

Introduction to Statistics - Young Researchers Fellowship Program

Lecture 3 - Introduction to Probability, Part 2

Daniel Sánchez Pazmiño

Table of contents

Probability distributions	1
Random variables	2
Probability distribution of an RV	2
Discrete probability distributions and the PDF	3
Cumulative distribution function	4
Expected value and variance of a probability distribution	5
Continuous probability distributions	6
The CDF of a continuous probability distribution	6
The normal distribution	7
CDF of the normal distribution	9
Standard normal distribution	10
The standard normal distribution table	11
The empirical rule or the 68-95-99.7 rule	12
Example exercises using the normal distribution and probability tables	13
Summary	18

Probability distributions

As mentioned before, we may want to understand a population through statistical inference based on a sample. We do not observe the population, so the sample is our best guess at how the population looks like. However, we do not know how good our guess is. We do not know how much the sample differs from the population. We do not know how much the sample differs from other samples. We do not know how much the sample differs from the population.

Because we cannot possibly know for sure, we need to make an *educated guess* about how we think the population looks like, with a *theoretical model* about how reality looks like. This theoretical model is knowledge about how we think the population looks like, which is called a *probability distribution*.

We can then use the sample to test our model, and see how well it fits the data. If the model fits the data well, we can be more confident that our model is a good representation of reality, and thus we will draw certain types of conclusions from the sample that we have. If the model does not fit the data well, we can be less confident that our model is a good representation of reality, and thus be a little bit more weary about drawing conclusions from the sample that we have.

Random variables

A random variable is a numerical description about the outcome of an “experiment”. In our case, the experiment is the process of sampling. So, for instance, the sample mean of a sample is a random variable since it is a numerical description about the outcome of the sampling process. Random variables (**RVs**) are random because they are subject to randomness. If we draw a sample many times, most likely the sample mean will be different each time.

There are two types of random variables, which are *discrete* and *continuous*. A discrete random variable is a random variable that can take on a finite number of values. For instance, the number of heads that we get when we flip a coin 10 times is a discrete random variable, since it can take on the values of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10. We don’t see decimals in discrete random variables.

A continuous random variable is a random variable that can take on an infinite number of values. For instance, the height of a person is a continuous random variable, since it can take on any value between 0 and infinity.

Probability distribution of an RV

We know that a random variable is subject to randomness, however, it being random doesn’t mean that it is not *predictable*: in statistics random does not mean unpredictable, simply, random denotes that something need not to be the exact same every time.

For instance, if we define an experiment which flips a coin two times, we know that every time the experiment is run, we might get different amounts of heads every time. However, we know that we can get at most 2 heads, and at least 0. And we also know that the probability of having each head is 0.5.

We may calculate the probability of getting 0 heads, 1 head and 2 heads by using the rules of probability. This table is called the *probability distribution* of the random variable of the number of heads that we get when we flip a coin two times. Because we can only get 0, 1 or

2 heads, the number of heads is a discrete number, and thus it follows a *discrete probability distribution*.

Discrete probability distributions and the PDF

When working with either discrete or continuous probability distributions, we need to know the difference between the probability density function for the random variable, and the cumulative density function of a random variable. Here, we explore the discrete case.

A probability distribution function (PDF) is a **table** that shows the probability of each outcome of an experiment. The sum of the probabilities of all possible outcomes is equal to 1. People often refer to the PDF as $f(x)$, where x is the outcome of the experiment and f is the associated probability in the table. This is weird notation, but it exists because of continuous probability distributions, which actually need equations for their PDFs. We will talk about those later.

The two coin tosses experiment is an example of a discrete probability distribution, because the number of heads that we can get is discrete, not continuous. The experiment has either 0 heads, 1 head or 2 heads. We cannot get 0.5 heads, or 1.5 heads. We can develop the PDF ourselves by using the basic rules of probability.

Number of heads	Probability
0	0.25
1	0.5
2	0.25

First of all, we must remember that getting one head (or zero heads) in the first stage of the experiment is fully independent of getting one head (or zero heads) in the second stage of the experiment. Thus, we can compute the probability of getting 0 heads in the first stage and 0 heads in the second stage by multiplying. We then define $f(x)$ as the probability of getting x heads in the experiment. The table goes below:

How to calculate the probability of getting 0 heads? We know that the probability of getting 0 heads in the first stage is 0.5. We also know that the probability of getting 0 heads in the second stage is 0.5. Thus, the probability of getting 0 heads in the first stage and 0 heads in the second stage is $0.5 * 0.5 = 0.25$.

For 1 head, we know that there are two ways of getting 1 head: either we get 1 head in the first stage and 0 heads in the second stage, or we get 0 heads in the first stage and 1 head in the second stage. We know that the probability of getting 1 head in the first stage is 0.5, and the probability of getting 0 heads in the second stage is 0.5. Thus, the probability of getting 1 head in the first stage and 0 heads in the second stage is $0.5 * 0.5 = 0.25$. We also know that the probability of getting 0 heads in the first stage is 0.5, and the probability of getting 1

head in the second stage is 0.5. Thus, the probability of getting 0 heads in the first stage and 1 head in the second stage is $0.5 * 0.5 = 0.25$. Thus, the probability of getting 1 head is $0.25 + 0.25 = 0.5$.

As you can see, all of the probabilities sum up to 1 and each individual probability is between 0 and 1. These, in fact, are the rules of all probability distributions.

Cumulative distribution function

The PDF is a table that shows the probability of each outcome of an experiment. However, we can also calculate the probability of an outcome falling between two values. For instance, we might want to know what is the probability of acquiring 0 to 1 heads. This, in fact, is the union of the event of getting 0 heads and the event of getting 1 head. Thus, we can calculate the probability of getting 0 to 1 heads by adding the probability of getting 0 heads and the probability of getting 1 head. Because we can either get 0 heads or 1 head, but not 0 and 1 and the same time (mutual exclusiveness), we can simply sum the probabilities.

A table that gives you the probabilities for intervals of the random variable is called the *cumulative distribution function* (**CDF**). The CDF is a table that shows the probability of the random variable falling between two values. The CDF is denoted as $F(x)$, (uppercase F) where x is the value of the random variable and F is the associated cumulative probability from 0 to x . Below, we show the CDF for the number of heads that we get when we flip a coin two times.

Number of heads	Cumulative Probability
0	0.25
1	0.75
2	1

You might remember the frequency table and the cumulative frequency tables from before. The probability density function $f(x)$ and the cumulative distribution function $F(x)$ mirror the frequency and cumulative frequency tables. The only difference is that the probability density function and the cumulative distribution function are for random variables, while the frequency and cumulative frequency tables are for data, which need not be random if they come from the population.

Let us interpret this table. The first entry in the table is similar to the probability density function, which means that the probability of getting heads is 0.25. The second entry in the table is the probability of getting 0 to 1 heads, which is $0.25 + 0.5 = 0.75$. Finally, the last entry is the probability of getting 0 to 2 heads. Since those are all of the possible results of the experiment, the probability of getting 0 to 2 heads is 1 (100%), but that is also $0.75 + 0.25 = 1$.

You may think: what is the probability of getting 1 to 2 heads? You can calculate directly through the union $P(H = 1 \cup H = 2) = 0.5 + 0.25 = 0.75$, but we should also be able to calculate it by using the CDF table. If we know that the probability of getting 0 to 2 heads is 1, and the probability of getting 0 heads is 0.25, we get the probability of getting 1 to 2 heads by subtracting the probability of getting 0 heads from the probability of getting 0 to 2 heads, which is $1 - 0.25 = 0.75$.

This is a general rule: if we want to know the probability of an RV falling between two values, we can calculate it by subtracting the probability of the lower value from the probability of the higher value. This applies to both discrete and continuous probability distributions, which we will cover later.

Expected value and variance of a probability distribution

Having understood what a probability distribution is, we can calculate some summary statistics of the probability distribution. The most common summary statistics are the *expected value* and the *variance*.

The expected value measures the center of the probability distribution. It is like the average of the probability distribution, and thus, it is a measure of a likely outcome of the RV that follows such probability distribution. It is computed as follows:

$$E(X) = \sum_{i=1}^n x_i f(x_i)$$

This formula means multiplying each value of the random variable (x_i) by its probability ($f(x)$), and then summing all of the values. For instance, if we want to know the expected value of the number of heads that we get when we flip a coin two times, we can compute it as follows:

$$E(X) = 0 * 0.25 + 1 * 0.5 + 2 * 0.25 = 1$$

The variance measures the spread of the probability distribution. It is computed as follows:

$$Var(X) = \sum_{i=1}^n (x_i - E(X))^2 f(x_i)$$

For the variance, we first need to compute the expected value. Then, we need to subtract the expected value from each value of the random variable, square the result, multiply it by the probability of that value, and then sum all of the values. For instance, if we want to know the variance of the number of heads that we get when we flip a coin two times, we can compute it as follows:

$$\text{Var}(X) = (0 - 1)^2 * 0.25 + (1 - 1)^2 * 0.5 + (2 - 1)^2 * 0.25 = 0.5$$

We can also calculate a standard deviation, which is the square root of the variance of the probability distribution. The standard deviation is a measure of spread, just like the variance.

Continuous probability distributions

When the RV is continuous, we cannot use the same approach as before. There is an added complication: how can we distribute the probability of an infinite number of values? A person can be 1.7 meters tall, 1.75, 1.7000001, and so on infinitely. So we cannot just take 100% of cases and distribute them among the possible values, because there are an infinite number of possible values!

We'll, the first principle of continuous probability distributions is that the probability of any single value is 0. This is because there are an infinite number of possible values, so we believe that the probability of any single, very specific value is 0. You might think that is impossible, but it is not, because in reality, while you might be told that you measure 1.60 meters, that is simply an estimation. It is an estimation because we cannot be possibly sure if you measure 1.600000000000 or 1.600000000001, as we will *never* have the measurement power to be so precise. So, as we cannot be sure of any single value, we assign a probability of 0 to any single value.

This might lead you to think that working with continuous probability distributions is useless, but it is not. What happens is that working with the PDF of a distribution is not useful, as it will always give us 0. However, we can work with the CDF of a distribution, which will give us the probability of an RV falling between two values. This is useful, because we can calculate the probability of an RV falling between two values, which is what we want to do in the first place.

The CDF of a continuous probability distribution

The PDF for a continuous variable is a function (a fancy word for an equation) that shows the probability of each value of the RV. For example, the PDF of the height of a person is a function that shows the probability of each height. The PDF is denoted as $f(x)$, where x is the value of the RV and f is the associated probability.

However, as you know, the PDF of a continuous variable will always spit out zero, because there are an infinite number of possible values. So, what is the point of the PDF? The point of the PDF is that we can build the CDF from it, which gives us more useful information: what is the probability that a person gives a height between 1.70 and 1.80 meters?

The technical details are above the scope of this course, but know that the area under the PDF will get us the probability of an RV falling between a certain *interval*. So, the CDF is a function or equation that will give a probability of an RV falling between two values. The CDF is denoted as $F(x)$, where x is the value of the RV and F is the associated cumulative probability from the interval of 0 to x .

When we say that what is the probability that we get a value between a given interval, what do we mean? For instance, we might have a list of students that come from a university. We believe that the way that this population of students in the university behaves according to a specific continuous probability distribution (meaning, a specific equation). From that information, we can get the probability of getting a student with a grade between 10 to 20 points.

The normal distribution

There are many different continuous probability distributions, and each one of them has different properties. We use different distributions for different purposes, and we need to be careful about which distribution we use. The normal distribution is the most common distribution, and it is the one that we will use the most. That is because of two reasons:

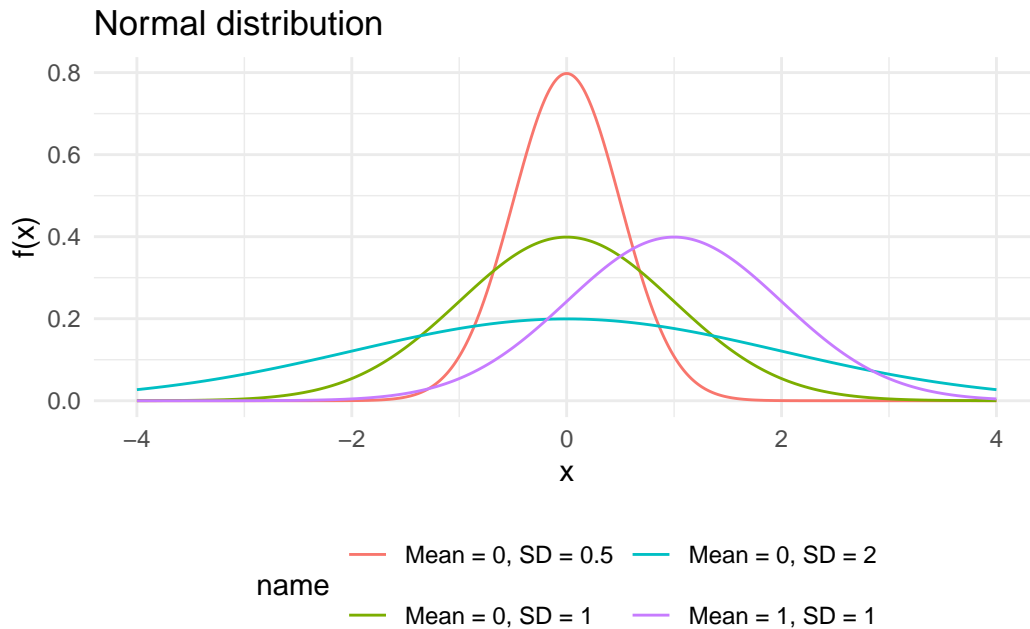
1. The normal distribution is a very flexible and easy to handle distribution, unlike other distributions which are more difficult and require a lot of mathematical knowledge to handle.
2. The normal distribution is a very common distribution in real life (something that we've confirmed using data), which means that many variables are distributed normally. So, it is not unrealistic to assume that a variable is distributed normally.

The normal distribution is a symmetric, bell-shaped distribution that is completely determined by two parameters: the mean and the standard deviation. The mean determines the center of the distribution, and the standard deviation determines the spread of the distribution. Everything other than that is given by the shape of the distribution, which is defined by the PDF of the distribution. This, as mentioned before, is a formula, which is the following:

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

The formula looks pretty scary. Thankfully, other people have done the mathematical hard work for us, so we don't need to worry about too much it. Notice how the equation uses the information about the mean and standard deviation of the distribution. The mean is denoted as μ (the Greek letter mu), and the standard deviation is denoted as σ (the Greek letter sigma).

Below, we graph the normal distribution for different values of the mean and the standard deviation. As you can see, the mean determines the center of the distribution, and the standard deviation determines the spread of the distribution.



It is important to notice that what this graph is doing is plotting a number which is the “density”, something like the frequency of the value. To get the probability, we would divide it by the “total number of cases”, but since in continuous distributions the “total” is infinity, we always would get 0.

Notice that the different curves are all normal distributions but with different values of the mean and the standard deviation. The mean determines the center of the distribution, and the standard deviation determines the spread of the distribution. However, all of the distributions are symmetric and bell-shaped.

There are some important things to remember about the normal distribution:

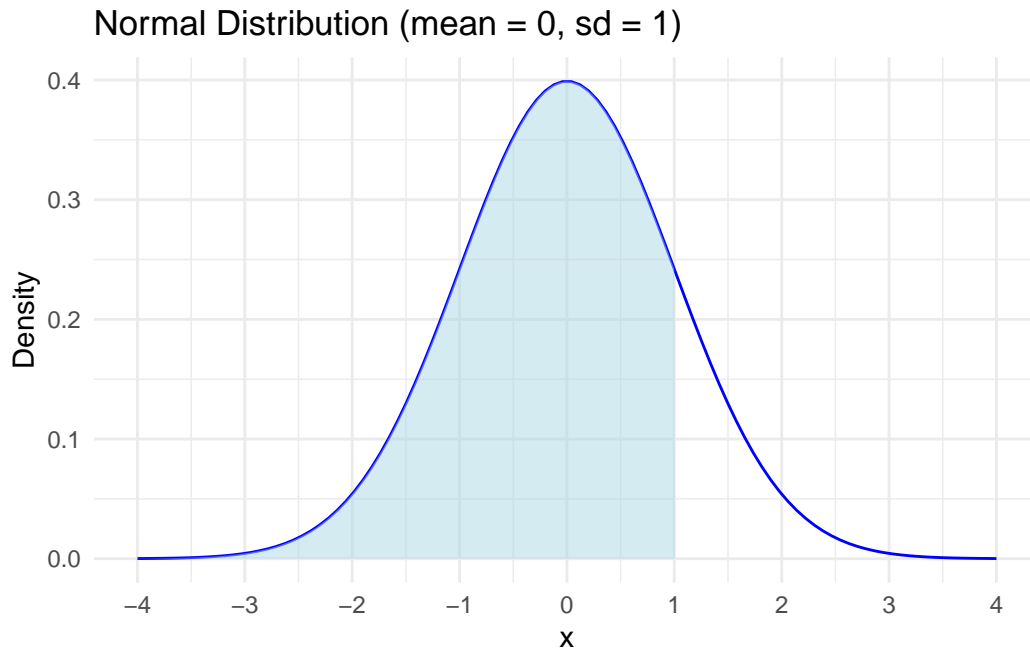
1. The mean, equal to median and mode is the highest point of the distribution. This is because the distribution is symmetric, so the mean, median and mode are all in the center of the distribution.
2. The mean can be any value, zero or negative. We’re not limited to positive values.
3. The standard deviation determines the spread of the distribution. The higher the standard deviation, the more spread out the distribution is. The lower the standard deviation, the less spread out the distribution is.
4. The standard deviation can only be positive. This is because the standard deviation is a measure of spread, so it cannot be negative.

5. The area under the curve is equal to 1. This is because the area under the curve is the probability of the RV falling between negative infinity and positive infinity, which is 1.
6. The tails of the distribution are asymptotic to the x-axis. This means that the tails of the distribution get closer and closer to the x-axis, but they never touch it. This is because the area under the curve is equal to 1, so the tails need to get closer and closer to the x-axis in order to keep the area under the curve equal to 1.
7. Since the distribution is symmetric, the probability left to the mean is equal to the probability right to the mean. This means that the probability of the RV falling between the mean and positive infinity is equal to the probability of the RV falling between the mean and negative infinity. This is because the area under the curve is equal to 1, so the area under the curve to the left of the mean needs to be equal to the area under the curve to the right of the mean, which is 0.5.

CDF of the normal distribution

In order to get the CDF of the distribution, which gives us the probability that the RV falls within an interval, we need to calculate the area below the curve that the PDF plots. This is a little complicated as it involves integral calculus but once again, thankfully other people have done the hard work for us. Computers can calculate the probability of an RV which is distributed normally by using the CDF of the normal distribution and the mean and standard deviation of the distribution, which we need to give.

Graphically, the CDF gives you the area under the curve of the PDF. The CDF is a function or equation that will give a probability of an RV falling between two values. This is how it looks like:



Using a computer, we calculate the area that is drawn to the left. This would be 0.8413447. This is the probability of the RV falling between negative infinity and 1. In interpretation, we would say that the probability of the random variable being less than 1 is 0.84.

Standard normal distribution

Up until this point, we know that we can acquire the probability of any given value of a normal distribution by using the mean and the standard deviation and a computer to calculate the area. However, it turns out that there is a little trick that we can make to make our lives easier, and not have to depend on a computer calculate the areas under the curve of infinitely many normal distributions.

Consider the case of a special normal distribution, that which has a mean of 0 and a standard deviation of 1. A normal distribution with a mean of 0 and a std. deviation of one is called the **standard normal distribution**. The graph above already showed the PDF for that distribution, and we calculated the area under the curve to the left of $x = 1$. We can also calculate the area under the curve to the left of 2, and the area under the curve to the left of 3. By using a computer, we can calculate the area under the curve to the left of any value. Further, remember that the normal distribution is symmetric, so the area under the curve to the right of any x is equal to the area under the curve to the left of $-x$, and viceversa. For example, the area under the curve to the left of 1, which is 0.84, is equal to the area to the right of -1, which is also 0.84.

Now, remember z -scores? z scores are the number of standard deviations away from the mean. The z -score of the mean is 0, which means that for *any* distribution, the mean will always have a z -score of 0. If we wanted to know what is the probability of a value being less or equal to the mean in any normal distribution we have two choices: we can either use the mean and the standard deviation of the distribution with a computer, or we can simply calculate the area under the curve to the left of 0, which is 0.5. We interpret this by saying that the probability of the random variable being less than or equal to the mean is 0.5. Note how this is always true regardless of what the actual value of the mean is.

We can repeat the same process with other values of the distribution. We know that a value which is one standard deviation above the mean will have a z -score of 1. So, what would the probability of a value being less than or equal to one standard deviation above the mean? We can either use the mean and the standard deviation of the distribution with a computer, or we can simply calculate the area under the curve to the left of 1, which is 0.84. We interpret this by saying that the probability of the random variable being less than or equal to one standard deviation above the mean is 0.84. Note how this is always true regardless of what the actual value of the mean is.

All values in a distribution have a z -score, which will be equivalent to the number of standard deviations from the mean. So, we can calculate what is the probability of a value being z distributions away from the mean through the probabilities of the standard normal distribution. This result is very powerful, as it allows us to use only one formula, which produces one graph and table with probabilities, and then use that graph and table to calculate the probabilities of any normal distribution. In exams where you are not allowed to use a computer, this is a very useful trick, because you only need to print out one table, which is the table that gives you the probabilities of getting a value less than or equal to z standard deviations away from the mean.

The standard normal distribution table

The standard normal distribution table is a table that gives you the probabilities for a random variable which falls on a given interval. An example table like that is given below.

Table 3: An extract of the standard normal distribution table

No. of std. deviations	Probability to the left	Probability to the right
-1.0	0.1587	0.8413
-0.9	0.1841	0.8159
-0.8	0.2119	0.7881
-0.7	0.2420	0.7580
-0.6	0.2743	0.7257
-0.5	0.3085	0.6915
-0.4	0.3446	0.6554

No. of std. deviations	Probability to the left	Probability to the right
-0.3	0.3821	0.6179
-0.2	0.4207	0.5793
-0.1	0.4602	0.5398
0.0	0.5000	0.5000
0.1	0.5398	0.4602
0.2	0.5793	0.4207
0.3	0.6179	0.3821
0.4	0.6554	0.3446
0.5	0.6915	0.3085
0.6	0.7257	0.2743
0.7	0.7580	0.2420
0.8	0.7881	0.2119
0.9	0.8159	0.1841
1.0	0.8413	0.1587

The table above is an extract of the complete standard normal distribution table. The first column gives you the number of standard deviations. The first table gives you the probability of getting a value less than or equal to the corresponding standard deviations away from the mean. The second column gives you the probability of getting a value greater than z standard deviations away from the mean.

In the Field textbook, the table looks much like this one. However, other textbooks might have a different table. It is common to *only* present the “left-tail” probabilities because the right-tail probabilities are simply the complement of the left-tail probabilities (subtract them from 1). For instance, if the probability of getting a value less than or equal to 1 standard deviation away from the mean is 0.84, the probability of getting a value greater than 1 standard deviation away from the mean is $1 - 0.84 = 0.16$.

The empirical rule or the 68-95-99.7 rule

The empirical rule is a rule that tells you how much of the data falls within a given number of standard deviations away from the mean. We’ve talked about it already, but here we finalise its discussion by saying that the empirical rule is a direct result from looking at the probability of getting a value within a given number of standard deviations away from the mean.

So, for any normal distribution, by using the standard normal distribution, we know that

1. Getting an observation within 1 standard deviation of the mean has a probability of approximately 0.68 (68% of the data is approximately within one std. deviation of the mean).

2. Getting an observation within 2 standard deviations of the mean has a probability of approximately 0.95 (95% of the data is approximately within two std. deviations of the mean).
3. Getting an observation within 3 standard deviations of the mean has a probability of approximately 0.997 (99.7% of the data is approximately within three std. deviations of the mean).

We will see how to get the numbers from this rule in the exercises below.

Example exercises using the normal distribution and probability tables

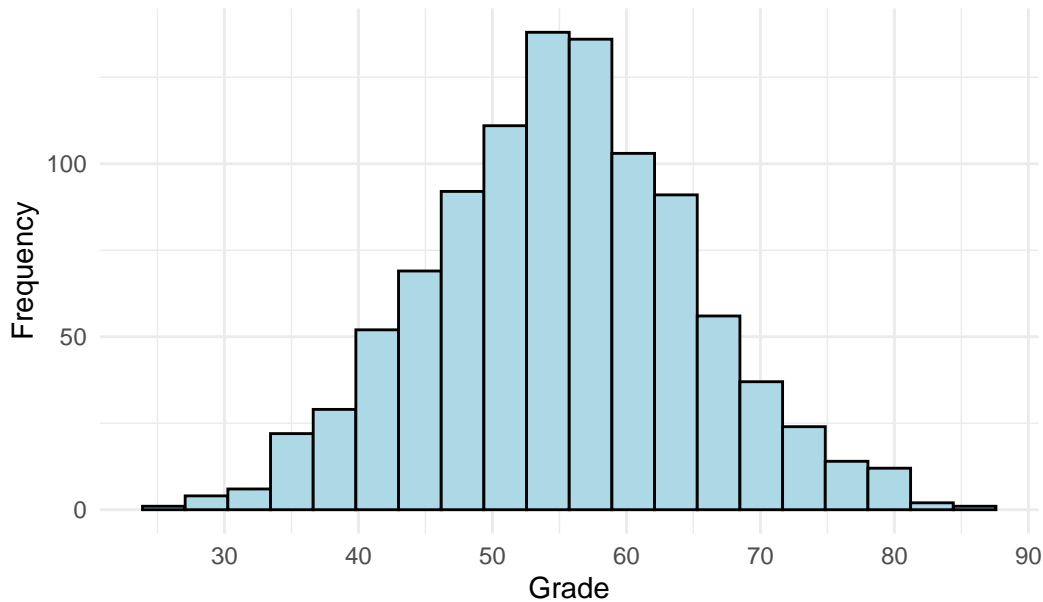
Here, I include a sample of exercises that require you to use a standard normal distribution table. I have faked some data for a probability experiment where you draw a student from a population of $N = 1000$. We can get their grades on a scale of 1 - 100 points. The mean grade is 55 points, and the standard deviation is 10 points. This means that $\mu_{grades} = 55$ and $\sigma_{grades} = 10$. Remember that this information is important to be able to calculate the z -scores for all of the observations. You can see an extract of the data below.

Table 4: An extract of the dataset of students

student	grade
1	49.40
2	52.70
3	70.59
4	55.71
5	56.29
6	72.15
7	59.61
8	42.35
9	48.13
10	50.54

Below, I also graph the complete dataset of students. As you can see, the distribution is normal, and the mean is 55 and the standard deviation is 10. More people have grades around the mean, and less people have grades that are further away from the mean.

Distribution of grades of students (20 bins)



Below, some exercises:

1. What is the probability of getting a student with 55 points?

We can quickly answer by saying the probability is exactly zero. This is because the probability of any single value is zero, because there are an infinite number of possible values.

2. What is the probability of getting a student with 55 points or less?

We can calculate the probability of getting a student with 55 points or less by using the mean and the standard deviation of the distribution with a computer, or use the standard deviation and the mean to calculate the z -score of 55 in this example, and then use the calculated probabilities of a standard normal table.

You should already know that the z -score of an observation equal to the mean is 0, but we can do it here for your reference.

$$z = \frac{x - \mu}{\sigma} = \frac{55 - 55}{10} = 0$$

The z -score of 55 is 0. We can then use the standard normal distribution table to calculate the probability of getting a value less than or equal to 0 standard deviations away from the mean. The probability of getting a value less than or equal to 0 standard deviations away from the mean is 0.5. We interpret this by saying that the probability of getting a student with 55 points or less is 0.5.

3. What is the probability of getting a student with 55 points or more?

Once again, we should use the standard normal distribution table to calculate the probability of getting a value greater than or equal to 0 standard deviations away from the mean. The fact that we're looking at a *greater than* probability makes us look at the *right-tail* probabilities, or the table column that gives us the probability of getting a value greater than or equal to z standard deviations away from the mean (second column in Table 3).

The probability of getting a value greater than or equal to 0 standard deviations away from the mean is 0.5. We interpret this by saying that the probability of getting a student with 55 points or more is 0.5.

Notice that the probability of getting a student with 55 points or less is the same as the probability of getting a student with 55 points or more. This is because the distribution is symmetric, so the probability of getting a value less than or equal to 0 standard deviations away from the mean is equal to the probability of getting a value greater than or equal to 0 standard deviations away from the mean.

Further, know that the probability of getting a student with a grade **over** 55 is equal to the complement of the probability of getting a student with a grade **less** than or equal to 55. This is because the probability of getting a student with a grade over 55 is equal to the probability of getting a student with a grade greater than or equal to 55, which is equal to 0.5. The complement of 0.5 is $1 - 0.5 = 0.5$, which is the probability of getting a student with a grade less than or equal to 55.

4. What is the probability of getting a student with a grade of 60 points or less?

Calculate the z -score of 60:

$$z = \frac{x - \mu}{\sigma} = \frac{60 - 55}{10} = 0.5$$

The z -score of 60 is 0.5. We can then use the standard normal distribution table to calculate the probability of getting a value less than or equal to 0.5 standard deviations away from the mean. The probability of getting a value less than or equal to 0.5 standard deviations away from the mean is 0.6915. We interpret this by saying that the probability of getting a student with 60 points or less is 0.6915.

5. What is the probability of getting a student with a grade of 70 points or more?

Calculate the z -score of 70:

$$z = \frac{x - \mu}{\sigma} = \frac{70 - 55}{10} = 1.5$$

The z -score of 70 is 1.5. We can then use the standard normal distribution table to calculate the probability of getting a value greater than or equal to 1.5 standard deviations away from the mean. The probability of getting a value greater than or equal to 1.5 standard deviations away from the mean is 0.0668. We interpret this by saying that the probability of getting a student with 70 points or more is 0.0668.

6. What is the probability of NOT getting a student with a grade of 60 points or less?

We already know that the probability of getting a student with a grade of 60 points or less is 0.6915. The probability of NOT getting a student with a grade of 60 points or less is the complement of 0.6915, which is $1 - 0.6915 = 0.3085$. We interpret this by saying that the probability of NOT getting a student with a grade of 60 points or less is 0.3085.

7. What is the probability of NOT getting a student with a grade of 70 points or more?

Proceed normally as in exercise 6. The probability of getting a student with a grade of 70 points or more is 0.0668. The probability of NOT getting a student with a grade of 70 points or more is the complement of 0.0668, which is $1 - 0.0668 = 0.9332$. We interpret this by saying that the probability of NOT getting a student with a grade of 70 points or more is 0.9332.

8. What is the probability of getting a student with a grade between 45 and 65 points?

No table gives you probabilities for intervals because there are far too many intervals. However, there is an easy way to calculate probabilities for intervals when we have a table that only gives us the probability of getting a value less or equal to a given number of standard deviations away from the mean.

First, calculate the z -score of both endpoints of the interval. For the lower endpoint, we have:

$$z = \frac{x - \mu}{\sigma} = \frac{45 - 55}{10} = -1$$

For the upper endpoint, we have:

$$z = \frac{x - \mu}{\sigma} = \frac{65 - 55}{10} = 1$$

Then, we use the standard normal distribution table to calculate the probabilities of getting a value less than or equal to the z -score of the lower endpoint and the z -score of the upper endpoint. The probability of getting a value less than or equal to -1 standard deviations away from the mean is 0.1587. The probability of getting a value less than or equal to 1 standard deviations away from the mean is 0.8413.

Finally, we make a subtraction using these probabilities. Subtract the lower endpoint probability from the upper endpoint probability. The probability of getting a value between -1

and 1 standard deviations away from the mean is $0.8413 - 0.1587 = 0.6826$. We interpret this by saying that the probability of getting a student with a grade between 45 and 65 points is 0.6826.

This number might remind you of something: the empirical rule!!! The empirical rule tells us that approximately 68% of the data falls within one standard deviation away from the mean. This is exactly what we got here! This is because the empirical rule is a direct result of the probabilities of the standard normal distribution, and because 45 and 65 are one standard deviation away from the mean.

10. What is the probability of NOT getting a student with a grade between 45 and 55 points?

We already know that the probability of getting a student with a grade within 45 and 55 points is 0.6826. The probability of NOT getting a student with a grade within 45 and 55 points is the complement of 0.6826, which is $1 - 0.6826 = 0.3174$. We interpret this by saying that the probability of NOT getting a student with a grade between 45 and 55 points is 0.3174.

This could also be done using the union of two events which are mutually exclusive. NOT getting a grade within one std. deviation of the mean is equal to **either** getting a grade less than 45 points **or** getting a grade greater than 55 points. Whenever the *either* and *or* words are involved, we know that we're talking about the union of two events. Because we know that a student either gets less than 45, or more than 55, but not both at the same time, we know that these two events are mutually exclusive.

The probability of getting a grade less than 45 points is 0.1587. The probability of getting a grade greater than 55 points is 0.1587. The probability of NOT getting a grade within one std. deviation of the mean is $0.1587 + 0.1587 = 0.3174$.

11. What is the probability of getting three students with a grade within 45 and 55 points?

This is a little bit more complicated because now we are not pulling one but three students. However, we already know that the probability of getting a student with a grade within 45 and 55 points is 0.6826. Further, the act of pulling out one student with a grade within that interval does not affect the probability of pulling out another student with a grade within that interval. This is because drawing one student is **independent** from the other students, so the probability of getting a student with a grade within 45 and 55 points is the same for all three students. We use the multiplication rule to get the intersection of three independent events:

$$P(A \cap B \cap C) = P(A) * P(B) * P(C)$$

So, if pulling out the first student is A, pulling out the second student is B, and pulling out the third student is C, we can calculate the probability of getting three students with a grade within 45 and 55 points as follows:

$$P(A \cap B \cap C) = P(A) * P(B) * P(C) = 0.6826 * 0.6826 * 0.6826 = 0.318$$

We interpret this by saying that the probability of getting three students with a grade within 45 and 55 points is 0.318.

12. What is the probability of not getting three students with a grade within 45 and 55 points?

We proceed as before, but now using the probability that we calculated in exercise 10, which is 0.3174. We use the multiplication rule to get the intersection of three independent events:

$$P'(A \cap B \cap C) = P(A) * P(B) * P(C) = 0.3174 * 0.3174 * 0.3174 = 0.32$$

We interpret this by saying that the probability of not getting three students with a grade within 45 and 55 points is 0.32.

Summary

- A random variable is a variable that can take on different values, and the value that it takes on is random.
- A discrete probability distribution is a probability distribution where the random variable can only take on a finite number of values.
- A continuous probability distribution is a probability distribution where the random variable can take on an infinite number of values.
- The expected value of a discrete probability distribution is the average of the probability distribution.
- The variance of a discrete probability distribution is a measure of spread of the probability distribution.
- The standard deviation of a discrete probability distribution is the square root of the variance of the probability distribution.
- The PDF of a continuous probability distribution is a function that shows the probability of each value of the random variable.
- The CDF of a continuous probability distribution is a function that shows the probability of the random variable falling between two values.
- The normal distribution is a symmetric, bell-shaped distribution that is completely determined by two parameters: the mean and the standard deviation.
- The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1.
- The standard normal distribution table is a table that gives you the probabilities for a random variable which falls on a given interval.

- Using the standard normal distribution table, we can calculate the probability of a value falling within a given number of standard deviations away from the mean.
- The empirical rule is a rule that tells you how much of the data falls within a given number of standard deviations away from the mean.
- The empirical rule is a direct result of the probabilities of the standard normal distribution.
- We can calculate the probability of a value falling within a given interval by calculating the lower tail probabilities of the lower endpoint and the upper endpoint, and then subtracting the lower endpoint probability from the upper endpoint probability.