Random Variables

Contents

2 Random Variables

For many cases where we are looking at random events, the result of the experiment being done is quantifiable; that is, they are measurable numbers.

In such a case we are dealing with a variable which is random in nature. Random Variables are the subject this section.
In the case of a Random Variable:

- The experiment is producing and then measuring the variable
- The event is the variable taking on a particular value, or range of values.

We will see that some of the examples we have already looked at can be recast as random variables.

2.1.1 Rolling Two dice – Revisit

In the case of rolling two dice, calculate the probability that the difference is 2. The possible outcomes which make up this event are:

(1,3), (2,4), (3,1), (3,5), (4,2), (4,6), (5,3), (6,4).

In this case:

- The experiment is the rolling of the dice, and seeing which numbers come up, and then finding the difference.
- The outcomes are whatever numbers show up.

- The Random variable is the difference of the two numbers, call it D.
- The event is the random variable taking on particular values or ranges of values, in this case, $D = 2$.

In this case, the variable D has been assigned to the difference between the two numbers. So the event in this case is

$$D = 2.$$

From counting the possible ways in which this comes about, exactly as before, the probability is

$$P[D = 2] = 8/36 = 2/9.$$

The key thing here is that we are treating the numbers that come up on the dice as such, that is, adding and subtracting with them.

2.1.2 Discrete and Continuous Variables

A random variable is called *discrete* if it takes on distinct values, such as integers or integer multiples of a real number.

A random variable which can take on *any* numerical value is called a continuous random variable. These are usually physical quantities, such as height, weight, distance, resistance, capacitance and times.

2.1.3 Definition – Discrete Random Variable

A Random Variable is said to be *discrete* if it can only take on values from a set of distinct numbers, such as integers or a finite subset of numbers.

2.1.4 Definition – Continuous Random Variable

A Random Variable is said to be *continuous* if it can take on any real numbers within a certain range as a value.

2.2 Distributions

Random variables often have a distribution; this is a law governing the probability of particular values of the variable coming up.

In the case of a discrete random variable, this could simply be a list of probabilities for each value. It can also be an equation giving the probability in terms of the value.

2.2.1 A Simple Example of a Distribution

Consider again the example of a random variable generated by rolling two dice and calculating the difference.

It is a discreet variable since the difference can only take on certain values, 0 up to 5.

They have been written as fractions of 18, the lowest common denominator:

$$P[D = 0] = 6/36 = 3/18.$$

$$P[D = 1] = 10/36 = 5/18.$$

$$P[D = 2] = 8/36 = 4/18.$$

$$P[D = 3] = 6/36 = 3/18.$$

$$P[D = 4] = 4/36 = 2/18.$$

$$P[D = 5] = 2/36 = 1/18.$$

2.2.2 Another Simple Distribution

Let X be a random variable, which takes on one of the ten discrete values

$$1, 2, \ldots, 9, 10.$$

This list is in fact the event space. This means that every event involving this variable is of the form:

$$X = n, \text{ where } n = 1, 2, \ldots, 9, 10.$$

Let us say that each value is equally likely to come up. Then:

$$P[X = n] = 0.1, \text{ where } n = 1, 2, \ldots, 9, 10.$$

This is the simplest possible distribution, where all the figures have the same probability.

The ten possible events, having the probability 0.1, means that if all probabilities are added up, we get 1, as should be the case:

$$\sum_{n=1}^{10} P[X = n] = 10 \times 0.1 = 1.$$

2.2.3 Another Discrete Random Variable

Consider the following random variable and its distribution. Let X be a variable which takes on any positive integer.

The event space is therefore the set

$$\{1, 2, 3, \ldots \infty\}$$

The probabilities are given by the following equation:

$$P[X = n] = \tfrac{1}{2}^{n}.$$

This means that the probability that the variable X takes on value n is simply ½ raised to that power. As an example:

$$P[X = 1] = \tfrac{1}{2}.$$

$$P[X = 2] = \tfrac{1}{4}.$$

All the probabilities must add up to 1; we will check this. The sum of all the probabilities is:

$$\sum_{n=1}^{\infty} P[X = n] = \sum_{n=1}^{\infty} \tfrac{1}{2}^{n} .$$

This is in fact a geometric series:

$$\tfrac{1}{2} + \tfrac{1}{4} + \ldots.$$

and it adds up to 1:

$$\sum_{n=1}^{\infty} \frac{1}{2}^{n} = 1.$$

2.2.4 The Exponential Distribution

Let X be a random variable which can take on any positive number, and let λ be a positive number. Let x represent an actual value of the variable which might come up. This random variable follows the distribution

$$P[X < x] = 1 - e^{-\lambda x}.$$

This distribution is called the Exponential Distribution.

It is important that for x being any positive number, the probability goes from 0 to 1, as we would expect.

So the probability we get a number less than 0 has to be 0:

$$P[X < 0] = 1 - e^{-\lambda 0} = 1 - e^0 = 1 - 1 = 0.$$

We can also see that as the possible value x gets larger and larger, the probability gets closer to 1, as the exponential term shrinks rapidly.

2.3 The Binomial Distribution

Consider the following problem – it is known that 11% of the Irish population is left-handed.

The following question is posed: in a randomly selected group of 12 people, find the probability of finding 3 people who are left-handed.

To answer this question, consider: in the 12 people, calculate the probability of the first 3 of the 12 being left-handed.

Each person being left or right-handed is an independent event.

The probabilities of these events for each person are therefore multiplied to find the overall probability, so the probability is

$$(0.11)^3.(0.89)^9.$$

But the first 3 being left-handed, and the next 9 being right-handed, is just one way of having the result in question, three people being left-handed.

But for each one, the probability of that selection coming up is the same as the value shown above. These possible outcomes are all mutually exclusive, so add on

$$(0.11)^3.(0.89)^9.$$

for each combination.

In other words, multiply by the number of combinations of 3 from 12 – this is ${}^{12}C_3$.

The probability of finding 5 left-handed people in a group of 30 is then

$$ {}^{12}C_3(0.11)^3(0.89)^9. $$

This is an example of what is called the Binomial distribution.

2.3.1 Definition – The Binomial Distribution

A trial is being repeated, with a possible result A:

- Each time the trial is done, the probability of result A turning up is p.
- The trial is repeated n times.

Let X be the random variable of the number of times event A comes up. The probability of getting r results from n trials is:

$$\mathrm{P}[X = r] = {}^{n}C_{r}\, p^{r}(1 - p)^{n - r}.$$

The case of the number of left-handed people fits into this pattern.

- The trial being repeated is checking whether or not a person is left-handed.
- In the example we studied, this is being repeated 12 times, so this means $n = 12$.
- The probability of this occurring for each person ‘tested’ is $p = 0.11$. This number comes from the proportion of 0.11 of the population being left-handed.

Let L be the random variable of the number of left-handers in the group. The probability we are looking at is then: P[L = 3].

Using the binomial distribution,

$$P[L = r] = {}^nC_r p^r(1 - p)^{n-r}.$$

$$P[L = 3] = {}^{12}C_3\,(0.11)^3(0.89)^9 =$$

$$= 220 \times 0.004913 \times 0.18694 =$$

$$= 0.20206.$$

2.3.2 Example – Defective Components

A factory is producing components, of which 1.5% are defective. They are packed in boxes, each containing 20 components. Calculate the probability that a box has 2 defective components.

This is a case of the binomial distribution.

- There are 20 components in each box, so the value of n is 20.

- Each component has a probability of 0.015 of being defective so $p = 0.015$.

To apply the binomial distribution, let X be the random variable of the number of defective components in a given box.

The event we are looking at is $X = 2$. So $r = 2$ in the equation.

Then

$$P[X = r] = {}^{n}C_{r}\, p^{r}(1 - p)^{n - r}:$$

For this case:

$$P[X = 2] = {}^{20}C_{2}(0.015)^{2}(1 - 0.015)^{18} =$$

$$= 190 x 0.015^{2} x 0.985^{18} = 0.033.$$

Thus of every 1000 boxes coming out of the factory, 33 would have two defectives.

2.3.3 More on this example

In this case, find the probability of getting

1. no defectives,
2. one defective,

and so find the probability of

3. less than 3 defectives.
4. 2 or more are defective

Part 1:

The probability of getting no defectives is

$$P[X = 0] = {}^{20}C_0\,(0.015)^0(1 - 0.015)^{20} =$$

$$= 1x1x0.985^{20} = 0.739.$$

Bear in mind that ${}^{20}C_{0} = 1$, this is because there is only one way of having no defectives in the box.

Part 2:

The probability of getting one defective is:

$$P[X = 1] = {}^{20}C_1(0.015)^1(1 - 0.015)^{19} =$$

$$= 20x0.015x0.985^{29} = 0.225.$$

Part 3:

To find the probability of getting less than three defectives, first look at exactly what this means. It means that

$$X = 0 \text{ or } X = 1 \text{ or } X = 2.$$

The probability of getting less than three defectives is then

$$P[X < 3] = P[X = 0 \text{ or } X = 1 \text{ or } X = 2].$$

Since these are mutually exclusive events, the probabilities can be added:

$$P[X < 3] = P[X = 0] + P[X = 1] + P[X = 2].$$

These are all probabilities we have worked out. So this means

$$P[X < 3] = 0.739 + 0.225 + 0.033 = 0.997.$$

Part 4:

To calculate the probability that in a box of 20, two or more are defective, we could add the probabilities for

$$X = 2,\ X = 3,\ X = 4 \text{ and so on.}$$

However, this would be simpler to use the fact that the event of getting two or more defectives is the converse of getting none or one. Thus

$$P[X \geq 2] = 1 - P[X = 0 \text{ or } X = 1].$$

Since these are mutually exclusive events, the probabilities can be added:

$$P[X = 0 \text{ or } X = 1] = P[X = 0] + P[X = 1].$$

So this means

$$P[X > 2] = 1 - \big(P[X = 0] + P[X = 1]\big) =$$

$$=1 - (0.739 + 0.225) = 1 - 0.964 = 0.036.$$

2.3.4 Example – Multiple Choice Exams

A classic example of a binomial distribution is the question of a student trying to pass a multiple choice examination by answering the questions at random.

A multiple choice exam has 20 questions, each one with a choice of five answers. A student chooses their answers at random. Find the probability they pass the exam, if 40% is the pass mark.

- A student is attempting 20 questions, and each time has a chance of getting it right. This is repeated 20 times, so $n = 20$.
- Each time a student answers a question, they have a chance of 0.2 of getting it right, so $p = 0.2$.

Let X be the random variable of the number of questions the student gets right. Then;

$$P[X = r] = {}^{20}\mathrm{C}_r 0.2^r\, 0.8^{20-r}.$$

To pass the exam, a student needs 8 questions or more right. The event is:

$$X \geq 8 \text{ which is } X = 8 \text{ or } X = 9 \text{ or} \ldots \text{ or } X = 20.$$

Since these are all mutually exclusive events, the probabilities are

$$P[X \geq 8] = P[X = 8] + P[X = 9] + \ldots + P[X = 20].$$

Each of these probabilities is calculated and the results summed up.

The results are;

$$P[X = 8] = {}^{20}C_8 0.2^8\ 0.8^{12} =$$

$$= 125{,}970 \text{x} 0.00000256 \text{x} 0.0687194 =$$

$$= 0.022161.$$

$$P[X = 9] = {}^{20}C_9 0.2^9\ 0.8^{11} =$$

$$= 167{,}960 \text{x} 0.00000512 \text{x}\ 0.08589934592$$

$$= 0.0073866.$$

Etc.

$$P[X = 12] = {}^{20}C_{12} 0.2^{12}\, 0.8^{8} =$$

$$125{,}970 \times 0.000000004096 \times 0.16777216$$

$$= 0.000086.$$

To three decimal places,

$$P[X \geq 8] = P[X = 8] + P[X = 9] + \ldots + P[X = 20] = 0.0321.$$

Thus the probability of passing is 0.0321, so the probability of failing is 0.9679.

2.4 The Normal Distribution

In the case of the binomial distribution, the random variable was a number of outcomes. This number, as a count, will always be an integer.

The next distribution we will look at is for continuous variables. These are variables which can take on any possible value, such as 1, 1.2, or 1.2345...

Most quantities in nature are like this, and very many of them follow the normal distribution.

When dealing with a continuous variable, we can no longer talk about the chances of the variable being equal to particular values; since there are an infinite number of possible values, the chances of one particular value coming up are zero.

Instead we talk about the probability of the variable being in a particular range. So if a variable r has a binomial distribution, the events were

$$r = 0;\ r = 1 \text{ or } r > 4.$$

With a continuous variable, say height H, the events are

$$H > 160\text{cm or } H < 170\text{cm}.$$

The Normal distribution for a particular variable is characterised by two numbers.

This is similar to the binomial distribution, where the value of the probabilities for the variable X is determined by the numbers n and p.

In the case of the normal distribution, the two numbers are

- The mean, or average, which is denoted μ, and
- The standard deviation, denoted σ.

Recall that for many cases of the use of the binomial distribution, the probability p of the event being counted often came from proportions of the population the sample came from.

The mean and the standard deviation may come from the analysis of a population. The ideas of the average and standard deviation are the same as those from the study of data presentation.

If a variable X is normally distributed, this means that if a large number of values are generated of the variable, then they will be more likely to be close to the mean μ, and unlikely to be far from it.

Just how ‘likely’ or ‘unlikely’ is determined by the standard deviation σ.

The probabilities for the Normal distribution are given by the following equation. Let X be a normally distributed variable, with mean μ and standard deviation σ.

Then the probability that it is less than a value a is

$$P[X < a] = \tfrac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^a e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x.$$

Now this is clearly a very complex integral, and in fact it is impossible to work out this integral using normal integration. Instead it must be calculated numerically using a computer.

This cannot be done for every possible choice of mean μ, and standard deviation σ. Instead it is done for one choice of μ and σ, and these are related to all others.

The table gives probabilities for a normally distributed random variable, with mean $\mu = 0$ and standard deviation $\sigma = 1$.

This is called the *standard* normal distribution. The probabilities given in the tables are that Z takes on a value greater than a given value a, in other words:

$$P[\, Z > a].$$

Its probabilities have been calculated by computer from the equation above.

2.4.1 Standardising

The Standard Normal Variable is a normal variable Z with mean 0 and standard deviation 1. These probabilities can then be used to find the probabilities of any normal variable because of the following property of the normal distribution.

Let the variable X be normally distributed with of mean μ, and standard deviation σ.

The variable Z, given by the relation

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

has mean 0 and standard deviation 1.

This comes from the defining integral equation for the distribution.

It can be shown that the event X > a is the same as the event

$$Z > \frac{a-\mu}{\sigma}.$$

This then means that the probabilities are the same:

$$P[X > a] = P\left[Z > \frac{a-\mu}{\sigma}\right].$$

An event involving *X* has been shown to be equivalent to an event involving *Z*, so if the probabilities for *Z* are known, then those for *X* are known too.

2.4.2 Summary of standardisation

If X is a normally distributed variable with mean μ and standard deviation σ, and Z is the normal variable with mean 0 and standard deviation 1, then

$$P[X > a] = P\left[Z > \frac{a-\mu}{\sigma}\right].$$

The normal variable can be reduced to one involving the standard normal variable.

What this means in practice is that to find $P[X > a]$, firstly work out the number:

$$\frac{a-\mu}{\sigma}$$

and then look it up in the table which has been calculated for the standard normal distribution.

It is then not necessary to recalculate the probabilities once they have been done for the standard variable.

2.4.3 Example of a standard normal distribution

The diameter of the ball bearings being produced in a factory is a normally distributed random variable, D with mean 4mm and standard deviation 0.01mm. What is the probability that a ball bearing chosen at random has a diameter greater than 4.015mm?

The question means we are calculating the probability

$$P[\, D > 4.015\text{mm}].$$

To find this probability, firstly work out the standardised number:

$$\frac{4.015-\mu}{\sigma} = \frac{4.015-4.0}{0.01} = 1.5.$$

Use the fact that

$$P[D > 4.015\text{mm}] = P[Z > 1.5]$$

and then look it up in the table.

It gives 0.0668.

So the result is that we now know that:

$$P[\, D > 4.015\text{mm}] = 0.0668.$$

The procedure here meant that we translated an event involving D to one involving Z.

2.4.4 Example

In the same situation, what is the probability that a ball bearing chosen at random has a diameter less than 4.0185 mm?

We are calculating $P[\, D < 4.0185\text{mm}]$. Standardise:

$$\frac{4.0185-\mu}{\sigma}=\frac{4.0185-4.0}{0.01}=1.85\,.$$

This means that

$$P[\, D < 4.0185] = P[\, Z < 1.85].$$

However, if we look this number 1.85 up in the table, the value given is the probability that Z is greater than 1.85:

$$P[\, Z > 1.85] = 0.0322.$$

To deal with this, use the fact that

$$P[\, Z < 1.85] = 1 - P[\, Z > 1.85].$$

Then

$$P[\, Z < 1.85] = 1 - 0.0322 = 0.9678.$$

So we have found that

$$P[\, D < 4.0185] = 0.9678.$$

To summarise these steps,

$$P[\ D < 4.0185] = P[\ Z < 1.85],$$

from standardising, and then

$$P[\ Z < 1.85] = 1 - P[\ Z > 1.85] = 1 - 0.0322 = 0.9678.$$

2.4.5 The Symmetry of the Standard Normal Distribution

The values in the log tables are probabilities $P[Z > a]$, for positive numbers a.

We must be able to extend this to probabilities like

$$P[Z < a]$$

$P[Z > a]$ for negative values of a,

ranges, $P[a < Z < b]$.

The standard normal distribution has some important properties arising from its definition as an integral which make it possible to calculate these probabilities.

To handle problems such as $P[Z < a]$, for some number a, use the fact that $Z < a$ and $Z > a$ cover all eventualities, and so their probabilities add up to 1. It then follows that

$$P[\ Z < a] = 1 - P[\ Z > \mathrm{a}].$$

To handle negative values of a, we use a property of the standard normal variable called *symmetry*. It means that

$$P[Z < -a] = P[Z > a],$$

$$P[Z > -a] = P[Z < a].$$

These two rules can be summarised by saying that to change the direction of the inequality, the sign must be changed also.

This also applies in reverse – to change the sign, the direction of the inequality must be changed too.

2.4.6 Example – symmetry

In the case of ball bearing production, calculate the probability that a ball bearing chosen at random has a diameter greater than 3.985mm.

Calculate:

$$P[\, D > 3.985].$$

When we standardise, we are left with a negative number:

$$\frac{3.985 - 4.0}{0.01} = -1.5 \; .$$

Applying symmetry to this example:

$$P[Z > -1.5] = P[Z < 1.5].$$

Now use the fact that

$$P[Z < 1.5] = 1 - P[Z > 1.5].$$

The last probability can be found in the tables, so:

$$P[D > 3.985\text{mm}] = 1 - 0.0668 = 0.9332.$$

To summarise, first standardise to get:

$$P[D > 3.985] = P[Z > -1.5].$$

Then apply symmetry:

$$P[Z > -1.5] = P[Z < 1.5],$$

and now the basic law:

$$P[Z < 1.5] = 1 - P[Z > 1.5] = 1 - 0.0668 = 0.9332.$$

2.4.7 Example – Heights

The height of men is normally distributed, with mean 1.71 metres and standard deviation 0.11 metres. Find the probability that the height of a man chosen at random is:

- Greater than 1.76 metres.
- Less than 1.74 metres.
- Less than 1.64 metres.
- Greater than 1.54 metres.

Let H be the random variable of height in men. The mean and standard deviation are $\mu = 1.71$m $\sigma = 0.11$m.

The first question here is straightforward. Standardise the value given, 1.76:

$$\frac{1.76-\mu}{\sigma}=\frac{1.76-1.71}{0.11}=0.45.$$

This means that

$$P[\,H > 1.76\text{m}] = P[\,Z > 0.45],$$

and this is then a matter of looking up the tables, to get 0.3264.

In summary

$$P[\,H > 1.76\text{m}] = P[\,Z > 0.45] = 0.3264.$$

For the second part, standardising gives

$$\frac{1.74 - 1.71}{0.11} = 0.27 .$$

It then follows that

$$P[\,H < 1.74\text{m}] = P[\,Z < 0.27] =$$

$$= 1 - P[\,Z > 0.27] =$$

$$= 1 - 0.3936 = 0.6064.$$

For the third part, first standardise to get:

$$\frac{1.64-1.71}{0.11}=-0.6\,,$$

so that $P[H < 1.64\text{m}] = P[Z < -0.64]$.

Then apply symmetry:

$$P[Z < -0.64] = P[Z > 0.64].$$

This is a straightforward case from the tables.

In summary:

$$P[H < 1.64] = P[Z < -0.64] = P[Z > 0.64] = 0.2611.$$

For the last part, first standardise to get:

$$\frac{1.54 - 1.71}{0.11} = -1.54,$$

so that $P[H > 1.54\text{m}] = P[Z > -1.54]$.

Then apply symmetry: $P[Z > -1.54] = P[Z < 1.54]$.

Using the tables:

$$P[Z < 1.54\ ] = 1 - P[Z > 1.54\ ] = 1 - 0.0618 = 0.9382.$$

In summary:

$$P[H > 1.54\text{m}] = P[Z > -1.54\ ] =$$

$$= P[Z < 1.54\ ] = 1 - P[Z > 1.54\ ] =$$

$$= 1 - 0.0618 = 0.9382.$$

2.4.8 Ranges

Recall the example of ball-bearings being produced in a factory – the diameter of the ball bearings is a normally distributed random variable, D with μ = 4mm, and σ = 0.01mm.

Consider the following question – calculate the probability that a ball bearing chosen at random has a diameter between 4.015mm and 4.02mm.

To address this problem, consider the following three possible events; the diameter is greater than 4.015mm, greater than 4.02mm and between the two.

These probabilities are

$$P[D > 4.015\text{mm}],\ P[D > 4.02\text{mm}],\ \text{and}$$

$$P[4.015\text{mm} < D < 4.02\text{mm}].$$

Since 4.015 is the lower number, the event '*D* between 4.015 and 4.02' and '*D* greater than 4.02', can be combined to be '*D* greater than 4.015'.

The three probabilities are connected by the following relation:

$$P[4.015 < D < 4.02] + P[D > 4.02] = P[D > 4.015].$$

Bringing one of the probabilities across the equals sign gives:

$$\mathrm{P}[4.015 < \mathrm{D} < 4.02] = \mathrm{P}[\mathrm{D} > 4.015] - \mathrm{P}[\mathrm{D} > 4.02].$$

Both of the probabilities on the RHS can be calculated in the usual way. Standardising and using the tables gives

$$P[D > 4.015\text{mm}] = P[Z > 1.5] = 0.0668,$$

and

$$P[D > 4.02\text{mm}] = P[Z > 2.0] = 0.0228.$$

The result of the calculation is:

$$P[4.015\text{mm} < D < 4.02\text{mm}] = 0.0668 - 0.0228 = 0.044.$$

The key step here was writing the probability of the range as

$$P[4.015 < D < 4.02] = P[D > 4.015] - P[D > 4.02].$$

2.4.9 Example of ranges

The height of men is normally distributed, with mean 1.71 metres and standard deviation 0.11 metres. Find the probability that the height of a man chosen at random is:

- Between 1.74 metres and 1.76 metres.
- Between 1.64 metres and 1.76 metres.

Let H be the random variable of height in men, with

$$\mu = 1.71\text{m and } \sigma = 0.11\text{m.}$$

For the first question, the probability is broken up as

$$P[1.74\text{m} < H < 1.76\text{m}] = P[H > 1.74\text{m}] - P[H > 1.76\text{m}].$$

Standardise the values 1.76 and 1.74 gives:

$$\frac{1.74\text{m} - 1.71\text{m}}{0.11\text{m}} = 0.27 \text{ and } \frac{1.76\text{m} - 1.71\text{m}}{0.11\text{m}} = 0.45.$$

This means that

$$P[H > 1.74\text{m}] = P[Z > 0.27],\ P[H > 1.76\text{m}] = P[Z > 0.45],$$

This means that

$$P[\, H > 1.74\text{m}] = P[\, Z > 0.27] = 0.3936.$$

and

$$P[\, H > 1.76\text{m}] = P[\, Z > 0.45] = 0.3264,$$

Then the final step is:

$$P[1.74\text{m} < H < 1.76\text{m}] = 0.3936 - 0.3264 = 0.0672.$$

For the second part, the probability is

$$P[1.64\text{m} < H < 1.76\text{m}] = P[H > 1.64\text{m}] - P[H > 1.76\text{m}].$$

Standardise 1.64, in the same way as above, to get:

$$P[H > 1.64\text{m}] = P[Z > -0.64].$$

Then apply symmetry:

$$P[Z > -0.64] = P[Z < 0.64],$$

and the laws of probability

$$P[Z < 0.64] = 1 - P[Z > 0.64] = 1 - 0.2611 = 0.7389.$$

From the first part of the question, we already know that

$$P[H > 1.76\text{m}] = P[Z > 0.45] = 0.3264.$$

So the overall result is

$$\text{P}[1.64\text{m} < H < 1.76\text{m}] = P[H > 1.64\text{m}] – P[H > 1.76\text{m}] =$$

$$= 0.7389 – 0.3264 = 0.4125.$$

2.4.10 Example – sales from a petrol station

The manager of a petrol station finds that the number of litres of petrol sold in a week, *S*, is a normally distributed random variable with mean 2,500L and standard deviation 240L.

Find the probability that they sell between 2,400L and 2,600L of petrol in a week.

The probability is a range, and is split up as

$$P[2{,}400 < S < 2{,}600] = P[S > 2{,}400] - P[S > 2{,}600].$$

Standardising the first, 2,400, gives –100/240 = –0.42. From this it follows that

$$P[S > 2{,}400] = P[Z > -0.42],$$

and applying symmetry gives

$$P[S > 2{,}400] = P[Z > -0.42] = P[Z < 0.42].$$

This is one step away from a probability in the tables:

$$P[Z < 0.42] = 1 - P[Z > 0.42] = 1 - 0.3372 = 0.6628.$$

The second is a straightforward case of standardising:

$$P[S > 2{,}600] = P[Z > 0.42] = 0.3372.$$

The overall result of this calculation is

$$\mathrm{P}[2{,}400 < S < 2{,}600] = P[S > 2{,}400] - P[S > 2{,}600] =$$

$$= 0.6628 - 0.3372 = 0.3256.$$

2.5 Working in Reverse

Consider the following example: A petrol station finds that the number of litres of petrol sold in a week, S, is a normally distributed random variable with mean 2,500L and standard deviation 200L.

Calculate how much petrol they should stock so the probability of running out is 0.02.

This is a new type of question – we have the probability, we need the figure such that the probability that the sales exceed it is 0.02.

The event we are looking at is: find a value a such that

$$P[S > a] = 0.02.$$

The sales of petrol is a normally distributed random variable, S, with mean and standard deviation

$$\mu = 2{,}500\text{L, and } \sigma = 200\text{L.}$$

Now find a value a such that $P[S > a] = 0.02$.

If we find the probability 0.02, within the body of the tables, we see that it came from the value 2.05. Mathematically,

$$P[Z > 2.05] = 0.02.$$

Thus we need the value of petrol sales, which was standardised to give 2.05, and then we will have our figure.

To standardise, we subtract the mean μ (= 2,500L), and divide by the standard deviation σ (= 200L).

To reverse this, we multiply by σ and add μ.

To look at this mathematically: We have, firstly,

$$P[S > a] = 0.02,$$

and then from the tables,

$$P[Z > 2.05] = 0.02.$$

The defining equation is

$$P[S > a] = P\left[Z > \frac{a - \mu}{\sigma}\right].$$

We know that the left-hand side is 0.02, and so comparing the right-hand side with the observation

$$P[Z > 2.05] = 0.02,$$

we can say that

$$\frac{a - \mu}{\sigma} = 2.0 .$$

This can be looked on as an equation to solve for a, yielding

$$a = 2.05\sigma + \mu.$$

Using the values we have,

$$a = 2.05\text{x}200\text{L} + 2{,}500\text{L} = 2{,}910\text{L}.$$

When we had the sales figures, we standardised to get the *Z* values, and now that we have the *Z* values, we reverse to get the sales figures.

2.5.1 Example – Ball Bearings

Recall the diameter of the ball bearings produced in a factory is a normally distributed random variable, D with mean and standard deviation $\mu = 4$mm, and $\sigma = 0.01$mm.

Find a value in mm such that there is only a 5% chance that a randomly chosen ball bearing diameter exceeds it.

Repeat for a 2.5% chance.

For the first question, we are looking for a number a, a diameter, such that

$$P[D > a] = 0.05.$$

If we find the probability 0.05 within the tables, we see it came from the number 1.65:

$$P[Z > 1.65] = 0.05.$$

We can look on this value 1.65 as the result of standardising a.

$$P[D > a] = P[Z > 1.65] = 0.05.$$

So we must ‘reverse standardise’ this figure.

This is simply:

$$a = 1.65 \text{x} 0.01\text{mm} + 4\text{mm} = 4.0165\text{mm}.$$

For the second part, we are looking for a number a, a diameter, such that

$$P[D > a] = 0.025.$$

If we find the probability 0.025 within the tables, we see it came from the number 1.96:

$$P[Z > 1.96] = 0.025.$$

So we must ‘reverse standardise’ this figure. This is simply:

$$a = 1.96 \times 0.01\text{mm} + 4\text{mm} = 4.0196\text{mm}.$$

2.5.2 Example – Heights

The height of men is normally distributed, with $\mu = 1.71$m and $\sigma = 0.11$m. Find heights such that the probability that a man chosen at random is taller is:

- 0.025.
- 0.1.

For the first question, look for a number a, a height, such that

$$P[H > a] = 0.025.$$

From the tables, the probability 0.025 came from the number 1.96:

$$P[Z > 1.96] = 0.025.$$

Reverse standardise this figure:

$$a = 1.96\text{x}0.11\text{m} + 1.71\text{m} = 1.9256\text{m}.$$

For the second part, we are looking for a height a, such that:

$$P[H > a] = 0.1.$$

The probability 0.1 comes from the number 1.28:

$$P[Z > 1.28] = 0.1.$$

Reverse standardise this figure:

$$a = 1.28\text{x}0.11\text{m} + 1.71\text{m} = 1.8508\text{m}.$$

So we can say that

$$P[H > 1.8508\text{m}] = 0.1.$$

Consider the following, slightly different case:

2.5.3 Example – Heights

For the example of height, find heights such that the probability that a man chosen at random is not as tall is

(i) 0.9,

(ii) 0.1.

Find a height such that the probability that a man chosen at random is taller is 0.95.

Let H be the random variable of height in men, with

$$\mu = 1.71\text{m and } \sigma = 0.11\text{m}.$$

For the first question, we are looking for a height a such that

$$P[H < a] = 0.9.$$

If we try and find the probability 0.9 in the tables, it is not there – the way to deal with this is simply change the question. Equivalently, we are looking for a height a such that

$$P[H > a] = 0.1.$$

This is a straightforward case:

$$P[Z > 1.28] = 0.1.$$

Reverse standardise this figure:

$$a = 1.28\text{x}0.11\text{m} + 1.71\text{mm} = 1.8508\text{m}.$$

For the second part, we are looking for a height a, such that:

$$P[H < a] = 0.1.$$

The probability 0.1 came from the number 1.28:

$$P[Z > 1.28] = 0.1.$$

The inequality signs go the wrong way. Use symmetry: if

$$P[Z > 1.28] = 0.1, \text{ then } P[Z < -1.28] = 0.1.$$

We can now just reverse standardise this figure:

$$a = -1.28\text{x}0.11\text{m} + 1.71\text{m} = 1.5692\text{m}.$$

Thus

$$P[\,H < 1.5692\text{m}] = 0.1.$$

For the third question, we are looking for a height a such that

$$P[\, H > a] = 0.95.$$

We must switch around to get probabilities less than 0.5:

$$P[\, H < a] = 0.05.$$

If we find the probability 0.05 within the tables, we see it came from the number 1.65. But the inequality goes the wrong way:

$$P[\, Z > 1.65] = 0.05.$$

Using symmetry

$$P[\,Z < -1.65] = 0.05.$$

‘Reverse standardise’ this figure:

$$a = -1.65 \text{x} 0.11\text{m} + 1.71\text{m} = 1.5285\text{m}.$$

2.5.4 Example – Middle Ranges

For the case of height in men, find the middle 90% of heights.

This means two values a and b such that $P[a < H < b] = 0.9$, but because it is the *middle* 90%, the probabilities of H being above or below the range are the same. This gives us a way of tackling the calculation, since we have two probabilities:

$$P[H < a] = 0.05 \text{ and } P[H > b] = 0.05.$$

The probabilities of these three possibilities add up to 1.

(The 0.05 comes from $0.05 = \frac{1}{2}(1 - 0.9)$.)

These cases are just like the problems above, and both involve the same z-value from the tables, except for a difference in sign.
We have done the second question already: $P[H > b] = 0.05$.
The value from the tables is 1.65, and this gives

$$b = 1.65 \text{x} 0.11\text{m} + 1.71\text{m} = 1.8915\text{m}.$$

The second part is done above, and involved the value –1.65; the calculation was

$$a = -1.65 \text{x} 0.11\text{m} + 1.71\text{m} = 1.5285\text{m}.$$

So once we have the *z*-value for one part, we have it for the other, with a change of sign.