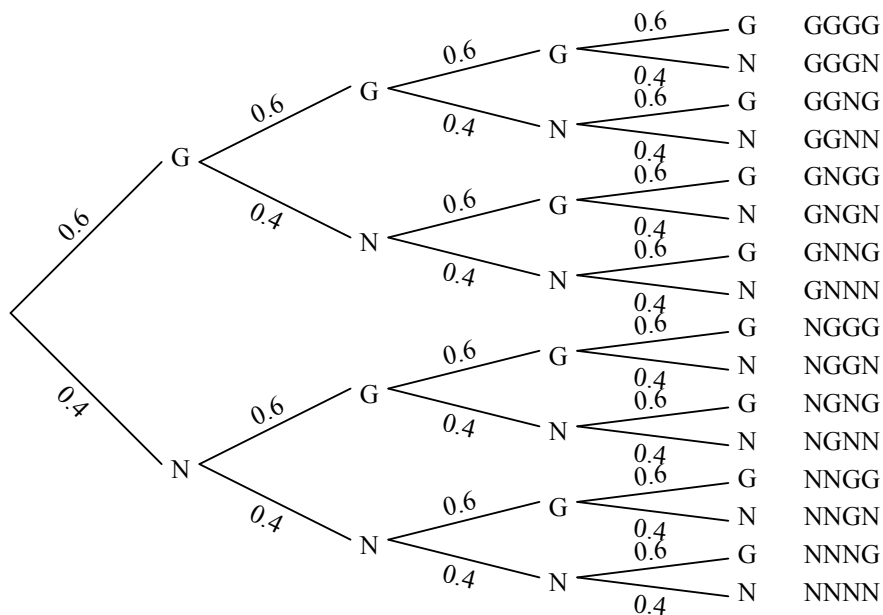
General features of the binomial distribution

We used the Bernoulli distribution to model a set of traffic lights:



What happens if I pass through, say, 4 sets of traffic lights (where the probability that each set is green is 0.6 and all sets are independent)?

To calculate the probabilities of getting 0, 1, 2, 3 or 4 green lights we draw a tree diagram:



Then to get the probability of, say, GGGN we multiply the probabilities on the appropriate branches:

$$P(GGGN) = 0.6 \times 0.6 \times 0.6 \times 0.4 = 0.0864$$

Well this isn't much of a random variable as we should be assigning probabilities to *numbers*. What we'll do is to count the number of 'successes' (*ie* green traffic lights) like we did for the Bernoulli distribution.

For our example, if X is the number of 'successes' for our 4 sets of traffic lights it will range from 0 (NNNN) up to 4 (GGGG).

So how can we work out probabilities for this distribution? Let's find out the probability of 1 'success' (*ie* 1 out the 4 sets of traffic lights were green) for our example. The options we require are:

| | |
|------|---|
| GNNN | $0.6 \times 0.4 \times 0.4 \times 0.4 = 0.0384$ |
| NGNN | $0.4 \times 0.6 \times 0.4 \times 0.4 = 0.0384$ |
| NNGN | $0.4 \times 0.4 \times 0.6 \times 0.4 = 0.0384$ |
| NNNG | $0.4 \times 0.4 \times 0.4 \times 0.6 = 0.0384$ |

Hence, the total probability is $4 \times 0.0384 = 0.1536$.

There is a quicker way of obtaining this result. We first note that each option has the same probability since 1 'success' consists of 1 green and 3 not green lights:

$$0.6 \times 0.4^3 = 0.0384$$

We then need to calculate how many different ways there are of choosing 1 green set from the 4 sets of traffic lights. This is:

$${}^4C_1 = 4$$

Hence, the probability is given by:

$$P(X=1) = {}^4C_1 \times 0.6 \times 0.4^3 = 4 \times 0.0384 = 0.1536$$

Question 1.7

Write down the expressions to calculate $P(X=2)$ and $P(X=3)$.

We call each set of traffic lights a **trial**.

What happens if I change the number of trials (*ie* sets of traffic lights) and the probability of success (*ie* probability of each set being green)?

For example, let's have 10 sets of lights. Now the probability that each set is green when I reach them is 0.8. Let X be the number of successes (*ie* the number of green sets).

To calculate the probability of 3 successes, we first note that we require all the options where we have 3 successes (green sets) and 7 failures (not green sets):

$$0.8^3 \times 0.2^7$$

The number of different ways of choosing 3 successes (green sets) from the 10 trials (*ie* sets of traffic lights) is:

$${}^{10}C_3$$

Hence, we have:

$$P(X=3) = {}^{10}C_3 \times 0.8^3 \times 0.2^7 = 0.000786 \text{ (3 SF)}$$

Question 1.8

X is the number of successes in 20 trials (*ie* sets of traffic lights). The probability of success (*ie* probability of each set being green) is 0.9. Calculate:

$$P(X=13)$$

In general, if we have n trials (*ie* sets of traffic lights) each of which has a probability of success (*ie* being green) of p , then the probability of x successes is:

$${}^nC_x \times p^x \times (1-p)^{n-x}$$

number of ways of choosing x successes from n 'trials' probability of x successes probability of $n-x$ failures

This is the **binomial distribution** with parameters n and p .

The probability function of the binomial distribution

The **binomial distribution** has two parameters: n (the number of trials) and p (the probability of success in each trial). Its probability function is:

$$P(X = x) = {}^n C_x p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n \quad 0 < p < 1$$

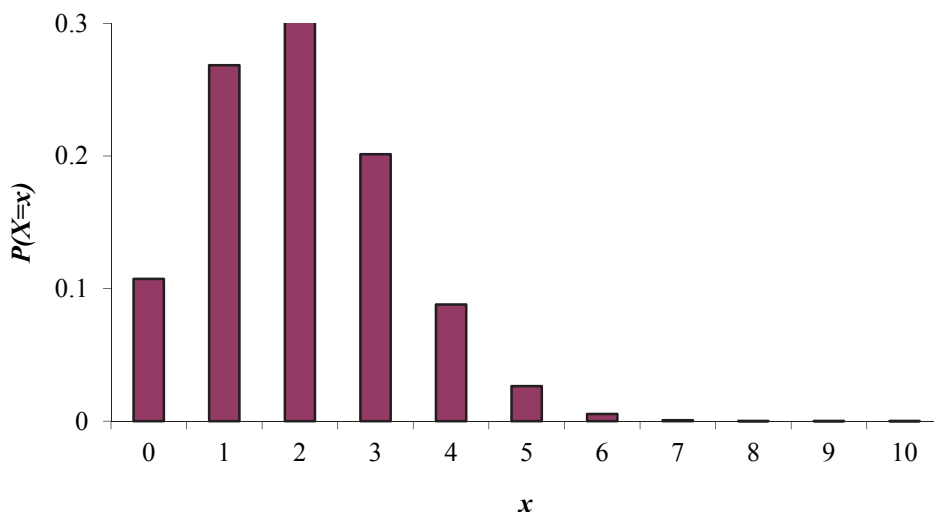
We can write ${}^n C_x$ as $\binom{n}{x}$. It means exactly the same; it's just an alternative way of writing combinations.

The shortcut way of writing ' X has a binomial distribution with parameters n and p ' is:

$$X \sim \text{Bin}(n, p)$$

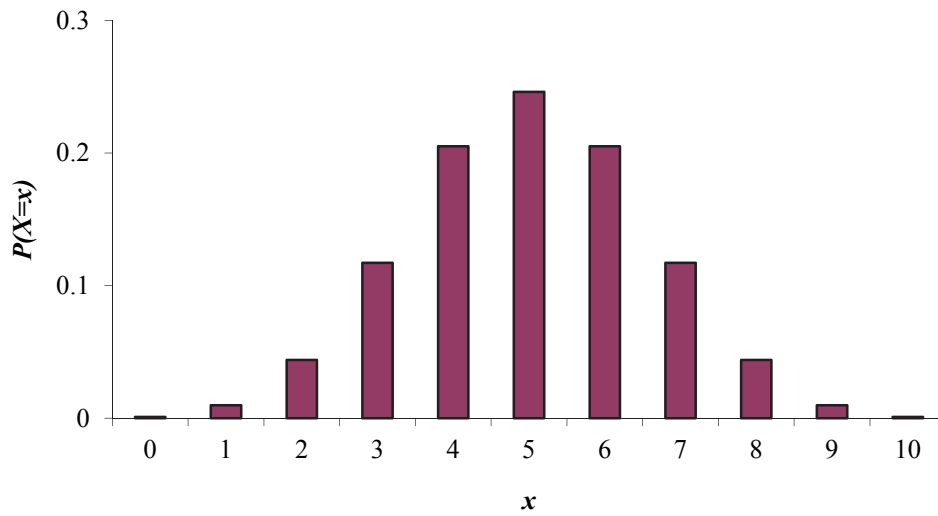
Remember that X is the random variable for the number of successes in n trials.

To get a better feel for this distribution, we'll look at a number of binomial distributions with different values for n and p .



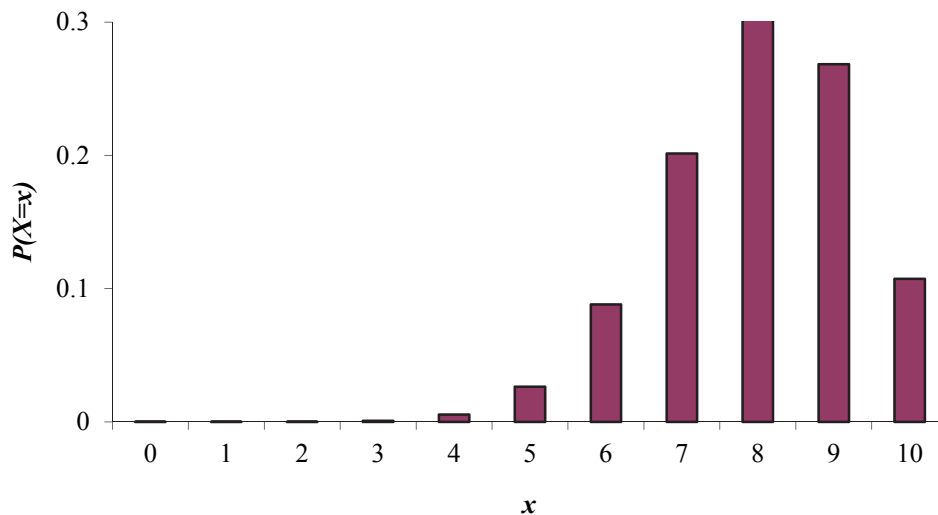
In the graph above, we have 10 trials and probability of success of 0.2. We can see that, since the probability of success is so low, we are most likely to get only a small number of successes (1, 2 or 3) out of our 10 trials.

What happens as we increase the probability of success to, say 0.5:



We can see that we are more likely to get a higher number of successes than before as we have a higher probability of success. We can also see that because $p=0.5$ the graph is symmetrical.

Increasing p to 0.8 we get:



We are now at the other extreme. Since the probability of success is high we are more likely to get successes from most of our 10 trials. Note also how $Bin(10,0.2)$ is a reflection of $Bin(10,0.8)$.

Moments of the binomial distribution

Mean

Finding the mean using $E(X) = \sum xP(X=x)$ is an algebraic nightmare best avoided.

Recall from earlier that the mean of a Bernoulli trial was p . The binomial distribution is the sum of n Bernoulli trials so we would expect the mean to be:

$$E(X) = np$$

This should make sense. If we have 10 traffic lights with a probability of success of 0.6, we would expect $10 \times 0.6 = 6$ of the lights to be green.

Variance

We can use a similar idea to calculate the variance of the binomial. Recall from earlier that the variance of a Bernoulli trial was pq . The binomial distribution is the sum of n Bernoulli trials so the variance is:

$$\text{var}(X) = npq$$

Question 1.9

The probability that an actuarial student studied for a maths degree is 0.7. An ActEd tutorial contains 12 students. State the:

- (i) expected number of students in a tutorial who have studied for a maths degree
- (ii) variance of the number of students who have studied for a maths degree.

Median

There is no easy way to get the median value other than counting through the probabilities to find which value is half way through the distribution. For example, when $X \sim \text{Bin}(6,0.4)$ we have:

| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|----------|-------|-------|-------|-------|-------|-------|-------|
| $P(X=x)$ | 0.047 | 0.187 | 0.311 | 0.276 | 0.138 | 0.037 | 0.004 |

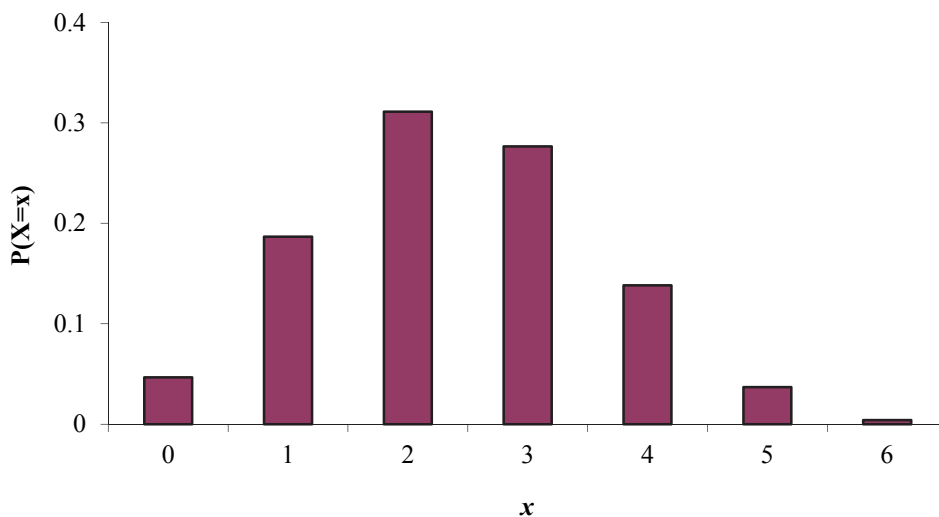
Annotations:

- 0.047 (pointing to $P(X=0)$)
- $0.047 + 0.187 = 0.234$ (pointing to $P(X=0)$ and $P(X=1)$)
- so 0.5 is in here! (pointing to $P(X=2)$)
- $0.047 + 0.187 + 0.311 = 0.545$ (pointing to $P(X=0)$, $P(X=1)$, and $P(X=2)$)

So we can see that the median is 2.

Mode

The mode is the value that has the greatest probability. Looking at either the probability distribution of $X \sim \text{Bin}(6,0.4)$ given above or its graph below:



It is easy to see that the mode is 2 as this has the greatest probability and so is most likely to occur.

Question 1.10

Calculate the mode and median of $X \sim \text{Bin}(3, 0.2)$.

In summary:

Binomial distribution, $\text{Bin}(n, p)$

$$P(X = x) = {}^n C_x p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n \quad 0 < p < 1$$

$$E(X) = np$$

$$\text{var}(X) = npq$$

Note that the binomial distribution with $n=1$ is the Bernoulli distribution.

Probabilities for the binomial distribution

We can use the probability function to calculate probabilities. Suppose $X \sim \text{Bin}(5, 0.7)$.

To calculate a single probability, $P(X=3)$, we just substitute the value of x in our probability function:

$$P(X=3) = {}^5 C_3 \times 0.7^3 \times 0.3^2 = 0.3087$$

What about calculating $P(X > 1)$? Since $X \sim \text{Bin}(5, 0.7)$ there are 5 trials, so the number of successes (ie the values that X can take) range from 0 to 5. So if $X > 1$, it can take the values 2, 3, 4 or 5. Hence:

$$P(X > 1) = P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

$$P(X=2) = {}^5 C_2 \times 0.7^2 \times 0.3^3 = 0.1323$$

$$P(X=3) = 0.3087 \quad \text{from above}$$

$$P(X=4) = {}^5 C_4 \times 0.7^4 \times 0.3^1 = 0.36015$$

$$P(X=5) = {}^5 C_5 \times 0.7^5 \times 0.3^0 = 0.16807$$

Hence:

$$P(X > 1) = 0.1323 + 0.3087 + 0.36015 + 0.16807 = 0.96922$$

However, there is a quicker way. Since the probabilities of all the values X can take sum to 1, we have:

$$P(X \leq 1) + P(X > 1) = 1 \Rightarrow P(X > 1) = 1 - P(X \leq 1)$$

But since X can only take the values 0, 1, 2, 3, 4 or 5, we have:

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

As there are fewer values to work out, this way will be (slightly) quicker:

$$P(X = 0) = {}^5C_0 \times 0.7^0 \times 0.3^5 = 0.00243$$

$$P(X = 1) = {}^5C_1 \times 0.7^1 \times 0.3^4 = 0.02835$$

Hence:

$$P(X > 1) = 1 - (0.00243 + 0.02835) = 0.96922$$

Question 1.11

Given that $X \sim \text{Bin}(9, 0.75)$, calculate:

- (i) $P(X = 6)$
- (ii) $P(X < 8)$
- (iii) $P(2 \leq X < 4)$.

Miscellaneous problems

Nearly all of the questions you come across will be presented 'in context' rather than stating that the random variable has a binomial distribution.

The key to spotting that we have a binomial distribution is that we are counting successes from a number of independent trials (and each trial can only give 1 success). For example the number of days it rains in 20 days, questions right out of a test of 30 questions, matches won in a league, deaths from 10 policyholders, or faulty components in a box of 50.

Question 1.12

An insurance company is assessing the complexity of its claims forms by checking how many of its forms are correctly completed.

The company picks a sample of 6 forms and counts the number of incorrect forms. It is believed that the probability of a form being incorrectly completed is 0.1.

Calculate the probability that the company discovers:

- (i) exactly two incorrect forms
- (ii) more than 4 incorrect forms

in a randomly picked sample.

The Poisson distribution

General features of the Poisson distribution

The exponential function can be expressed as a series expansion. For example:

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots$$

We can use this to get a probability distribution. Dividing both sides by e^{λ} (or equivalently multiply by $e^{-\lambda}$) we get:

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

This is the same as:

$$1 = \frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} + \frac{\lambda^2}{2!} e^{-\lambda} + \frac{\lambda^3}{3!} e^{-\lambda} + \dots$$

Since the terms add up to 1 (and are all positive) they could be our probabilities. We just need to write out a general expression for each term. This is:

$$\frac{\lambda^x}{x!} e^{-\lambda}$$

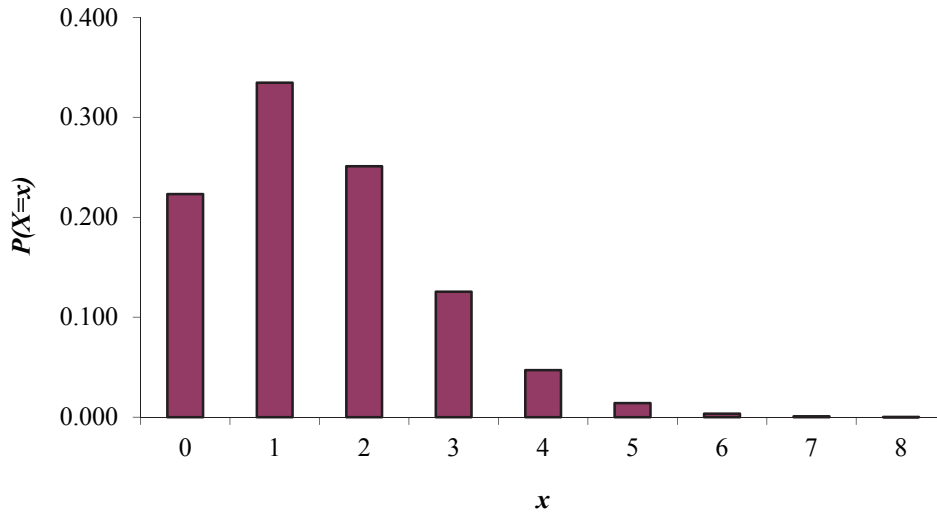
where the x can take the values 0, 1, 2, 3, ...

So we have:

$$P(X=x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x=0,1,2,3,\dots$$

This is called the **Poisson distribution** (after its creator Siméon Poisson).

Now at the moment we have a distribution that probably means very little to you. Let's take a quick look at a Poisson distribution with $\lambda = 1.5$:



This gives us a positively skewed distribution.

But what is it useful for? It can be shown that this is great for modelling events that occur at a rate of λ per unit of time or length.

For example, it can be used to model the number of claims per month, deaths per week, people joining a queue per minute, flaws per metre in a piece of metal, or accidents per mile on a road.

This makes the Poisson distribution exceptionally useful in general insurance for modelling the number of claims received by an insurance office. It also has applications for modelling the number of deaths for pensions and life companies.

The probability function of the Poisson distribution

In general, for events occurring at a (mean) rate of λ per unit, we get:

$$P(X=x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x=0,1,2,3,\dots$$

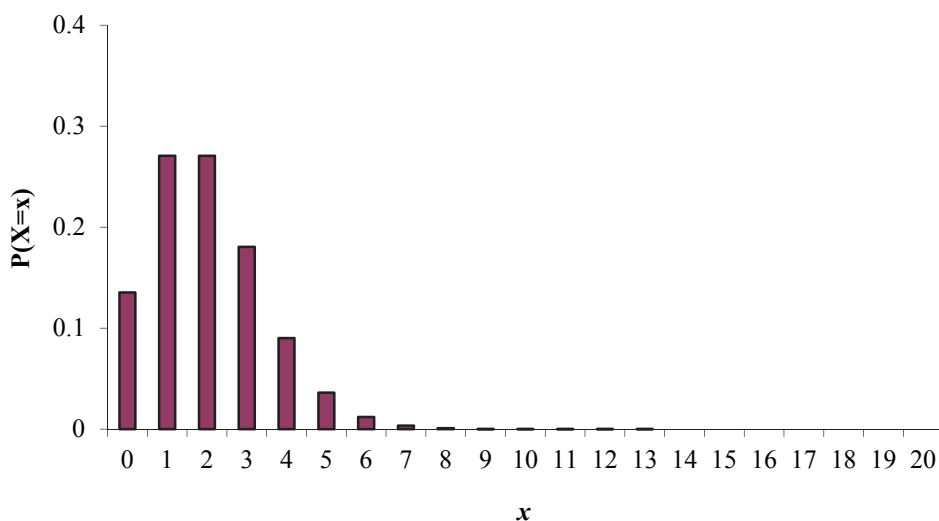
This distribution is called the **Poisson distribution** with parameter λ .

The shortcut way of writing 'X has a Poisson distribution with parameter λ ' is:

$$X \sim \text{Poi}(\lambda)$$

Some textbooks use μ instead of λ to remind us that the parameter gives us the *mean* rate of occurrences.

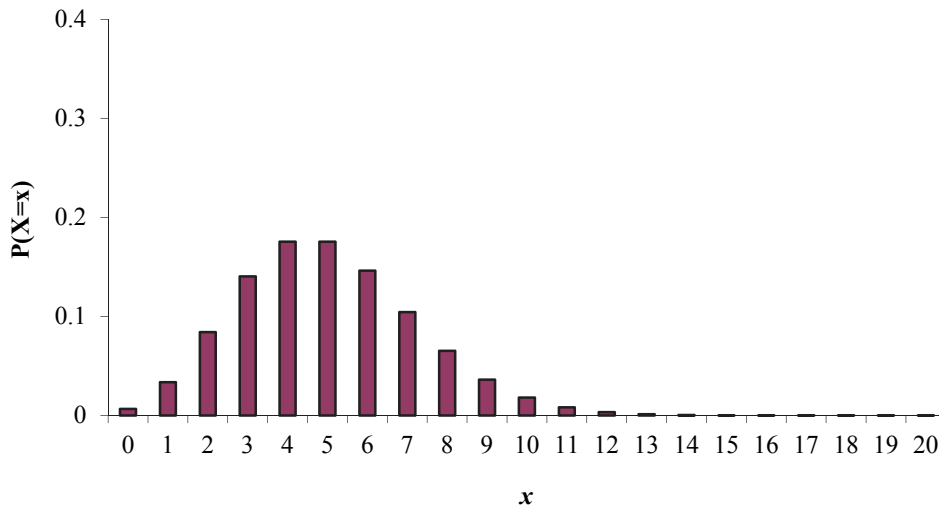
To get a better feel for this distribution, we'll look at a number of Poisson distributions with different values for λ .



In the graph above, we have $\lambda = 2$, that is we have a mean rate of occurrences of 2. We can see that we are most likely to get about 2 occurrences.

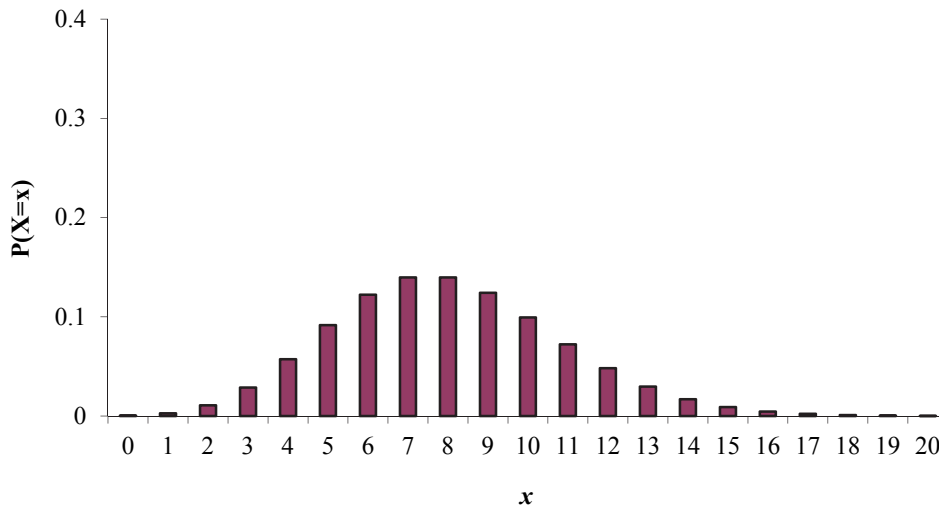
We can see that the Poisson distribution is positively skew.

What happens as we increase λ to, say 5?



Since we now have a mean rate of occurrences of 5, we can see that we are most likely to get about 5 occurrences. Since λ is greater than before, we are more likely to get more occurrences.

Increasing λ to 8 we get:



Since we now have a mean rate of occurrences of 8, we can see that we are most likely to get about 8 occurrences. Since λ is greater than before, we are more likely to get more occurrences.

Notice how as λ increases the distribution becomes more symmetrical.

Moments of the Poisson distribution

Mean

We can find the mean by using the formula:

$$E(X) = \sum xP(X = x)$$

and by using the series expansion of e^x :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Question 1.13

The Poisson distribution has probability function:

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

Use the formulae above to show that the mean is λ .

This answer should be obvious: a Poisson distribution with a *mean* rate λ has a mean of λ .

Variance

We can find the variance using the formulae:

$$\text{var}(X) = E(X^2) - E^2(X) \quad \text{where} \quad E(X^2) = \sum x^2 P(X = x)$$

and by using the series expansion of e^x :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

This gives:

$$\text{var}(X) = \lambda$$

The proof is messy but can be found in Appendix B.

Question 1.14

The number of radioactive particles emitted per hour has a Poisson distribution with $\lambda = 25$.

Find the mean and standard deviation of the number of particles emitted in an hour.

Median

There is no easy way to get the median value other than counting through the probabilities to find which value is half way through the distribution. For example, when $X \sim Poi(2)$ we have:

| x | 0 | 1 | 2 | 3 | 4 | 5 | ... |
|----------|-------|-------|-------|-------|-------|-------|-----|
| $P(X=x)$ | 0.135 | 0.271 | 0.271 | 0.180 | 0.090 | 0.036 | ... |

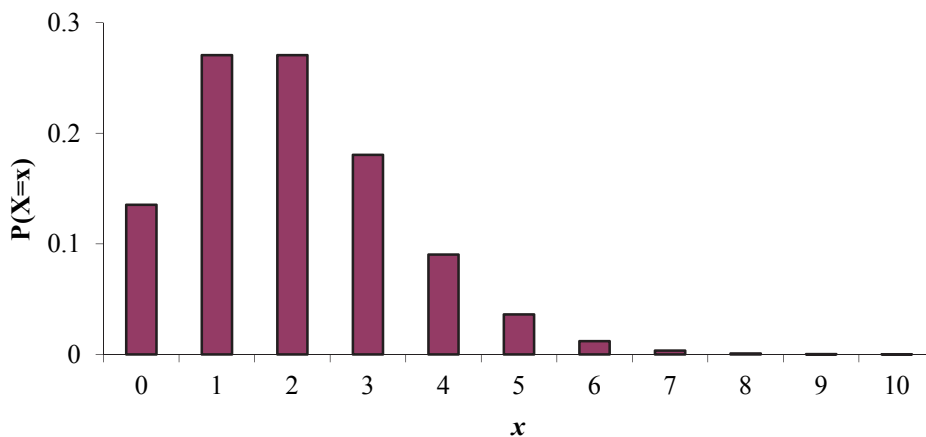
Annotations:

- A box containing 0.135 has an arrow pointing to the probability for $x=0$.
- A box containing $0.135 + 0.271 = 0.406$ has an arrow pointing to the cumulative probability between $x=0$ and $x=1$.
- A box containing $0.135 + 0.271 + 0.271 = 0.677$ has an arrow pointing to the cumulative probability between $x=0$ and $x=2$.
- A box containing "so 0.5 is in here!" has an arrow pointing to the probability for $x=2$.

So we can see that the median is 2.

Mode

The mode is the value that has the greatest probability. Looking at either the probability distribution of $X \sim Poi(2)$ given above or its graph below:



It is easy to see that 1 and 2 are both modes as these have the greatest probability of occurring.

Question 1.15

Calculate the median and mode of $X \sim Poi(1)$.

In summary:

Poisson distribution, $Poi(\lambda)$

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

$$E(X) = \lambda \quad \text{var}(X) = \lambda$$

Probabilities of a Poisson distribution

We can use the probability function to calculate probabilities. Suppose $X \sim Poi(3)$, so that our probability function is:

$$P(X = x) = \frac{3^x}{x!} e^{-3} \quad x = 0, 1, 2, \dots$$

To calculate a single probability, $P(X = 2)$, we just substitute the value into our probability function:

$$P(X = 2) = \frac{3^2}{2!} e^{-3} = \frac{9}{2} e^{-3} = 0.22404$$

What about calculating $P(X \geq 2)$? Since $X \sim Poi(3)$ can take values $0, 1, 2, \dots$ this means:

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + \dots$$

There's no way we can work it out this way. So we need to use the fact that probabilities of all the values X can take sum to 1:

$$P(X < 2) + P(X \geq 2) = 1 \Rightarrow P(X \geq 2) = 1 - P(X < 2)$$

Since X can take the values $0, 1, 2, 3, \dots$ we have:

$$P(X < 2) = P(X = 0) + P(X = 1)$$

All we have to do is work out $P(X = 0)$ and $P(X = 1)$:

$$P(X = 0) = \frac{3^0}{0!} e^{-3} = e^{-3} = 0.04979$$

$$P(X = 1) = \frac{3^1}{1!} e^{-3} = 3e^{-3} = 0.14936$$

Hence:

$$P(X < 2) = 0.04979 + 0.14936 = 0.19915$$

$$\Rightarrow P(X \geq 2) = 1 - 0.19915 = 0.80085$$

Question 1.16

Given that $X \sim \text{Poi}(1.8)$, calculate:

- (i) $P(X = 4)$
- (ii) $P(X \geq 1)$
- (iii) $P(2 \leq X < 5)$.

Miscellaneous problems

'In context' questions

Many of the questions you come across will be presented 'in context' rather than stating that the random variable has a Poisson distribution.

The key to spotting that we have a Poisson distribution (as opposed to a binomial distribution) is that we can have an infinite number of occurrences per unit time/length.

The question will refer to a rate (or mean number) of occurrences per unit time/length. This rate (or the mean) gives us the value of λ for that unit time/length.

For example, if we are told that telephone calls to a switchboard arrive at a rate of 4 per hour this means that $\lambda = 4$. We then use the Poisson distribution to work out probabilities of calls *per hour*.

Question 1.17

The mean number of accidents (in a year) on a twenty-mile stretch of motorway is 5. Calculate the probability that:

- (i) there are exactly 7 accidents on the twenty miles of road in any given year
- (ii) there are more than 3 accidents on the twenty miles of road in any given year.

Conditional probabilities

We can work out conditional probabilities involving the Poisson distribution using the conditional probability formula:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

For example, to calculate $P(X=1|X<2)$ where $X \sim Poi(3)$:

$$P(X=1|X<2) = \frac{P(X=1 \text{ and } X<2)}{P(X<2)} = \frac{P(X=1)}{P(X<2)}$$

Now:

$$P(X<2) = P(X=0) + P(X=1) = 0.04979 + 0.14936 = 0.19915$$

since:

$$P(X=0) = \frac{3^0}{0!} e^{-3} = 0.04979$$

$$P(X=1) = \frac{3^1}{1!} e^{-3} = 0.14936$$

So:

$$P(X=1|X<2) = \frac{P(X=1)}{P(X<2)} = \frac{0.14936}{0.19915} = 0.750$$

Changing the units

Let's look at our telephone example again. We had telephone calls to a switchboard arriving at a rate of 4 per hour. Our distribution was:

$$X \sim \text{Poi}(4) \quad \text{calls per hour}$$

A common trick used by the examiners is to change the unit. For example, they could ask us to calculate the probability that we have 10 calls in *5 hours*.

Our unit has changed from 1 hour to 5 hours. This means our distribution needs to change as well.

A bit of common sense tells us that if we have calls arriving at 4 per hour on average, then we will have 20 calls per *5 hours* on average. So to calculate probabilities for 5 hours, we use:

$$X \sim \text{Poi}(20) \quad \text{calls per 5 hours}$$

Question 1.18

- (i) A small company receives home insurance claims at a rate of 3 per month. Calculate the probability that the company receives 20 claims in a year.
- (ii) The same company receives car insurance claims at a rate of 4.8 per working week. Calculate the probability no claims are received in a day.

Adding Poisson distributions

Suppose a company receives on average 2 travel insurance claims in a week and on average 1 pet insurance claims in a week. So we have:

$X \sim Poi(2)$ travel insurance claims per week

$Y \sim Poi(1)$ pet insurance claims per week

How can we work out the probability that the company has a total of 4 claims of any type in a week?

Well, a bit of common sense tells us that since we have 2 travel and 1 pet claim per week on average, we will have a total of 3 claims of any type per week on average:

$Poi(3)$ total number of claims of any type per week

What we have shown is that we can add Poisson distributions together:

$$' Poi(2) + Poi(1) \sim Poi(3) '$$

Question 1.19

Car insurance policyholders make claims at a rate of 0.8 per year. Calculate the probability that 5 policyholders make a total of 3 claims in a year.

Appendix A – mean and variance for $P(X = x) = \frac{1}{k}$

For the discrete uniform distribution the probability function is:

$$P(X = x) = \frac{1}{k} \quad x = 1, 2, \dots, k$$

Mean

The mean can be found using:

$$E(X) = \sum xP(X = x)$$

This gives:

$$\begin{aligned} E(X) &= \left(1 \times \frac{1}{k}\right) + \left(2 \times \frac{1}{k}\right) + \dots + \left(k \times \frac{1}{k}\right) \\ &= \frac{1}{k}(1 + 2 + \dots + k) \end{aligned}$$

We now use the pure maths result that:

$$1 + 2 + \dots + n = \frac{1}{2}n(n + 1)$$

This gives:

$$E(X) = \frac{1}{k} \times \frac{1}{2}k(k + 1) = \frac{1}{2}(k + 1)$$

An alternative method is to note that the mean is the middle value (ie the median) due to symmetry. The median is the $\frac{1}{2}(n + 1)$ th value. Now since the values are just 1, 2, ... the $\frac{1}{2}(n + 1)$ th value is $\frac{1}{2}(n + 1)$.

Variance

The variance is found using:

$$\text{var}(X) = E(X^2) - E^2(X)$$

where:

$$E(X^2) = \sum x^2 P(X = x)$$

This gives:

$$\begin{aligned} E(X^2) &= \left(1^2 \times \frac{1}{k}\right) + \left(2^2 \times \frac{1}{k}\right) + \dots + \left(k^2 \times \frac{1}{k}\right) \\ &= \frac{1}{k}(1^2 + 2^2 + \dots + k^2) \end{aligned}$$

We now use the pure maths result that:

$$1^2 + 2^2 + \dots + n^2 = \frac{1}{6}n(n+1)(2n+1)$$

This gives:

$$E(X^2) = \frac{1}{k} \times \frac{1}{6}k(k+1)(2k+1) = \frac{1}{6}(k+1)(2k+1)$$

Substituting, we get:

$$\begin{aligned} \text{var}(X) &= E(X^2) - E^2(X) \\ &= \frac{1}{6}(k+1)(2k+1) - \left[\frac{1}{2}(k+1)\right]^2 \\ &= \frac{1}{6}(k+1)(2k+1) - \frac{1}{4}(k+1)^2 \\ &= \frac{1}{6}(2k^2 + 3k + 1) - \frac{1}{4}(k^2 + 2k + 1) \\ &= \frac{1}{3}k^2 + \frac{1}{2}k + \frac{1}{6} - \frac{1}{4}k^2 - \frac{1}{2}k - \frac{1}{4} \\ &= \frac{1}{12}k^2 - \frac{1}{12} \\ &= \frac{1}{12}(k^2 - 1) \end{aligned}$$

Appendix B – variance of $X \sim \text{Poi}(\lambda)$

For the Poisson distribution the probability function is:

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

The variance is found using:

$$\text{var}(X) = E(X^2) - E^2(X) \quad \text{where} \quad E(X^2) = \sum x^2 P(X=x)$$

This gives:

$$\begin{aligned} E(X^2) &= (0^2 \times e^{-\lambda}) + (1^2 \times \lambda e^{-\lambda}) + \left(2^2 \times \frac{\lambda^2}{2!} e^{-\lambda}\right) + \left(3^2 \times \frac{\lambda^3}{3!} e^{-\lambda}\right) + \dots \\ &= \lambda e^{-\lambda} + \frac{4\lambda^2}{2!} e^{-\lambda} + \frac{9\lambda^3}{3!} e^{-\lambda} + \frac{16\lambda^4}{4!} e^{-\lambda} + \dots \\ &= \lambda e^{-\lambda} \left\{ 1 + \frac{4\lambda}{2!} + \frac{9\lambda^2}{3!} + \frac{16\lambda^3}{4!} + \dots \right\} \\ &= \lambda e^{-\lambda} \left\{ 1 + 2\lambda + \frac{3\lambda^2}{2!} + \frac{4\lambda^3}{3!} + \dots \right\} \end{aligned}$$

Recall that the exponential series is:

$$e^\lambda = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots$$

The part in the brackets isn't quite the exponential series – so we'll split it up into the exponential series and a bit left over:

$$\begin{aligned} E(X^2) &= \lambda e^{-\lambda} \left\{ 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right. \\ &\quad \left. + \lambda + \frac{2\lambda^2}{2!} + \frac{3\lambda^3}{3!} + \dots \right\} \end{aligned}$$

The first series is just e^λ . We can now simplify the left over bit.

$$\begin{aligned}
E(X^2) &= \lambda e^{-\lambda} \left\{ e^{\lambda} + \lambda + \frac{2\lambda^2}{2!} + \frac{3\lambda^3}{3!} + \dots \right\} \\
&= \lambda e^{-\lambda} \left\{ e^{\lambda} + \lambda + \lambda^2 + \frac{\lambda^3}{2!} + \dots \right\} \\
&= \lambda e^{-\lambda} \left\{ e^{\lambda} + \lambda \left(1 + \lambda + \frac{\lambda^2}{2!} + \dots \right) \right\} \\
&= \lambda e^{-\lambda} \left\{ e^{\lambda} + \lambda e^{\lambda} \right\}
\end{aligned}$$

Multiplying out the bracket gives:

$$\begin{aligned}
E(X^2) &= \lambda e^0 + \lambda^2 e^0 \\
&= \lambda + \lambda^2
\end{aligned}$$

Therefore the variance is given by:

$$\begin{aligned}
\text{var}(X) &= E(X^2) - E^2(X) \\
&= (\lambda + \lambda^2) - \lambda^2 \\
&= \lambda
\end{aligned}$$

Extra practice questions

P1.1 In a particular road it is estimated that there is a 25% chance that any specified house will be burgled over a period of two years, independently for each house. There are six houses in the road.

Calculate the probability that fewer than two houses will be burgled over the period. [3]

P1.2 The probability of suffering a side effect from a certain flu vaccine is 0.005. If 1,000 persons are inoculated, calculate the probability that at most 1 person suffers a side effect. [2]

P1.3 Insurance policies of a particular type covering risks related to accidents are such that each policy, independently of all others, has probability 0.025 of giving rise to a claim exceeding £100,000 in a year.

Consider the claims experience of a random sample of 500 such policies over a year.

Determine the most likely number of policies in the sample to give rise to claims exceeding £100,000.

P1.4 A very crude model for the distribution of claim size, X , in a particular situation represents X as a discrete random variable, which takes the values £5,000, £10,000, and £20,000 with probabilities 0.4, 0.5, and 0.1 respectively.

Calculate the probability that of five randomly selected claims, three are for £5,000 each and the other two are for larger amounts. [2]

P1.5 In a sampling inspection scheme 3 bottles in every box of 12 are examined. If all 3 selected are found to be faulty, the remaining 9 bottles are examined. If at least one of the first 3 is satisfactory, no further bottles from the box are examined. Assume that each bottle examined has the same chance independently of being faulty, equal to 0.4.

When a box is inspected under this sampling scheme, work out the probability that 5 faulty bottles are found. [3]

P1.6 If X has a Poisson distribution with mean 5, what is the value of $E(X^2)$? [2]

P1.7 For a certain class of non-life insurance business, the number of claims per policy in a year has a Poisson distribution with mean 0.2.

What is the probability that there are a total of exactly 3 claims in a year under a group of 12 identical and unrelated such policies? [2]

P1.8 For a certain type of insurance business, the number of claims per policy in a year has a Poisson distribution with mean 0.4.

Consider a policy, which you know, has given rise to at least one claim in the last year. What is the probability that this policy has in fact given rise to exactly two claims in the last year? [3]

P1.9 The number of demands made on a service team each day has a Poisson distribution with mean 2. Under current arrangements the service team can handle, at most, 3 demands per day and no demands are carried forward.

Calculate the mean number of demands handled by the service team per day. [3]

Summary

The binomial distribution

If X has a **binomial distribution** with parameters n and p then we write $X \sim \text{Bin}(n, p)$. It models the number of 'successes' from n trials where each trial has a probability p of success. It has probability function:

$$P(X = x) = {}^n C_x p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n \quad 0 < p < 1$$

The moments are:

$$E(X) = np \quad \text{var}(X) = npq$$

Probabilities can be found by using the probability function, $P(X = x)$.

The Poisson distribution

If X has a **Poisson distribution** with parameter λ then we write $X \sim \text{Poi}(\lambda)$. It models the number of occurrences. It has probability function:

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, 2, 3, \dots$$

The moments are:

$$E(X) = \lambda \quad \text{var}(X) = \lambda$$

Probabilities can be found by using the probability function, $P(X = x)$.

If the unit changes, we can change the Poisson distribution.

$$\text{eg} \quad \text{Poi}(5) = \text{claims per month} \Rightarrow \text{Poi}(60) = \text{claims per year}$$

We can also add Poisson distributions:

$$\text{eg} \quad \text{Poi}(3) + \text{Poi}(5) \sim \text{Poi}(8)$$

Solutions

Solution 1.1

We have:

$$P(X = x) = \frac{1}{5} \quad x = 1, 2, 3, 4, 5$$

So the mean is given by:

$$\begin{aligned} E(X) &= \sum xP(X = x) \\ &= \left(1 \times \frac{1}{5}\right) + \left(2 \times \frac{1}{5}\right) + \left(3 \times \frac{1}{5}\right) + \left(4 \times \frac{1}{5}\right) + \left(5 \times \frac{1}{5}\right) \\ &= \frac{15}{5} = 3 \end{aligned}$$

Solution 1.2

We have:

$$P(X = x) = \frac{1}{5} \quad x = 1, 2, 3, 4, 5$$

The variance is given by:

$$\text{var}(X) = E(X^2) - E^2(X)$$

So we need $E(X^2)$:

$$\begin{aligned} E(X^2) &= \sum x^2 P(x = x) \\ &= \left(1^2 \times \frac{1}{5}\right) + \left(2^2 \times \frac{1}{5}\right) + \left(3^2 \times \frac{1}{5}\right) + \left(4^2 \times \frac{1}{5}\right) + \left(5^2 \times \frac{1}{5}\right) \\ &= \frac{55}{5} = 11 \end{aligned}$$

$$\begin{aligned} \Rightarrow \text{var}(X) &= E(X^2) - E^2(X) \\ &= 11 - 3^2 \\ &= 2 \end{aligned}$$

Solution 1.3

The probabilities are:

| | | | | | |
|-----------|---------------|---------------|---------------|---------------|---------------|
| x | 1 | 2 | 3 | 4 | 5 |
| P (X = x) | $\frac{1}{5}$ | $\frac{1}{5}$ | $\frac{1}{5}$ | $\frac{1}{5}$ | $\frac{1}{5}$ |

Counting through these gives:

| | | | | | |
|-----------|---------------|---------------|---------------|---------------|---------------|
| x | 1 | 2 | 3 | 4 | 5 |
| P (X = x) | $\frac{1}{5}$ | $\frac{1}{5}$ | $\frac{1}{5}$ | $\frac{1}{5}$ | $\frac{1}{5}$ |

$\frac{1}{5} = 0.2$

$\frac{1}{5} + \frac{1}{5} = \frac{2}{5} = 0.4$

so 0.5 is in here!

$\frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5} = 0.6$

We can see that the median is 3.

Solution 1.4

The probability distribution is:

$$P(X = x) = \frac{1}{20} \quad x = 1, 2, 3, \dots, 20$$

(i) The probability is:

$$\begin{aligned} P(X = \text{even}) &= P(X = 2, 4, 6, \dots, 20) \\ &= P(X = 2) + P(X = 4) + \dots + P(X = 20) \\ &= \frac{1}{20} + \frac{1}{20} + \dots + \frac{1}{20} \\ &= \frac{10}{20} \\ &= \frac{1}{2} \end{aligned}$$

(ii) The probability is:

$$\begin{aligned} P(X > 18) &= P(X = 19, 20) \\ &= P(X = 19) + P(X = 20) \\ &= \frac{1}{20} + \frac{1}{20} \\ &= \frac{2}{20} = 0.1 \end{aligned}$$

Solution 1.5

We have:

$$P(X = 1) = p$$

$$P(X = 0) = q$$

The mean is given by:

$$\begin{aligned} E(X) &= \sum xP(X = x) \\ &= (1 \times p) + (0 \times q) \\ &= p \end{aligned}$$

Solution 1.6

We have:

$$P(X = 1) = p$$

$$P(X = 0) = q$$

The variance is given by:

$$\text{var}(X) = E(X^2) - E^2(X)$$

So we need $E(X^2)$:

$$\begin{aligned} E(X^2) &= \sum x^2 P(X = x) \\ &= (1^2 \times p) + (0^2 \times q) \\ &= p \end{aligned}$$

Hence:

$$\begin{aligned} \text{var}(X) &= E(X^2) - E^2(X) \\ &= p - p^2 \\ &= p(1 - p) \\ &= pq \end{aligned}$$

Solution 1.7

For 2 successes we have 2 green and 2 not green lights with probability:

$$0.6^2 \times 0.4^2$$

We then need to calculate how many different ways there are of choosing 2 green sets from the 4 sets of traffic lights:

$4C_2$

Hence, the probability is given by:

$$P(X=2) = {}^4C_2 \times 0.6^2 \times 0.4^2$$

Similarly, we get:

$$P(X=4) = {}^4C_3 \times 0.6^3 \times 0.4^1$$

Solution 1.8

We want 13 green lights out of 20 lights. This gives a probability of:

$$P(X=13) = {}^{20}C_{13} \times 0.9^{13} \times 0.1^7 = 0.00197$$

Solution 1.9

We have 12 students (trials) with a probability of studying maths (success) of 0.7. So we have:

$$X \sim \text{Bin}(12, 0.7)$$

(i) The mean is given by $E(X) = np$, therefore:

$$E(X) = 12 \times 0.7 = 8.4$$

(ii) The variance is given by $\text{var}(X) = npq$, therefore:

$$\text{var}(X) = 12 \times 0.7 \times 0.3 = 2.52$$

Solution 1.10

First, we need to calculate the probabilities:

$$P(X=0) = {}^3C_0 \times 0.2^0 \times 0.8^3 = 0.512$$

$$P(X=1) = {}^3C_1 \times 0.2^1 \times 0.8^2 = 0.384$$

$$P(X=2) = {}^3C_2 \times 0.2^2 \times 0.8^1 = 0.096$$

$$P(X=3) = {}^3C_3 \times 0.2^3 \times 0.8^0 = 0.008$$

The mode is the value with the greatest probability, which is 0.

To find the median, we need to count through the probabilities until we get to 0.5. We can see that we are there in the very first value. So the median is also 0.

Solution 1.11

We have $X \sim \text{Bin}(9, 0.75)$, so our probability function is:

$$P(X = x) = {}^9C_x \times 0.75^x \times 0.25^{9-x} \quad x = 0, 1, \dots, 9$$

(i) Substituting directly into the probability function, we get:

$$P(X = 6) = {}^9C_6 \times 0.75^6 \times 0.25^3 = 0.23360$$

(ii) Now since X can take the values from 0 to 9, it will take too long to calculate $P(X < 8)$ using:

$$P(X < 8) = P(X = 0) + P(X = 1) + \dots + P(X = 7)$$

So we use the fact that all the probabilities sum to 1:

$$P(X < 8) = 1 - P(X \geq 8)$$

Now:

$$\begin{aligned} P(X \geq 8) &= P(X = 8) + P(X = 9) \\ &= 0.22525 + 0.07508 = 0.3003 \end{aligned}$$

where:

$$\begin{aligned} P(X = 8) &= {}^9C_8 \times 0.75^8 \times 0.25^1 = 0.22525 \\ P(X = 9) &= {}^9C_9 \times 0.75^9 \times 0.25^0 = 0.07508 \end{aligned}$$

Hence:

$$P(X < 8) = 1 - P(X \geq 8) = 1 - 0.3003 = 0.6997$$

(iii) $P(2 \leq X < 4) = P(X = 2) + P(X = 3)$

$$= 0.00124 + 0.00865 = 0.00989$$

where:

$$\begin{aligned} P(X = 2) &= {}^9C_2 \times 0.75^2 \times 0.25^7 = 0.00124 \\ P(X = 3) &= {}^9C_3 \times 0.75^3 \times 0.25^6 = 0.00865 \end{aligned}$$

Solution 1.12

We have 6 forms (trials) each with a 0.1 chance of being incorrectly filled in (success). So we have $X \sim \text{Bin}(6, 0.1)$ with probability function is:

$$P(X = x) = {}^6C_x \times 0.1^x \times 0.9^{6-x} \quad x = 0, 1, \dots, 6$$

(i) Substituting directly into the probability function, we get:

$$P(X = 2) = {}^6C_2 \times 0.1^2 \times 0.9^4 = 0.098415$$

(ii) Now since X can take the values from 0 to 6, we have:

$$\begin{aligned} P(X > 4) &= P(X = 5) + P(X = 6) \\ &= 0.000054 + 0.000001 \\ &= 0.000055 \end{aligned}$$

where:

$$P(X = 5) = {}^6C_5 \times 0.1^5 \times 0.9^1 = 0.000054$$

$$P(X = 6) = {}^6C_6 \times 0.1^6 \times 0.9^0 = 0.000001$$

Solution 1.13

For the Poisson distribution the probability function is:

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

The mean is found using:

$$E(X) = \sum xP(X=x)$$

This gives:

$$\begin{aligned} E(X) &= (0 \times e^{-\lambda}) + (1 \times \lambda e^{-\lambda}) + \left(2 \times \frac{\lambda^2}{2!} e^{-\lambda}\right) + \left(3 \times \frac{\lambda^3}{3!} e^{-\lambda}\right) + \dots \\ &= \lambda e^{-\lambda} + \lambda^2 e^{-\lambda} + \frac{\lambda^3}{2!} e^{-\lambda} + \frac{\lambda^4}{3!} e^{-\lambda} + \dots \\ &= \lambda e^{-\lambda} \left\{ 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right\} \end{aligned}$$

Recall that the exponential series is:

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots$$

So we have:

$$\begin{aligned} E(X) &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda \end{aligned}$$

Solution 1.14

We have:

$$X \sim \text{Poi}(25)$$

So:

$$E(X) = \lambda = 25$$

$$\text{var}(X) = \lambda = 25$$

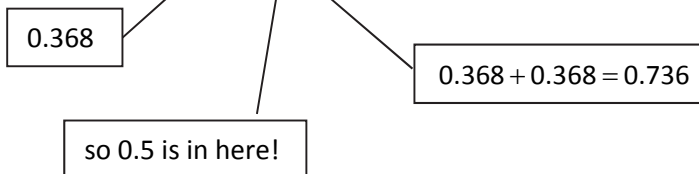
So the standard deviation is:

$$\sqrt{25} = 5$$

Solution 1.15

The probabilities are:

| | | | | | | | |
|------------------|-------|-------|-------|-------|-------|-------|-----|
| x | 0 | 1 | 2 | 3 | 4 | 5 | ... |
| P (X = x) | 0.368 | 0.368 | 0.184 | 0.061 | 0.015 | 0.003 | ... |



So the median is 1.

The mode is the value with the highest probability. Here we can see that both 0 and 1 are the mode.

Solution 1.16

(i) Using the probability function:

$$P(X=4) = \frac{1.8^4}{4!} e^{-1.8} = 0.07230$$

(ii) We have:

$$P(X \geq 1) = 1 - P(X=0) = 1 - \frac{1.8^0}{0!} e^{-1.8} = 0.83470$$

(iii) We have:

$$\begin{aligned} P(2 \leq X < 5) &= P(X=2) + P(X=3) + P(X=4) \\ &= \frac{1.8^2}{2!} e^{-1.8} + \frac{1.8^3}{3!} e^{-1.8} + \frac{1.8^4}{4!} e^{-1.8} \\ &= 0.26778 + 0.16067 + 0.07230 \\ &= 0.5008 \end{aligned}$$

Solution 1.17

The number of accidents in a year on the twenty-mile stretch of road can be modelled as a Poisson distribution with parameter 5.

(i)
$$P(X=7) = \frac{5^7}{7!} e^{-5} = 0.10444$$

(ii)
$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) \\ &= 1 - (P(X=0) + P(X=1) + P(X=2) + P(X=3)) \\ &= 1 - (0.00674 + 0.03369 + 0.08422 + 0.14037) \\ &= 0.7350 \end{aligned}$$

Solution 1.18

- (i) The number of insurance claims per month has a Poisson distribution with parameter 3.

Therefore the number of insurance claims per year has a Poisson distribution with parameter $3 \times 12 = 36$. So:

$$P(X = 20) = \frac{36^{20}}{20!} e^{-36} = 0.00127$$

- (ii) The number of car insurance claims per week has a Poisson distribution with parameter 4.8.

Therefore the number of insurance claims per day has a Poisson distribution with parameter $4.8 \div 5 = 0.96$. So:

$$P(X = 0) = \frac{0.96^0}{0!} e^{-0.96} = 0.38289$$

Solution 1.19

The number of claims per year for each policyholder has a Poisson distribution with parameter 0.8.

The number of claims per year for five policyholders has a Poisson distribution with parameter 4.

We have:

$$P(X = 3) = \frac{4^3}{3!} e^{-4} = 0.19537$$

Solutions to extra practice questions

P1.1 Let X be the number of houses burgled over the period. Then $X \sim \text{Bin}(6, 0.25)$.

$$\begin{aligned} P(X < 2) &= P(X = 0) + P(X = 1) \\ &= {}^6C_0 \times 0.25^0 \times 0.75^6 + {}^6C_1 \times 0.25^1 \times 0.75^5 \\ &= 0.17798 + 0.35596 = 0.534 \end{aligned}$$

P1.2 Let X be the number of people suffering a side effect. Then $X \sim \text{Bin}(1000, 0.005)$.

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= {}^{1,000}C_0 \times 0.005^0 \times 0.995^{1,000} + {}^{1,000}C_1 \times 0.005^1 \times 0.995^{999} \\ &= 0.0066540 + 0.033437 = 0.040 \end{aligned}$$

P1.3 Let X be the number of policies giving rise to a claim exceeding £100,000 in a year. Then $X \sim \text{Bin}(500, 0.025)$.

The most likely number is the number of policies that has the highest probability of occurring. Therefore we need to work out the probability of observing the values given in the question:

$$\begin{aligned} P(X = 11) &= {}^{500}C_{11} \times 0.025^{11} \times 0.975^{489} \\ &= 0.109648 \end{aligned}$$

$$\begin{aligned} P(X = 12) &= {}^{500}C_{12} \times 0.025^{12} \times 0.975^{488} \\ &= 0.114568 \end{aligned}$$

$$\begin{aligned} P(X = 13) &= {}^{500}C_{13} \times 0.025^{13} \times 0.975^{487} \\ &= 0.110275 \end{aligned}$$

$$\begin{aligned} P(X = 14) &= {}^{500}C_{14} \times 0.025^{14} \times 0.975^{486} \\ &= 0.098359 \end{aligned}$$

The most likely number of policies is 12.

P1.4 Let X be the number of claims for £5,000 (as opposed to over £5,000). Then $X \sim \text{Bin}(5, 0.4)$.

$$\begin{aligned} P(X=3) &= {}^5C_3 \times 0.4^3 \times 0.6^2 \\ &= 0.2304 \end{aligned}$$

P1.5 The probability that the first three are faulty is 0.4^3 . Otherwise we won't continue the inspection and won't find 5. We then require that 2 out of the remaining 9 bottles are faulty. The number of faulty bottles out of 9 has a binomial distribution with parameters $n=9$ and $p=0.4$.

The required probability is then $0.4^3 \times {}^9C_2 \times 0.4^2 \times 0.6^7 = 0.0103$.

P1.6 Using the formula for the variance backwards:

$$E(X^2) = \text{var}(X) + E^2(X)$$

But the mean and variance of the *Poisson*(5) distribution are both 5. So:

$$E(X^2) = 5 + 5^2 = 30$$

P1.7 For one policy, the number of claims has a Poisson distribution with mean 0.2. This means that for 12 policies, the number of claims has a Poisson distribution with mean 2.4.

Let X be the number of claims, then we want $P(X=3)$:

$$P(X=3) = \frac{e^{-2.4} 2.4^3}{3!} = 0.209$$

P1.8 Let X be the number of claims per policy in a year. We want $P(X=2|X \geq 1)$:

$$P(X=2|X \geq 1) = \frac{P(X=2 \text{ and } X \leq 1)}{P(X \geq 1)} = \frac{P(X=2)}{P(X \geq 1)}$$

We have:

$$P(X=2) = \frac{e^{-0.4} 0.4^2}{2!} = 0.053626$$

$$P(X \geq 1) = 1 - P(X=0) = 1 - \frac{e^{-0.4} 0.4^0}{0!} = 1 - 0.670320 = 0.329680$$

Therefore:

$$P(X=2|X \geq 1) = \frac{0.053626}{0.329680} = 0.1627$$

P1.9 The number of demands handled by the team each day is X , where:

$$P(X=0) = e^{-2} \quad P(X=1) = 2e^{-2}$$

$$P(X=2) = 2e^{-2} \quad P(X=3) = 1 - 5e^{-2}$$

So the mean of this random variable is:

$$E[X] = 0 \times e^{-2} + 1 \times 2e^{-2} + 2 \times 2e^{-2} + 3(1 - 5e^{-2}) = 3 - 9e^{-2} = 1.782$$

Continuous distributions

Introduction

Here we will be looking at two continuous distributions.

The **uniform distribution**, which assumes that all values have an equal chance of occurring.

The **exponential distribution**, which is useful for modelling lifetimes of equipment, claims amounts and the waiting times between events.

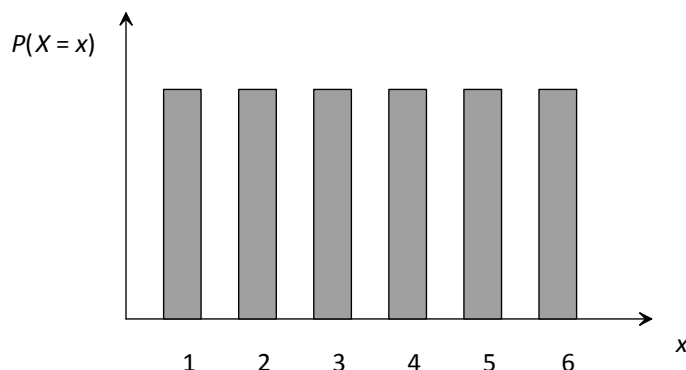
For each of these distributions we will look at graphs of the probability distribution functions, derive the basic results (mean, median, mode and variance) and find probabilities.

The uniform distribution

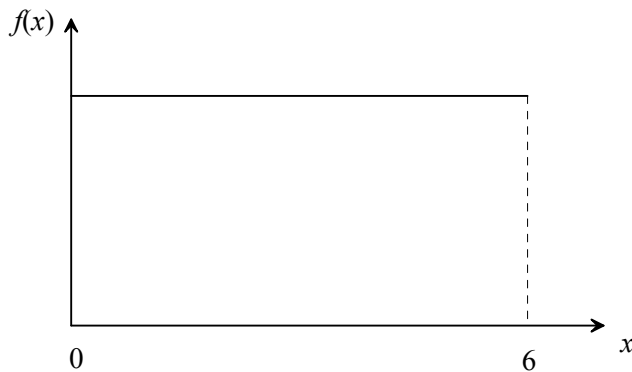
General features of the uniform distribution

Do you remember school uniform? Everyone was supposed to wear the *same* clothes. Well this is the idea behind the uniform distribution – all numbers have the *same* probability of occurring.

In the *discrete* uniform distribution each whole number has the same probability of occurring. For example, if we roll a fair die there are 6 numbers, each of which has the same probability ($\frac{1}{6}$) of occurring.



The continuous uniform extends this by allowing *any* number in a range to occur:



How can we work out the PDF for any uniform distribution? Recall that one of the properties of a PDF is that the area under its graph equals 1:

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Since the area under the graph is just a rectangle it's easy to find the height, $f(x)$, from the area (which must be 1) and the width using:

$$\text{height} = \frac{\text{area}}{\text{width}}$$

So for our distribution from 0 to 6 the width is $6-0=6$. Then our PDF is:

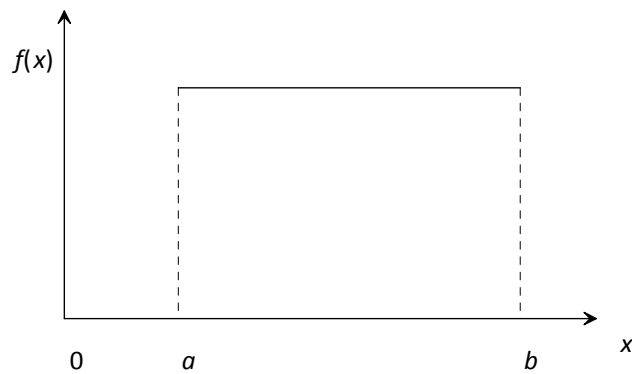
$$f(x) = \text{height} = \frac{\text{area}}{\text{width}} = \frac{1}{6}$$

Question 1.1

What would be the PDF for a continuous uniform distribution between 2 and 6?

The PDF of the uniform distribution

As we have seen all we need to calculate the PDF for a continuous uniform distribution is the values between which it lies, denoted a and b . So what is the PDF for a general continuous uniform distribution between a and b ?



Using the fact that the area under the graph is 1, we get:

$$f(x) = \text{height} = \frac{\text{area}}{\text{width}} = \frac{1}{b-a}$$

So we have:

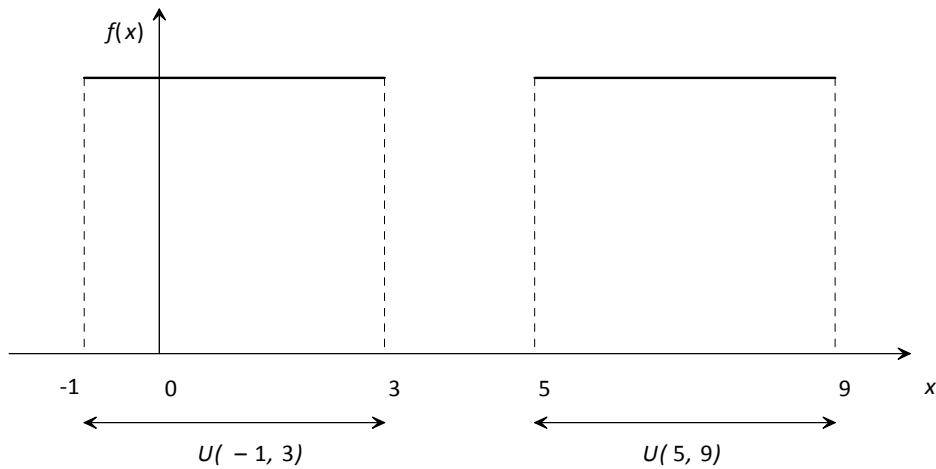
$$f(x) = \frac{1}{b-a} \quad a < x < b$$

The shortcut way of writing 'X has a uniform distribution between a and b ' is:

$$X \sim U(a, b)$$

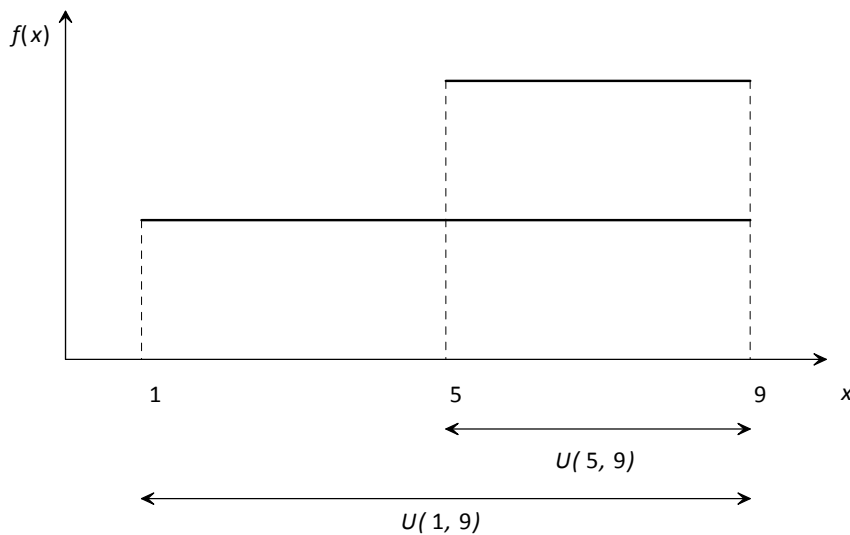
To get a better 'feel' for this distribution, we'll look at a number of continuous uniform distributions with different values for a and b .

First, let's look at $U(-1,3)$ and $U(5,9)$:



Both of these uniform distributions have the same width – so they will both have the same height. Only the position is different.

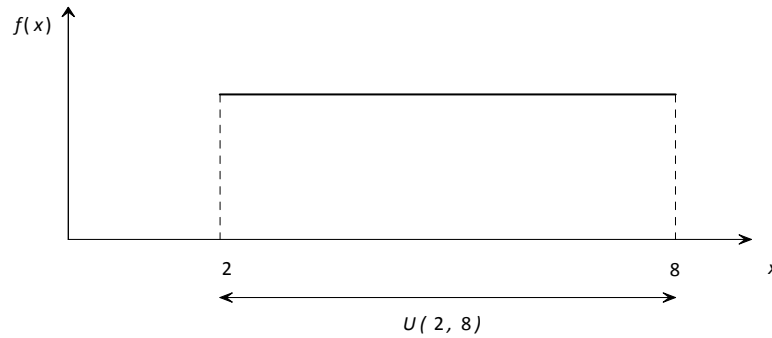
Next we'll look at what happens when we make the width bigger, say doubling the width from a $U(5,9)$ to a $U(1,9)$ distribution.



We can see that when the width has doubled (from 4 to 8), the height has halved (from $\frac{1}{4}$ to $\frac{1}{8}$). Since the distribution is spread over twice as many values, the probability of any particular value occurring must be half as much.

Moments of the uniform distribution

To find the moments for the continuous uniform distribution let's look at $X \sim U(2, 8)$:



Mean

Since the distribution is symmetrical, the mean is in the middle:

$$E(X) = \frac{2+8}{2} = 5$$

We could also have found the mean by using the formula:

$$E(X) = \int xf(x) dx$$

Question 1.2

Use integration to prove that the mean of the $U(2,8)$ distribution is 5.

Variance

There's no easy way to get the variance except to use the formula:

$$\text{var}(X) = E(X^2) - E^2(X)$$

Remembering that:

$$E(X^2) = \int x^2 f(x) dx$$

Question 1.3

Use the above formulae to show that the variance of $U(2,8)$ is 3.

Median

Since the graph of the PDF is symmetrical the median is also in the middle and is the same as the mean. Alternatively, we could prove this using the formula:

$$P(X \leq M) = \int_{\text{start}}^M f(x) dx = \frac{1}{2}$$

Question 1.4

Use the above formula to show that the median of the $U(2,8)$ distribution is 5 (same as $E(X)$).

Mode

Since $f(x)$ is the same for all values in the range there is no mode, as no value occurs with a greater probability than the rest.

In general, for a $U(a,b)$ distribution the results are:

Continuous uniform distribution, $X \sim U(a,b)$

$$f(x) = \frac{1}{b-a} \quad a < x < b$$

$$E(X) = \frac{1}{2}(a+b)$$

$$\text{var}(X) = \frac{1}{12}(b-a)^2$$

Probabilities for a uniform distribution

There are two ways of calculating the probabilities for a continuous uniform distribution: integrating the PDF or using the cumulative distribution function.

Calculating probabilities using integration

Recall that to calculate probabilities for a continuous random variable we just integrate the PDF:

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x) dx$$

This is easy to do for the continuous uniform distribution.

For example, to calculate $P(X > 8)$ where $X \sim U(7,12)$ we first need the PDF:

$$f(x) = \frac{1}{b-a} = \frac{1}{12-7} = \frac{1}{5}$$

Hence:

$$P(X > 8) = \int_8^{12} f(x) dx = \int_8^{12} \frac{1}{5} dx = \left[\frac{1}{5}x \right]_8^{12} = \frac{12}{5} - \frac{8}{5} = \frac{4}{5}$$

Question 1.5

If $X \sim U(13,30)$, calculate:

(i) $P(X < 24)$

(ii) $P(15 < X < 18.2)$.

Calculating probabilities using $F(x)$

Recall that the cumulative distribution function, $F(x)$, is:

$$F(x) = P(X < x) = \int_{\text{start}}^x f(t) dt$$

For the uniform distribution $U(a, b)$ distribution it is:

$$F(x) = \int_a^x \frac{1}{b-a} dt = \left[\frac{t}{b-a} \right]_a^x = \frac{x-a}{b-a}$$

We can use this to calculate any probabilities. For example:

$$X \sim U(7,12) \Rightarrow F(x) = \frac{x-7}{5}$$

$$P(X < 10) = F(10) = \frac{10-7}{5} = 0.6$$

$$P(X > 8.5) = 1 - P(X < 8.5) = 1 - F(8.5) = 1 - \frac{8.5-7}{5} = 0.7$$

$$\begin{aligned} P(8.4 < X < 10.2) &= P(X < 10.2) - P(X < 8.4) \\ &= F(10.2) - F(8.4) \\ &= \frac{10.2-7}{5} - \frac{8.4-7}{5} \\ &= 0.64 - 0.28 = 0.36 \end{aligned}$$

Question 1.6

Write down the cumulative distribution function, $F(x)$ for $X \sim U(-2,6)$ and use it to calculate:

- (i) $P(X < 5)$
- (ii) $P(X > 1.7)$
- (iii) $P(-1 < X < 5.5)$.

Miscellaneous problems

We can also work out conditional probabilities. For example, $P(X > 3 | X > 2)$ where $X \sim U(1,5)$. Using the conditional probability formula:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

So:

$$P(X > 3 | X > 2) = \frac{P(X > 3 \text{ and } X > 2)}{P(X > 2)} = \frac{P(X > 3)}{P(X > 2)} = \frac{0.5}{0.75} = \frac{2}{3}$$

Other problems that could be asked are to calculate the values of a and b given some information such as probabilities, the mean or the variance. For example, find b if $X \sim U(7,b)$ and $E(X) = 14$:

$$E(X) = \frac{7+b}{2} = 14 \Rightarrow b = 21$$

Question 1.7

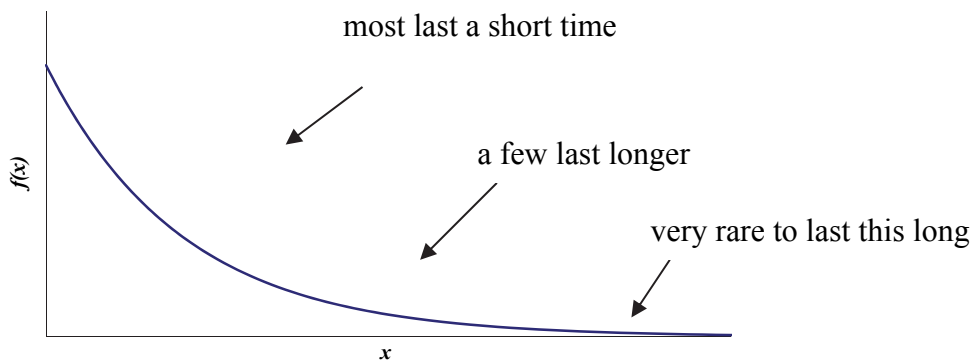
The random variable X has the distribution $X \sim U(a,b)$. The mean of X is 10 and the standard deviation is 4. Calculate the values of a and b .

The exponential distribution

General features of the exponential distribution

Consider the lifetime of a light bulb. Most last a short while until they break. A few bulbs last a long time and very rarely a bulb lasts for absolutely ages.

If we were to draw the PDF for the lifetime of the bulb, it would look something like this:



This looks rather like the graph of the exponential function $f(x) = e^{-x}$ where $x > 0$. Is this a PDF? Recall that a PDF must satisfy:

$$f(x) \geq 0 \text{ for all } x$$

$$\int f(x) dx = 1$$

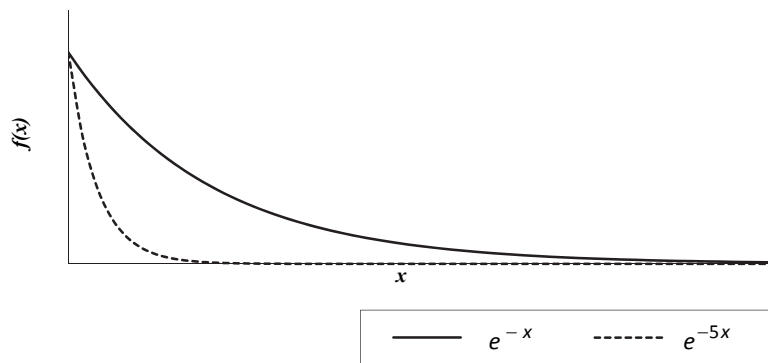
The graph is clearly non-negative so all we need to do is check that the area under the graph is 1:

$$\int_0^{\infty} e^{-x} dx = \left[-e^{-x} \right]_0^{\infty} = 0 - (-e^0) = 1$$

We have a PDF that can be used to model lifetimes. Our next question is how can we generalise this? For example, suppose we want an exponential with a shorter tail (they could be cheap light bulbs which don't last as long on average). We could choose:

$$f(x) = e^{-5x} \quad x > 0$$

The graph for this function looks like:



Is this a PDF as well? Well it's clearly non-negative, but since the area under e^{-x} is 1 we can see that the area under e^{-5x} must be less than this.

Maybe we could multiply the function by a constant, k , to scale the area up to 1:

$$f(x) = ke^{-5x} \quad x > 0$$

We'll now find out what this constant must be so that we *do* have a PDF:

$$\begin{aligned} \int_0^{\infty} ke^{-5x} dx &= 1 \\ \Rightarrow \left[-\frac{k}{5}e^{-5x} \right]_0^{\infty} &= 1 \\ 0 - \left(-\frac{k}{5} \right) &= 1 \Rightarrow k = 5 \end{aligned}$$

So the following function *is* a PDF:

$$f(x) = 5e^{-5x} \quad x > 0$$

Similarly, $f(x) = 3e^{-3x}$ will be a PDF and so on. In general, we have:

$$f(x) = \lambda e^{-\lambda x} \quad x > 0$$

Since these PDFs have the exponential function in them we call them **exponential distributions**.

The PDF of the exponential distribution

The **exponential distribution** has one parameter, λ . Its PDF is given by:

$$f(x) = \lambda e^{-\lambda x} \quad x > 0$$

The shortcut way of writing 'X has an exponential distribution with parameter λ ' is:

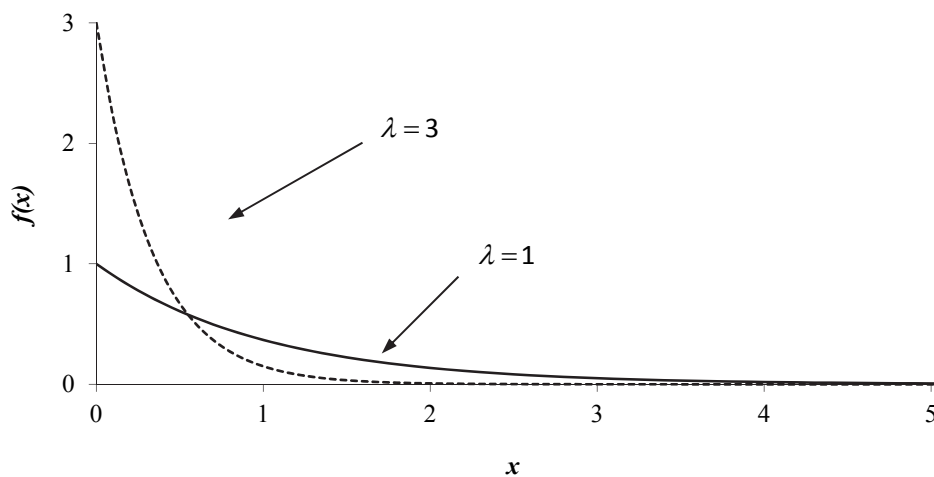
$$X \sim \text{Exp}(\lambda)$$

Question 1.8

Prove that:

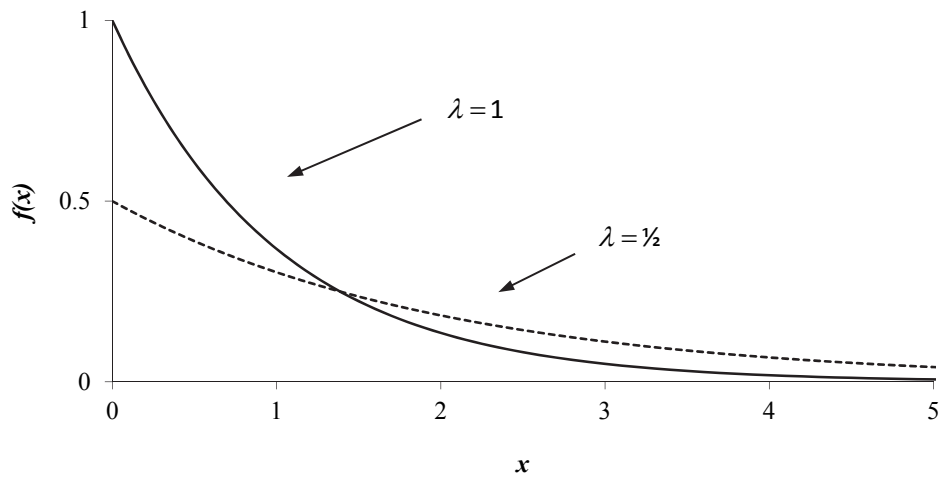
$$f(x) = \lambda e^{-\lambda x} \quad x > 0$$

is a PDF (*ie* the area under this function is 1).



The first thing to notice is that the graph always crosses the vertical axis at λ . Increasing the value of λ 'squashes' the graph towards to vertical axis. This means that as λ increases, the exponential distribution is more likely to take lower values.

A larger value of λ corresponds to a 'cheap' light bulb that has a shorter lifetime on average. As we decrease the value of λ , the graph becomes more 'stretched out':



With a smaller value of λ the exponential distribution is less likely to take smaller values (we can see the graph is lower at the start) and more likely to take higher values (the graph is higher towards the end).

A smaller value of λ would correspond to a 'deluxe' light bulb that has a longer lifetime on average.

Moments of the exponential distribution

Mean

We can find the mean by using the formula:

$$E(X) = \int xf(x) dx$$

Question 1.9

Use this formula to show that the mean of the exponential distribution is:

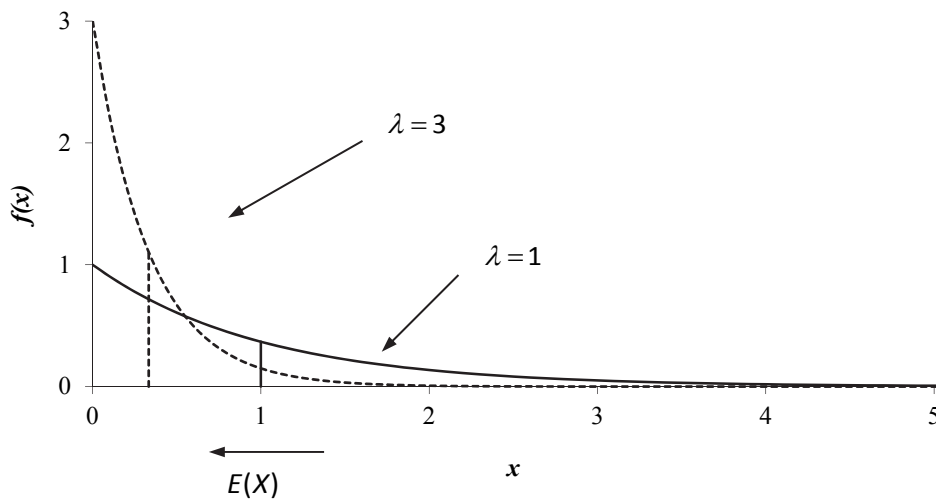
$$E(X) = \frac{1}{\lambda}$$

How does this result tie in with our graphs of the PDFs? Let's take a look at the means of two exponential distributions:

$$\lambda = 1 \quad E(X) = 1$$

$$\lambda = 3 \quad E(X) = \frac{1}{3}$$

The bigger the value of lambda, the smaller the mean.



We can see on the graph that as λ gets bigger the graph is 'squashed' more towards the vertical axis and so the mean will be smaller. Similarly as λ gets smaller the graph is more 'stretched out' and so the mean will be bigger.

Variance

The formula for the variance is:

$$\text{var}(X) = E(X^2) - E^2(X)$$

where:

$$E(X^2) = \int x^2 f(x) dx$$

Question 1.10

Use the above formulae to show that the variance of the exponential distribution is:

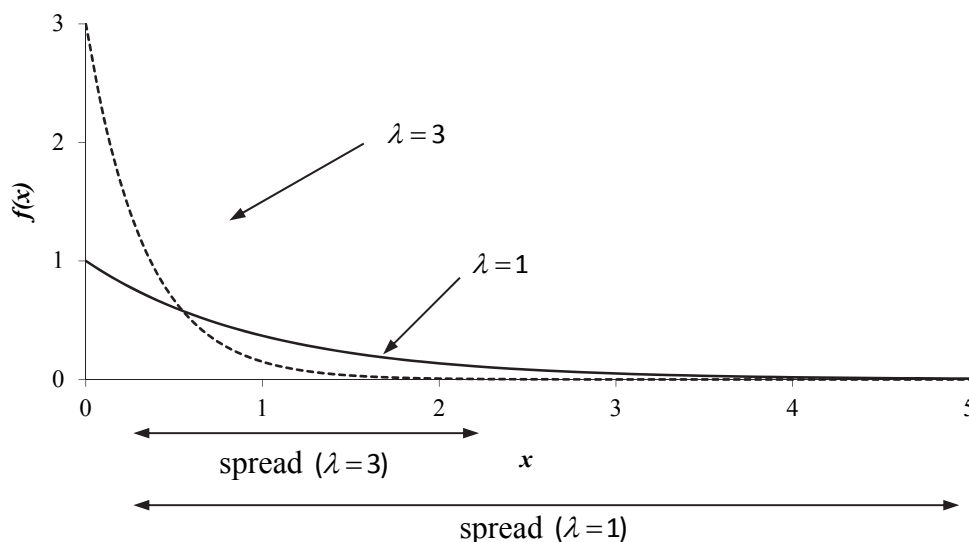
$$\text{var}(X) = \frac{1}{\lambda^2}$$

Now let's look at the variances of two exponential distributions:

$$\lambda = 1 \quad \text{var}(X) = 1$$

$$\lambda = 3 \quad \text{var}(X) = \frac{1}{9}$$

The bigger the value of λ , the smaller the variance (*ie* the smaller the spread).



We can see that (whilst the graphs continue indefinitely) as λ gets bigger the graph is 'squashed' more towards the vertical axis and so the spread of values is smaller (*ie* more of the values are bunched towards the start). Similarly as λ gets smaller the graph is more 'stretched out' and so the spread of values is greater.

Median

To find the median we use the formula:

$$\int_{\text{start}}^M f(x) dx = \frac{1}{2}$$

Question 1.11

Use this formula to show that the median of the exponential distribution is:

$$M = \frac{1}{\lambda} \ln 2$$

Mode

The graph of the exponential PDF is greatest when $x=0$ so this is the mode. However, technically it never quite reaches this point as x can only take positive values ($x > 0$).

In summary:

Exponential distribution, $Exp(\lambda)$

$$f(x) = \lambda e^{-\lambda x} \quad x > 0$$

$$E(X) = \frac{1}{\lambda}$$

$$\text{var}(X) = \frac{1}{\lambda^2}$$

Probabilities for an exponential distribution

There are two ways of calculating the probabilities for an exponential distribution: integrating the PDF or using the cumulative distribution function.

Calculating probabilities using integration

Recall that to calculate probabilities for a continuous random variable we just integrate the PDF:

$$P(a < X < b) = \int_a^b f(x) dx$$

This is easy to do for the exponential distribution.

For example, let's calculate $P(X > 5)$ where $X \sim Exp(0.1)$. First, we need the PDF:

$$f(x) = 0.1e^{-0.1x}$$

Next, we integrate this from 5 upwards:

$$P(X > 5) = \int_5^{\infty} f(x) dx = \int_5^{\infty} 0.1e^{-0.1x} dx = \left[-e^{-0.1x} \right]_5^{\infty} = 0 - (-e^{-0.5}) = 0.60653$$

Question 1.12

Claim amounts for an insurer are exponentially distributed with parameter 0.005. Calculate the probability that a claim:

- (i) exceeds 210
- (ii) is less than 100
- (iii) is between 300 and 500.

Calculating probabilities using $F(x)$

Recall that the cumulative distribution function, $F(x)$, is:

$$F(x) = P(X < x) = \int_{\text{start}}^x f(t) dt$$

For the exponential distribution, $X \sim \text{Exp}(\lambda)$, it is:

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t} \right]_0^x = 1 - e^{-\lambda x}$$

We can use this to calculate any probabilities. For example:

$$X \sim \text{Exp}(0.1) \Rightarrow F(x) = 1 - e^{-0.1x}$$

$$P(X < 5) = F(5) = 1 - e^{-0.5} = 0.39347$$

$$P(X > 2) = 1 - P(X < 2) = 1 - F(2) = 1 - (1 - e^{-0.2}) = e^{-0.2} = 0.81873$$

$$\begin{aligned} P(3.1 < X < 4.7) &= P(X < 4.7) - P(X < 3.1) \\ &= F(4.7) - F(3.1) \\ &= (1 - e^{-0.47}) - (1 - e^{-0.31}) \\ &= 0.37500 - 0.26655 = 0.10844 \end{aligned}$$

Question 1.13

Write down the cumulative distribution function, $F(x)$ for $X \sim \text{Exp}(0.5)$ and use it to calculate:

- (i) $P(X < 3)$
- (ii) $P(X > 0.5)$
- (iii) $P(2 < X < 8)$.

Miscellaneous problems

Given the mean of the distribution

Questions may give the **mean**, μ , of the exponential distribution rather than the parameter, λ .

Therefore we will have to calculate λ first *before* we can write down the PDF and calculate probabilities. To do this we will need to use:

$$\mu = E(X) = \frac{1}{\lambda}$$

For example, suppose we are given an exponential distribution with mean 100:

$$\mu = 100 = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{100} = 0.01$$

We can now write down the PDF:

$$f(x) = 0.01e^{-0.01x} \quad x > 0$$

This can be used to calculate probabilities.

Question 1.14

The lifetime of a certain battery is exponentially distributed with mean 500 hours. Calculate the probability that the battery lasts:

- (i) more than 700 hours
- (ii) between 400 and 600 hours.

Common Error:

Students often assume that the parameter λ is given without reading the question carefully to see whether the parameter or the mean is given.

Conditional probabilities

We can work out conditional probabilities involving the exponential distribution using the conditional probability formula:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

For example, to calculate $P(X < 1 | X < 2)$ where $X \sim \text{Exp}(2)$:

$$P(X < 1 | X < 2) = \frac{P(X < 1 \text{ and } X < 2)}{P(X < 2)} = \frac{P(X < 1)}{P(X < 2)} = \frac{1 - e^{-2}}{1 - e^{-4}} = 0.88080$$

'Backwards' problems

We could be asked to calculate the values of λ given a probability. For example, find λ if $X \sim \text{Exp}(\lambda)$ and $P(X < 10) = 0.05$:

$$P(X < 10) = \int_0^{10} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_0^{10} = 1 - e^{-10\lambda}$$

But since $P(X < 10) = 0.05$, we have:

$$1 - e^{-10\lambda} = 0.05 \Rightarrow e^{-10\lambda} = 0.95 \Rightarrow \lambda = -\frac{1}{10} \ln 0.95 = 0.00513$$

Question 1.15

Given that $X \sim \text{Exp}(\lambda)$ and $P(X > 40) = 0.7$, calculate:

- (i) λ_X
- (ii) $P(X > 105 | X > 80)$

Waiting time for a Poisson distribution

The Poisson distribution is excellent for modelling the number of insurance claims and is one of the most useful distributions for actuaries. We will now show that, if events occur as a Poisson distribution (with parameter λ), the waiting time between events has an exponential distribution (with parameter λ).

For example, suppose the number of claims arriving per hour has a Poisson distribution with parameter λ :

$Poi(\lambda)$ claims per hour

If X is the number of claims arriving in t hours, we would have:

$X \sim Poi(\lambda t)$ claims per t hours

So the probability function for the number of claims arriving in t hours is:

$$P(X=x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \quad x=0,1,2,3,\dots$$

Let T be the number of hours that we have been waiting since the last claim. The probability that we have been waiting for more than t hours for a claim is:

$$P(T > t)$$

Since we are still waiting, this means that we have had no claims in t hours:

$$P(T > t) = P(0 \text{ claims in } t \text{ hours})$$

We can find this probability using $P(X=0)$:

$$\begin{aligned} P(T > t) &= P(0 \text{ claims in } t \text{ hours}) \\ &= P(X=0) \quad \text{where } X \sim \text{Poi}(\lambda t) \\ &= \frac{(\lambda t)^0}{0!} e^{-\lambda t} \\ &= e^{-\lambda t} \end{aligned}$$

Since we know $P(T > t)$, we can find $P(T < t)$ as probabilities sum to 1:

$$P(T < t) = 1 - P(T > t) = 1 - e^{-\lambda t}$$

But this is the definition of the cumulative distribution function of T :

$$F(t) = P(T < t) = 1 - e^{-\lambda t}$$

Now either we can recognise that this is the cumulative distribution function of $\text{Exp}(\lambda)$ or we can use it to find the PDF:

$$f(t) = F'(t) = \lambda e^{-\lambda t}$$

This is the PDF of the $\text{Exp}(\lambda)$ distribution. So the waiting time (in hours) has an exponential distribution with parameter λ .

Question 1.16

The number of calls arriving at a switchboard each minute has a Poisson distribution with mean 5.

- (i) Write down the distribution for the time between calls.
- (ii) Calculate the probability that the time from one call to the next exceeds 1 minute.

Appendix A – proof of mean and variance for $U(a,b)$

For $X \sim U(a,b)$ the PDF is $f(x) = \frac{1}{b-a}$.

Mean

The mean is given by:

$$E(X) = \int xf(x) dx$$

This gives:

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \left[\frac{\frac{1}{2}x^2}{b-a} \right]_a^b = \frac{\frac{1}{2}(b^2 - a^2)}{b-a}$$

To simplify this expression, we use the difference of two squares result:

$$(x-y)(x+y) = x^2 - y^2$$

This gives:

$$E(X) = \frac{\frac{1}{2}(b^2 - a^2)}{b-a} = \frac{\frac{1}{2}(b-a)(b+a)}{b-a} = \frac{1}{2}(a+b)$$

Variance

The variance is found using:

$$\text{var}(X) = E(X^2) - E^2(X)$$

where:

$$E(X^2) = \int x^2 f(x) dx$$

This gives:

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \left[\frac{\frac{1}{3}x^3}{b-a} \right]_a^b = \frac{\frac{1}{3}(b^3 - a^3)}{b-a}$$

To simplify the expression, we use the fact that:

$$(b-a)(a^2 + ab + b^2) = b^3 - a^3$$

This gives:

$$E(X^2) = \frac{\frac{1}{3}(b^3 - a^3)}{b-a} = \frac{\frac{1}{3}(b-a)(a^2 + ab + b^2)}{b-a} = \frac{1}{3}(a^2 + ab + b^2)$$

Substituting, we get:

$$\begin{aligned} \text{var}(X) &= E(X^2) - E^2(X) \\ &= \frac{1}{3}(a^2 + ab + b^2) - \left[\frac{1}{2}(a+b) \right]^2 \\ &= \frac{1}{3}(a^2 + ab + b^2) - \frac{1}{4}(a+b)^2 \\ &= \frac{1}{3}(a^2 + ab + b^2) - \frac{1}{4}(a^2 + 2ab + b^2) \\ &= \frac{1}{12}a^2 - \frac{1}{6}ab + \frac{1}{12}b^2 \\ &= \frac{1}{12}(a^2 - 2ab + b^2) \\ &= \frac{1}{12}(b-a)^2 \end{aligned}$$

Extra practice questions**P1.1** If $X \sim U(15,21)$:

- (i) Sketch the PDF of X .
- (ii) Calculate:
 - (a) $P(X > 18)$
 - (b) $P(19.5 < X < 20.3)$.
- (iii) Calculate the mean and variance of X .

P1.2 The random variable X is distributed uniformly over $[2, 4]$.

- (i) State the value of the population mean $E(X)$. [1]
- (ii) Show that $\text{var}(X) = \frac{1}{3}$. [2]

[Total 3]

P1.3 A random variable, Y , is uniformly distributed with mean 7 and standard deviation 5.

- (i) Obtain the PDF of Y .
- (ii) Calculate $P(Y > 5 | Y > 0)$.

P1.4 Claim amounts are modelled as an exponential random variable with mean £1,000.

- (i) Calculate the probability that one such claim amount is greater than £5,000. [1]
- (ii) Calculate the probability that a claim amount is greater than £5,000 given that it is greater than £1,000. [2]

[Total 3]

P1.5 Claim sizes are modelled by an exponential distribution with mean £50. Determine the probabilities that an individual claim:

- (i) exceeds £50 [1]
- (ii) exceeds £200 [1]
- (iii) exceeds £200, given that it exceeds £50. [2]

[Total 4]

P1.6 The median, m , of the distribution of a continuous random variable X is defined by the relation $P(X < m) = 0.5$.

Show that the median of an exponential random variable is about 0.7 of the mean. [4]

P1.7 The number of accidents in a factory is represented by a Poisson distribution averaging 2 accidents per 5 days. Evaluate the probability that the time from one accident to the next exceeds 3 days. [2]

Summary

The uniform distribution

If X has a **uniform distribution** between a and b then we write $X \sim U(a, b)$. It has PDF:

$$f(x) = \frac{1}{b-a} \quad a < x < b$$

The moments are:

$$E(X) = \frac{1}{2}(a + b)$$

$$\text{var}(X) = \frac{1}{12}(b - a)^2$$

The median is the same as the mean (by symmetry). There is no mode.

Probabilities can be found by integrating the PDF or using $F(x) = \frac{x-a}{b-a}$.

The exponential distribution

If X has an **exponential distribution** with parameter λ then we write $X \sim \text{Exp}(\lambda)$. It models the waiting time between events occurring in a Poisson process. It has PDF:

$$f(x) = \lambda e^{-\lambda x} \quad x > 0$$

The moments are:

$$E(X) = \frac{1}{\lambda}$$

$$\text{var}(X) = \frac{1}{\lambda^2}$$

The median can be found by integration or using $F(x)$ and is $\frac{1}{\lambda} \ln 2$. The mode is 0.

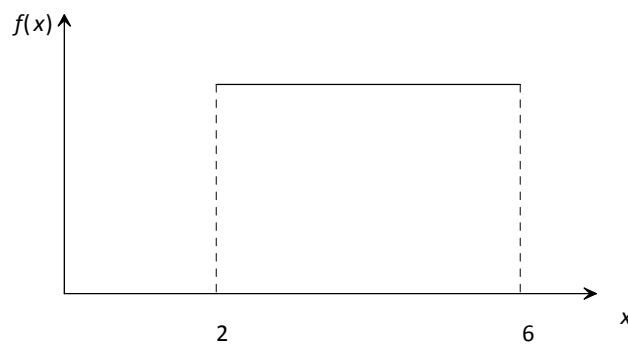
Probabilities can be found by integrating the PDF or using $F(x) = 1 - e^{-\lambda x}$.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Solutions

Solution 1.1

The graph of the PDF would be:



The area under this graph must be 1 (since it's a PDF). Our distribution is from 2 to 6 so the width is $6 - 2 = 4$. Therefore the PDF is:

$$f(x) = \text{height} = \frac{\text{area}}{\text{width}} = \frac{1}{4}$$

Solution 1.2

The PDF of $U(2,8)$ is:

$$f(x) = \frac{1}{8-2} = \frac{1}{6} \quad 2 < x < 6$$

So:

$$\begin{aligned} E(X) &= \int xf(x) dx \\ &= \int_2^8 x \frac{1}{6} dx \\ &= \left[\frac{1}{12} x^2 \right]_2^8 \\ &= \frac{1}{12} (8^2 - 2^2) \\ &= 5 \end{aligned}$$

Solution 1.3

The PDF of the $U(2,8)$ distribution is:

$$f(x) = \frac{1}{8-2} = \frac{1}{6} \quad 2 < x < 8$$

So:

$$\begin{aligned} E(X^2) &= \int x^2 f(x) dx \\ &= \int_2^8 x^2 \frac{1}{6} dx \\ &= \left[\frac{1}{18} x^3 \right]_2^8 \\ &= \frac{1}{18} (8^3 - 2^3) \\ &= 28 \end{aligned}$$

This gives:

$$\begin{aligned} \text{var}(X) &= E(X^2) - E^2(X) \\ &= 28 - 5^2 \\ &= 3 \end{aligned}$$

Solution 1.4

The PDF of the $U(2,8)$ distribution is:

$$f(x) = \frac{1}{8-2} = \frac{1}{6} \quad 2 < x < 6$$

So the median, M , is given by:

$$\int_2^M \frac{1}{6} dx = \frac{1}{2}$$

Solving this gives:

$$\left[\frac{1}{6}x \right]_2^M = \frac{1}{2}$$

$$\frac{1}{6}(M-2) = \frac{1}{2}$$

$$M = 5$$

Solution 1.5

The PDF of the $X \sim U(13,30)$ distribution is:

$$f(x) = \frac{1}{17} \quad 13 < x < 30$$

Integrating the PDF to find the probabilities gives:

$$\begin{aligned} \text{(i)} \quad P(X < 24) &= \int_{13}^{24} \frac{1}{17} dx \\ &= \left[\frac{1}{17}x \right]_{13}^{24} \\ &= \frac{1}{17}(24 - 13) \\ &= \frac{11}{17} \end{aligned}$$

$$\begin{aligned}
 \text{(ii)} \quad P(15 < X < 18.2) &= \int_{15}^{18.2} \frac{1}{17} dx \\
 &= \left[\frac{1}{17} x \right]_{15}^{18.2} \\
 &= \frac{1}{17}(18.2 - 15) \\
 &= 0.18824
 \end{aligned}$$

Solution 1.6

The cumulative distribution function of $X \sim U(-2, 6)$ is:

$$F(x) = P(X < x) = \frac{x - a}{b - a} = \frac{x + 2}{8}$$

$$\text{(i)} \quad P(X < 5) = F(5) = \frac{5 + 2}{8} = \frac{7}{8} = 0.875$$

$$\begin{aligned}
 \text{(ii)} \quad P(X > 1.7) &= 1 - P(X < 1.7) && \text{since probabilities sum to 1} \\
 &= 1 - F(1.7) \\
 &= 1 - \frac{3.7}{8} \\
 &= 0.5375
 \end{aligned}$$

$$\begin{aligned}
 \text{(iii)} \quad P(-1 < X < 5.5) &= P(X < 5.5) - P(X < -1) \\
 &= F(5.5) - F(-1) \\
 &= \frac{7.5}{8} - \frac{1}{8} \\
 &= 0.8125
 \end{aligned}$$

Solution 1.7

For $X \sim U(a, b)$ we have:

$$E(X) = \frac{1}{2}(a + b) = 10$$

$$\text{var}(X) = \frac{1}{12}(b - a)^2 = 4^2$$

Rearranging the mean gives $b = 20 - a$. Substituting this into the variance:

$$\frac{1}{12}(20 - a - a)^2 = 16$$

$$(20 - 2a)^2 = 192$$

Solving this:

$$20 - 2a = \pm\sqrt{192}$$

$$a = \frac{20 \pm \sqrt{192}}{2}$$

$$= 16.93 \text{ or } 3.07$$

$$\Rightarrow b = 20 - 3.07 = 16.93 \text{ or } b = 20 - 16.93 = 3.07$$

Since $a < b$, we have $a = 3.07$ and $b = 16.93$.

Solution 1.8

The area under the graph of $f(x) = \lambda e^{-\lambda x}$ is given by:

$$\int_0^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_0^{\infty} = 0 - (-e^0) = 1$$

Since the area under this graph is 1 and $f(x) > 0$, $f(x)$ is a PDF.

Solution 1.9

Using the formula for the mean, we get:

$$E(X) = \int xf(x) dx = \int_0^{\infty} x\lambda e^{-\lambda x} dx$$

We need to use integration by parts. The formula is:

$$\int_a^b u \frac{dv}{dx} dx = [uv]_a^b - \int_a^b v \frac{du}{dx} dx$$

Setting $u = x$ and $\frac{dv}{dx} = \lambda e^{-\lambda x}$ we get:

$$\begin{aligned} \int_0^{\infty} x\lambda e^{-\lambda x} dx &= \left[-xe^{-\lambda x} \right]_0^{\infty} - \int_0^{\infty} -e^{-\lambda x} dx \\ &= 0 + \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^{\infty} \\ &= 0 - \left(-\frac{1}{\lambda} e^0 \right) \\ &= \frac{1}{\lambda} \end{aligned}$$

Solution 1.10

The second moment, $E(X^2)$, is given by:

$$E(X^2) = \int x^2 f(x) dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx$$

We need to use integration by parts. The formula is:

$$\int_a^b u \frac{dv}{dx} dx = [uv]_a^b - \int_a^b v \frac{du}{dx} dx$$

Setting $u = x^2$ and $\frac{dv}{dx} = \lambda e^{-\lambda x}$ we get:

$$\begin{aligned} \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx &= \left[-x^2 e^{-\lambda x} \right]_0^{\infty} - \int_0^{\infty} -2x e^{-\lambda x} dx \\ &= 0 + \int_0^{\infty} 2x e^{-\lambda x} dx = \int_0^{\infty} 2x e^{-\lambda x} dx \end{aligned}$$

We now need to use integration by parts to calculate this integral. Setting $u = 2x$ and

$\frac{dv}{dx} = e^{-\lambda x}$ we get:

$$\begin{aligned} \int_0^{\infty} 2x e^{-\lambda x} dx &= \left[-2x \frac{1}{\lambda} e^{-\lambda x} \right]_0^{\infty} - \int_0^{\infty} -2 \frac{1}{\lambda} e^{-\lambda x} dx \\ &= 0 + \left[-2 \frac{1}{\lambda^2} e^{-\lambda x} \right]_0^{\infty} \\ &= 0 - \left(-\frac{2}{\lambda^2} e^0 \right) \\ &= \frac{2}{\lambda^2} \end{aligned}$$

Hence:

$$\text{var}(X) = E(X^2) - E^2(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}$$

Solution 1.11

Using:

$$\int_{\text{start}}^M f(x) dx = \frac{1}{2}$$

we get:

$$\int_0^M \lambda e^{-\lambda x} dx = \frac{1}{2}$$

$$\Rightarrow \left[-e^{-\lambda x} \right]_0^M = \frac{1}{2}$$

$$1 - e^{-\lambda M} = \frac{1}{2}$$

$$e^{-\lambda M} = \frac{1}{2}$$

$$-\lambda M = \ln \frac{1}{2}$$

$$\lambda M = -\ln \frac{1}{2}$$

Now using the fact that $a \ln x = \ln x^a$ gives $-\ln \frac{1}{2} = \ln \left(\frac{1}{2} \right)^{-1} = \ln 2$. So:

$$\Rightarrow \lambda M = \ln 2$$

$$M = \frac{1}{\lambda} \ln 2$$

Solution 1.12

The PDF of $X \sim \text{Exp}(0.005)$ is:

$$f(x) = 0.005e^{-0.005x} \quad x > 0$$

Using integration:

$$\begin{aligned} \text{(i)} \quad P(X > 210) &= \int_{210}^{\infty} 0.005e^{-0.005x} dx \\ &= \left[-e^{-0.005x} \right]_{210}^{\infty} \\ &= 0 - (-e^{-1.05}) \\ &= 0.34994 \end{aligned}$$

$$\begin{aligned} \text{(ii)} \quad P(X < 100) &= \int_0^{100} 0.005e^{-0.005x} dx \\ &= \left[-e^{-0.005x} \right]_0^{100} \\ &= -e^{-0.5} - (-e^0) \\ &= 1 - e^{-0.5} = 0.39347 \end{aligned}$$

$$\begin{aligned} \text{(iii)} \quad P(300 < X < 500) &= \int_{300}^{500} 0.005e^{-0.005x} dx \\ &= \left[-e^{-0.005x} \right]_{300}^{500} \\ &= -e^{-2.5} - (-e^{-1.5}) \\ &= e^{-1.5} - e^{-2.5} = 0.14105 \end{aligned}$$

Solution 1.13

The cumulative distribution function for $X \sim \text{Exp}(0.5)$ is:

$$F(x) = P(X < x) = 1 - e^{-0.5x}$$

(i) $P(X < 3) = F(3)$

$$= 1 - e^{-1.5}$$

$$= 0.77687$$

(ii) $P(X > 0.5) = 1 - P(X < 0.5)$

$$= 1 - F(0.5)$$

$$= 1 - (1 - e^{-0.25})$$

$$= e^{-0.25}$$

$$= 0.77880$$

(iii) $P(2 < X < 8) = P(X < 8) - P(X < 2)$

$$= F(8) - F(2)$$

$$= (1 - e^{-4}) - (1 - e^{-1})$$

$$= e^{-1} - e^{-4}$$

$$= 0.34956$$

Solution 1.14

We have an exponential distribution with mean 500 hours:

$$\mu = 500 = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{500} = 0.002$$

The PDF for this exponential distribution is:

$$f(x) = 0.002e^{-0.002x} \quad x > 0$$

We can now find the required probabilities:

$$\begin{aligned} \text{(i)} \quad P(X > 700) &= \int_{700}^{\infty} 0.002e^{-0.002x} dx \\ &= \left[-e^{-0.002x} \right]_{700}^{\infty} \\ &= 0 - (-e^{-1.4}) \\ &= 0.24660 \end{aligned}$$

or using the cumulative distribution function:

$$P(X > 700) = 1 - P(X < 700) = 1 - F(700) = 1 - (1 - e^{-1.4}) = e^{-1.4} = 0.24660$$

$$\begin{aligned} \text{(ii)} \quad P(400 < X < 600) &= \int_{400}^{600} 0.002e^{-0.002x} dx \\ &= \left[-e^{-0.002x} \right]_{400}^{600} \\ &= -e^{-1.2} - (-e^{-0.8}) \\ &= e^{-0.8} - e^{-1.2} = 0.14813 \end{aligned}$$

or using the cumulative distribution function:

$$\begin{aligned} P(400 < X < 600) &= P(X < 600) - P(X < 400) = F(600) - F(400) \\ &= (1 - e^{-1.2}) - (1 - e^{-0.8}) = e^{-0.8} - e^{-1.2} = 0.14813 \end{aligned}$$

Solution 1.15

(i) We are given:

$$P(X > 40) = \int_{40}^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_{40}^{\infty} = 0 - (-e^{-40\lambda}) = e^{-40\lambda}$$

or using the cumulative distribution function:

$$P(X > 40) = 1 - F(40) = 1 - (1 - e^{-40\lambda}) = e^{-40\lambda}$$

But since $P(X > 40) = 0.7$, we have:

$$e^{-40\lambda} = 0.7 \Rightarrow \lambda = -\frac{1}{40} \ln 0.7 = 0.00892$$

(ii) Using the conditional probability formula, we get:

$$\begin{aligned} P(X > 105 | X > 80) &= \frac{P(X > 105 \text{ and } X > 80)}{P(X > 80)} \\ &= \frac{P(X > 105)}{P(X > 80)} \end{aligned}$$

So we need to calculate $P(X > 80)$ and $P(X > 105)$:

$$P(X > 80) = \int_{80}^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_{80}^{\infty} = 0 - (-e^{-0.00892 \times 80}) = 0.49$$

$$P(X > 105) = \int_{105}^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_{105}^{\infty} = 0 - (-e^{-0.00892 \times 105}) = 0.392$$

Hence:

$$P(X > 105 | X > 80) = \frac{0.392}{0.49} = 0.8$$

Solution 1.16

- (i) The number of calls per minute arriving at the switchboard has the distribution:

$$X \sim \text{Poi}(5)$$

So the waiting time between calls, in minutes, has the distribution:

$$T \sim \text{Exp}(5)$$

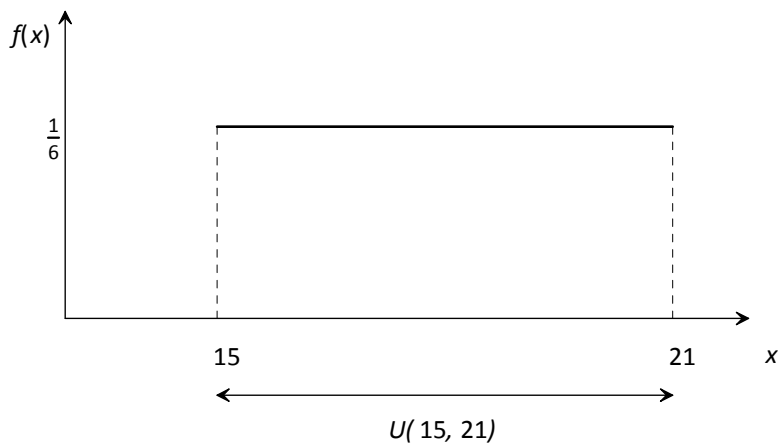
(ii)
$$P(T > 1) = \int_1^{\infty} 5e^{-5t} dt = \left[-e^{-5t} \right]_1^{\infty} = 0 - (-e^{-5}) = 0.00674$$

Solutions to extra practice questions

P1.1 (i) The PDF for $X \sim U(15, 21)$ is:

$$f(x) = \frac{1}{6} \quad 15 \leq x \leq 21$$

The graph is:



(ii) (a)
$$P(X > 18) = \int_{18}^{21} \frac{1}{6} dx = \left[\frac{1}{6}x \right]_{18}^{21} = \frac{1}{6}(21 - 18) = \frac{1}{2}$$

Alternatively, using the cumulative distribution function gives:

$$P(X > 18) = 1 - P(X < 18) = 1 - F(18) = 1 - \frac{18-15}{21-15} = \frac{1}{2}$$

(b)
$$P(19.5 < X < 20.3) = \int_{19.5}^{20.3} \frac{1}{6} dx = \left[\frac{1}{6}x \right]_{19.5}^{20.3} = \frac{1}{6}(20.3 - 19.5) = 0.1\dot{3}$$

Alternatively, using the cumulative distribution function gives:

$$\begin{aligned} P(19.5 < X < 20.3) &= P(X < 20.3) - P(X < 19.5) \\ &= F(20.3) - F(19.5) \\ &= \frac{20.3-15}{21-15} - \frac{19.5-15}{21-15} = 0.1\dot{3} \end{aligned}$$

(iii) Using our formulae:

$$E(X) = \frac{1}{2}(15 + 21) = 18$$

$$\text{var}(X) = \frac{1}{12}(21 - 15)^2 = 3$$

Alternatively, from first principles:

$$E(X) = \int_{15}^{21} x \frac{1}{6} dx = \left[\frac{1}{12} x^2 \right]_{15}^{21} = \frac{1}{12}(21^2 - 15^2) = 18$$

$$E(X^2) = \int_{15}^{21} x^2 \frac{1}{6} dx = \left[\frac{1}{18} x^3 \right]_{15}^{21} = \frac{1}{18}(21^3 - 15^3) = 327$$

$$\Rightarrow \text{var}(X) = E(X^2) - E^2(X) = 327 - 18^2 = 3$$

P1.2 The PDF of $X \sim U(2, 4)$ is:

$$f(x) = \frac{1}{4-2} = \frac{1}{2} \quad 2 < x < 4$$

(i) $E(X) = 3$ by symmetry.

$$(ii) \quad E(X^2) = \int_2^4 x^2 \frac{1}{2} dx = \left[\frac{1}{6} x^3 \right]_2^4 = \frac{1}{6}(4^3 - 2^3) = 9\frac{1}{3}$$

$$\Rightarrow \text{var}(X) = E(X^2) - E^2(X) = 9\frac{1}{3} - 3^2 = \frac{1}{3}$$

P1.3 (i) For $Y \sim U(a, b)$ we have:

$$E(X) = \frac{1}{2}(a + b) = 7$$

$$\text{var}(X) = \frac{1}{12}(b - a)^2 = 5^2$$

Rearranging the mean gives $b = 14 - a$. Substituting this into the variance:

$$\frac{1}{12}(14 - a - a)^2 = 25 \Rightarrow (14 - 2a)^2 = 300$$

Solving this:

$$14 - 2a = \pm\sqrt{300}$$

$$a = \frac{14 \pm \sqrt{300}}{2}$$

$$= -1.66 \text{ or } 15.66$$

$$\Rightarrow b = 14 - (-1.66) = 15.66 \text{ or } b = 14 - 15.66 = -1.66$$

Since $a < b$, we have $a = -1.66$ and $b = 15.66$. Hence:

$$f(y) = \frac{1}{17.32} \quad -1.66 < y < 15.66$$

(ii) Using the conditional probability formula, we get:

$$P(Y > 5 | Y > 0) = \frac{P(Y > 5 \text{ and } Y > 0)}{P(Y > 0)} = \frac{P(Y > 5)}{P(Y > 0)}$$

$$P(Y > 5) = \int_5^{15.66} \frac{1}{17.32} dx = \left[\frac{1}{17.32} x \right]_5^{15.66} = 0.61547$$

$$P(Y > 0) = \int_0^{15.66} \frac{1}{17.32} dx = \left[\frac{1}{17.32} x \right]_0^{15.66} = 0.90415$$

Hence:

$$P(Y > 5 | Y > 0) = \frac{0.61547}{0.90415} = 0.681$$

P1.4 Be careful. The **mean** is £1,000, so:

$$\mu = \frac{1}{\lambda} = 1,000 \Rightarrow \lambda = \frac{1}{1,000} = 0.001$$

This gives a PDF of:

$$f(x) = 0.001e^{-0.001x} \quad x > 0$$

(i) We find the probability by integrating the PDF:

$$\begin{aligned} P(X > 5,000) &= \int_{5,000}^{\infty} 0.001e^{-0.001x} dx = \left[-e^{-0.001x} \right]_{5,000}^{\infty} \\ &= 0 - (-e^{-5}) = 0.00674 \end{aligned}$$

or by using the cumulative distribution function:

$$P(X > 5,000) = 1 - P(X < 5,000) = 1 - F(5,000) = e^{-5} = 0.00674$$

(ii) Using the conditional probability formula, we get:

$$\begin{aligned} P(X > 5,000 | X > 1,000) &= \frac{P(X > 5,000 \text{ and } X > 1,000)}{P(X > 1,000)} \\ &= \frac{P(X > 5,000)}{P(X > 1,000)} \end{aligned}$$

So we need to calculate $P(X > 1,000)$ by integrating the PDF:

$$\begin{aligned} P(X > 1,000) &= \int_{1,000}^{\infty} 0.001e^{-0.001x} dx = \left[-e^{-0.001x} \right]_{1,000}^{\infty} \\ &= 0 - (-e^{-1}) = 0.36788 \end{aligned}$$

or by using the cumulative distribution function:

$$P(X > 1,000) = 1 - P(X < 1,000) = 1 - F(1,000) = e^{-1} = 0.36788$$

Substituting this into our conditional probability formula gives:

$$P(X > 5,000 | X > 1,000) = \frac{0.00674}{0.36788} = 0.0183$$

P1.5 Be careful. The **mean** is £50, so:

$$\mu = \frac{1}{\lambda} = 50 \Rightarrow \lambda = \frac{1}{50}$$

This gives a PDF of:

$$f(x) = \frac{1}{50} e^{-\frac{1}{50}x} \quad x > 0$$

(i) We find the probability by integrating the PDF:

$$P(X > 50) = \int_{50}^{\infty} \frac{1}{50} e^{-\frac{1}{50}x} dx = \left[-e^{-\frac{1}{50}x} \right]_{50}^{\infty} = 0 - (-e^{-1}) = 0.36788$$

or by using the cumulative distribution function:

$$P(X > 50) = 1 - P(X < 50) = 1 - F(50) = e^{-1} = 0.36788$$

(ii) We find the probability by integrating the PDF:

$$P(X > 200) = \int_{200}^{\infty} \frac{1}{50} e^{-\frac{1}{50}x} dx = \left[-e^{-\frac{1}{50}x} \right]_{200}^{\infty} = 0 - (-e^{-4}) = 0.01832$$

or by using the cumulative distribution function:

$$P(X > 200) = 1 - P(X < 200) = 1 - F(200) = e^{-4} = 0.01832$$

(iii) Using the conditional probability formula:

$$\begin{aligned} P(X > 200 | X > 50) &= \frac{P(X > 200 \text{ and } X > 50)}{P(X > 50)} = \frac{P(X > 200)}{P(X > 50)} \\ &= \frac{0.01832}{0.36788} = 0.0498 \end{aligned}$$

P1.6 Using $P(X < m) = 0.5$, we get:

$$P(X < m) = \int_0^m \lambda e^{-\lambda x} dx = 0.5$$

$$\Rightarrow \left[-e^{-\lambda x} \right]_0^m = \frac{1}{2}$$

$$1 - e^{-\lambda m} = \frac{1}{2}$$

$$e^{-\lambda m} = \frac{1}{2}$$

$$-\lambda m = \ln \frac{1}{2}$$

$$\lambda m = \ln 2$$

$$m = \frac{1}{\lambda} \ln 2$$

Since the mean is $\mu = \frac{1}{\lambda}$, we can see the median is $\ln 2 = 0.693$ times the mean.

P1.7 2 accidents per 5 days is 0.4 accidents per day. So, if X is the number of claims per day we have:

$$X \sim \text{Poi}(0.4)$$

This means the waiting time, T , between these claims is:

$$T \sim \text{Exp}(0.4)$$

The probability that the waiting time is greater than 3 days can be found by integration:

$$P(T > 3) = \int_3^{\infty} 0.4 e^{-0.4t} dt = \left[-e^{-0.4t} \right]_3^{\infty} = 0 - (-e^{-1.2}) = 0.30119$$

Alternatively, we could use the cumulative distribution function:

$$P(T > 3) = 1 - P(T < 3) = 1 - F(3) = e^{-1.2} = 0.30119$$

The normal distribution

Introduction

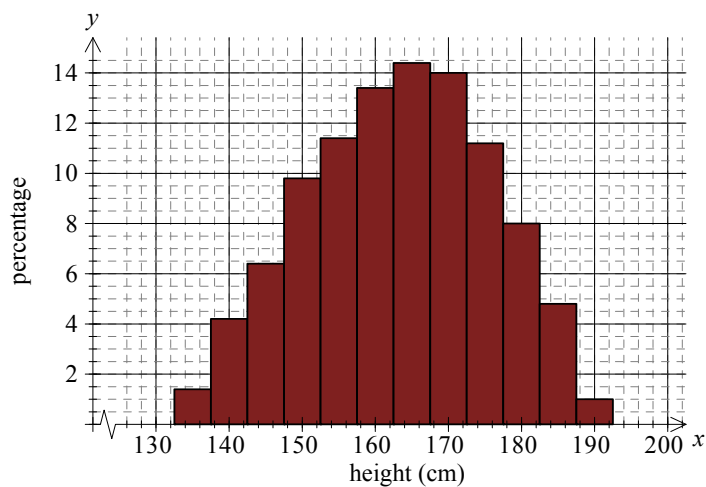
Here we look at the normal distribution. This is a continuous distribution that occurs naturally, for example the weights of babies at birth, the time taken to get to work every day, or the height of adult males.

These variables all follow the same pattern. There will be some low values and some high values, but the majority of values will lie somewhere in the 'middle'.

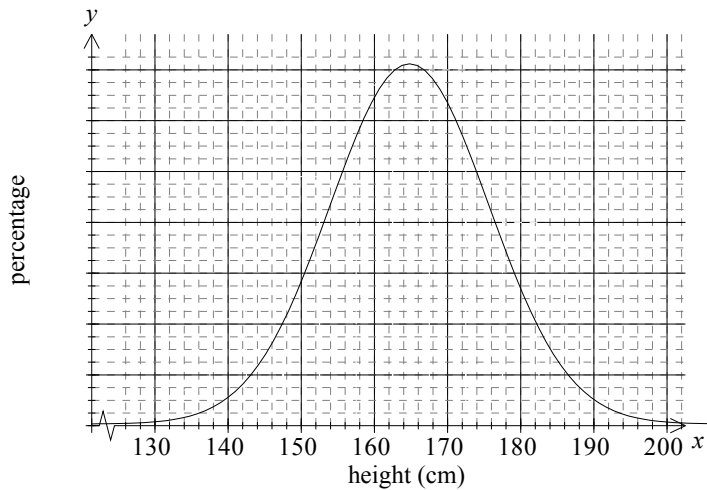
Features of the normal distribution

General features

If we were to plot a histogram of the heights of all female student actuaries, we would probably end up with a histogram like this:



As we sample a larger and larger group (and use smaller classes) we will approach the following graph:



The distribution is symmetrical, with most heights around the average of 165cm and fewer and fewer people as we approach the extremes. This symmetrical bell-shaped distribution is called the **normal distribution**. It occurs naturally in many other areas, for example: weights, IQ's, exam scores, and so on.

The PDF of the normal distribution

The normal distribution depends only on two parameters, μ and σ^2 , which you will recognise as the mean and variance. So μ and σ^2 are the only parameters that appear in the probability density function (PDF):

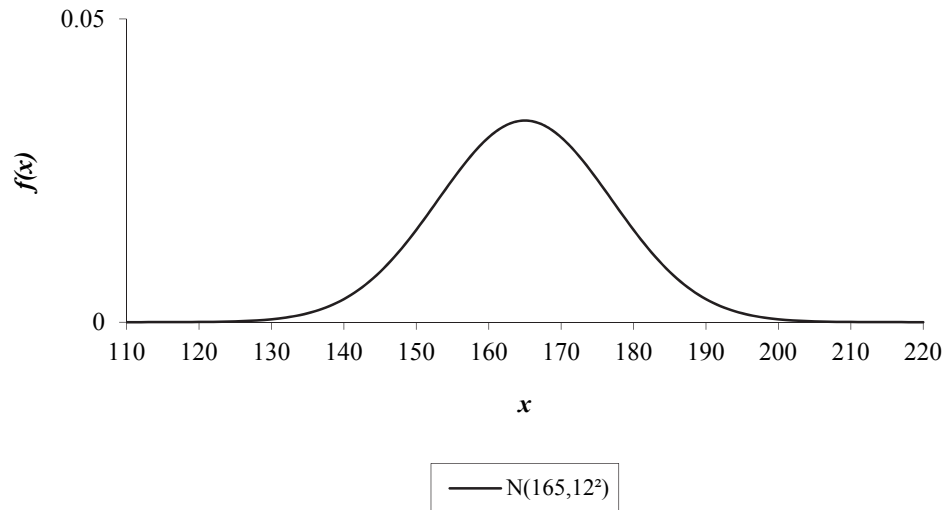
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

The shortcut way of writing X has a normal distribution with mean μ and variance σ^2 is:

$$X \sim N(\mu, \sigma^2)$$

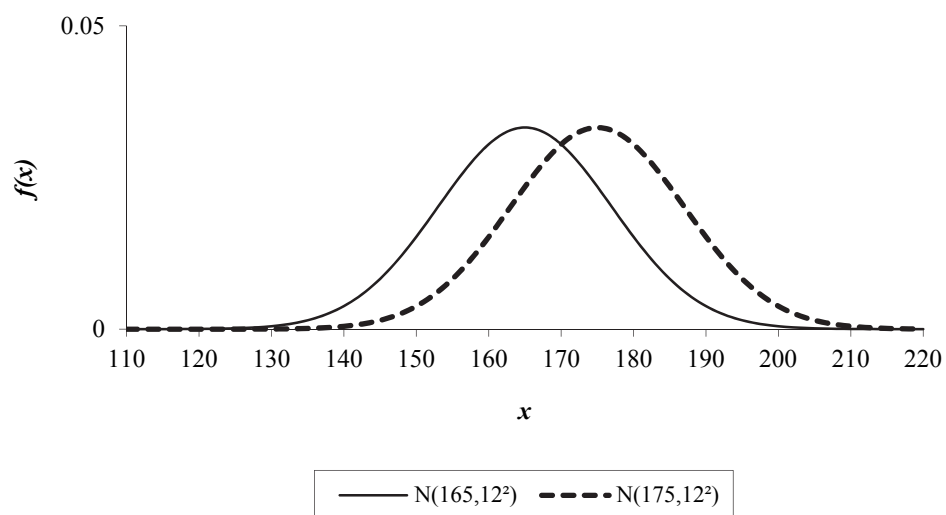
Now the PDF is grotty, but it does give the lovely symmetrical bell-shaped curve that we saw earlier. To get a better 'feel' for the shape of the PDF, we'll look at a number of normal distributions with different values of μ and σ^2 .

Let's start with the distribution of the heights of our female actuaries, which had a mean of 165cm and a standard deviation of 12cm (so variance of 12^2 cm^2), ie $X \sim N(165, 12^2)$:



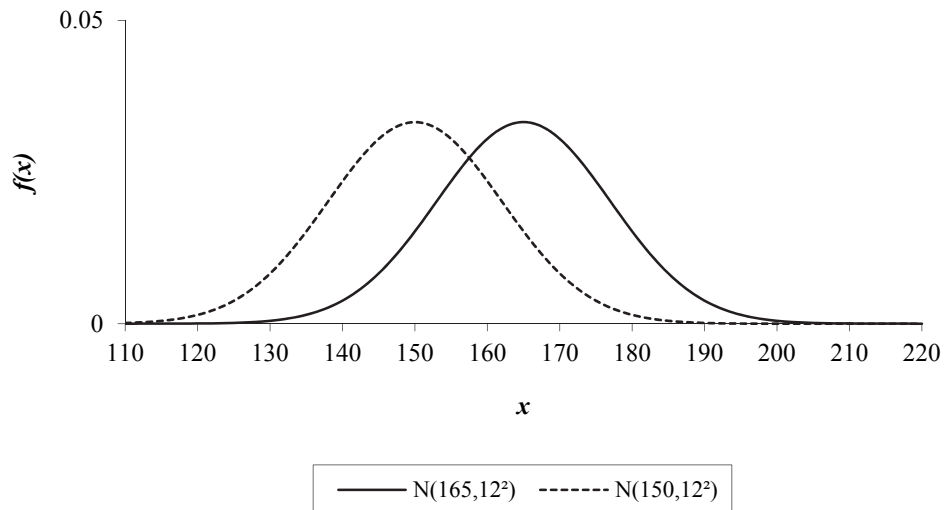
We can see that the PDF is greatest around the mean of 165cm and decreases as it moves further and further away from the mean.

So what happens if we increase the mean from 165cm to 175cm? Well, we would expect the female actuaries to be taller on average:



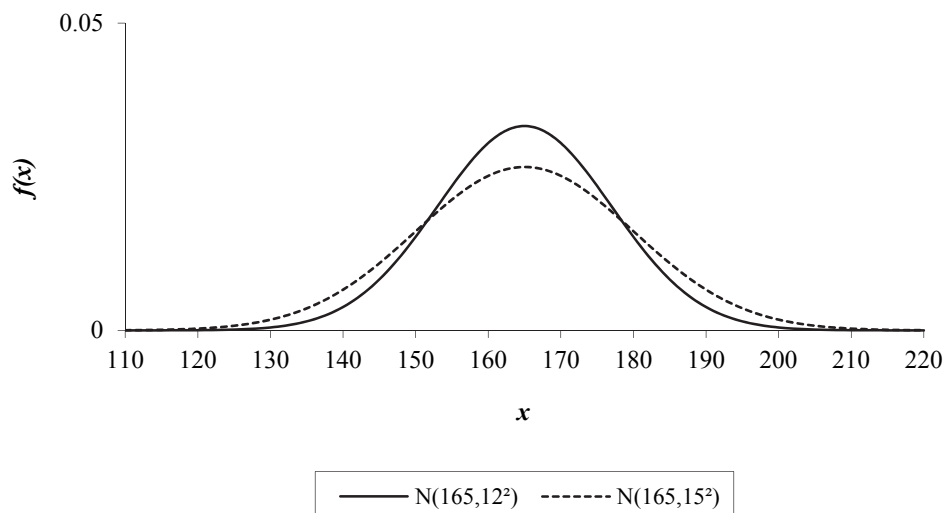
As you can see, the graph is simply shifted to the right, as all female actuaries are now 10 cm taller on average. Since we didn't alter the spread (the variance) the shape is still the same.

Similarly, if we were to reduce the mean from 165 cm to 150 cm we would see that the graph shifts to the left. All the female actuaries are 15 cm shorter on average. Again, the shape is the same, as we didn't change the variance.



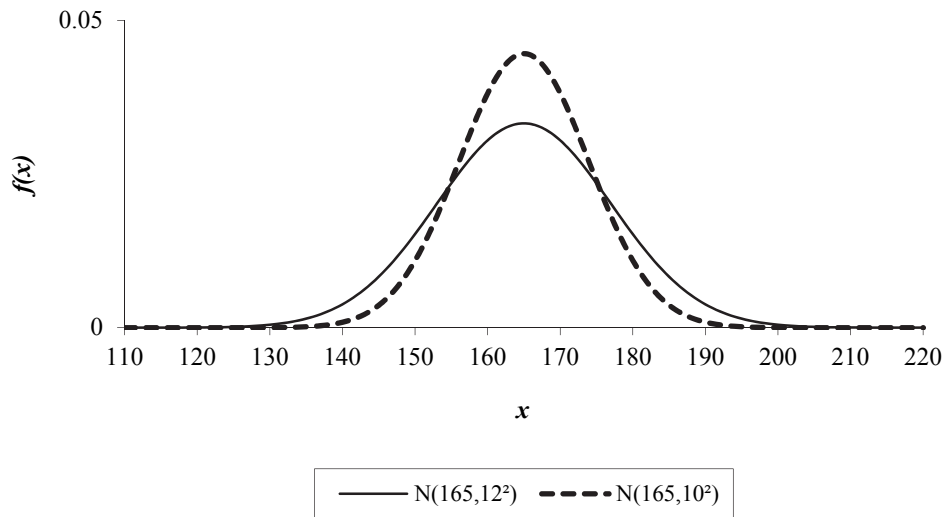
In summary, changing the mean, μ , simply moves the position of the normal PDF along the x -axis – it doesn't change the shape of the PDF.

So what happens if we increase the variance from 12^2 cm^2 to 15^2 cm^2 ? Well, we would expect the heights of the female actuaries to be more spread out:



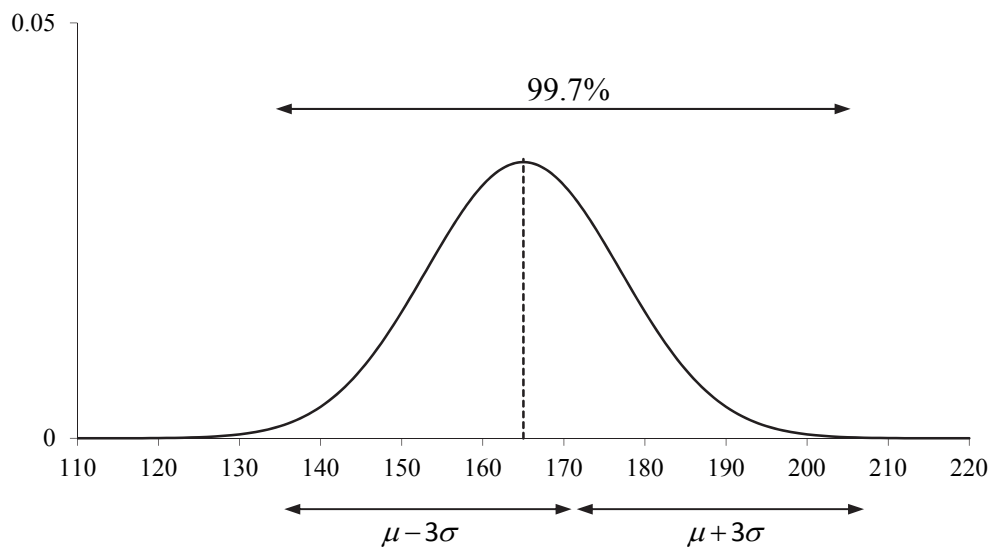
Here we can see that the heights of the actuaries are spread over more values than before. Consequently, there is a smaller percentage of people in the middle due to the greater variety in the heights.

If we now reduce the variance from 12^2 cm^2 to 10^2 cm^2 , we should see a smaller range of heights. Consequently, there will be a larger percentage of people in the middle as more people have a similar height:



In summary, changing the variance, σ^2 , alters the shape of the PDF of the normal distribution. A smaller variance ‘squashes’ it, whereas a larger variance ‘stretches’ it.

One last thing to note is that nearly all (in fact 99.7%) of the results are within 3 standard deviations of the mean. So for our female actuaries, with the $N(165, 12^2)$ distribution, nearly all the heights are between $(165 - 3 \times 12, 165 + 3 \times 12) = (129, 201)$:



Question 1.1

For each part sketch the two PDF graphs on the same diagram:

(i) $X \sim N(50, 5^2)$ and $Y \sim N(50, 10^2)$

(ii) $X \sim N(50, 5^2)$ and $W \sim N(60, 5^2)$.

Recall that one of the properties of a PDF is that the area under its graph equals to 1:

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

We will not prove this result here as the proof is too complex.

Moments of the normal distribution

The moments are easily defined for $X \sim N(\mu, \sigma^2)$ as the mean is just μ and the variance is just σ^2 .

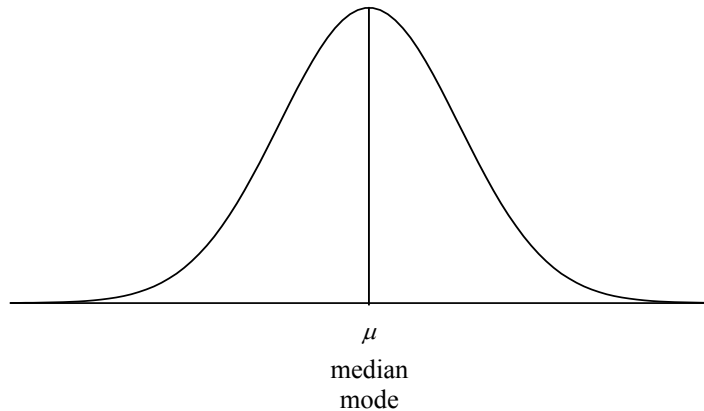
$$E(X) = \mu$$

$$\text{var}(X) = \sigma^2$$

Question 1.2

Determine $E(X^2)$ for $X \sim N(\mu, \sigma^2)$.

From the graph of the PDF of a normal distribution, it is clear that the mode is the same as the mean. Also, by symmetry the median is the same as the mean.



Probabilities of a normal distribution

To calculate probabilities for a continuous random variable we just integrate the PDF:

$$P(a < X < b) = \int_a^b f(x) dx$$

However, as mentioned earlier, integrating the PDF of a normal distribution is a nightmare, so we need another way. We could use statistical tables, but with so many combinations of μ and σ^2 which one do we tabulate? It turns out that we only need $N(0,1)$.

The standard normal distribution

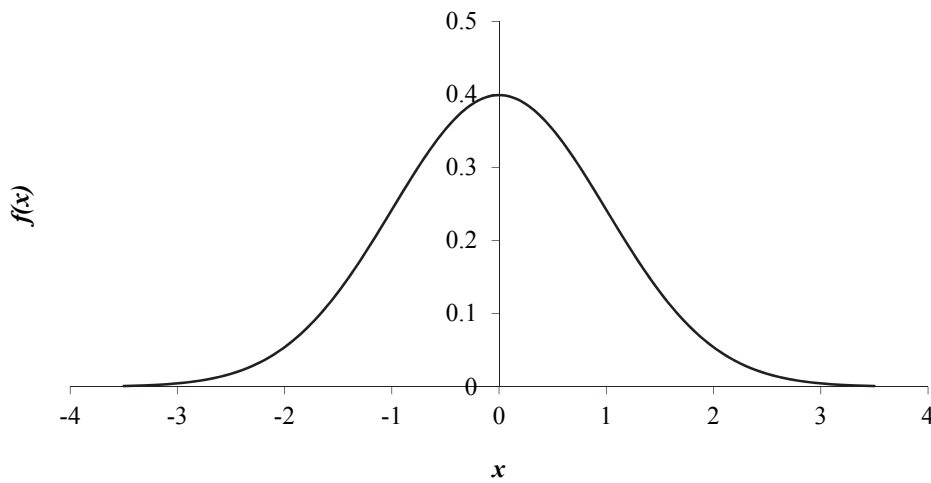
The **standard normal distribution** is a special case of the normal distribution with a mean of 0 and a variance of 1. We use the letter Z to stand for a random variable with a standard normal distribution:

$$Z \sim N(0,1)$$

The PDF of the standard normal distribution is obtained by simply substituting $\mu = 0$ and $\sigma^2 = 1$ into the normal PDF. This gives:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad -\infty < z < \infty$$

The standard normal PDF is often denoted $\phi(z)$. It has the following shape:



We can see that the PDF is symmetrical about zero and that nearly all of the values lie within 3 standard deviations of the mean, *ie* between $(0-3\times 1, 0+3\times 1) = (-3, 3)$.

The moments are:

$$E(Z) = 0$$

$$\text{var}(Z) = 1$$

The median is the same as the mean (by symmetry), as is the mode.

Probabilities of the standard normal distribution

The standard normal distribution is the only normal distribution that is tabulated. In this section, we will look at how to find probabilities using this table.

Simple probabilities

The cumulative distribution function of the standard normal is included in Appendix A. In the ACET exam you will be given the relevant values.

Recall that the cumulative density function, $F(x)$, is defined to be:

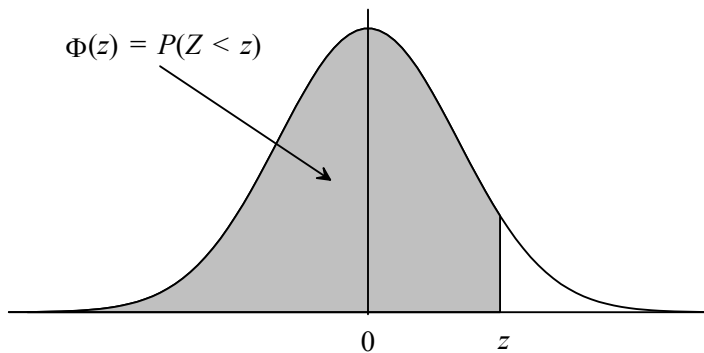
$$F(x) = P(X < x)$$

ie the probability that the random variable is less than x .

The standard normal cumulative density function is often denoted $\Phi(z)$:

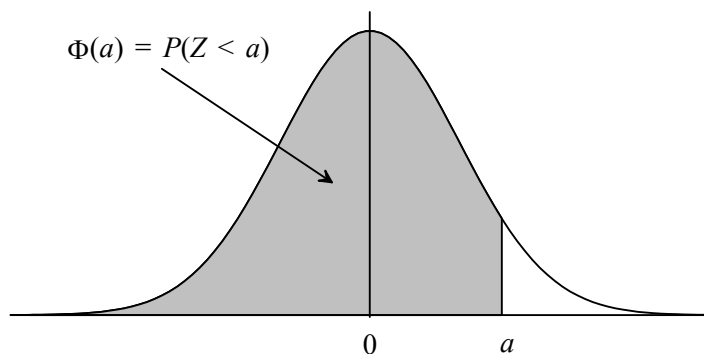
$$\Phi(z) = P(Z < z)$$

Notice how the table only gives 'less than' probabilities for positive values of z . Since the area under the curve represents the probability, this is shown as:



Calculating $P(Z < a)$

Firstly, we'll find the probability that Z (the standard normal random variable) is less than a positive number, say, a :



We simply read the values from the table. For example:

$$P(Z < 1.39) = 0.91774$$

1.39 is the value of x given (in bold) in the columns of the table in Appendix A. The probability is given by $\Phi(x)$ ie the next column (which is not bold).

Question 1.3

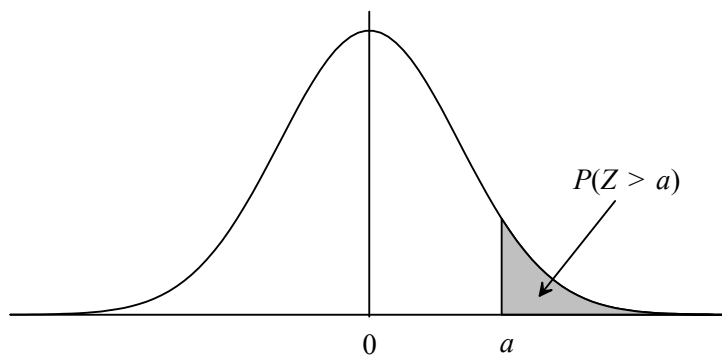
Find these probabilities:

(i) $P(Z < 1.23)$

(ii) $P(Z < 2.725)$.

Calculating $P(Z > a)$

We also need to find the probability that Z (the standard normal random variable) is more than a positive number, say, a :



The whole area under the graph (ie the total probability) is 1. This gives the following relationship:

$$P(Z < a) + P(Z > a) = 1$$

Therefore:

$$P(Z > a) = 1 - P(Z < a)$$

So all we do is look up the 'less than' probability given in the standard normal tables and then subtract it from 1. For example:

$$P(Z > 0.26) = 1 - P(Z < 0.26) = 1 - 0.60257 = 0.39743$$

Question 1.4

Find these probabilities:

(i) $P(Z > 2.17)$

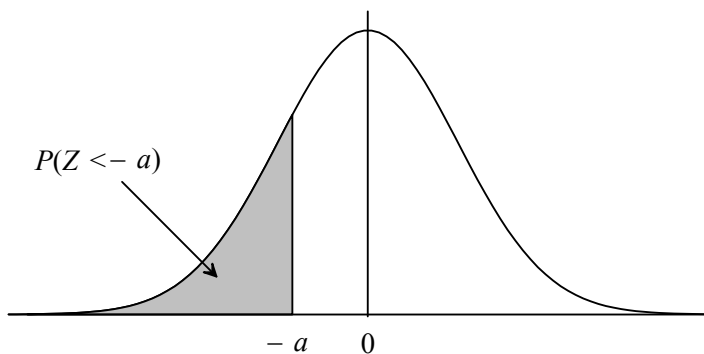
(ii) $P(Z > 0.08)$.

Using the notation $\Phi(z) = P(Z < z)$, we could write 'more than' probabilities as:

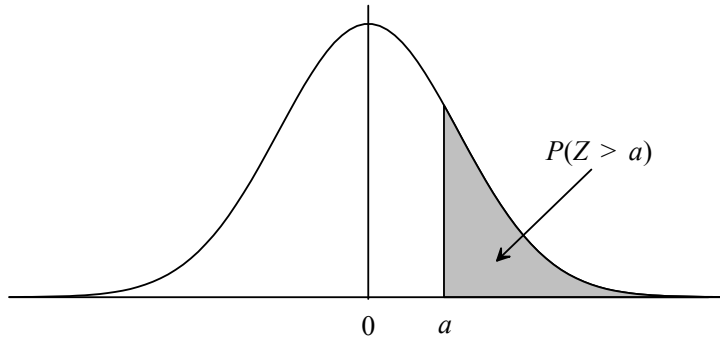
$$P(Z > a) = 1 - \Phi(a)$$

Calculating $P(Z < -a)$

Now we'll find the probability that Z (the standard normal random variable) is less than a negative number, say, $-a$:



The problem is that negative values are not tabulated. So how do we do it? We use the fact that the normal distribution is symmetrical. By symmetry, the area shaded in the diagram above is *exactly* the same as the area shaded in the diagram below:



This area corresponds to the probability that Z (the standard normal random variable) is greater than the positive number a :

$$P(Z < -a) = P(Z > a)$$

We just worked out how to find 'more than' probabilities on the previous page. So:

$$P(Z < -a) = P(Z > a) = 1 - P(Z < a)$$

For example, to find:

$$P(Z < -3.08)$$

We first use symmetry:

$$P(Z < -3.08) = P(Z > 3.08)$$

We now use the method for calculating 'more than' probabilities:

$$P(Z > 3.08) = 1 - P(Z < 3.08) = 1 - 0.99896 = 0.00104$$

Perhaps the easiest way to remember the symmetry result (other than drawing a diagram) is to note that we swap the sign of the inequality (from 'less than' to 'more than') and we swap the sign of the number (from negative to positive). So in short we '*swap the sign and swap the sign*'.

Question 1.5

Find these probabilities:

(i) $P(Z < -1.50)$

(ii) $P(Z < -0.21)$

(iii) $P(Z < -2.05)$.

Using the notation $\Phi(z) = P(Z < z)$, we could write these probabilities as:

$$P(Z < -a) = \Phi(-a)$$

We can then write the relationship as:

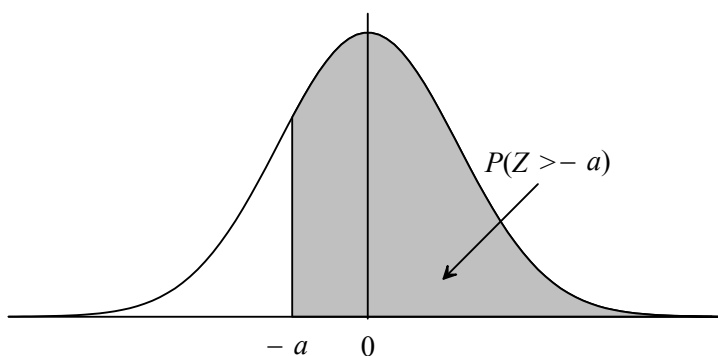
$$\Phi(-a) = 1 - \Phi(a)$$

ie:

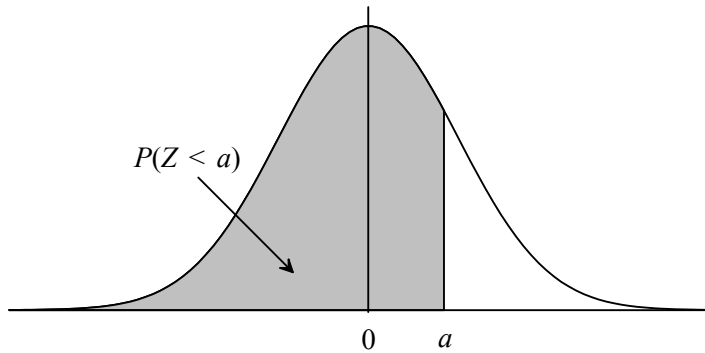
$$P(Z < -a) = 1 - \Phi(a)$$

Calculating $P(Z > -a)$

Finally we'll find the probability that Z (the standard normal random variable) is greater than a negative number, say, $-a$:



Again, since negative values are not tabulated we use the fact that the normal distribution is symmetrical. By symmetry, the area shaded in the diagram above is *exactly* the same as the area shaded in the diagram below:



This area corresponds to the probability that Z (the standard normal random variable) is less than the positive number a :

$$P(Z > -a) = P(Z < a)$$

We can just read this probability from the standard normal table in Appendix A.

For example, to find:

$$P(Z > -2.62)$$

First we use symmetry:

$$P(Z > -2.62) = P(Z < 2.62)$$

We now simply read this probability from the table in Appendix A:

$$P(Z < 2.62) = 0.99560$$

Once again we can remember the symmetry result by noting that we swap the sign of the inequality (from 'more than' to 'less than') and we swap the sign of the number (from negative to positive). So in short we *'swap the sign and swap the sign'*.

Question 1.6

Find these probabilities:

(i) $P(Z > -3.94)$

(ii) $P(Z > -0.73)$.

Using the notation $\Phi(z) = P(Z < z)$, we could write these probabilities as:

$$P(Z > -a) = \Phi(a)$$

Summary

We have now met all four of the probabilities that you could be asked:

Important result

$P(Z < a)$ read off the standard normal table.

$P(Z > a)$ $1 - P(Z < a)$ since area (and total probability) equals 1.

$P(Z < -a)$ $P(Z > a)$ using symmetry (*swap the sign and swap the sign*).

$= 1 - P(Z < a)$ since area (and total probability) equals 1.

$P(Z > -a)$ $P(Z < a)$ using symmetry (*swap the sign and swap the sign*).

Since we will be using the normal distribution more than any other, it is vital that you can calculate these probabilities quickly. Take the time to learn the methods above. Try to do the following question *without* referring to the summary above.

Question 1.7

Find these probabilities:

(i) $P(Z > 3.6)$

(ii) $P(Z > -0.76)$

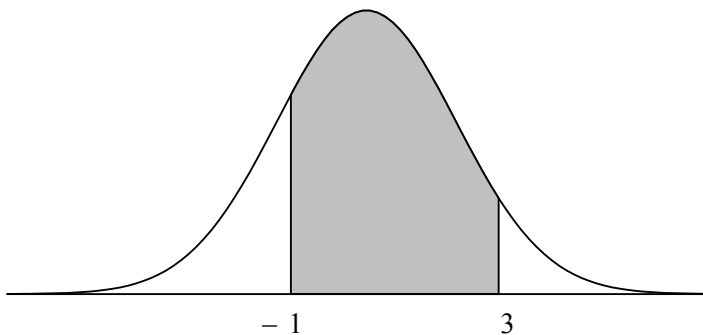
(iii) $P(Z < 1.98)$

(iv) $P(Z < -2.41)$.

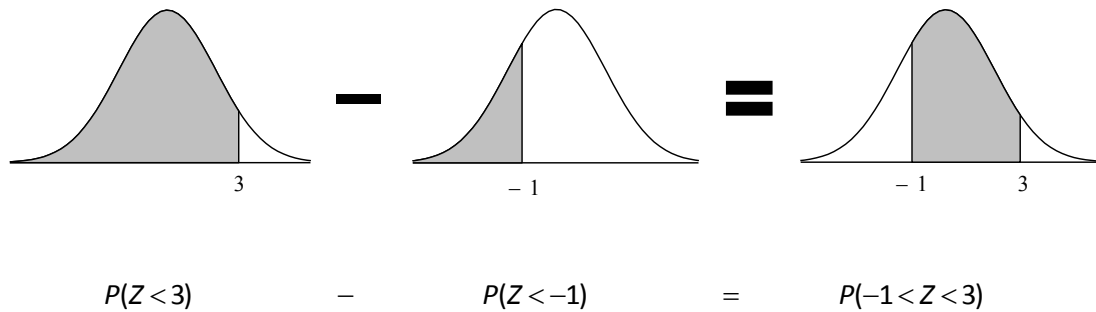
Compound probabilities

We can now calculate 'single' probabilities, like $P(Z < 3)$ or $P(Z > -1)$, so how can we use this to calculate 'compound' probabilities like $P(-1 < Z < 3)$?

The easiest way to see how to calculate these probabilities is to look at a diagram:



The probability $P(-1 < Z < 3)$ is the shaded area between -1 and 3 . How can we get this from 'single' probabilities? Imagine that the normal distribution is a piece of paper that we are cutting out and we want to just leave the shaded part. It can be made by starting with the $P(Z < 3)$ part and subtracting the $P(Z < -1)$:



So, in general we have:

Important result

$$P(a < Z < b) = P(Z < b) - P(Z < a) \quad \text{for } a < b$$

So to calculate a 'compound' probability, we first split it up into two 'single' probabilities. We then calculate each of those 'single' probabilities as before.

Completing our example, we get:

$$P(-1 < Z < 3) = P(Z < 3) - P(Z < -1)$$

We can read $P(Z < 3)$ directly from the standard normal table:

$$P(Z < 3) = 0.99865$$

To get $P(Z < -1)$, we need to use symmetry to get a positive value and then with a little rearranging we get $P(Z < 1)$ which we can read from the standard normal table:

$$\begin{aligned}
 P(Z < -1) &= P(Z > 1) \\
 &= 1 - P(Z < 1) \\
 &= 1 - 0.84134 = 0.15866
 \end{aligned}$$

Now we substitute these two answers back into our original equation:

$$\begin{aligned}P(-1 < Z < 3) &= P(Z < 3) - P(Z < -1) \\ &= 0.99865 - 0.15866 \\ &= 0.83999\end{aligned}$$

These questions *are* common, so it is vital that you can calculate these probabilities quickly.

Question 1.8

Find these probabilities:

- (i) $P(1.24 < Z < 2.19)$
- (ii) $P(-0.92 < Z < 0.83)$
- (iii) $P(-2.92 < Z < -1.67)$.

Common Error:

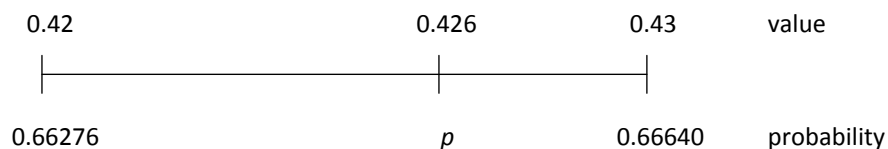
Many students calculate $P(a < Z < b) = P(Z < b) - P(Z > a)$ instead.

Probabilities involving interpolation

We will look at one last thing before we move on to how we can calculate probabilities for any normal distribution *other* than the standard normal. How do we calculate probabilities like $P(Z < 0.426)$?

The table in Appendix A only has $P(Z < 0.42)$ and $P(Z < 0.43)$ tabulated. So how do we get a probability for a value that is in between these? We can use **linear interpolation**.

Linear interpolation assumes that the probabilities increase *linearly* between values. This means that we can use proportions to find our 'in between' probability. Consider a number line:



The proportion of the 'length' that 0.426 is between 0.42 and 0.43:

$$\frac{0.426 - 0.42}{0.43 - 0.42} = 0.6$$

The proportion of the 'length' that the probability, p , is between 0.66278 and 0.66640:

$$\frac{p - 0.66276}{0.66640 - 0.66276}$$

For linear interpolation, we assume that the probabilities are spread out linearly between values. This means that the proportions will be equal. Hence:

$$\frac{p - 0.66276}{0.66640 - 0.66276} = 0.6$$

Rearranging this we get:

$$p = 0.66276 + 0.6 \times (0.66640 - 0.66276) = 0.66494$$


You may think 'is it worth it?' Well if we had just rounded 0.426 to 0.43, our answer would have been 66.64% instead of 66.49%. This is fairly serious.

So we need a quick way of getting to the last line. The first thing to notice is that the proportion 0.6 is just the third decimal place of 0.436. So had we been finding the probability for 1.284 we would have used a proportion of 0.4.

In general, we have:

$$p = (\text{start probability}) + 0.\dots \times (\text{difference between probabilities})$$

proportion:
just use 3rd decimal
place (onwards)



We'll use this shortcut method to calculate a slightly harder probability:

$$P(Z < -1.793)$$

First we need to change it into a probability we can read from the standard normal table. Using the 'swap the sign, swap the sign' symmetry rule and the 'more than' rule gives:

$$P(Z < -1.793) = P(Z > 1.793) = 1 - P(Z < 1.793)$$

Now we look up the probabilities each side of 1.793:

$$P(Z < 1.79) = 0.96327 \text{ and } P(Z < 1.80) = 0.96407$$

Now we use our linear interpolation rule:

$$P(Z < 1.793) = 0.96327 + 0.3 \times (0.96407 - 0.96327) = 0.96351$$

So we have:

$$P(Z < -1.793) = 1 - 0.96351 = 0.03649$$

Question 1.9

Calculate these probabilities using linear interpolation:

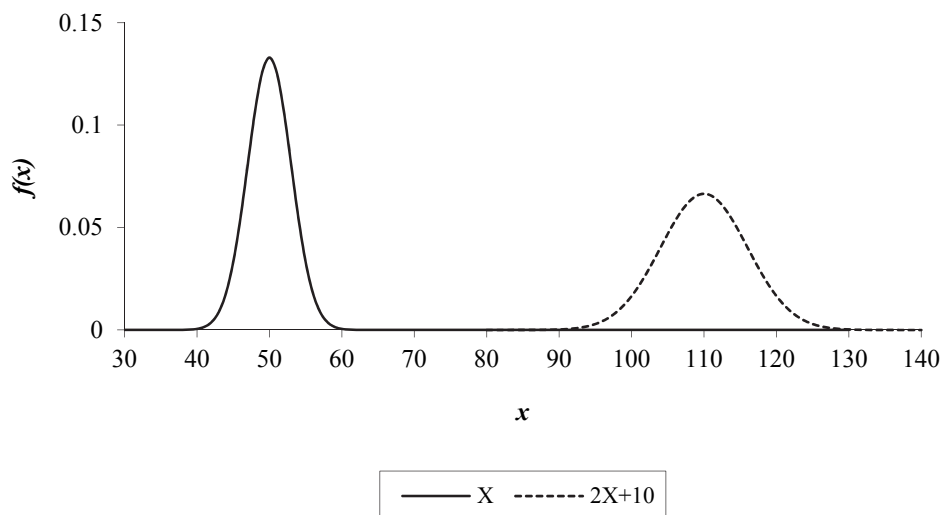
- (i) $P(Z < 1.048)$
- (ii) $P(Z > 0.271)$
- (iii) $P(Z > -2.389)$
- (iv) $P(-0.704 < Z < 0.897)$.

Probabilities for any normal distribution

So far we have only calculated probabilities for the standard normal distribution, $Z \sim N(0,1)$. Now we need to calculate probabilities for any other normal distribution.

Transforming normal distributions

We use the fact that a linear function of a normal random variable has a normal distribution. For example, if X has a normal distribution, say $X \sim N(50, 3^2)$, then $2X + 10$ also has a normal distribution:



All we need to do now is to identify the new mean and variance of our transformed function.

First let's look at how the mean has changed. Recall the following result for *any* distribution:

$$E(aX + b) = aE(X) + b$$

ie if we multiply the random variable by a and then add b , the mean of the random variable is also multiplied by a with b added on.

Now we had $X \sim N(50, 3^2)$ with $E(X) = 50$, so $2X + 10$ should have a mean of:

$$2E(X) + 10 = 2 \times 50 + 10 = 110$$

We can see on the diagram that $2X + 10$ does indeed have a mean of 110.

Next let's look at how the variance has changed. Recall the following result for *any* distribution:

$$\text{var}(aX + b) = a^2 \text{var}(X)$$

ie if we multiply the random variable by a and then add b , the variance of the random variable is multiplied by a^2 .

Now we had $X \sim N(50, 3^2)$ with $\text{var}(X) = 3^2$, so $2X + 10$ should have a variance of:

$$2^2 \text{var}(X) = 2^2 \times 3^2 = 6^2$$

We can see this on the diagram (remembering that the majority of the values of the normal distribution are spread over 3 standard deviations each side of the mean).

$2X + 10$ is spread over $(110 - 3 \times 6, 110 + 3 \times 6) = (92, 128)$.

So in general, we have:

Important result

If $X \sim N(\mu, \sigma^2)$ then $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

We will not formally prove this result here.

Question 1.10

If $X \sim N(100, 4^2)$, write down the distribution of:

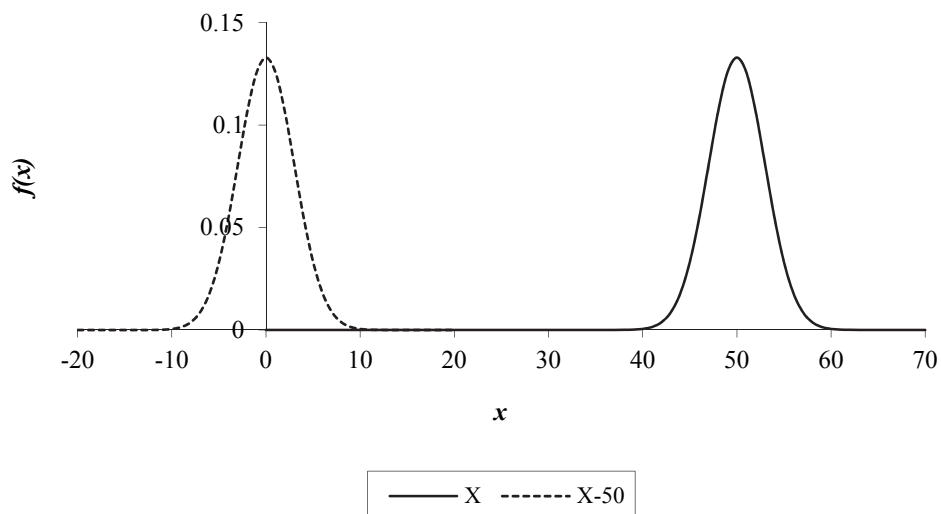
- (i) $3X + 5$
- (ii) $\frac{1}{2}X - 20$
- (iii) $\frac{1}{4}(X - 100)$.

Standardising

We now know how to change one type of normal distribution into another normal distribution. So how does this help us find probabilities for any normal distribution?

What we are going to do is transform our normal distribution into the standard normal distribution. We can then look up the probabilities in the standard normal table in Appendix A. So how do we do it?

Let's take our normal distribution from before, $X \sim N(50, 3^2)$. We want to change this into the standard normal $Z \sim N(0, 1)$. First we want to change the mean from 50 to 0. The easiest way to do this is to subtract 50:



We had $X \sim N(50, 3^2)$ with $E(X) = 50$, so $X - 50$ has a mean of:

$$E(X - 50) = E(X) - 50 = 50 - 50 = 0$$

Next we want to change the variance from 3^2 to 1. Since only multiplying or dividing affects the variance, we need to divide the random variable by 3:

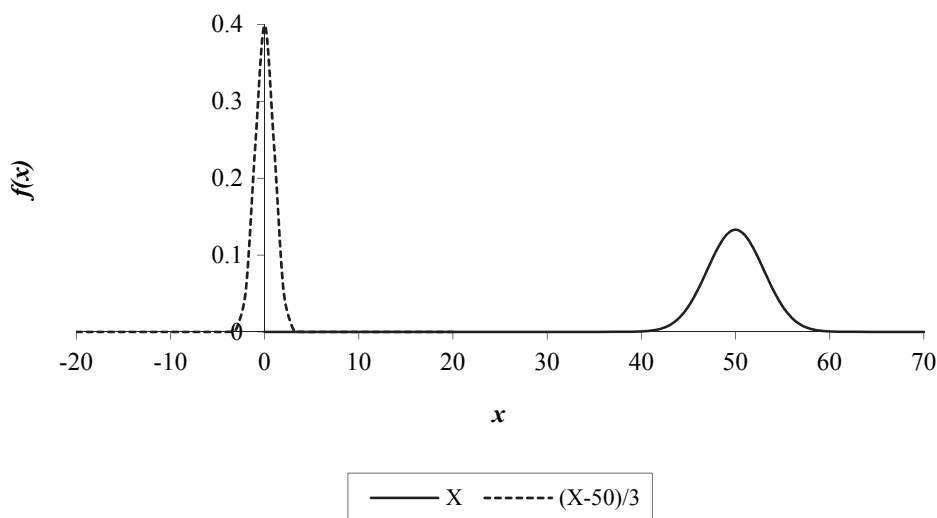
$$\text{var}\left(\frac{1}{3}X\right) = \frac{1}{3^2} \text{var}(X) = \frac{1}{3^2} 3^2 = 1$$

Now do we divide by the 3 before or after subtracting the 50? If we did it before we would have a mean of:

$$E\left(\frac{1}{3}X - 50\right) = \frac{1}{3}E(X) - 50 = \frac{1}{3}50 - 50 = -33\frac{1}{3}$$

Whereas if we did it after we would have a mean of:

$$E\left[\frac{1}{3}(X - 50)\right] = \frac{1}{3}E(X - 50) = \frac{1}{3}[E(X) - 50] = \frac{1}{3}(50 - 50) = 0$$



Although it's squashed because of the scale, we can see that we have now transformed our normal distribution into the standard normal distribution. This process is called **standardising** the normal distribution.

How do we do this in general? Well for $X \sim N(50, 3^2)$ we took away the mean, 50, first and *then* we divided it by 3 which was the standard deviation.

So for $X \sim N(\mu, \sigma^2)$ we would subtract the mean, μ , first and then divide by the standard deviation, σ , second:

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

This gives us the standard normal distribution, $Z \sim N(0,1)$.

Standardising

If $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X - \mu}{\sigma}$ then $Z \sim N(0,1)$.

Calculating probabilities for any normal distribution

We can now use this idea of subtracting the mean and dividing by the standard deviation to change the probability for *any* normal distribution into a probability for a standard normal distribution.

Taking our normal distribution of $X \sim N(50, 3^2)$, let's find $P(X > 56)$.

First we subtract the mean of 50 from *both* sides:

$$P(X > 56) = P(X - 50 > 6)$$

Next we divide *both* sides by the standard deviation of 3:

$$P(X > 56) = P\left(\frac{X - 50}{3} > 2\right)$$

Now we use the fact that when we subtract the mean and divide by the standard deviation we get a standard normal distribution:

$$P(X > 56) = P(Z > 2)$$

We can now look this probability up in our standard normal table:

$$P(Z > 2) = 1 - P(Z < 2) = 1 - 0.97725 = 0.02275$$

So we have:

$$P(X > 56) = 0.02275$$

This may seem a little longwinded at the moment (and more so when we need to use interpolation) but it does get quicker with a little practice. Usually we jump straight to:

$$P(X > 56) = P\left(Z > \frac{56 - 50}{3}\right) = P(Z > 2)$$

Because this is *the* method for calculating probabilities, we'll work through another example just to make sure it's totally clear:

If $X \sim N(15, 25)$, calculate:

$$(i) \quad P(X < 18) \qquad (ii) \quad P(X > 11.8) \qquad (iii) \quad P(9.8 < X < 18.2)$$

Working through each of these in turn:

$$(i) \quad P(X < 18) = P\left(Z < \frac{18 - 15}{\sqrt{25}}\right) = P(Z < 0.6) = 0.72575$$

This value is read directly from the standard normal table.

$$(ii) \quad P(X > 11.8) = P\left(Z > \frac{11.8 - 15}{\sqrt{25}}\right) = P(Z > -0.64)$$

Since we have a negative number we use '*swap the sign, swap the sign*', we can then read the resulting probability directly from the standard normal table:

$$P(X > 11.8) = P(Z > -0.64) = P(Z < 0.64) = 0.73891$$

(iii) We first need to split up this 'compound' probability:

$$P(9.8 < X < 18.2) = P(X < 18.2) - P(X < 9.8)$$

Now we standardise and calculate each part:

$$P(X < 18.2) = P\left(Z < \frac{18.2 - 15}{\sqrt{25}}\right) = P(Z < 0.64) = 0.73891$$

This value is read directly from the standard normal table.

$$P(X < 9.8) = P\left(Z < \frac{9.8 - 15}{\sqrt{25}}\right) = P(Z < -1.04)$$

Using the '*swap the sign, swap the sign*' rule and the fact that less than and more than probabilities sum to 1 we get:

$$\begin{aligned} P(Z < -1.04) &= P(Z > 1.04) \\ &= 1 - P(Z < 1.04) \\ &= 1 - 0.85083 = 0.14917 \end{aligned}$$

Hence:

$$P(9.8 < X < 18.2) = 0.73891 - 0.14917 = 0.58974$$

Question 1.11

If $X \sim N(100, 16)$, calculate these probabilities using standardisation:

(i) $P(X > 110)$

(ii) $P(X > 87)$

(iii) $P(95 < X < 107)$.

Common Error:

Many students divide by the variance instead of the standard deviation.

This next question involves interpolation as well as standardising:

Question 1.12

The heights of male actuaries are normally distributed with mean 178 cm and variance 250 cm². Calculate the probability that a randomly chosen male actuary has height:

- (i) less than 186 cm
- (ii) more than 160 cm
- (iii) between 150 cm and 190 cm.

Miscellaneous problems

This last section covers two types of question involving normal distribution probabilities.

Conditional probabilities

We can work out conditional probabilities using the conditional probability formula:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

For example, to calculate $P(X < 3 | X < 6)$ where $X \sim N(5, 4)$, we get:

$$P(X < 3 | X < 6) = \frac{P(X < 3 \text{ and } X < 6)}{P(X < 6)} = \frac{P(X < 3)}{P(X < 6)}$$

We now calculate these probabilities as before.

$$\begin{aligned}
 P(X < 3) &= P\left(Z < \frac{3-5}{\sqrt{4}}\right) \text{ standardising} \\
 &= P(Z < -1) \\
 &= P(Z > 1) \quad \text{by symmetry} \\
 &= 1 - P(Z < 1) \quad \text{using area} = 1 \\
 &= 1 - 0.84134 \quad \text{from tables} \\
 &= 0.15866
 \end{aligned}$$

$$\begin{aligned}
 P(X < 6) &= P\left(Z < \frac{6-5}{\sqrt{4}}\right) \text{ standardising} \\
 &= P(Z < 0.5) \\
 &= 0.69146 \quad \text{from tables}
 \end{aligned}$$

Hence:

$$P(X < 3 | X < 6) = \frac{0.15866}{0.69146} = 0.2295$$

Modulus

The modulus function, $|x|$, gives the positive value of x :

$$|x| = \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}$$

So if we had:

$$|x| < 3$$

this means that either:

$$x < 3 \quad \text{if } x \text{ was positive}$$

or:

$$-x < 3 \Rightarrow x > -3 \quad \text{if } x \text{ was negative}$$

So $|x| < 3$ corresponds to the following range:

$$-3 < x < 3$$

If we are asked to calculate a probability involving a modulus, we rewrite it as a range:

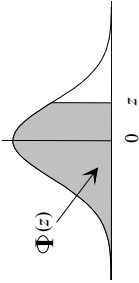
$$P(|X| < 5) = P(-5 < X < 5)$$

$$\begin{aligned} P(|X - 3| < 10) &= P(-10 < X - 3 < 10) \\ &= P(-7 < X < 13) \end{aligned}$$

We can then calculate these in the normal (no pun intended) way.

Question 1.13

Calculate $P(|X - 5| < 2)$ where $X \sim N(6, 1)$.



Appendix A: Probabilities for the standard normal distribution

| z | $\Phi(z)$ | z | $\Phi(z)$ | z | $\Phi(z)$ | z | $\Phi(z)$ | z | $\Phi(z)$ | z | $\Phi(z)$ | z | $\Phi(z)$ | z | $\Phi(z)$ |
|------|-----------|------|-----------|------|-----------|------|-----------|------|-----------|------|-----------|------|-----------|------|-----------|
| 0.00 | 0.50000 | 0.40 | 0.65542 | 0.80 | 0.78814 | 1.20 | 0.88493 | 1.60 | 0.94520 | 2.00 | 0.97725 | 2.40 | 0.99180 | 2.80 | 0.99744 |
| 0.01 | 0.50399 | 0.41 | 0.65910 | 0.81 | 0.79103 | 1.21 | 0.88686 | 1.61 | 0.94630 | 2.01 | 0.97778 | 2.41 | 0.99202 | 2.81 | 0.99752 |
| 0.02 | 0.50798 | 0.42 | 0.66276 | 0.82 | 0.79389 | 1.22 | 0.88877 | 1.62 | 0.94738 | 2.02 | 0.97831 | 2.42 | 0.99224 | 2.82 | 0.99760 |
| 0.03 | 0.51197 | 0.43 | 0.66640 | 0.83 | 0.79673 | 1.23 | 0.89065 | 1.63 | 0.94845 | 2.03 | 0.97882 | 2.43 | 0.99245 | 2.83 | 0.99767 |
| 0.04 | 0.51595 | 0.44 | 0.67003 | 0.84 | 0.79955 | 1.24 | 0.89251 | 1.64 | 0.94950 | 2.04 | 0.97932 | 2.44 | 0.99266 | 2.84 | 0.99774 |
| 0.05 | 0.51994 | 0.45 | 0.67364 | 0.85 | 0.80234 | 1.25 | 0.89435 | 1.65 | 0.95053 | 2.05 | 0.97982 | 2.45 | 0.99286 | 2.85 | 0.99781 |
| 0.06 | 0.52392 | 0.46 | 0.67724 | 0.86 | 0.80511 | 1.26 | 0.89617 | 1.66 | 0.95154 | 2.06 | 0.98030 | 2.46 | 0.99305 | 2.86 | 0.99788 |
| 0.07 | 0.52790 | 0.47 | 0.68082 | 0.87 | 0.80785 | 1.27 | 0.89796 | 1.67 | 0.95254 | 2.07 | 0.98077 | 2.47 | 0.99324 | 2.87 | 0.99795 |
| 0.08 | 0.53188 | 0.48 | 0.68439 | 0.88 | 0.81057 | 1.28 | 0.89973 | 1.68 | 0.95352 | 2.08 | 0.98124 | 2.48 | 0.99343 | 2.88 | 0.99801 |
| 0.09 | 0.53586 | 0.49 | 0.68793 | 0.89 | 0.81327 | 1.29 | 0.90147 | 1.69 | 0.95449 | 2.09 | 0.98169 | 2.49 | 0.99361 | 2.89 | 0.99807 |
| 0.10 | 0.53983 | 0.50 | 0.69146 | 0.90 | 0.81594 | 1.30 | 0.90320 | 1.70 | 0.95543 | 2.10 | 0.98214 | 2.50 | 0.99379 | 2.90 | 0.99813 |
| 0.11 | 0.54380 | 0.51 | 0.69497 | 0.91 | 0.81859 | 1.31 | 0.90490 | 1.71 | 0.95637 | 2.11 | 0.98257 | 2.51 | 0.99396 | 2.91 | 0.99819 |
| 0.12 | 0.54776 | 0.52 | 0.69847 | 0.92 | 0.82121 | 1.32 | 0.90658 | 1.72 | 0.95728 | 2.12 | 0.98300 | 2.52 | 0.99413 | 2.92 | 0.99825 |
| 0.13 | 0.55172 | 0.53 | 0.70194 | 0.93 | 0.82381 | 1.33 | 0.90824 | 1.73 | 0.95818 | 2.13 | 0.98341 | 2.53 | 0.99430 | 2.93 | 0.99831 |
| 0.14 | 0.55567 | 0.54 | 0.70540 | 0.94 | 0.82639 | 1.34 | 0.90988 | 1.74 | 0.95907 | 2.14 | 0.98382 | 2.54 | 0.99446 | 2.94 | 0.99836 |
| 0.15 | 0.55962 | 0.55 | 0.70884 | 0.95 | 0.82894 | 1.35 | 0.91149 | 1.75 | 0.95994 | 2.15 | 0.98422 | 2.55 | 0.99461 | 2.95 | 0.99841 |
| 0.16 | 0.56356 | 0.56 | 0.71226 | 0.96 | 0.83147 | 1.36 | 0.91308 | 1.76 | 0.96080 | 2.16 | 0.98461 | 2.56 | 0.99477 | 2.96 | 0.99846 |
| 0.17 | 0.56749 | 0.57 | 0.71566 | 0.97 | 0.83398 | 1.37 | 0.91466 | 1.77 | 0.96164 | 2.17 | 0.98500 | 2.57 | 0.99492 | 2.97 | 0.99851 |
| 0.18 | 0.57142 | 0.58 | 0.71904 | 0.98 | 0.83646 | 1.38 | 0.91621 | 1.78 | 0.96246 | 2.18 | 0.98537 | 2.58 | 0.99506 | 2.98 | 0.99856 |
| 0.19 | 0.57535 | 0.59 | 0.72240 | 0.99 | 0.83891 | 1.39 | 0.91774 | 1.79 | 0.96327 | 2.19 | 0.98574 | 2.59 | 0.99520 | 2.99 | 0.99861 |
| 0.20 | 0.57926 | 0.60 | 0.72575 | 1.00 | 0.84134 | 1.40 | 0.91924 | 1.80 | 0.96407 | 2.20 | 0.98610 | 2.60 | 0.99534 | 3.00 | 0.99865 |
| 0.21 | 0.58317 | 0.61 | 0.72907 | 1.01 | 0.84375 | 1.41 | 0.92073 | 1.81 | 0.96485 | 2.21 | 0.98645 | 2.61 | 0.99547 | 3.01 | 0.99869 |
| 0.22 | 0.58706 | 0.62 | 0.73237 | 1.02 | 0.84614 | 1.42 | 0.92220 | 1.82 | 0.96562 | 2.22 | 0.98679 | 2.62 | 0.99560 | 3.02 | 0.99874 |
| 0.23 | 0.59095 | 0.63 | 0.73565 | 1.03 | 0.84849 | 1.43 | 0.92364 | 1.83 | 0.96638 | 2.23 | 0.98713 | 2.63 | 0.99573 | 3.03 | 0.99878 |
| 0.24 | 0.59483 | 0.64 | 0.73891 | 1.04 | 0.85083 | 1.44 | 0.92507 | 1.84 | 0.96712 | 2.24 | 0.98745 | 2.64 | 0.99585 | 3.04 | 0.99882 |
| 0.25 | 0.59871 | 0.65 | 0.74215 | 1.05 | 0.85314 | 1.45 | 0.92647 | 1.85 | 0.96784 | 2.25 | 0.98778 | 2.65 | 0.99598 | 3.05 | 0.99886 |
| 0.26 | 0.60257 | 0.66 | 0.74537 | 1.06 | 0.85543 | 1.46 | 0.92785 | 1.86 | 0.96856 | 2.26 | 0.98809 | 2.66 | 0.99609 | 3.06 | 0.99889 |
| 0.27 | 0.60642 | 0.67 | 0.74857 | 1.07 | 0.85769 | 1.47 | 0.92922 | 1.87 | 0.96926 | 2.27 | 0.98840 | 2.67 | 0.99621 | 3.07 | 0.99893 |
| 0.28 | 0.61026 | 0.68 | 0.75175 | 1.08 | 0.85993 | 1.48 | 0.93056 | 1.88 | 0.96995 | 2.28 | 0.98870 | 2.68 | 0.99632 | 3.08 | 0.99896 |
| 0.29 | 0.61409 | 0.69 | 0.75490 | 1.09 | 0.86214 | 1.49 | 0.93189 | 1.89 | 0.97062 | 2.29 | 0.98899 | 2.69 | 0.99643 | 3.09 | 0.99900 |
| 0.30 | 0.61791 | 0.70 | 0.75804 | 1.10 | 0.86433 | 1.50 | 0.93319 | 1.90 | 0.97128 | 2.30 | 0.98928 | 2.70 | 0.99653 | 3.10 | 0.99903 |
| 0.31 | 0.62172 | 0.71 | 0.76115 | 1.11 | 0.86650 | 1.51 | 0.93448 | 1.91 | 0.97193 | 2.31 | 0.98956 | 2.71 | 0.99664 | 3.11 | 0.99906 |
| 0.32 | 0.62552 | 0.72 | 0.76424 | 1.12 | 0.86864 | 1.52 | 0.93574 | 1.92 | 0.97257 | 2.32 | 0.98983 | 2.72 | 0.99674 | 3.12 | 0.99910 |
| 0.33 | 0.62930 | 0.73 | 0.76730 | 1.13 | 0.87076 | 1.53 | 0.93699 | 1.93 | 0.97320 | 2.33 | 0.99010 | 2.73 | 0.99683 | 3.13 | 0.99913 |
| 0.34 | 0.63307 | 0.74 | 0.77035 | 1.14 | 0.87286 | 1.54 | 0.93822 | 1.94 | 0.97381 | 2.34 | 0.99036 | 2.74 | 0.99693 | 3.14 | 0.99916 |
| 0.35 | 0.63683 | 0.75 | 0.77337 | 1.15 | 0.87493 | 1.55 | 0.93943 | 1.95 | 0.97441 | 2.35 | 0.99061 | 2.75 | 0.99702 | 3.15 | 0.99918 |
| 0.36 | 0.64058 | 0.76 | 0.77637 | 1.16 | 0.87698 | 1.56 | 0.94062 | 1.96 | 0.97500 | 2.36 | 0.99086 | 2.76 | 0.99711 | 3.16 | 0.99921 |
| 0.37 | 0.64431 | 0.77 | 0.77935 | 1.17 | 0.87900 | 1.57 | 0.94179 | 1.97 | 0.97558 | 2.37 | 0.99111 | 2.77 | 0.99720 | 3.17 | 0.99924 |
| 0.38 | 0.64803 | 0.78 | 0.78230 | 1.18 | 0.88100 | 1.58 | 0.94295 | 1.98 | 0.97615 | 2.38 | 0.99134 | 2.78 | 0.99728 | 3.18 | 0.99926 |
| 0.39 | 0.65173 | 0.79 | 0.78524 | 1.19 | 0.88298 | 1.59 | 0.94408 | 1.99 | 0.97670 | 2.39 | 0.99158 | 2.79 | 0.99736 | 3.19 | 0.99929 |
| 0.40 | 0.65542 | 0.80 | 0.78814 | 1.20 | 0.88493 | 1.60 | 0.94520 | 2.00 | 0.97725 | 2.40 | 0.99180 | 2.80 | 0.99744 | 3.20 | 0.99931 |

***This page has been left blank so that you can pull out
the normal tables for reference.***

Extra practice questions

P1.1 Sketch a graph of $N(100,100)$.

P1.2 Find:

(i) $P(Z > 1.6)$

(ii) $P(Z < 2.84)$

(iii) $P(Z < -0.42)$

(iv) $P(Z > -2.61)$

(v) $P(-1.69 < Z < 0.52)$

(vi) $P(-3.05 < Z < -1.7)$.

P1.3 Calculate the following using interpolation:

(i) $P(Z < 0.371)$

(ii) $P(Z > 2.598)$

(iii) $P(Z > -0.904)$

(iv) $P(Z < -2.319)$

(v) $P(1.572 < Z < 3.087)$

(vi) $P(-1.382 < Z < 0.493)$.

P1.4 Calculate the following, using interpolation:

(i) $P(X > 117)$ where $X \sim N(128, 8^2)$

(ii) $P(Y < 30)$ where $Y \sim N(23, 5^2)$

(iii) $P(W > 297)$ where $W \sim N(285, 34)$

(iv) $P(V < 90)$ where $V \sim N(96, 12)$

(v) $P(3 < T < 8)$ where $T \sim N(5, 2)$.

P1.5 Given that $X \sim N(17, 5)$, evaluate:

(i) $P(|X - 17| < 2)$

(ii) $P(X < 15 | X < 16)$.

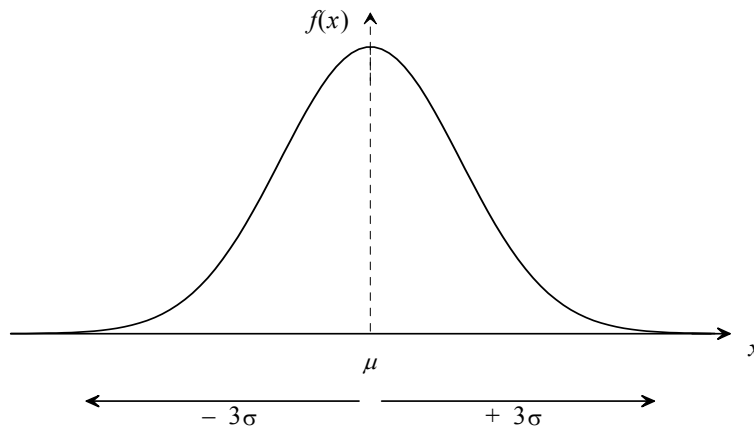
P1.6 The sizes of claims, which arise under policies of a certain type, are normally distributed with mean $\mu = \text{£}4,000$ and standard deviation $\sigma = \text{£}600$. The size of a particular claim is known to be greater than $\text{£}3,400$.

What is the probability that this claim size is greater than $\text{£}4,000$, the average of all claims sizes? [3]

Summary

The normal distribution

If X has a **normal distribution** with mean μ and variance σ^2 then we write $X \sim N(\mu, \sigma^2)$. This distribution occurs naturally and has a symmetrical, bell-shaped PDF:



The PDF for $X \sim N(\mu, \sigma^2)$ is:

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

The moments are:

$$E(X) = \mu$$

$$\text{var}(X) = \sigma^2$$

The median is the same as the mean (by symmetry) as is the mode.

Probabilities can only be found by **standardising** the normal distribution (*ie* transforming it into a standard normal distribution) using:

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

We then use the standard normal distribution tables.

The standard normal distribution

This is a normal distribution with mean 0 and variance 1. We write:

$$Z \sim N(0,1)$$

It has PDF:

$$\phi(z) = f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

The cumulative density function $\Phi(z) = P(Z < z)$ is given in the table in Appendix A.

We can use this to calculate probabilities as follows:

$$P(Z < a) \quad \text{read off the tables}$$

$$P(Z > a) = 1 - P(Z < a) \quad \text{using area (and total probability) equals 1}$$

$$P(Z < -a) = P(Z > a) \quad \text{using symmetry}$$

$$= 1 - P(Z < a) \quad \text{using area (and total probability) equals 1}$$

$$P(Z > -a) = P(Z < a) \quad \text{using symmetry}$$

'Compound' probabilities can be found using:

$$P(a < X < b) = P(X < b) - P(X < a)$$

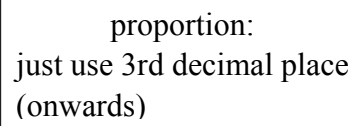
Probabilities involving a modulus can be rewritten as a range:

$$P(|X| < a) = P(-a < X < a)$$

Linear interpolation

It is expected that standard normal values are given to *at least* 3 decimal places and a more accurate answer is obtained using linear interpolation between the tabulated values:

$$p = (\text{start probability}) + 0.\dots \times (\text{difference between probabilities})$$

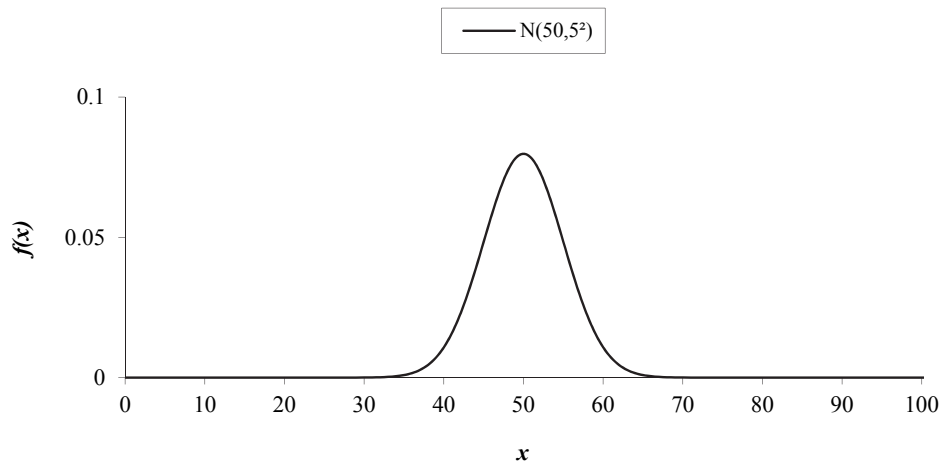


proportion:
just use 3rd decimal place
(onwards)

Solutions

Solution 1.1

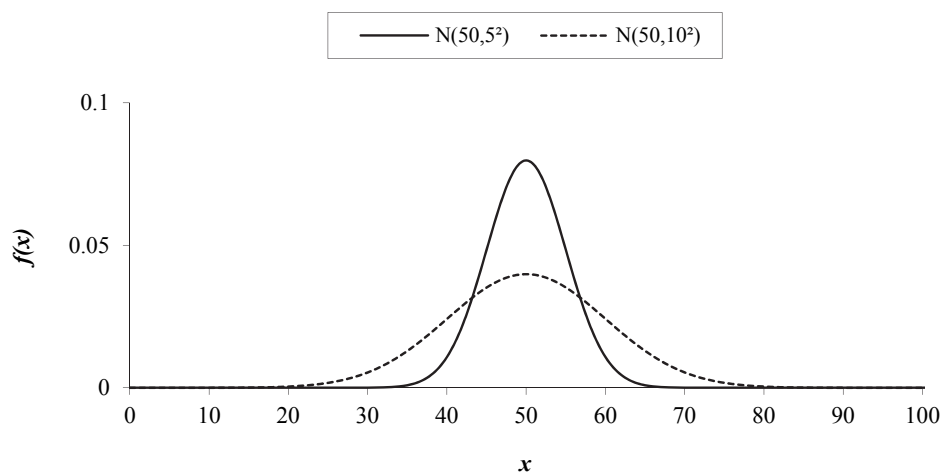
(i) The graph for $X \sim N(50, 5^2)$ is:



The majority of the values lie between $(50 - 3 \times 5, 50 + 3 \times 5) = (35, 65)$.

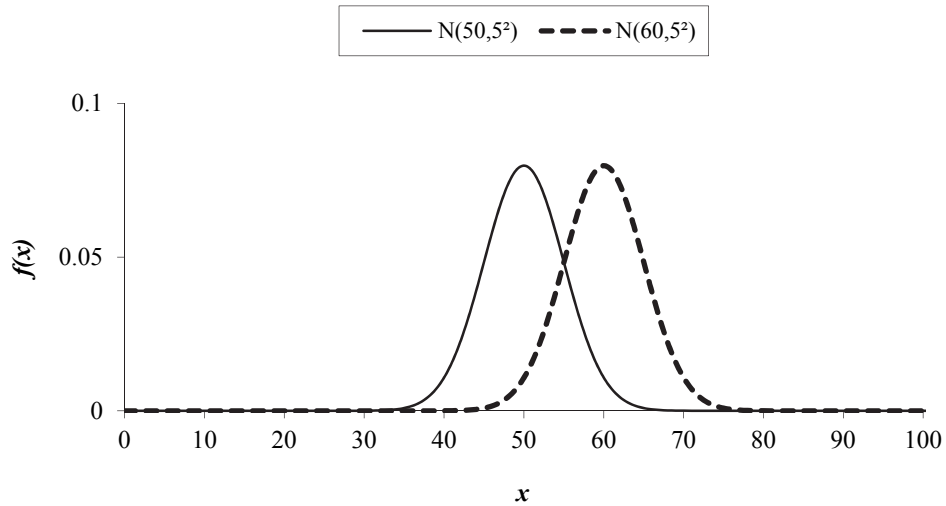
The height can be found by $f(50) = \frac{1}{\sqrt{2\pi}5^2} e^0 = 0.080$.

Now $Y \sim N(50, 10^2)$ has the same mean but a larger variance, so the PDF will have the same position/centre but will be more spread out.



The majority of the values lie between $(50 - 3 \times 10, 50 + 3 \times 10) = (20, 80)$. Since it's twice as spread out, it should be half as high.

- (ii) Compared to $N(50,5^2)$, the mean of $N(60,5^2)$ has increased by 10 but the variance is the same, so the PDF will have the same shape/spread but will be shifted 10 to the right.



Solution 1.2

Since:

$$\text{var}(X) = E(X^2) - E^2(X)$$

This gives:

$$\begin{aligned} E(X^2) &= \text{var}(X) + E^2(X) \\ &= \sigma^2 + \mu^2 \end{aligned}$$

Solution 1.3

- (i) Reading from the standard normal table:

$$P(Z < 1.23) = 0.89065$$

- (ii) 2.725 is halfway between 2.72 and 2.73. Reading from the standard normal table:

$$P(Z < 2.72) = 0.99674$$

$$P(Z < 2.73) = 0.99683$$

Since 2.725 is halfway between these it will be:

$$P(Z < 2.725) = \frac{0.99674 + 0.99683}{2} = 0.996785$$

Solution 1.4

- (i) Using the fact that the area under the graph is 1, we have:

$$P(Z > 2.17) = 1 - P(Z < 2.17)$$

Now reading $P(Z < 2.17)$ from the standard normal table:

$$P(Z > 2.17) = 1 - 0.98500 = 0.01500$$

- (ii) Similarly:

$$\begin{aligned} P(Z > 0.08) &= 1 - P(Z < 0.08) && \text{using area} = 1 \\ &= 1 - 0.53188 && \text{from tables} \\ &= 0.46812 \end{aligned}$$

Solution 1.5

(i) Using our 'swap the sign, swap the sign' symmetry result:

$$P(Z < -1.50) = P(Z > 1.50)$$

Now using the fact that the area under the graph is 1:

$$P(Z < -1.50) = P(Z > 1.50) = 1 - P(Z < 1.50)$$

Reading the value of $P(Z < 1.50)$ from the standard normal table:

$$P(Z < -1.50) = 1 - 0.93319 = 0.06681$$

(ii) Similarly,

$$\begin{aligned} P(Z < -0.21) &= P(Z > 0.21) && \text{by symmetry} \\ &= 1 - P(Z < 0.21) && \text{using area} = 1 \\ &= 1 - 0.58317 && \text{from tables} \\ &= 0.41683 \end{aligned}$$

(iii) We get:

$$\begin{aligned} P(Z < -2.05) &= P(Z > 2.05) && \text{by symmetry} \\ &= 1 - P(Z < 2.05) && \text{using area} = 1 \\ &= 1 - 0.97982 && \text{from tables} \\ &= 0.02018 \end{aligned}$$

Solution 1.6

- (i) Using our 'swap the sign, swap the sign' symmetry result:

$$P(Z > -3.94) = P(Z < 3.94)$$

We can read this value directly from the standard normal table:

$$P(Z > -3.94) = P(Z < 3.94) = 0.99996$$

- (ii) Similarly,

$$\begin{aligned} P(Z > -0.73) &= P(Z < 0.73) && \text{by symmetry} \\ &= 0.76730 && \text{from tables} \end{aligned}$$

Solution 1.7

$$\begin{aligned} \text{(i)} \quad P(Z > 3.6) &= 1 - P(Z < 3.6) && \text{using area} = 1 \\ &= 1 - 0.99984 && \text{from tables} \\ &= 0.00016 \end{aligned}$$

$$\begin{aligned} \text{(ii)} \quad P(Z > -0.76) &= P(Z < 0.76) && \text{by symmetry} \\ &= 0.77637 && \text{from tables} \end{aligned}$$

$$\text{(iii)} \quad P(Z < 1.98) = 0.97615 \quad \text{from tables}$$

$$\begin{aligned} \text{(iv)} \quad P(Z < -2.41) &= P(Z > 2.41) && \text{by symmetry} \\ &= 1 - P(Z < 2.41) && \text{using area} = 1 \\ &= 1 - 0.99202 && \text{from tables} \\ &= 0.00798 \end{aligned}$$

Solution 1.8

(i) Splitting up the 'compound' probability:

$$P(1.24 < Z < 2.19) = P(Z < 2.19) - P(Z < 1.24)$$

Both of these probabilities can be read directly from the standard normal table:

$$P(1.24 < Z < 2.19) = 0.98574 - 0.89251 = 0.09323$$

(ii) Similarly:

$$P(-0.92 < Z < 0.83) = P(Z < 0.83) - P(Z < -0.92)$$

Now:

$$P(Z < 0.83) = 0.79673 \quad \text{from tables}$$

$$P(Z < -0.92) = P(Z > 0.92) \quad \text{by symmetry}$$

$$= 1 - P(Z < 0.92) \quad \text{using area} = 1$$

$$= 1 - 0.82121 \quad \text{from tables}$$

$$= 0.17879$$

So this gives:

$$P(-0.92 < Z < 0.83) = 0.79673 - 0.17879 = 0.61794$$

(iii) Splitting up the 'compound' probability:

$$P(-2.92 < Z < -1.67) = P(Z < -1.67) - P(Z < -2.92)$$

Now:

$$\begin{aligned} P(Z < -1.67) &= P(Z > 1.67) && \text{by symmetry} \\ &= 1 - P(Z < 1.67) && \text{using area} = 1 \\ &= 1 - 0.95254 && \text{from tables} \\ &= 0.04746 \end{aligned}$$

$$\begin{aligned} P(Z < -2.92) &= P(Z > 2.92) && \text{by symmetry} \\ &= 1 - P(Z < 2.92) && \text{using area} = 1 \\ &= 1 - 0.99825 && \text{from tables} \\ &= 0.00175 \end{aligned}$$

This gives:

$$P(-2.92 < Z < -1.67) = 0.04746 - 0.00175 = 0.04571$$

Solution 1.9

(i) To calculate $P(Z < 1.048)$ we first look up the tabulated probabilities either side:

$$P(Z < 1.04) = 0.85083$$

$$P(Z < 1.05) = 0.85314$$

Using linear interpolation:

$$P(Z < 1.048) = 0.85083 + 0.8 \times (0.85314 - 0.85083) = 0.85268$$

Note that this figure is given to only 5 DP as the tables are only accurate to 5DP.

(ii) Now:

$$P(Z > 0.271) = 1 - P(Z < 0.271) \quad \text{using area} = 1$$

The tabulated probabilities either side of 0.271 are:

$$P(Z < 0.27) = 0.60642$$

$$P(Z < 0.28) = 0.61026$$

Using linear interpolation:

$$P(Z < 0.271) = 0.60642 + 0.1 \times (0.61026 - 0.60642) = 0.60680$$

So:

$$P(Z > 0.271) = 1 - 0.60680 = 0.39320$$

(iii) Now:

$$P(Z > -2.389) = P(Z < 2.389) \quad \text{by symmetry}$$

The tabulated probabilities either side of 2.389 are:

$$P(Z < 2.38) = 0.99134$$

$$P(Z < 2.39) = 0.99158$$

Using linear interpolation:

$$P(Z < 2.389) = 0.99134 + 0.9 \times (0.99158 - 0.99134) = 0.99156$$

So:

$$P(Z > -2.389) = 0.99156$$

(iv) First we split up the 'compound' probability:

$$P(-0.704 < Z < 0.897) = P(Z < 0.897) - P(Z < -0.704)$$

For $P(Z < 0.897)$ the tabulated probabilities either side are:

$$P(Z < 0.89) = 0.81327$$

$$P(Z < 0.90) = 0.81594$$

Using linear interpolation:

$$P(Z < 0.897) = 0.81327 + 0.7 \times (0.81594 - 0.81327) = 0.81514$$

Now:

$$\begin{aligned} P(Z < -0.704) &= P(Z > 0.704) \quad \text{by symmetry} \\ &= 1 - P(Z < 0.704) \quad \text{using area} = 1 \end{aligned}$$

The tabulated probabilities either side of $P(Z < 0.704)$ are:

$$P(Z < 0.70) = 0.75804$$

$$P(Z < 0.71) = 0.76115$$

Using linear interpolation:

$$P(Z < 0.704) = 0.75804 + 0.4 \times (0.76115 - 0.75804) = 0.75928$$

This gives:

$$P(Z < -0.704) = 1 - 0.75928 = 0.24072$$

Hence:

$$P(-0.704 < Z < 0.897) = 0.81514 - 0.24072 = 0.57442$$

Solution 1.10

$$(i) \quad 3X + 5 \sim N(3 \times 100 + 5, 3^2 \times 4^2) = N(305, 12^2)$$

$$(ii) \quad \frac{1}{2}X - 20 \sim N(\frac{1}{2} \times 100 - 20, \frac{1}{2}^2 \times 4^2) = N(30, 2^2)$$

$$(iii) \quad \frac{1}{4}(X - 100) \sim N\left(\frac{1}{4} \times (100 - 100), \frac{1}{4}^2 \times 4^2\right) = N(0, 1)$$

*We have transformed $N(100, 4^2)$ into the standard normal distribution. This process is called **standardising**. We shall be using it to calculate probabilities for any normal distribution.*

Solution 1.11

$$\begin{aligned}
 \text{(i)} \quad P(X > 110) &= P\left(Z > \frac{110-100}{\sqrt{16}}\right) \\
 &= P(Z > 2.5) \\
 &= 1 - P(Z < 2.5) && \text{using area} = 1 \\
 &= 1 - 0.99379 = 0.00621 && \text{from tables}
 \end{aligned}$$

(ii) Following the same steps:

$$\begin{aligned}
 P(X > 87) &= P\left(Z > \frac{87-100}{\sqrt{16}}\right) \\
 &= P(Z > -3.25) \\
 &= P(Z < 3.25) && \text{by symmetry} \\
 &= 0.99942 && \text{from tables}
 \end{aligned}$$

$$\text{(iii)} \quad P(95 < X < 107) = P(X < 107) - P(X < 95)$$

Now:

$$\begin{aligned}
 P(X < 107) &= P\left(Z < \frac{107-100}{\sqrt{16}}\right) && \text{standardising} \\
 &= P(Z < 1.75) \\
 &= 0.95994 && \text{from tables}
 \end{aligned}$$

$$\begin{aligned}
 P(X < 95) &= P\left(Z < \frac{95-100}{\sqrt{16}}\right) && \text{standardising} \\
 &= P(Z < -1.25) \\
 &= P(Z > 1.25) && \text{by symmetry} \\
 &= 1 - P(Z < 1.25) && \text{using area} = 1 \\
 &= 1 - 0.89435 && \text{using tables} \\
 &= 0.10565
 \end{aligned}$$

Therefore:

$$P(95 < X < 107) = 0.95994 - 0.10565 = 0.85429$$

Solution 1.12

Let X be the height of a male actuary, so we have $X \sim N(178, 250)$

$$\begin{aligned} \text{(i)} \quad P(X < 186) &= P\left(Z < \frac{186 - 178}{\sqrt{250}}\right) && \text{standardising} \\ &= P(Z < 0.506) && \text{3 DP} \end{aligned}$$

The tabulated probabilities either side of 0.506 are:

$$P(Z < 0.50) = 0.69146$$

$$P(Z < 0.51) = 0.69497$$

Using linear interpolation:

$$P(Z < 0.506) = 0.69146 + 0.6 \times (0.69497 - 0.69146) = 0.69357$$

Hence:

$$P(X < 186) = 0.69357$$

Note that we rounded the standardised value to 3 DP. We could have kept the complete figure of 0.5059644..., in which case we would have obtained:

$$\begin{aligned} P(X < 0.5059644\dots) &= 0.69146 + 0.59644\dots \times (0.69497 - 0.69146) \\ &= 0.69355 \end{aligned}$$

This does make a difference. We will give both the rounded and full solutions from now on.

$$\begin{aligned} \text{(ii)} \quad P(X > 160) &= P\left(Z > \frac{160 - 178}{\sqrt{250}}\right) && \text{standardising} \\ &= P(Z > -1.138) && \text{3 DP} \\ &= P(Z < 1.138) && \text{by symmetry} \end{aligned}$$

The tabulated probabilities either side of 1.138 are:

$$P(Z < 1.13) = 0.87076$$

$$P(Z < 1.14) = 0.87286$$

Using linear interpolation:

$$P(Z < 1.138) = 0.87076 + 0.8 \times (0.87286 - 0.87076) = 0.87244$$

So:

$$P(X > 160) = 0.87244$$

Using the complete standardised figure would give an answer of 0.87253.

$$\begin{aligned}
 \text{(iii)} \quad P(150 < X < 190) &= P(X < 190) - P(X < 150) \\
 &= P\left(Z < \frac{190 - 178}{\sqrt{250}}\right) - P\left(Z < \frac{150 - 178}{\sqrt{250}}\right) \text{ standardising} \\
 &= P(Z < 0.759) - P(Z < -1.771)
 \end{aligned}$$

$P(Z < 0.759)$ can be found directly from the standard normal table using interpolation. The tabulated probabilities either side of 0.759 are:

$$P(Z < 0.75) = 0.77337$$

$$P(Z < 0.76) = 0.77637$$

This gives:

$$P(Z < 0.759) = 0.77337 + 0.9 \times (0.77637 - 0.77337) = 0.77607$$

Now:

$$\begin{aligned}
 P(Z < -1.771) &= P(Z > 1.771) \quad \text{by symmetry} \\
 &= 1 - P(Z < 1.771) \quad \text{using area} = 1
 \end{aligned}$$

The tabulated probabilities either side of 1.771 are:

$$P(Z < 1.77) = 0.96164$$

$$P(Z < 1.78) = 0.96246$$

This gives:

$$P(Z < 1.771) = 0.96164 + 0.1 \times (0.96246 - 0.96164) = 0.96172$$

So:

$$P(Z < -1.771) = 1 - 0.96172 = 0.03828$$

Therefore:

$$P(150 < X < 190) = 0.77607 - 0.03828 = 0.73779$$

Using the complete standardised figures would give 0.73777.

Solution 1.13

Rewriting this probability as a range:

$$\begin{aligned} P(|X - 5| < 2) &= P(-2 < X - 5 < 2) \\ &= P(3 < X < 7) \\ &= P(X < 7) - P(X < 3) \end{aligned}$$

$$\begin{aligned} P(X < 7) &= P\left(Z < \frac{7-6}{\sqrt{1}}\right) \text{ standardising} \\ &= P(Z < 1) \quad \text{from tables} \\ &= 0.84134 \end{aligned}$$

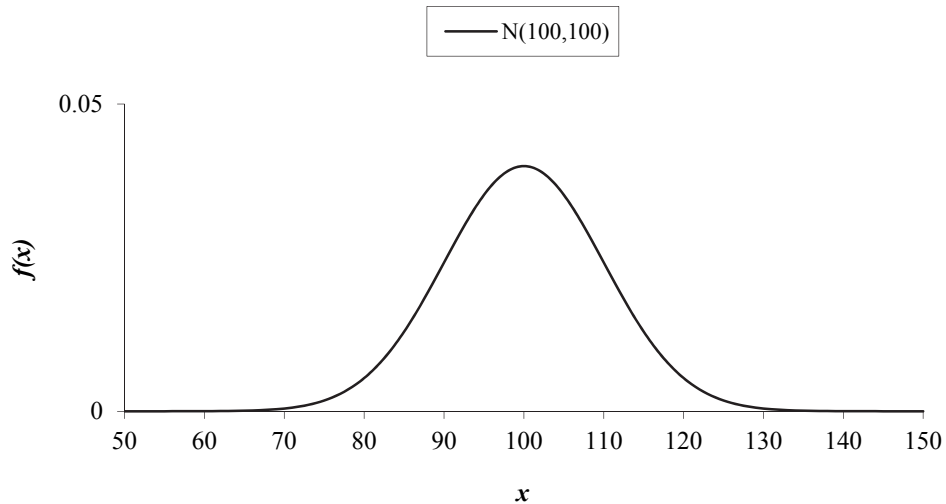
$$\begin{aligned} P(X < 3) &= P\left(Z < \frac{3-6}{\sqrt{1}}\right) \text{ standardising} \\ &= P(Z < -3) \\ &= P(Z > 3) \quad \text{by symmetry} \\ &= 1 - P(Z < 3) \quad \text{using area} = 1 \\ &= 1 - 0.99865 \quad \text{from tables} \\ &= 0.00135 \end{aligned}$$

Hence:

$$P(|X - 5| < 2) = 0.84134 - 0.00135 = 0.83999$$

Solutions to extra practice questions

P1.1 The graph for $X \sim N(100,100)$ is:



The majority of the values are between $(100 - 3 \times 10, 100 + 3 \times 10) = (70, 130)$.

The height can be found from $f(100) = \frac{1}{\sqrt{2\pi \cdot 100}} e^0 = 0.040$.

- P1.2**
- (i) $P(Z > 1.6) = 1 - P(Z < 1.6)$ using area = 1
 $= 1 - 0.94520$ from tables
 $= 0.05480$
- (ii) $P(Z < 2.84) = 0.99774$ from tables
- (iii) $P(Z < -0.42) = P(Z > 0.42)$ by symmetry
 $= 1 - P(Z < 0.42)$ using area = 1
 $= 1 - 0.66276$ from tables
 $= 0.33724$
- (iv) $P(Z > -2.61) = P(Z < 2.61)$ by symmetry
 $= 0.99547$ from tables

$$(v) \quad P(-1.69 < Z < 0.52) = P(Z < 0.52) - P(Z < -1.69)$$

Now:

$$P(Z < 0.52) = 0.69847 \quad \text{from tables}$$

$$\begin{aligned} P(Z < -1.69) &= P(Z > 1.69) && \text{by symmetry} \\ &= 1 - P(Z < 1.69) && \text{using area} = 1 \\ &= 1 - 0.95449 && \text{from tables} \\ &= 0.04551 \end{aligned}$$

Hence:

$$P(-1.69 < Z < 0.52) = 0.69847 - 0.04551 = 0.65296$$

$$(vi) \quad P(-3.05 < Z < -1.7) = P(Z < -1.7) - P(Z < -3.05)$$

Now:

$$\begin{aligned} P(Z < -1.7) &= P(Z > 1.7) && \text{by symmetry} \\ &= 1 - P(Z < 1.7) && \text{using area} = 1 \\ &= 1 - 0.95543 && \text{from tables} \\ &= 0.04457 \end{aligned}$$

$$\begin{aligned} P(Z < -3.05) &= P(Z > 3.05) && \text{by symmetry} \\ &= 1 - P(Z < 3.05) && \text{using area} = 1 \\ &= 1 - 0.99886 && \text{from tables} \\ &= 0.00114 \end{aligned}$$

Hence:

$$P(-3.05 < Z < -1.7) = 0.04457 - 0.00114 = 0.04343$$

P1.3 (i) The tabulated probabilities either side of $P(X < 0.371)$ are:

$$P(Z < 0.37) = 0.64431$$

$$P(Z < 0.38) = 0.64803$$

Using linear interpolation:

$$P(Z < 0.371) = 0.64431 + 0.1 \times (0.64803 - 0.64431) = 0.64468$$

(ii) $P(Z > 2.598) = 1 - P(Z < 2.598)$ using area = 1

The tabulated probabilities either side of $P(Z < 2.598)$ are:

$$P(Z < 2.59) = 0.99520$$

$$P(Z < 2.60) = 0.99534$$

Using linear interpolation:

$$P(Z < 2.598) = 0.99520 + 0.8 \times (0.99534 - 0.99520) = 0.99531$$

This gives:

$$P(Z > 2.598) = 1 - 0.99531 = 0.00469$$

(iii) $P(Z > -0.904) = P(Z < 0.904)$ by symmetry

The tabulated probabilities either side of $P(Z < 0.904)$ are:

$$P(Z < 0.90) = 0.81594$$

$$P(Z < 0.91) = 0.81859$$

Using linear interpolation:

$$P(Z < 0.904) = 0.81594 + 0.4 \times (0.81859 - 0.81594) = 0.81700$$

This gives:

$$P(Z > -0.904) = 0.8170$$

$$\begin{aligned}
 \text{(iv)} \quad P(Z < -2.319) &= P(Z > 2.319) \quad \text{by symmetry} \\
 &= 1 - P(Z < 2.319) \quad \text{using area} = 1
 \end{aligned}$$

The tabulated probabilities either side of $P(Z < 2.319)$ are:

$$P(Z < 2.31) = 0.98956$$

$$P(Z < 2.32) = 0.98983$$

Using linear interpolation:

$$P(Z < 2.319) = 0.98956 + 0.9 \times (0.98983 - 0.98956) = 0.98980$$

This gives:

$$P(Z < -2.319) = 1 - 0.98980 = 0.01020$$

$$\text{(v)} \quad P(1.572 < Z < 3.087) = P(Z < 3.087) - P(Z < 1.572)$$

For $P(Z < 3.087)$ the tabulated probabilities either side are:

$$P(Z < 3.08) = 0.99896$$

$$P(Z < 3.09) = 0.99900$$

Using linear interpolation:

$$P(Z < 3.087) = 0.99896 + 0.7 \times (0.99900 - 0.99896) = 0.99899$$

For $P(Z < 1.572)$ the tabulated probabilities either side are:

$$P(Z < 1.57) = 0.94179$$

$$P(Z < 1.58) = 0.94295$$

Using linear interpolation:

$$P(Z < 1.572) = 0.94179 + 0.2 \times (0.94295 - 0.94179) = 0.94202$$

Hence:

$$P(1.572 < Z < 3.087) = 0.99899 - 0.94202 = 0.05697$$

(vi) $P(-1.382 < Z < 0.493) = P(Z < 0.493) - P(Z < -1.382)$

For $P(Z < 0.493)$ the tabulated probabilities either side are:

$$P(Z < 0.49) = 0.68793$$

$$P(Z < 0.50) = 0.69146$$

Using linear interpolation:

$$P(Z < 0.493) = 0.68793 + 0.3 \times (0.69146 - 0.68793) = 0.68899$$

Now:

$$\begin{aligned} P(Z < -1.382) &= P(Z > 1.382) && \text{by symmetry} \\ &= 1 - P(Z < 1.382) && \text{using area} = 1 \end{aligned}$$

The tabulated probabilities either side of $P(Z < 1.382)$ are:

$$P(Z < 1.38) = 0.91621$$

$$P(Z < 1.39) = 0.91774$$

Using linear interpolation:

$$P(Z < 1.382) = 0.91621 + 0.2 \times (0.91774 - 0.91621) = 0.91652$$

This gives:

$$P(Z < -1.382) = 1 - 0.91652 = 0.08348$$

Hence:

$$P(-1.382 < Z < 0.493) = 0.68899 - 0.08348 = 0.60551$$

$$\begin{aligned}
 \mathbf{P1.4} \quad (i) \quad P(X > 117) &= P\left(Z > \frac{117 - 128}{8}\right) && \text{standardising} \\
 &= P(Z > -1.375) \\
 &= P(Z < 1.375) && \text{by symmetry}
 \end{aligned}$$

The tabulated probabilities either side of $P(Z < 1.375)$ are:

$$P(Z < 1.37) = 0.91466$$

$$P(Z < 1.38) = 0.91621$$

Using linear interpolation (or just finding the average):

$$P(Z < 1.375) = 0.91466 + 0.5 \times (0.91621 - 0.91466) = 0.91544$$

Hence:

$$P(X > 117) = 0.91544$$

$$\begin{aligned}
 (ii) \quad P(Y < 30) &= P\left(Z < \frac{30 - 23}{5}\right) && \text{standardising} \\
 &= P(Z < 1.4) \\
 &= 0.91924
 \end{aligned}$$

$$\begin{aligned}
 (iii) \quad P(W > 297) &= P\left(Z > \frac{297 - 285}{\sqrt{34}}\right) && \text{standardising} \\
 &= P(Z > 2.058) && \text{3 DP} \\
 &= 1 - P(Z < 2.058) && \text{using area} = 1
 \end{aligned}$$

The tabulated probabilities either side of $P(Z < 2.058)$ are:

$$P(Z < 2.05) = 0.97982$$

$$P(Z < 2.06) = 0.98030$$

Using linear interpolation:

$$P(Z < 2.058) = 0.97982 + 0.8 \times (0.98030 - 0.97982) = 0.98020$$

Hence:

$$P(W > 297) = 1 - 0.98020 = 0.01980$$

$$\begin{aligned} \text{(iv)} \quad P(V < 90) &= P\left(Z < \frac{90 - 96}{\sqrt{12}}\right) && \text{standardising} \\ &= P(Z < -1.732) && \text{3 DP} \\ &= P(Z > 1.732) && \text{by symmetry} \\ &= 1 - P(Z < 1.732) && \text{using area} = 1 \end{aligned}$$

The tabulated probabilities either side of $P(Z < 1.732)$ are:

$$P(Z < 1.73) = 0.95818$$

$$P(Z < 1.74) = 0.95907$$

Using linear interpolation:

$$P(Z < 1.732) = 0.95818 + 0.2 \times (0.95907 - 0.95818) = 0.95836$$

Hence:

$$P(V < 90) = 1 - 0.95836 = 0.04164$$

Using the complete standardised figure would give:

$$P(V < 90) = 1 - 0.95833 = 0.04167$$

$$\begin{aligned}
 \text{(v)} \quad P(3 < T < 8) &= P(T < 8) - P(T < 3) \\
 &= P\left(Z < \frac{8-5}{\sqrt{2}}\right) - P\left(Z < \frac{3-5}{\sqrt{2}}\right) && \text{standardising} \\
 &= P(Z < 2.121) - P(Z < -1.414) && 3 \text{ DP}
 \end{aligned}$$

For $P(Z < 2.121)$ the tabulated probabilities either side are:

$$P(Z < 2.12) = 0.98300$$

$$P(Z < 2.13) = 0.98341$$

Using linear interpolation:

$$P(Z < 2.121) = 0.98300 + 0.1 \times (0.98341 - 0.98300) = 0.98304$$

Now:

$$\begin{aligned}
 P(Z < -1.414) &= P(Z > 1.414) && \text{by symmetry} \\
 &= 1 - P(Z < 1.414) && \text{using area} = 1
 \end{aligned}$$

The tabulated probabilities either side of $P(Z < 1.414)$ are:

$$P(Z < 1.41) = 0.92073$$

$$P(Z < 1.42) = 0.92220$$

Using linear interpolation:

$$P(Z < 1.414) = 0.92073 + 0.4 \times (0.92220 - 0.92073) = 0.92132$$

This gives:

$$P(Z < -1.414) = 1 - 0.92132 = 0.07868$$

Therefore:

$$P(3 < T < 8) = 0.98304 - 0.07868 = 0.90436$$

Using the complete standardised figures would have given:

$$P(3 < T < 8) = 0.98305 - 0.07865 = 0.90440$$

P1.5 (i) Rewriting this probability as a range:

$$\begin{aligned}
 P(|X - 17| < 2) &= P(-2 < X - 17 < 2) \\
 &= P(15 < X < 19) \\
 &= P(X < 19) - P(X < 15) \\
 &= P\left(Z < \frac{19 - 17}{\sqrt{5}}\right) - P\left(Z < \frac{15 - 17}{\sqrt{5}}\right) && \text{standardising} \\
 &= P(Z < 0.894) - P(Z < -0.894) && 3 \text{ DP}
 \end{aligned}$$

For $P(Z < 0.894)$ the tabulated probabilities either side are:

$$P(Z < 0.89) = 0.81327$$

$$P(Z < 0.90) = 0.81594$$

Using linear interpolation:

$$P(Z < 0.894) = 0.81327 + 0.4 \times (0.81594 - 0.81327) = 0.81434$$

Now:

$$\begin{aligned}
 P(Z < -0.894) &= P(Z > 0.894) && \text{by symmetry} \\
 &= 1 - P(Z < 0.894) && \text{using area} = 1
 \end{aligned}$$

Since we have just calculated $P(Z < 0.894)$ we get:

$$P(Z < -0.894) = 1 - 0.81434 = 0.18566$$

Therefore:

$$P(|X - 17| < 2) = 0.81434 - 0.18566 = 0.62868$$

Using the complete figures for interpolation would give:

$$P(|X - 17| < 2) = 0.81445 - 0.18555 = 0.62890$$

(ii) Using the conditional probability formula:

$$P(X < 15 | X < 16) = \frac{P(X < 15 \text{ and } X < 16)}{P(X < 16)} = \frac{P(X < 15)}{P(X < 16)}$$

$$P(X < 15) = P\left(Z < \frac{15-17}{\sqrt{5}}\right) \quad \text{standardising}$$

$$= P(Z < -0.894) \quad \text{3 DP}$$

$$= P(Z > 0.894) \quad \text{by symmetry}$$

$$= 1 - P(Z < 0.894) \quad \text{using area} = 1$$

$$= 0.18566 \quad \text{from part (i)}$$

$$P(X < 16) = P\left(Z < \frac{16-17}{\sqrt{5}}\right) \quad \text{standardising}$$

$$= P(Z < -0.447) \quad \text{3 DP}$$

$$= P(Z > 0.447) \quad \text{by symmetry}$$

$$= 1 - P(Z < 0.447) \quad \text{using area} = 1$$

The tabulated probabilities either side of $P(Z < 0.447)$ are:

$$P(Z < 0.44) = 0.67003$$

$$P(Z < 0.45) = 0.67364$$

Using linear interpolation:

$$P(Z < 0.447) = 0.67003 + 0.7 \times (0.67364 - 0.67003) = 0.67256$$

This gives:

$$P(X < 16) = 1 - 0.67256 = 0.32744$$

Hence:

$$P(X < 15 | X < 16) = \frac{0.18566}{0.32744} = 0.5670$$

Using the complete standardised figures would give 0.5668.

P1.6 Let X be the sizes of claims, then $X \sim N(4000, 600^2)$.

We want:

$$P(X > 4,000 | X > 3,400) = \frac{P(X > 4,000 \text{ and } X > 3,400)}{P(X > 3,400)} = \frac{P(X > 4,000)}{P(X > 3,400)}$$

Now:

$$\begin{aligned} P(X > 4,000) &= P\left(Z > \frac{4,000 - 4,000}{600}\right) && \text{standardising} \\ &= P(Z > 0) \\ &= 1 - P(Z < 0) && \text{using area} = 1 \\ &= 1 - 0.5 && \text{from tables} \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} P(X > 3,400) &= P\left(Z > \frac{3,400 - 4,000}{600}\right) && \text{standardising} \\ &= P(Z > -1) \\ &= P(Z < 1) && \text{by symmetry} \\ &= 0.84134 && \text{from tables} \end{aligned}$$

Therefore:

$$P(X > 4,000 | X > 3,400) = \frac{0.5}{0.84134} = 0.59429$$