Probability: Random Variables

Normal Distributions

The normal probability distribution forms the foundation of many statistical methods. In this document, we will show properties of normal probability density functions.
The normal probability density has the familiar bell-curve shape.

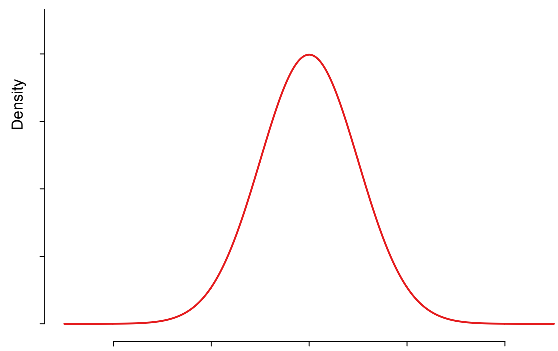

Figure 1: Shape of Normal probability density function

Suppose we have a random variable X whose probabilities are found as areas under a normal probability density function. We then say that X has a normal distribution. The normal distribution is sometimes also called the 'Gaussian' distribution. If the random variable X has a normal distribution, then X is a continuous random variable. The interval of values this random variable can take is anything from negative infinity to positive infinity. However most of the probability is concentrated in the middle.

Observe that the normal distribution is symmetric, unimodal and every normal probability density function has a bell-shape, but this bell shape can be more narrow or more spread out:

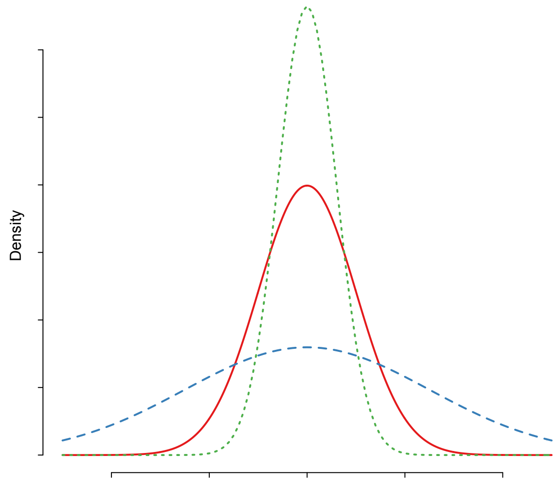

Figure 2: Normal densities with different spreads

The peak of the bell can also be at any particular value:

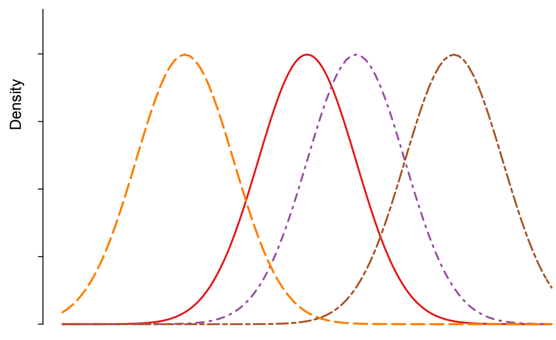

Figure 3: Normal densities with different centers

Like all density curves, the total area under each of these has to be 1. The actual value of the density function is not very important; areas under these densities are used to calculate probabilities. In order to describe a particular normal distribution fully, we need two parameters.

The first one is the location of the peak. We use a Greek letter μ ('mu') for location of the centre of the bell curve. Note that μ is also the expected value or mean of a random variable X with this normal distribution. We will almost always call this parameter, the **mean** of a normal distribution. And because the normal distribution is symmetric, μ is also the median of this distribution. The second parameter tells us how spread out the bell curve is. The spread is reflected by the standard deviation of the random variable X, and its symbol is the Greek letter σ ('sigma'). The larger σ is, the more spread out the normal curve is.

For a random variable X that has normal distribution with mean μ and standard deviation σ, we write

$$X \sim N(\mu, \sigma)$$

Here $\sim$ means X has the distribution, N stands for Normal and then the values of 2 parameters in brackets.

On a normal distribution, about 68% of the probability is within one standard deviation of the mean:

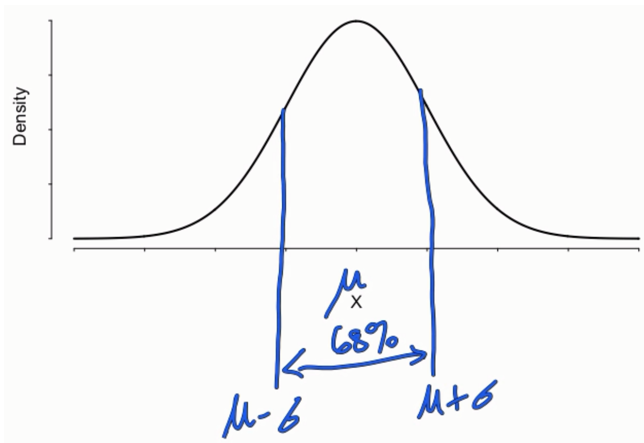

Figure 4: Normal density: Mean +/− one Standard deviation

Therefore this middle area from the mean minus one standard deviation to the mean plus one standard deviation is 68%. If we go out to 2 standard deviations from the mean we cover 95% of the area. And area +/− 3 standard deviations from the mean covers 99.7% (almost all of the area). Note that if we observe data whose possible values follow a normal distribution, we expect that almost all of the data will be within 3 standard deviations from the mean. It is possible for observations to be 4 or 5 standard deviations from the mean, but they are rare.

The **Standard Normal Distribution** has a mean of 0 and a standard deviation of 1.

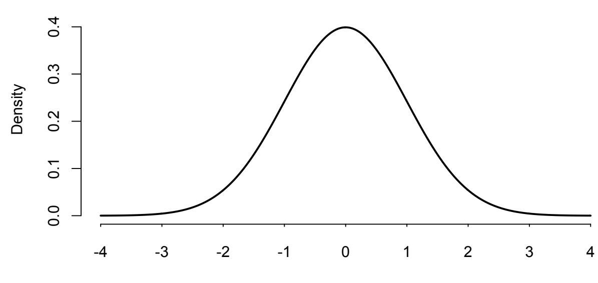

Figure 5: Standard normal density function

It is usual practice to use the variable name Z for a random variable that has a standard normal distribution.

$$Z \sim N(0, 1)$$

Any normal random variable X can be transformed to a standard normal random variable by subtracting its mean, and dividing the result by its standard deviation. So that if

$$X \sim N(\mu, \sigma)$$

then

$$\frac{X - \mu}{\sigma} \sim N(0, 1)$$

Random variables with normal distributions have one more interesting property that only holds for some probability distributions. Suppose we have two independent random variables that have normal distributions ($X_1 \sim N(\mu_1, \sigma_1)$ and $X_2 \sim N(\mu_2, \sigma_2)$), then their sum also has a normal distribution. If we multiply a normal random variable by a constant (a), the result still has a normal distribution:

- $X_1 + X_2$ has a Normal distribution
- aX_1 has a Normal distribution

A consequence of the above results is that when we average normal random variables together, the average has a normal distribution. We still need to be careful with the mean and standard deviation however.