# 5. An Introduction to Probability

Ali Ataullah and Hang Le

01/04/2020

## Contents

# 1 Introduction

The number of assets (financial and non-financial) available to investors is substantial. Although these assets differ in their characteristics, they all have one thing in common: the pay-offs associated with them are uncertain. For example, although stocks and bonds differ in terms of their claims on issuing firms' assets, both classes of assets have uncertain return. However, the level of uncertainty associated with pay-offs of different assets is not the same. Some assets (e.g. short-term bonds by the US government) are considered *low risk*, while others (e.g. stocks of poorly performing firms) are considered *high risk*.

Broadly speaking, the level of risk associated with an asset captures our uncertainty (subjective or objective) about the pay-offs from the asset. But, how do we quantify the level of risk? The theory and practice of finance provide various measures to quantify risk. Most of these measures rely on the **probability theory** to describe and quantify uncertainty associated with future pay-offs. This section provides a short introduction to some key concepts in the probability theory. Wasserman (2013) provides an excellent but succinct review of the probability theory. For a more detailed study of the probability theory, see Larsen and Marx (2005).

# 2 Intended Learning Outcomes

By the end of this session, students should be able to

1. recognise probability as a representation of uncertainty,
2. understand a random variable and its probability distribution, and
3. describe key features of a widely used probability distribution (normal distribution).

# 3 Probability, A Representation of Uncertainty

When we flip a coin, the outcome is uncertain; when we buy shares of a public listed company, the price at which we will be able to sell shares in the future is uncertain; when one company acquires another, the impact of acquisition on future cash flow of the combined entity is uncertain; when a patient receives a vaccine for a viral infection, the effectiveness of vaccination is uncertain. These examples highlight that many (perhaps most) of our decisions lead to outcomes that are uncertain (i.e. the outcomes are unknown when we make our decision).

Thus, to facilitate decision-making, we seek a *representation of uncertainty* that enables us to compare costs and benefits associated with different decisions. For example, we may like to know whether the uncertainty associated positive return from the shares of Apple is higher than that of Netflix; we may like to know whether positive outcome associated with a new vaccine is more certain than the possible side effects.

The most widely used representation of uncertainty is *probability.* If an expert analysts tells us that the probability of a stock market crash next year is 0.8, then we know what this number mean. However, there are several very difficult questions that we can ask about such a number as a measure of uncertainty. For example, where does probability come from? Is it a reliable measure of uncertainty? Is probability a person's subjective belief or can we obtain objective probabilities? We make no effort to address these questions and encourage interested students to consult Halpern (2017), which provides an excellent book-length review of uncertainty. Our modest objective is to provide a very brief overview of a few key ideas from the axiomatic probability theory.

# 4 Basics of Probability

## 4.1 Sample Space, A Set of Possible Outcomes

The probability representation of uncertainty begins with a **sample space**, denoted by $\Omega$, which is a set of all possible outcomes associated with a given situation. For example, suppose we are interested in capturing uncertainty associated with the outcome of a flip of two coins. Our sample space, the set of all possible outcomes, could be

$$\Omega = \{(HH), (HT), (TH), (TT)\}$$

In the above set, $(HH)$ represent the outcome that we observe two heads, $(TT)$ represent the outcome that we observe two tails, and so forth.

Alternatively, suppose we have a portfolio that contains stocks of two companies, Dorian plc and Mixolydian plc. Suppose the current price of Dorian is £100 per share, while the current price of Mixolydian is £200 per share. This can be represented by a price vector **p** as follows:

$$\mathbf{p}_{today} = \begin{bmatrix} 100 \\ 200 \end{bmatrix}$$

We do not know what the prices will be tomorrow, but suppose we do know that tomorrow we will end up in one of the 4 possible **states of the world**, denoted by $s_1, s_2, s_3, s_4$. Each state will produce a particular price vector for Dorian and Mixolydian. For example, in state $s_1$, we will observe

$$\mathbf{p}_{tomorrow,\ state\ 1} = \begin{bmatrix} 120 \\ 210 \end{bmatrix}$$

in state $s_2$ we may observe

$$\mathbf{p}_{tomorrow,\ state\ 2} = \begin{bmatrix} 120 \\ 190 \end{bmatrix}$$

in state $s_3$ we may observe

$$\mathbf{p}_{tomorrow,\ state\ 3} = \begin{bmatrix} 80 \\ 210 \end{bmatrix}$$

and in state $s_4$ we may observe

$$\mathbf{p}_{tomorrow,\ state\ 4} = \begin{bmatrix} 80 \\ 190 \end{bmatrix}$$

Overall, the possible prices can be specified as a vector $\mathbf{p}_{tomorrow}$:

$$\mathbf{p}_{tomorrow} = \begin{bmatrix} p_{\ tomorrow,Dorian} \\ p_{\ tomorrow,Mixolydian} \end{bmatrix}$$

where $p_{\ tomorrow,Dorian} \in \{120, 80\}$ and $p_{\ tomorrow,Mixolydian} \in \{210, 190\}$. We call an object like $\mathbf{p}_{tomorrow}$ a **random vector**.

## 4.2 Events

The subsets of our sample sample are called **events**. For example, in the example of two coin flips, $A_1 = \{(TT)\}$ is an event that we observe no heads, while $A_2 = \{(HH)\}$ is an event that we observe two heads. Let $\mathcal{F}$ denote the set of events that we are interested in. It is important to note that specifying the relevant set of events is not always easy, especially when our sample space contains infinite elements. Interested reader may consider reading a brief discussion on sigma-algebras in Stachurski (2016). But this is not essential for our purpose.

## 4.3 Probability Measure

Given the set of events that we are interested in, we next have to specify a **probability measure**, $\mathbb{P}$, which assigns a number in the interval $[0, 1]$ to each element of $\mathcal{F}$. That is,

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1] \tag{1}$$

For example, in our example of two coin flips, we may assume that for each coin, heads and tails are equally likely, and the outcome of the first coin flip does not have any impact on the second coin flip. Thus, each outcome in the sample space is equally likely. Then, our probability measure tells us that $\mathbb{P}\{(HH)\} = \frac{1}{4}$, $\mathbb{P}\{(HH)(HT)\} = \frac{2}{4}$, and so forth.

## 4.4 Probability Space

The set $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**. Within the probability theory, we represent our uncertainty about a situation by specifying a relevant probability space that describes set of possible outcomes and probabilities associated with different outcomes (and events).

# 5 Different Approaches to Probability

## 5.1 The Classical Approach

What is the probability of a 6 when we roll a "fair" dice? According to the *classical approach*, given the finite number of possible outcomes, each outcome is equally likely. So, the probability of 6 is $\frac{1}{6}$. In `R`, we can use the function `sample()` to randomly choose from a vector of data as shown in the code below.

```
# The rep() function generates repeated observations
# For example, below we use it to generate 6 observations of 1/6.

sample(c(1,2,3,4,5,6), size = 1, prob = c(rep(1/6,6)))
```

```
## [1] 2
```

```
# Try changing the prob argument above. For example,

sample(c(1,2,3,4,5,6), size = 1, prob = c(0.1,0.3,0.2,0,0,0.4))
```

```
## [1] 2
```

```
# Now 6 is the most likely outcome
```

## 5.2 The Relative Frequency Approach

Richard Von Mises provided the foundations for the *relative frequency* approach to understand probabilities (Larsen and Marx, 2005, p.17). This approach is empirical in the sense that it relies on the idea that we can repeat an experiment that lead to uncertain outcomes a very large number of time. The probability of an outcome from the set of possible outcomes is the relative frequency of that outcome in a large number of identical repetitions of the experiment. For example, the probability of 6 when we roll a dice is the frequency of 6 relative to all possible outcomes in a very large number of rolls of a dice. Figure 1 shows 40 (top) and 1000 (bottom) rolls of a dice. It is clear that for a large number of rolls, the relative frequency of 6 converges to $\frac{1}{6}$.

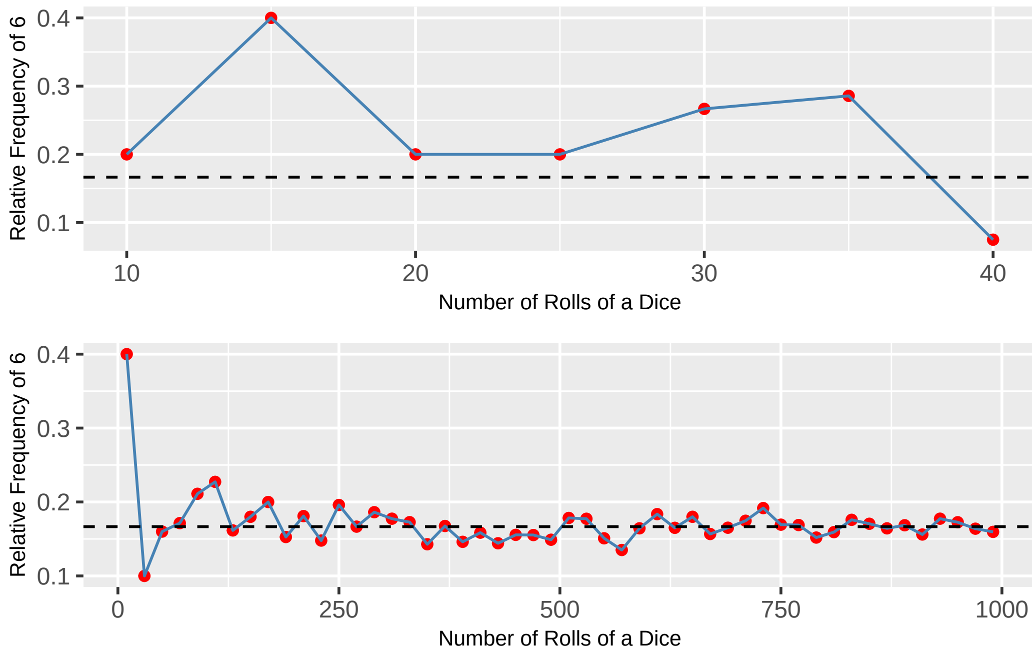


Figure 1: Relative Frequency for a Dice Roll

### 5.3 The Kolmogorov Axioms

Perhaps the most widely used view of probability is based on the following **Kolmogorov Axioms**. First, the probability of an event (a set of possible outcome) is greater than or equal to 0. Second, the probability of all possible outcome is 1. Third, if we have a set of disjoint events (i.e. events that cannot occur together like 3 and 6 for a roll of a dice), then the probability that one of them will occur is the sum of their individual probabilities (Wasserman, 2013, p. 5). But are these axioms justified? One approach to justify these axioms is to view that they emerge because people have **coherent beliefs**. It can be shown that people hold behave as if they believe in these axioms in order to avoid a sure loss. We will not say more about different approaches to understand probability. Interested readers should consult (Kadane, 2011, Chapter 1).

## 6 Random Variables and Probability Distributions

### 6.1 Random Variable as a Map

For most practical purposes, we do not explicitly talk about sample spaces. Instead, we work with random variables and their probability distributions. Many natural and social world problems are modelled as random variables. Examples include the spread of diseases, the amount of rainfall, the growth of population, traffic on a bridge, the number of patients admitted to hospitals, number of fatalities due to a particular virus, and the occurrence of natural disasters.

A random variable maps the elements of the sample space to the real numbers. More formally, $X$ is a random variable for a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$X : \Omega \to \mathbb{R} \tag{2}$$

As an example, consider the probability space for the toss of two fair coins. We know from our earlier discussion, $\Omega = \{(HH), (HT), (TH), (TT)\}$. Let us define a random variable $X$ that is equal to the number of heads when we toss two coins (Wasserman, 2013, pp. 20-22) . Then, $X(HH) = 2$, $X(HT) = H(TH) = 1$, and $X(TT) = 0$. Thus, $X$ is a random variable as it assigns a real number to each outcome in the sample space.

## 6.2 Discrete vs Continuous Random Variable

Random variables may be discrete or continuous. **Discrete random variables** assume countable values (e.g. number of heads when tossing two coins or the number of deaths due to an infectious disease), while **continuous random variable** can assume uncountable number of values (e.g. the rate of return on an asset or change in global temperature).

## 6.3 Probability Distribution

A **probability mass function** (PMF) for a discrete random variable gives the probability that the random variable will take a particular value. For example, for our example of two coin flips, our probability mass function $f_X(x)$ is as follows:

$$f_X(x) = \begin{cases} 1/4 & \text{if } x = 0 \\ 1/2 & \text{if } x = 1 \\ 1/4 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

If $X$ is a continuous random variable, then it has infinite possible outcomes. For example, when we model returns on a stock as a random variable, then return could be any number ranging from $-100$ to $\infty$. In this context, the probability of a continuous random variable to assume a particular value is 0. Instead, we may **probability density function** (PDF) $f_X(x)$ that enables us to compute the probability that our random variable will be within a specific interval $A$ as follows

$$\int_A f_X(x)dx$$

## 6.4 Mean and Variance

In most decision-making situations in which uncertainty is modelled as a random variable, we seek to find key features of the probability distribution of our random variables. Two most important features or summaries of a probability distribution are its expected value and variance.

Let $X$ be a discrete random variable with PMF $f(x)$ and let $Y$ be a continuous random variable with PDF $f(y)$. The **expected values** or the means of these random variables are:

$$\mathbb{E}(X) = \mu_X = \sum x f(x) \tag{4}$$

$$\mathbb{E}(Y) = \mu_Y = \int f(y)dy \tag{5}$$

You can think about the expected value as the weighted mean of the variable, where probabilities as weights.

The **variance** of a random variable is

$$\mathbb{V}ar(X) = \sigma_X^2 = \sum (x - \mu_X)^2 f(x) \tag{6}$$

$$\mathbb{V}ar(Y) = \sigma_Y^2 = \int (y - \mu_y)^2 f(y)dy \tag{7}$$

Note that the variance is also a kind of expected value - now we have weighted mean of squared deviations from the expected value. The square root of the variance is called the standard deviation and is denoted by $\sigma$. This is an important measure of risk in finance.

The expected value is also called the **first moment**, while the variance is the **second central moment**. The expected value is a measure of the centre of the probability distribution, while the variance is a measure of the spread of a probability distribution.

The probability distribution of a random describes the process that generates the data that we observe. Means and variance describe two summary features of the probability distribution. For example, consider the flip of two coins that we considered earlier with the sample space

$$\Omega = \{(HH), (HT), (TH), (TT)\}$$

Define a random variable $X$ that is equal to the number of heads. The probability distribution of $X$ is

$$f_X(x) = \begin{cases} 1/4 & \text{if } x = 0 \\ 1/2 & \text{if } x = 1 \\ 1/4 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Instead of looking at the whole distribution, we can compute its mean and variance. Let us do this using `R`.

```
x <- c(0,1,2)
p <- c(1/4,1/2,1/4)

mean_x <- sum(x * p)

mean_x
```

```
## [1] 1
```

```
var_x <- sum((x-mean_x)^2*p)

var_x
```

```
## [1] 0.5
```

As noted above, we use features like mean and variance of probability distribution to make decisions. For example, in the traditional portfolio theory, assets' returns are considered as random variables. Assets are then chosen on the basis of their means and variances (and covariances/correlations - see below).

### 6.5 Joint Distributions

In most real-world scenarios, we wish to measure uncertainty associated with more than one random variable. For example, investors have the option to invest in more than one asset. So, they will be interested in random variation in the return of more than one assets. In such situations, we need the **joint distribution** of the random variables that we are interested in. For an example of joint distribution, consider the prices for Dorian and Mixolydian. Suppose the joint probability distribution is given in Table 1. The table shows that the probability that Mixolydian's price will be 210 and Dorian's price will be 120 is 0.25,the probability that Mixolydian's price will be 210 and Dorian's price will be 80 is 0.25, and so forth.

Table 1: Joint Probability Distribution for Prices of Two Stocks

| | **Mixolydian = 190** | **Mixolydian = 210** |
|---|---|---|
| **Dorian = 80** | 0.25 | 0.25 |
| **Dorian = 120** | 0.25 | 0.25 |

### 6.6 Covariance and Correlation

For a joint probability distribution, covariance determines the *co-movement* between random variables. The covariance between two random variables $X$ and $Y$ is

$$\mathbb{C}ov(X,Y) = \sigma_{X,Y} = \mathbb{E}\Big((X-\mu_X)(Y-\mu_Y)\Big) \tag{9}$$

It is usually more useful to divide the covariance between two random variable by their standard deviations to obtain the **correlation coefficient** $\rho_{X,Y}$. That is,

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \tag{10}$$

# 7 Models of Random Variables

The probability theory contains a large number of mathematical models in the form of probability distributions that provide good approximations to the random variations in different situations. We describe the normal probability distribution, one of the most widely used models. Our aim is to describe the key features of a probability distributions including, expected value, variance and percentiles. We will also look at various functions in `R` that can be used to handle probability models.

## 7.1 Normal Probability Distribution

Perhaps the most well-known probability distribution is the **Normal (Gaussian) Distribution**. A random variable $Y$ has a normal probability distribution with parameters $\mu$ and $\sigma$ if it has the following PDF (Wasserman, 2013, p. 28):

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right) \tag{11}$$

A normally distributed random variable can assume any value ranging from $-\infty$ to $\infty$. A normally distributed random variable $Y$ with $\mu_Y$ and $\sigma_Y$ differs from another normally distributed random variable $X$ with $\mu_X$ and $\sigma_X$. For example, we may assume that log returns on portfolios are normally distributed. Suppose we have two portfolios, Portfolio Y and Portfolio X. We assume that the log return on Portfolio 1 is normally distributed with $\mu_Y = 5$ and $\sigma_Y = 4$, while the log return on Portfolio 2 is normally distributed with $\mu_X = 3$ and $\sigma_X = 2$.

Figure 2 shows the probability distributions of Portfolio Y and Portfolio X. Portfolio Y is centred at $\mu_Y = 5$, while Portfolio X and Portfolio X. Portfolio X is centred at $\mu_X = 3$. We say that Portfolio Y has the higher expected return compared to Portfolio X. However, notice that the dispersion of return around the mean for Portfolio Y is greater than that of Portfolio X. This larger dispersion is captured by $\sigma_Y = 4$ compared to $\sigma_X = 2$.

```
df <- tibble(x = seq(-10,25, length.out = 500)) %>%
  mutate(y1 = dnorm(x, mean = 5,sd = 4),
         y2 = dnorm(x, mean = 2,sd = 2))
```

```
n1 <- ggplot(df) +
  geom_path(aes(x,y1, color = "Portfolio Y")) +
  geom_path(aes(x, y2, color = "Portfolio X")) +
  labs(x = "Log Return Y", y = "f(y)") +
  theme(axis.title=element_text(size=8),
        legend.title=element_blank())

n1
```

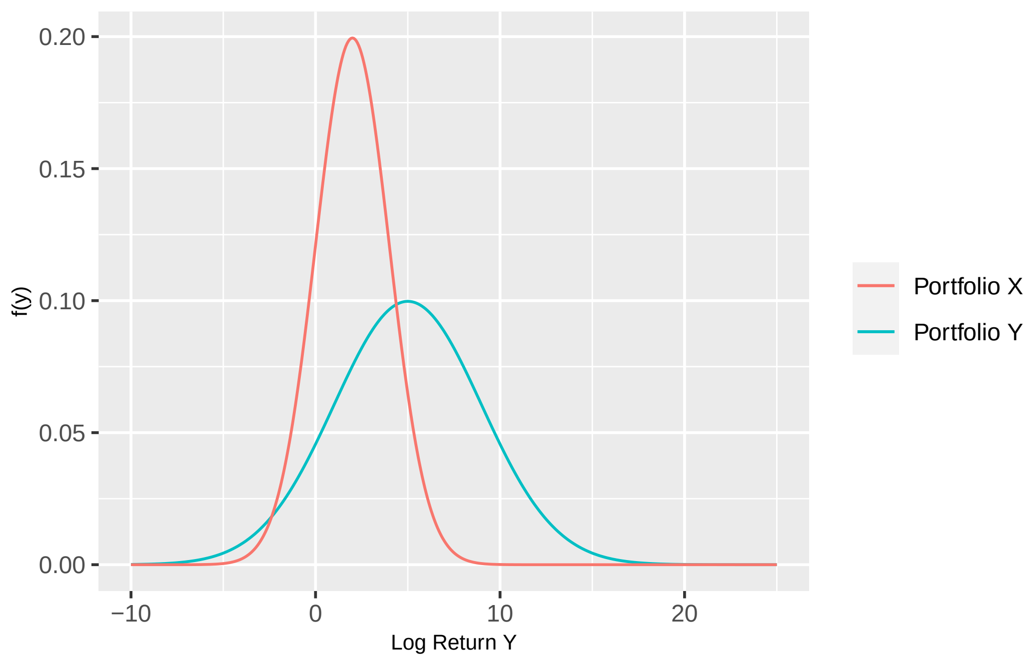


Figure 2: Normal Distribution and Log Return

```
rm(df, n1)
```

### 7.2 `R` Functions for Normal Distributions

This section describes four functions in `R` for normal distributions. These are `rnorm()`, `dnorm()`, `pnorm()` and `qnorm()`. We use `rnorm()` to draw a random sample from a normal distribution. Let us draw a random sample of size 5 from a normal distribution with $\mu = 0$ and $\sigma = 1$.

```r
# Mean 0, sigma 1


rnorm(n = 5, mean = 0, sd = 1)
```

```
## [1] -1.714505656 -0.081296040 -0.413535210 -0.008566746 -0.402758490
```

What is the probability that a normal variable will take a value be within a specific interval? For example, if $X$ is a normal random variable with $\mu_X = 0$ and $\sigma_X = 1$, then what is $Probability(X < 0) = \mathbb{P}(X < 0)$?. We can compute this probability using the `pnorm()` function in `R` as follows.

```r
pnorm(q = 0, mean = 0,sd = 1)
```

```
## [1] 0.5
```

The `qnorm()` is the quantile function for normal distribution and it does the opposite of `pnorm()`. To understand a quantile, suppose $X$ is a normal random variable with $\mu_X = 0$ and $\sigma_X = 1$, then then the 0.5 quantile $q_{0.5}$ of $X$ is the value of $X$ such that $\mathbb{P}(X < q_{0.5}) = 0.5$. The 0.5 quantile is also called the *50th percentile$* or *second quartile* or the **median** of a distribution. The 0.25 quantile is called the **first quartile**, and the 0.75 quartile is called the **third quartile**. Let us compute the 0.25 quantile of a normal distribution with mean 1 and standard deviation 1 using `qnorm()`. We sometimes use the term **percentile**. The 0.25 quantile is the 25th percentile, and the 0.75 quantile is the 75th percentile.

```r
qnorm(p = 0.25, mean = 1, sd = 1)
```

```
## [1] 0.3255102
```

```r
# Composite function
pnorm(qnorm(p = 0.25, mean = 1, sd = 1))
```

```
## [1] 0.6276025
```

The `dnorm()` function determines the PDF of a normal random variable. To understand this, we now plot the standard normal variable that has $\mu = 0$ and $\sigma = 1$. We first create a standard normal variable $\epsilon$ that ranges from $-5$ to 5. We then use `dnorm()` to create the graph of this variable over the specified range. The graph is shown in Figure 3 shows a graph of a standard normal distribution (i.e. normal distribution with $\mu = 0$ and $\sigma = 1$). If has standard normal distribution, then we write it as follows: $\epsilon \sim N(0, 1)$.

```r
df <- tibble(epsilon = seq(-5,5, length.out = 500)) %>%
  mutate(f = dnorm(epsilon, mean = 0,sd = 1))



n1 <- ggplot(df, aes(epsilon,f)) +
  geom_path(color="steelblue") +
  xlab("Standard Normal Random Variable"~""*epsilon*"") +
  ylab(expression("f"~"("*epsilon*")")) +
  theme(axis.title=element_text(size=8))
```

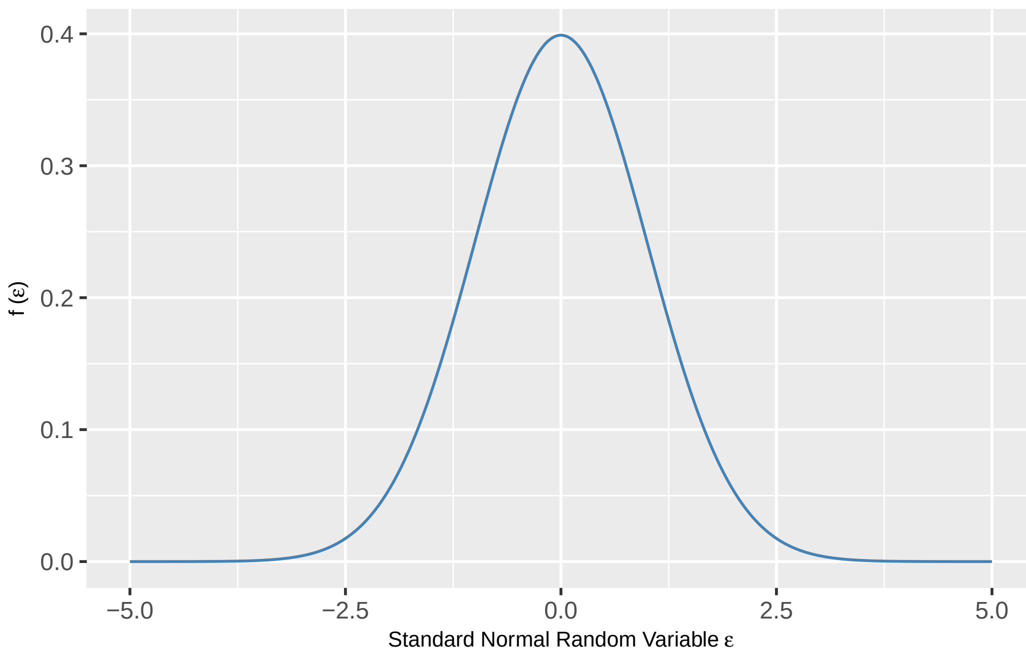


Figure 3: Standard Normal Distribution

### 7.3 Exercise

1.Draw a random sample of size 5 from a normal distribution with $\mu = 5$ and $\sigma = 10$.

```r
# Your code
# Mean 5, sigma 10


rnorm(n = 5, mean = 5, sd = 1)
```

```
## [1] 5.503193 4.044363 5.964346 3.132470 4.867827
```

2. Compute the 0.75 quantile of a normal distribution with mean 0 and standard deviation 1.

```
# Your code
```

3. What is the probability that a standard normal variable will take a value in the interval $[-2, 2]$?

```
# Your code
```

4. What is the probability that a standard normal variable will take a value in the interval $[-3, 3]$?

```
# Your code
```

5. Suppose $X \sim N(5, 1)$. Compute the probability that $X$ lies in the interval $[4, 6]$.

```
# Your code
```

## 7.4 Important Facts about Standard Normal Distribution

Normal distribution is symmetrical and has the skewness of 0. Also, for $\epsilon \sim N(\mu, \sigma^2)$,

$$\mathbb{P}(\mu - 1\sigma < \epsilon < \mu + 1\sigma) \approx 0.683$$

$$\mathbb{P}(\mu - 2\sigma < \epsilon < \mu + 2\sigma) \approx 0.95$$

$$\mathbb{P}(\mu - 3\sigma < \epsilon < \mu + 3\sigma) \approx 0.997$$

These important areas are shown in Figure 4 below.

```
# In situations where we repeat code, it is better to create a function.
# Writing functions is an important skills for all data analysts.

df <- tibble(epsilon = seq(-5,5, length.out = 500)) %>%
  mutate(f = dnorm(epsilon, mean = 0,sd = 1))



n1 <- ggplot(df, aes(epsilon,f)) +
  geom_path(color="steelblue") +
```

```
  xlab("Standard Normal Random Variable"~""*epsilon*"") +
  ylab(expression("f"~"("*epsilon*")")) +
  theme(axis.title=element_text(size=8)) +
  geom_segment(aes(x = -1, y = 0, xend = -1, yend = dnorm(-1)), color = "red") +
  geom_segment(aes(x = 1, y = 0, xend = 1, yend = dnorm(1)), color = "red") +
  geom_area(stat = "function", fun = dnorm, fill = "lavenderblush", xlim = c(-1, 1)) +
  annotate("text", x = -1, y = -0.02, label = "-1", size = 2) +
  annotate("text", x = 1, y = -0.02, label = "1", size = 2)

n2 <- ggplot(df, aes(epsilon,f)) +
  geom_path(color="steelblue") +
  xlab("Standard Normal Random Variable"~""*epsilon*"") +
  ylab(expression("f"~"("*epsilon*")")) +
  theme(axis.title=element_text(size=8)) +
  geom_segment(aes(x = -2, y = 0, xend = -2, yend = dnorm(-2)), color = "red") +
  geom_segment(aes(x = 2, y = 0, xend = 2, yend = dnorm(2)), color = "red") +
  geom_area(stat = "function", fun = dnorm, fill = "thistle", xlim = c(-2, 2)) +
  annotate("text", x = -2, y = -0.02, label = "-2", size = 2) +
  annotate("text", x = 2, y = -0.02, label = "2", size = 2)



n3 <- ggplot(df, aes(epsilon,f)) +
  geom_path(color="steelblue") +
  xlab("Standard Normal Random Variable"~""*epsilon*"") +
  ylab(expression("f"~"("*epsilon*")")) +
  theme(axis.title=element_text(size=8)) +
  geom_segment(aes(x = -3, y = 0, xend = -3, yend = dnorm(-3)), color = "red") +
  geom_segment(aes(x = 3, y = 0, xend = 3, yend = dnorm(3)), color = "red") +
  geom_area(stat = "function", fun = dnorm, fill = "antiquewhite", xlim = c(-3, 3)) +
  annotate("text", x = -3, y = -0.02, label = "-3", size = 2) +
  annotate("text", x = 3, y = -0.02, label = "3", size = 2)
```

```r
grid.arrange(n1,n2,n3)
```

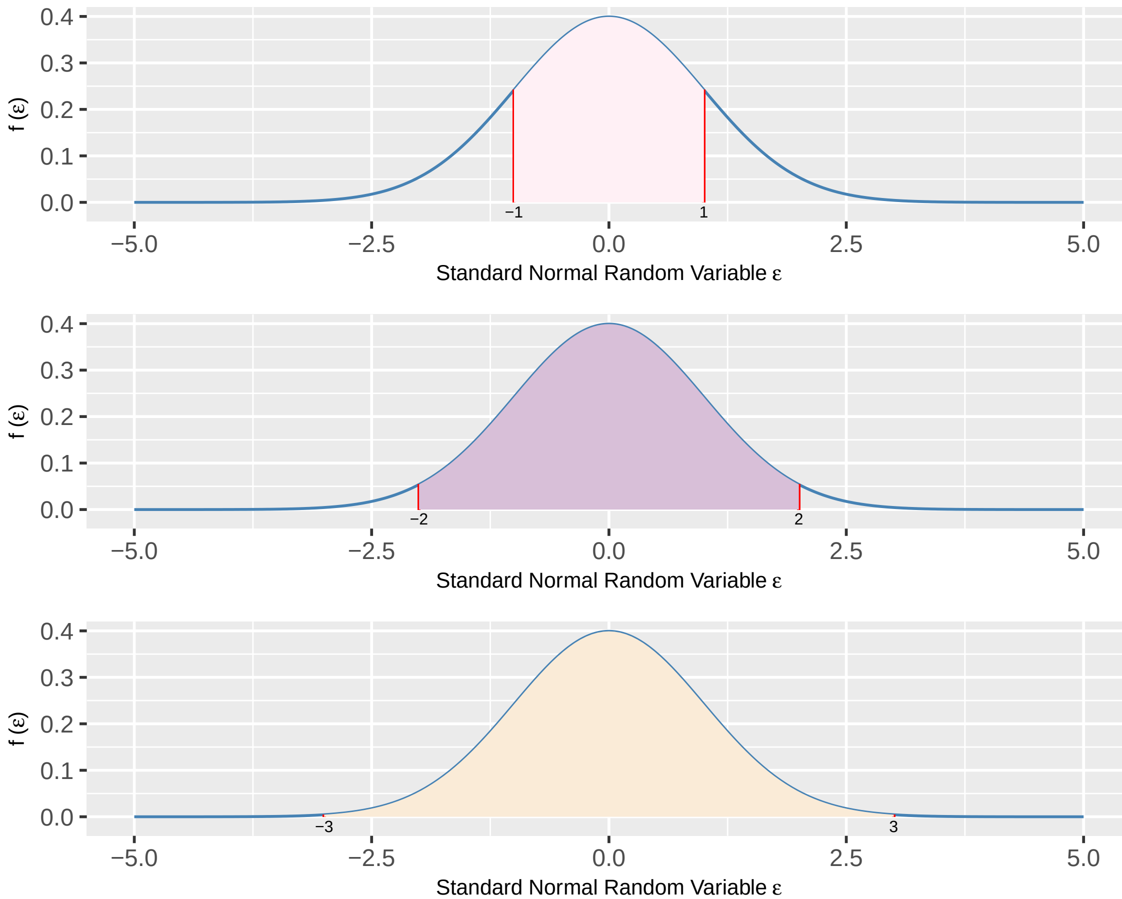


Figure 4: Important Intervals for Standard Normal Variable

```r
rm(list = ls())
```

## 7.5 Other Distributions

There are a large number of probability models. These models are used to capture uncertainty associated with a variety of phenomenon. For example, another important probability distribution is the **t distribution**. A random variable $X$ has t distribution with $\nu$ degrees of freedom if its probability function is

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \frac{1}{(1+\frac{x^2}{\nu})^{(\nu+1)/2})} \tag{12}$$

where $\Gamma(.)$ is called the gamma function. The t distribution is symmetrical like the normal distribution.

However, it has much thicker tails. This means that it is a more suitable model for phenomenon that can exhibit extreme outcomes. `R` also contains functions to compute probabilities and quantiles for random variables that have t distribution. For example, we have `rt()` to draw random sample from a t distribution, `pt()` to compute probabilities for a t variable to be in an interval, and `qt()` as the quantile function.

Some other widely used probability models include Binomial distribution, Poisson distribution, exponential distribution, beta distribution, and chi-squared distribution. These and many other probability models can be handled easily in `R`.

# 8 Next Step

This session provides a very quick overview of some concepts in the probability theory. Students interested in the theory and practice are strongly encouraged to invest some time in studying the probability theory as it forms the basis not only for financial modelling but almost all quantitative subjects. Once again, we highly recommend Wasserman (2013).

## References


Halpern, J. Y. (2017). *Reasoning About Uncertainty.* MIT press.

Kadane, J. B. (2011). *Principles of Uncertainty.* CRC Press.

Larsen, R. J. and Marx, M. L. (2005). *An Introduction to Mathematical Statistics.* prentice Hall.

Stachurski, J. (2016). *A Primer in Econometric Theory.* Mit Press.

Wasserman, L. (2013). *All of Statistics: A Concise Course in Statistical Inference.* Springer Science & Business Media.