3.5 The Normal Distribution

To end this course, we consider one of most useful types of continuous random variable, the Gaussian Random Variable, also known as the normal distribution. The normal distribution arises as a result of a sampling from certain large data sets. Some examples of the normal distribution:

(i) Suppose we have a school of 1000 students and we measure their height. Let H be the height of a randomly chosen student, then H has a normal distribution.

(ii) In a factory, a machine is filling bottles of orange juice. The machine is aiming to fill each bottle with 500ml of juice. Let O be the amount of juice inside a random bottle, then O has a normal distribution.

Below we provide a definition for the normal distribution including its pdf, we remark however that it is generally quite hard to work with. Therefore many calculations involving the normal distribution are completed using known values and probabilities. In this section we look at the shape of the pdf, and how the pdf changes as we vary different parameters.

Definition 3.5.1. Let $\sigma > 0$ and $\mu \in \mathbb{R}$ then we say the continuous real-valued random variable X is normally distributed with mean μ and variance σ^2, written $X \sim \mathrm{N}(\mu, \sigma^2)$, if X has the following pdf:

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

Really the main point to take from this definition, is that every normal distribution is defined by two parameters, a mean μ and a variance σ^2.

3.5.1 Visualising the Normal Distribution

A type of normal random variable which is of central importance, is that of the standard normal distribution $X \sim \mathrm{N}(0, 1)$, which has a mean of 0, and a variance of 1. We include of sketch of its pdf below.

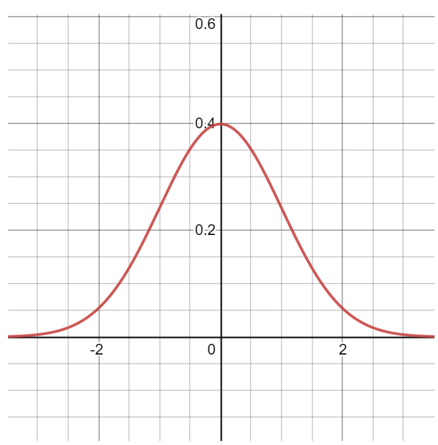

Figure 3.3: The pdf of $X \sim \mathrm{N}(0, 1)$.

Some main properties to point out, the peak of a normal variable always occurs at $x = \mu$. In this case we can see the peak occurs at $x = 0$. The pdf is symmetric around the mean, and then has a bell like shape. If we change the mean, but keep the variance the same, then shifts the curve so that the peak is now at $x = \mu$. For example, suppose $Y \sim \mathrm{N}(2, 1)$ then if we sketch the pdf of Y in blue, we can see the peak of Y occurs at $x = 2$. However, both curves have the same shape.

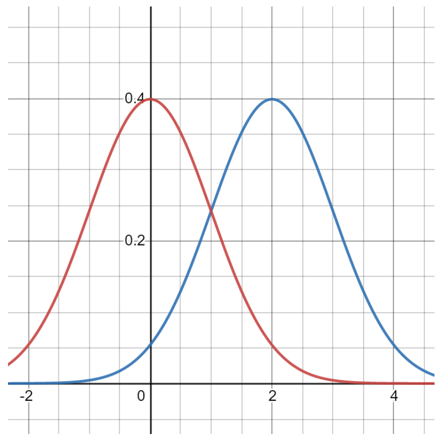

Figure 3.4: The pdf of X in red, and the pdf of $Y \sim \mathrm{N}(2, 1)$ in blue.

Changing the variance on the other hand changes the shape of the curve. If the variance is decreased, more of probability is focused around the mean. This causes the height of the peak to increase, and bell shape becomes much thinner. In blue we have the random variable $Z \sim \mathrm{N}(0, 0.5)$:

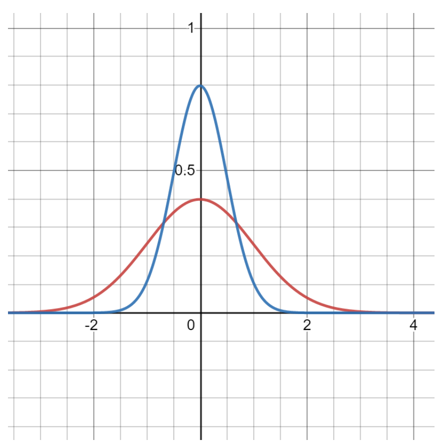

Figure 3.5: The pdf of X in red, and the pdf of $Z \sim \mathrm{N}(0, 0.5)$ in blue.

Note that X and Z both have a mean of zero, so their peaks occurs in the same place, but the shape and height of the peaks are different. Similarly if we increase the variance, the probability becomes more spread out. This causes the height of the peak to decrease, and the bell shape to widen out. Here we sketch $W \sim \mathrm{N}(0, 2)$ in blue:

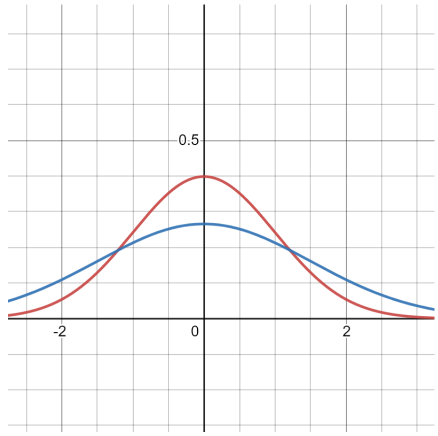

Figure 3.6: The pdf of X in red, and the pdf of $W \sim \mathrm{N}(0, 2)$ in blue.

The reason the shape of the curve changes is due to the fact that the area under the curve is always equal to one, for any choice of μ and σ. Hence if the peak height decreases, then the surrounding parts of the curve must increase to accommodate this change.