6 Standard Statistical Distributions

In everyday life, we encounter different types of data such as data related to stocks, data of students' performance in class, weather data, census data, socio-economic data of different households and many more. While plotting the distribution of say, heights of women in a country you would encounter a bell-shaped curve, but financial returns (such as returns on a benchmark index like S&P500) have fatter tails and the census data may show an entirely different pattern.

While conducting research and in business, we often need to model the distribution of data. As different datasets give rise to different distribution shapes, we can't have a "one distribution fits all" approach. So do we have a better approach then to deal with this problem? The answer is yes, by using distribution patterns. In fact, we have a number of standard distribution types that are used as models to fit our datasets.

Each type of distribution has specific properties and says something useful about the data. Thus, by classifying the data as per its nearest distribution type, we can deduce its various properties based on the distribution pattern that it follows.

In the last chapter, we have already learnt about two standard statistical distributions namely, the discrete and continuous uniform distributions. In this chapter, we will discuss some other standard statistical distributions that we often encounter in finance such as the normal, t, exponential and the log-normal distributions. Let us start with the exponential distribution.

6.1 Exponential Distribution

Exponential distribution is a continuous distribution that is often used to model the expected time one needs to wait before the occurrence of an event. In other words, it determines the probability distribution of times between random occurrences.

For example, how long will a trader need to wait to get a profitable trade, how long a shopkeeper needs to wait until a customer enters the shop, how much time to wait before the occurrence of next flood etc. can be modelled using exponential distribution.

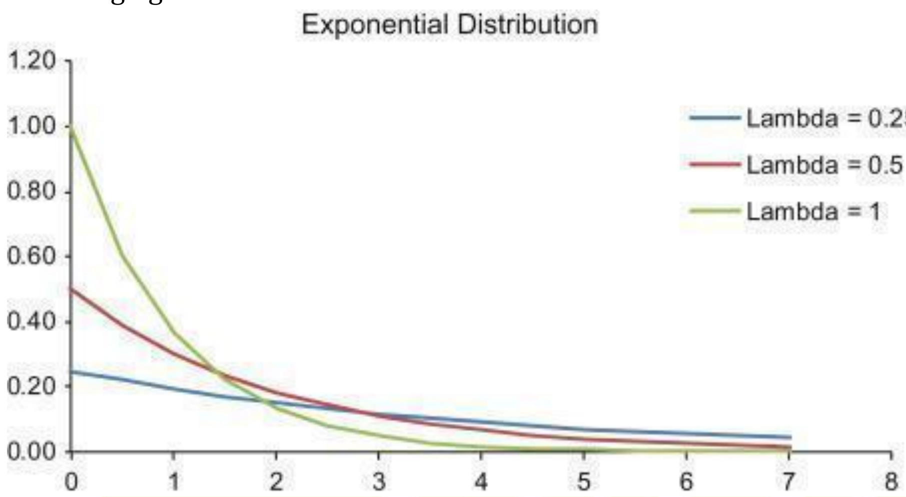
In these events, waiting time is unknown, and it can be thought of as a random variable following an exponential distribution. If a random variable X has exponential distribution, we write **X ~ Exp(λ)** and its Probability distribution function (PDF) is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } x \geq 0 \\ 0, & \text{for } x < 0 \end{cases}$$

Exponential random variable X can be used to model time intervals as the probability density is 0 for negative values(and we know that time can't be a negative quantity).

The **lambda parameter (λ) is called the rate parameter**. It is the inverse of the expected duration. Expected duration represents the amount of time you expect to wait. So, if the expected duration is 10 minutes, then the rate parameter is 0.1.

Probability distribution function (PDF) for different values of lambda (λ) is shown in the following figure:



Here, we note that as the value of lambda decreases, the curve approaches zero and becomes flatter.

6.2 Normal Distribution

The most well-known statistical distribution is the normal distribution, and not without a reason. Many variables in nature including human characteristics such as weight, height, speed, blood pressure, length, IQ and years of life expectancy approximate to a normal distribution. Even many variables in industry and business also follow normal distribution such as annual cost of household insurance, cost per square foot of renting a warehouse etc.

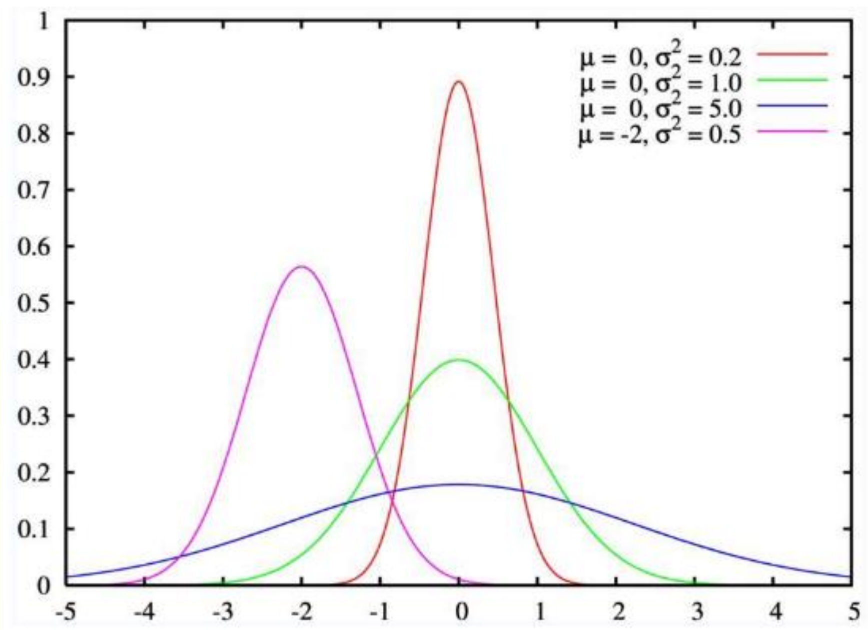
Normal distribution was discovered by the mathematician and astronomer named Karl Gauss (1777-1855), who discovered that errors of repeated measurements of certain objects are normally distributed. He identified that the errors of measurement follow a bell-shaped curve which he termed as the normal curve of error. Therefore, it is also referred to as **Gaussian distribution**.

Shape of the curve of normal distribution (i.e. its PDF) looks similar to the general **bell curve** and is symmetrically distributed around its mean, which is shown in the following figure:

Properties of Normal Distribution

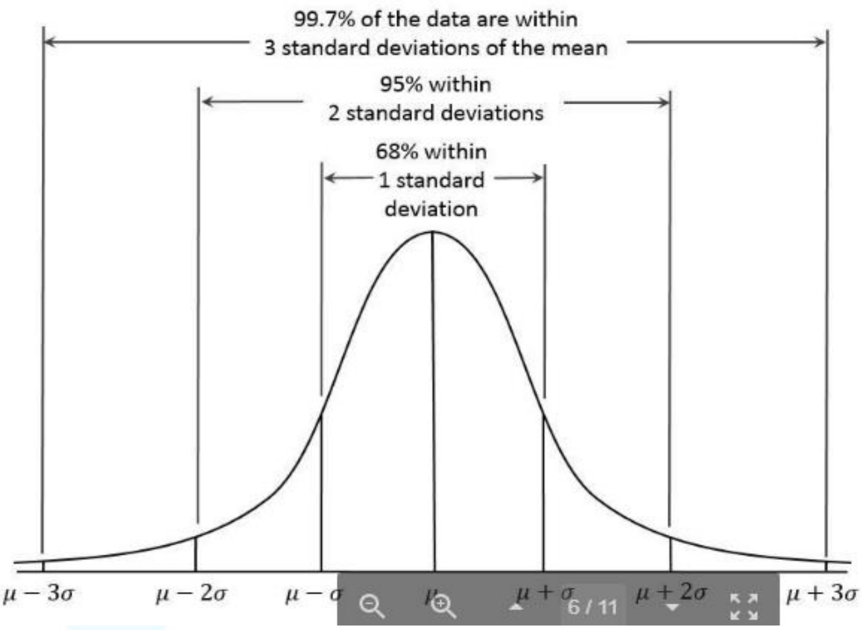
- It is a continuous distribution which is completely described by two parameters: mean & variance/ standard deviation.

- It has symmetrical distribution around the mean (area of distribution on each side of its mean is 0.5). Total area under the curve is one, like for any continuous distribution.
- It is asymptotic to the horizontal axis (Doesn't touch the x axis and goes forever in each direction)
- It is unimodal (most of the values lie around the centre of the curve)
- Mean, median and mode of normal distribution are equal
- It is a family of curves (Every unique value of mean and standard deviation will give different normal curves) as seen in the following figure:



Noteworthy Points About the Normal Distribution

- 68% of the data falls within one standard deviation of the mean ($\mu \pm \sigma$)
- 95% of the data falls within two standard deviations of the mean ($\mu \pm 2*\sigma$)
- 99.7 % of the data falls within three standard deviation of the mean ($\mu \pm 3*\sigma$)
- For a given mean, a larger value of σ would give a flatter curve while a smaller value of σ would give a narrower curve



Probability Density Function (PDF) of Normal Distribution

If a random variable X follows normal distribution with the parameters describing it being mean (μ) and standard deviation (σ), it is represented as **X ~ N(μ, σ)** and its Probability distribution function (PDF) is given by:

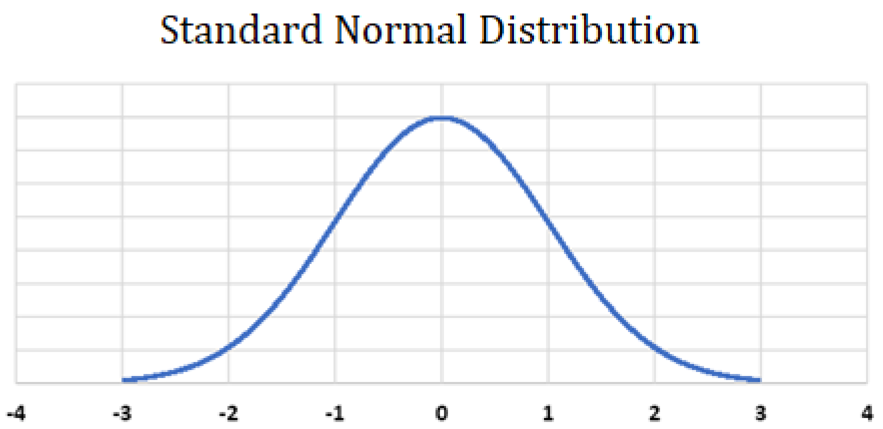
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , \text{for } -\infty \leq X \leq \infty$$

This probably looks daunting, especially if you are looking at it for the first time. However, it is more important to understand the underlying idea and not get bogged down by this monstrous expression.

Standard Normal Distribution

The shape of the curve depends on the values of parameters μ and σ and it is cumbersome to compare distributions with different values of these parameters. A nifty trick that we use is to convert them into a standardized distribution called the standard normal distribution.

The standard normal distribution is a special type of normal distribution where mean (μ) is equal to zero, and standard deviation (σ) is equal to one. This type of distribution is also known as the Z distribution.



Mean = 0
Variance = Standard Deviation = 1

Any random variable X, such that **X ~ N(μ, σ)** can be **converted into a standard normal variable Z**, using the following transformation:

$$Z = \frac{Normal\ variable - mean}{standard\ deviation} = \frac{X - \mu}{\sigma}, \quad \text{for } \sigma \neq 0$$

The X-axis of the standard normal distribution is known as the standard score or **z-score**.
For any specific value of X, say x, the z-score can be calculated as:

$$z = \frac{x - \mu}{\sigma}$$

When the value of x is higher than mean, then z score is positive, and if it is less than mean, then the z-score is negative.

Noteworthy Points about the Standard Normal Distribution:

For a standard normal variable,

- 68.2% of the values lie within one standard deviation of the mean i.e. z values between ± 1
- 95.4% of the values lie within two standard deviation of the mean i.e. z values between ± 2
- 99.7% of the values lie between three standard deviations of the mean i.e. z values between ± 3

Application of Normal Distribution in Finance & Trading

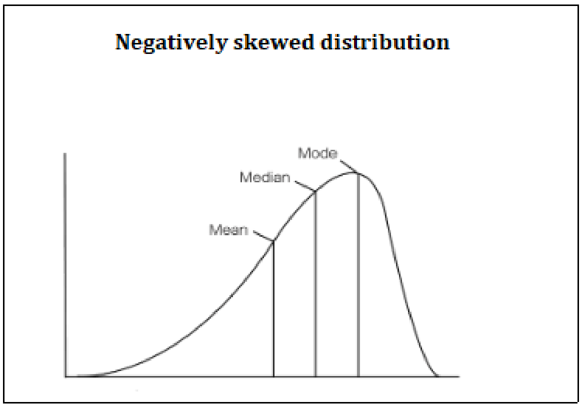
It has been empirically observed that the returns of assets such as stocks tend to approximately follow the normal distribution pattern. If we treat the returns of a particular stock as a normally distributed random variable, its mean (μ) represents the average return and standard deviation represents the volatility of stock returns (a proxy for riskiness of stock). Thus, the assumption of normality helps us to model the return and risk of an asset .

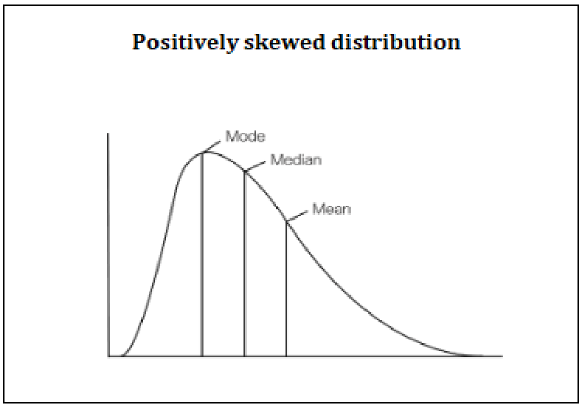
Skewness

Skewness measures the extent of asymmetry in a distribution compared to a normal distribution. To be clear, a normal distribution is perfectly symmetrical with no skew. Its mean, mode and median are equal.

A left-skewed distribution is also called a negatively skewed distribution and has a long left tail. In this case, the mean is less than the mode.

A right-skewed distribution is also called a positively skewed distribution and has a long right tail. In this, the mean is more than the mode.





Kurtosis

Kurtosis is a measure that defines how heavy the tails of a distribution are compared to that of a normal distribution. Higher kurtosis implies 'fatter tails' i.e. more probability density for extreme values.

Based on kurtosis, distributions can be classified into the following three categories:

Mesokurtic Distribution

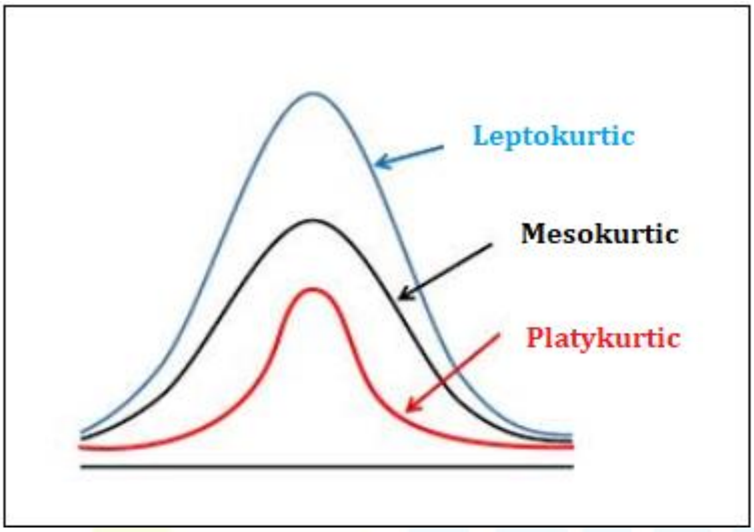
A distribution that has tails shaped in roughly the same way as any normal distribution is said to be mesokurtic. The kurtosis of a mesokurtic distribution is neither high nor low.

Leptokurtic Distribution

A leptokurtic distribution has greater kurtosis than a mesokurtic distribution. Leptokurtic distributions are sometimes identified by peaks that are thin and tall. The tails of these distributions, to both the right and the left, are thick and heavy. One of the most well-known leptokurtic distributions is the student's t distribution (which we discuss later in this chapter).

Platykurtic Distribution

Platykurtic distributions are those that have lean tails. Many times they possess a peak lower than a mesokurtic distribution. All uniform distributions are platykurtic as they have no outliers which can cause fat tails.



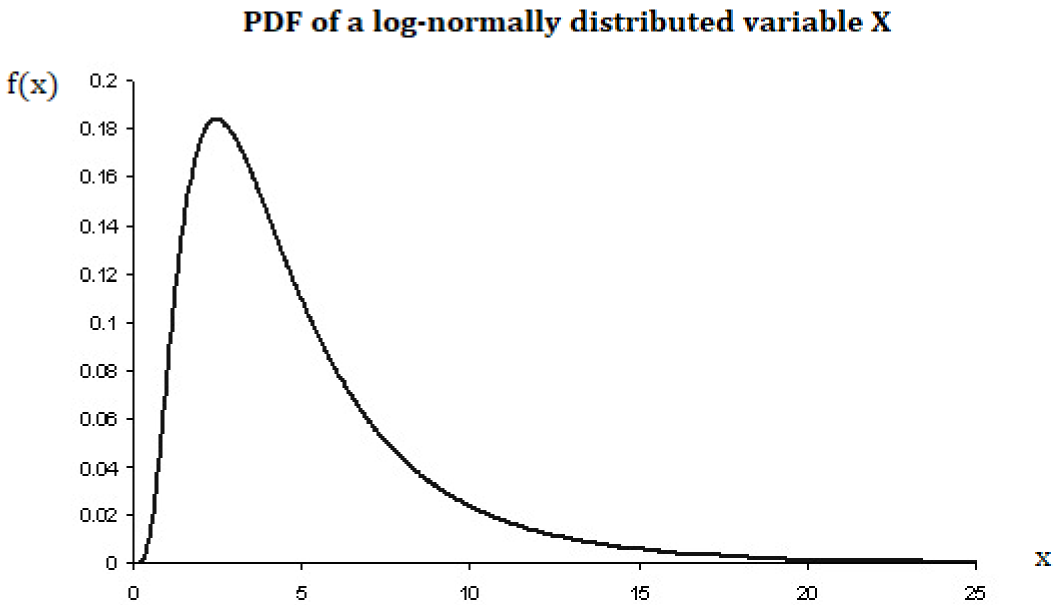
6.3 Log-Normal Distribution

In the previous section, we discussed that asset returns tend to be normally distributed. What about asset prices? To model asset prices, we use another standard statistical distribution called the log-normal distribution.

It is a continuous probability distribution of a random variable whose natural logarithm is normally distributed. In other words, a variable X is said to be lognormally distributed if Y=ln(X) is normally distributed, where 'ln' stands for natural logarithm.

A random variable which is log-normally distributed takes only positive real values and hence is ideal for modelling non-negative variables such as asset prices.

Other examples of log-normally distributed variables include the income of people, fuel consumption of a bus, size distribution of rainfall droplets etc. Log-normal distribution is presented as shown in the following figure:

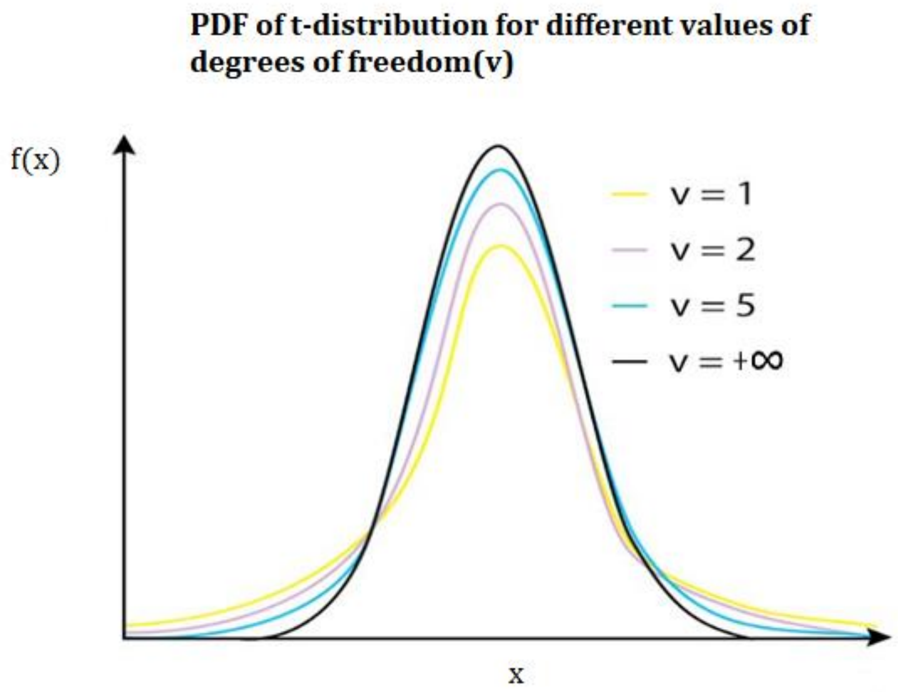


Here, you can see that log-normal distribution has long-right tails i.e. are positively skewed, unlike the normal distribution which is perfectly symmetric.

6.4 Student's T- Distribution

The t-distribution (also called student's t distribution) is very similar to normal distribution, in that it is perfectly symmetric, but it has 'fatter tails' (i.e. higher kurtosis). This means that a random variable following a t-distribution has higher probability of taking more extreme values compared to a normally distributed variable. This makes it a very good candidate for modelling financial risk, as it gives a higher probability to more extreme losses.

The shape of the PDF curve for a t-distribution is controlled by a parameter called 'degrees of freedom', which refers to the number of independent observations in a set of data. This can be seen in the following figure:



For higher values of degrees of freedom (>=30), the t-distribution starts resembling the normal curve.

The t-distribution is also used heavily in conducting statistical tests such as hypothesis testing, which we will discuss in this primer later.

6.5 Other Distributions

In the last two chapters, we have discussed only a handful of very important continuous distributions such as the uniform, normal, log-normal and t- distributions.

Hopefully you have gained a good intuition of how these standard distributions help us in modelling different types of data. However, we have only scratched the surface and hopefully gotten you curious enough to learn about other discrete and continuous distributions such as the binomial, chi-squared and F distributions. Towards the end, we will point you to other sources in case you want to learn more about it.

Note: Refer to appendices 1 and 2 at the end of this course to see the detailed procedure for calculating probabilities for the standard normal and the t- distribution in Excel.