3.6 Computations with the Normal Distribution

In this final section, we look at how we compute probabilities of normally distributed variables. We recall the pdf of the normal distribution, if $X \sim \mathrm{N}(\mu, \sigma)$ then we have that the pdf of X is:

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

Unlike previous distributions we have considered, even with knowledge of integration, computing an expression for the integral of the pdf is not possible. This leads us to a problem when we wish to compute:

$$\mathbb{P}(X \leq a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

3.6.1 The Standard Normal Distribution

In order to approach the problem of finding probabilities for the normal distribution, we first look at a simpler case. In this section $Z \sim \mathrm{N}(0, 1)$ will always denote the standard normal random variable. As we have previously mentioned, the pdf of this variable can not be easily integrated. However techniques have been employed so that we know the value of this integral for the standard normal variable. We define the function Φ, read (Phi) as:

$$\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \mathrm{d}x$$

We note that $\Phi(a) = \mathbb{P}(X \leq a)$. Mathematicians have computed the value of $\Phi(a)$ for a wide range of values for a, these are usually compiled within statistical tables. As a geometrical interpretation, the blue region below indicates $\Phi(1) = \mathbb{P}(Z \leq 1)$:

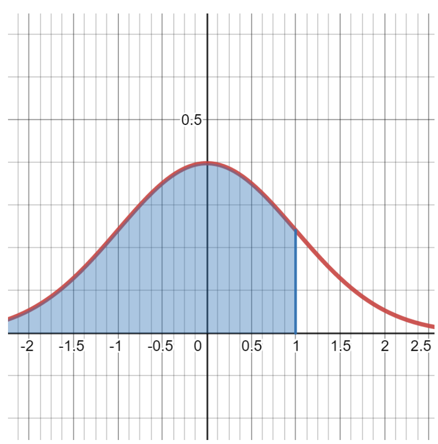


Figure 3.7: The pdf of Z in red, and the region described by $\Phi(1)$ in blue.

From statistical tables we know that $\Phi(1) = 0.8413$. Thus for Z, the standard normal variable, $\mathbb{P}(Z \leq 1) = 0.8413$. Again we can apply ideas that we had from the continuous random variable section about combining known probabilities.

Lemma 3.6.1. For a standard random normal variable $Z \sim \mathrm{N}(0, 1)$ we have for any $a \leq b$ that:

$$\mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

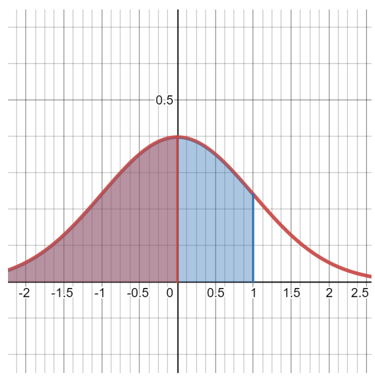
We can see this in the following example:

Example 3.6.1. Suppose $X \sim \mathrm{N}(0, 1)$ what is $\mathbb{P}(0 \leq Z \leq 1)$?

This question is entirely covered by Lemma 3.6.1 so we have that:

$$\mathbb{P}(0 \leq Z \leq 1) = \Phi(1) - \Phi(0) = 0.8413 - 0.5 = 0.3413.$$

You should ask yourself why, without calculation, that $\Phi(0) = 0.5$. We can also approach this from first principles: We are looking for area under the normal curve from $x = 0, x = 1$. If we draw the situation,



We can see that the area we are interested in can be made by taking the area from $x = -\infty$ to $x = 1$, and then subtracting it from the area from $x = -\infty$ to $x = 0$. The former value is $\Phi(1)$, while the later is $\Phi(0)$. Hence the $\mathbb{P}(0 \leq Z \leq 1) = \Phi(1) - \Phi(0)$. This reasoning extends to any pair a and b.

We also recall that the area under the entire is curve is equal to one, as it is a probability distribution. Therefore it follows that,

$$\mathbb{P}(Z \geq z) = 1 - \mathbb{P}(Z \leq z) = 1 - \Phi(z).$$

For example suppose we wish to find the $\mathbb{P}(Z \geq -1)$. Then we have that

$$\mathbb{P}(Z \geq -1) = 1 - \mathbb{P}(Z \leq -1) = 1 - \Phi(-1) = 0.8413.$$

Again we illustrate this region below:

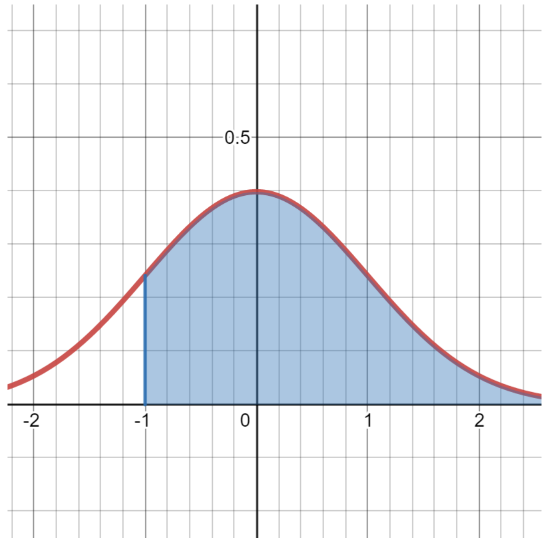


Figure 3.8: The region given by $\mathbb{P}(Z \geq -1)$.