Kurtosis : Diving in its Controversy

Subhajit Karmakar (2nd year), Anik Chakraborty (3rd year)

The Controversies . . .

"Kurtosis measures the degree of 'peakedness' of a distribution…"

"Kurtosis tells us nothing about the 'peak' of a distribution, it gives an idea whether there are outliers…"

"Kurtosis can be interpreted as the degree of deviation from normality in comparison to the 'peakedness'…"

"Kurtosis measures the probability concentration inside the range $(\mu \pm \sigma)$*…"*

Textbooks and Journals : What they say?

Textbooks and journals describe the kurtosis as, *"The distribution with higher peak than normal distribution has positive kurtosis (*$\gamma_2 > 0$*) and negative kurtosis indicates that the distribution has a lower peak than normality."*

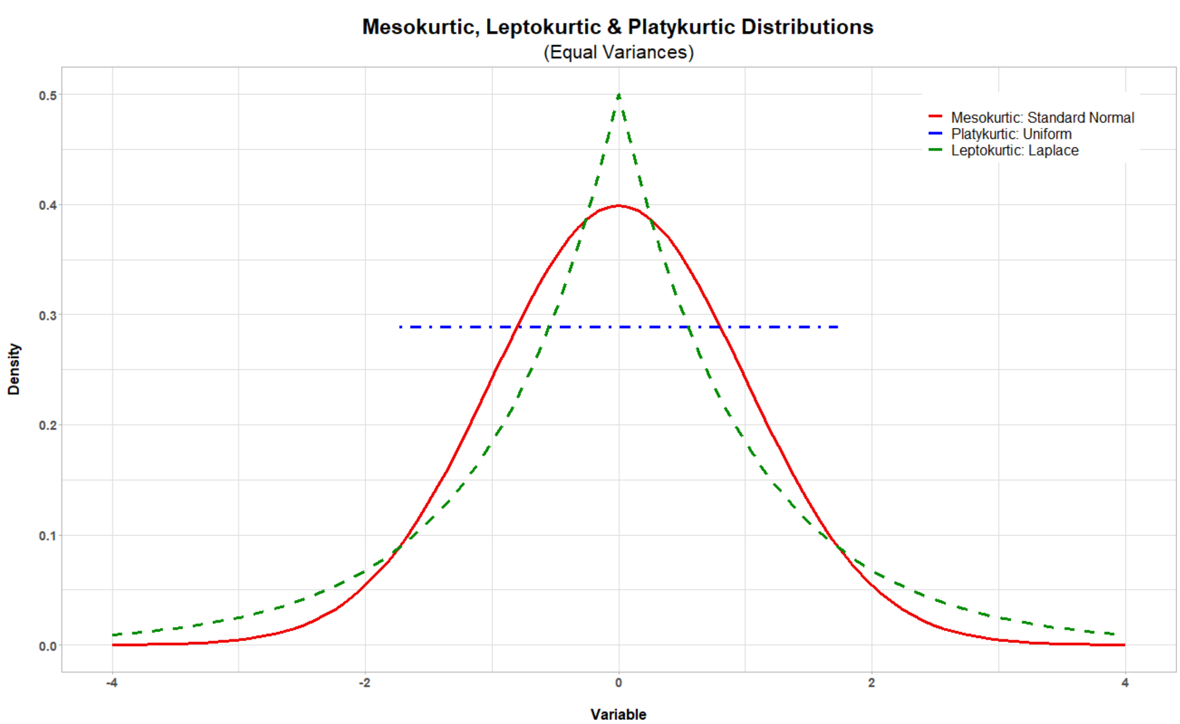


Figure 1 : **Peakedness & Kurtosis**

Revealing the Truth . . .

Suppose X is our variable of interest with $E(X) = \mu$ and σ is the standard deviation. If the kurtosis of the distribution is 'γ_2' then,

$$\gamma_2 = \frac{E(X-\mu)^4}{\sigma^4} ,$$

Now, for a given distribution σ is a constant. Hence, the value of 'γ_2' will depend only on the expression $E(X - \mu)^4$.

What does the expression $E(X - \mu)^4$ try to tell us?

Note that, $(X - \mu)$ gives the deviation of the values of the random variable X from its mean. This deviation will be larger in magnitude if the values of X are too far away from μ (i.e., outliers) and will be smaller for values which are nearer to μ.

Now, as we raise the values $(X - \mu)$ to their 4^{th} power, the larger values of the deviation will become more larger and smaller values (i.e., values nearer to 0) will become smaller. As a result of which, we observe that on an average the extreme values (outliers) of X contribute more to $E(X - \mu)^4$ and for the central part of the distribution the contribution is almost nil.

Hence, we observed that, **kurtosis gives an idea of presence of outliers or *'tailedness'* rather than anything about *'peakedness'* of a distribution.**

<u>Sample Counterpart</u>

The sample analogue of the measure of kurtosis for n observations $(x_1, x_2, \ldots, x_n)$ is given by,

$$g_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 , \qquad where\ s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

We have $E(g_2) \approx \gamma_2$, for large 'n'

Essentially, γ_2 measures the propensity to produce outliers from the probability distribution and g_2 measures whether the sample contains any outlier or not.

Tailedness & Peakedness : Work Together!

Let, us visualize the theoretical concept discussed above through diagrams to get a better grip of the discussion.

We consider here the ***Student's t*** distribution.

For this distribution the population kurtosis is,

$$k = \frac{6}{m-4}, m > 4$$

where, **m** is the degrees of freedom of the **t distribution**.

From the theoretical expression it can be readily seen that, as m increases, kurtosis of $\boldsymbol{t_m}$ **distribution** decreases. We now choose three values of m = 5, 15, 25 and plot their respective theoretical densities to verify the result *graphically*.

Case I : m = 5

Here, for m = 5, $\gamma_2 = \frac{6}{5-4} = 6$

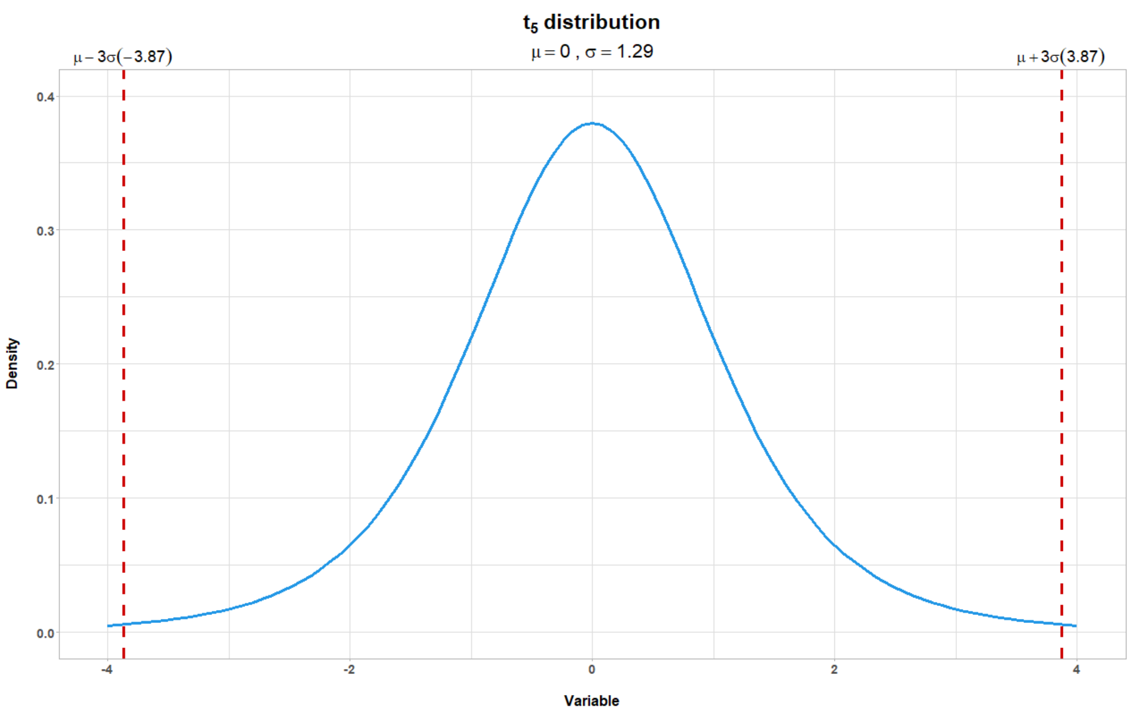


Figure 2 : **PDF of t_5 distribution**

Case II : m = 15

Here, for m = 15, $\gamma_2 = \frac{6}{15-4} = 0.55$

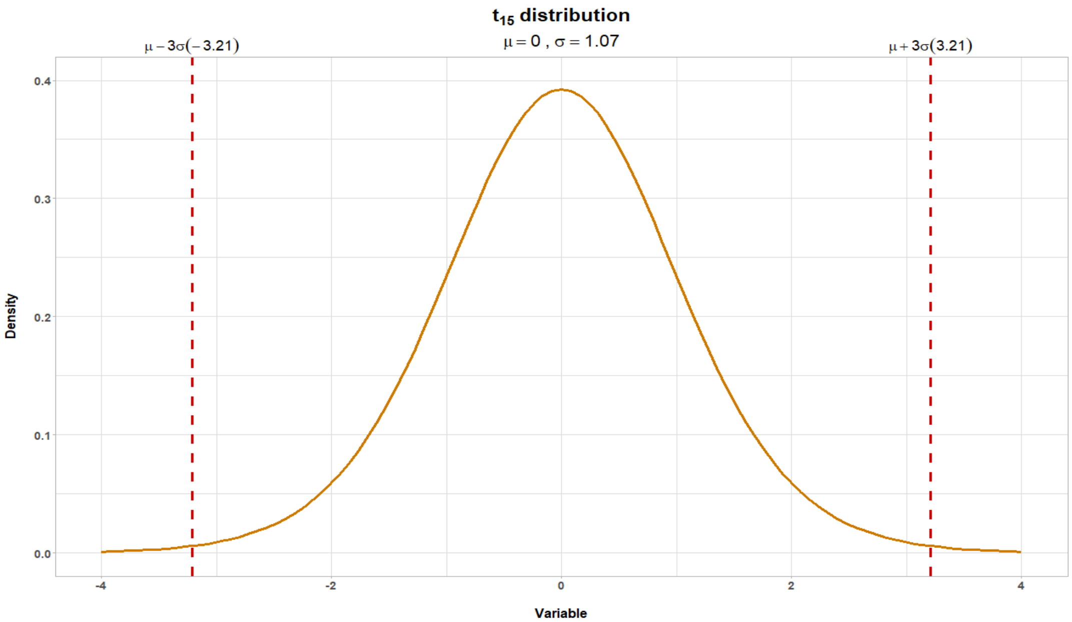


Figure 3 : **PDF of t_{15} distribution**

Case III : m = 25

Here, for m = 25, $\gamma_2 = \frac{6}{25-4} = 0.286$

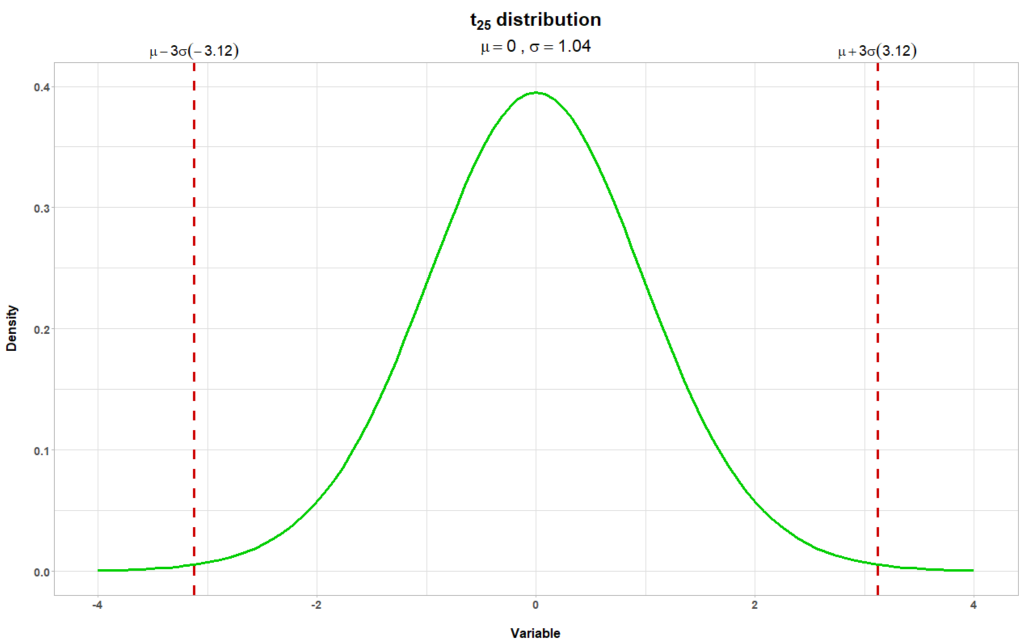


Figure 4 : **PDF of t_{25} distribution**

Observations

From 'Figure 2' we observe that, there are too many extreme values in the distribution i.e., the tail of the distribution can be considered as ***'thick'***.

Note that, in 'Figure 3' we can observe the presence of extreme values in t_{15} distribution but the densities at extremities are less than that of t_5 distribution.

Lastly, in 'Figure 4' fewer extreme values can be observed than both t_5 and t_{15} distributions.

Hence, from the above three cases, it seems that ***as the degrees of freedom increases, tail of the distribution become lighter and kurtosis also decreases***.

Observe that, the **peakedness** also decreases as the curve tends to bell shaped for large degrees of freedom.

Normal & t_5 : A Comparison

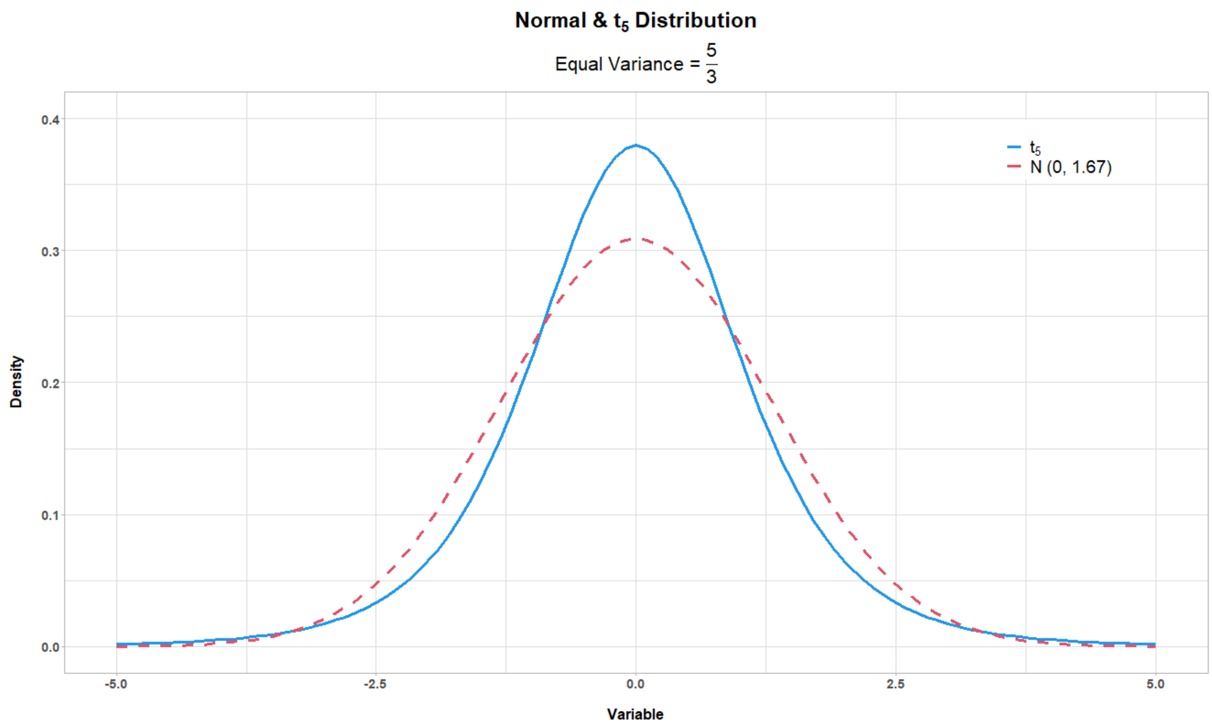


Figure 5: **Comparing the Kurtosis**

Note that the t_5 distribution has a variance of $\frac{5}{3}$, and the normal distribution shown in the 'Figure 5' is scaled to also have a variance of $\frac{5}{3}$.

Hence, from the above theoretical density plots (Figure 5), we can observe that t_5 has a thicker tail and a higher peak as compared to $N(0,\frac{5}{3})$ distribution. Thus, it matches with both the theory of *"tailedness"* and *"peakedness"* as the kurtosis is higher in case of t_5 distribution than $N(0,\frac{5}{3})$.

Lawrence T. DeCarlo in his article *'On the Meaning and Use of Kurtosis'* told us, ***"...the t_5 distribution crosses the normal twice on each side of the mean, that is, the density shows a pattern of higher-lower-higher on each side, which is a common characteristic of distributions with excess kurtosis."***

Hence, it can be said that both the concepts, *"tailedness"* and *"peakedness"* coexist here. In general, according to Karl Pearson, for the symmetric bell-shaped curves the idea of "peakedness" go with the kurtosis of a distribution.

The Parting of the Ways . . .

Let us now lean onto some cases where *"peakedness"* and kurtosis do not go hand in hand. If there is negative excess kurtosis for a probability distribution, then from the concept of *"peakedness"*, we are supposed to conclude that the distribution is flat-topped but that is not the case always.

Let us go through some examples –

1. In a short note Kaplansky (1945a) drew attention to four examples of distributions with different values of kurtosis and behaviour not consistent with the interrelation between kurtosis and peakedness.
2. Westfall (2014) showed examples of some distributions where this well-known connection between *"peakedness"* and *"kurtosis"* was quite doubtful.

We consider here, two distributions to visualize this –

- **Beta (0.5, 1)**

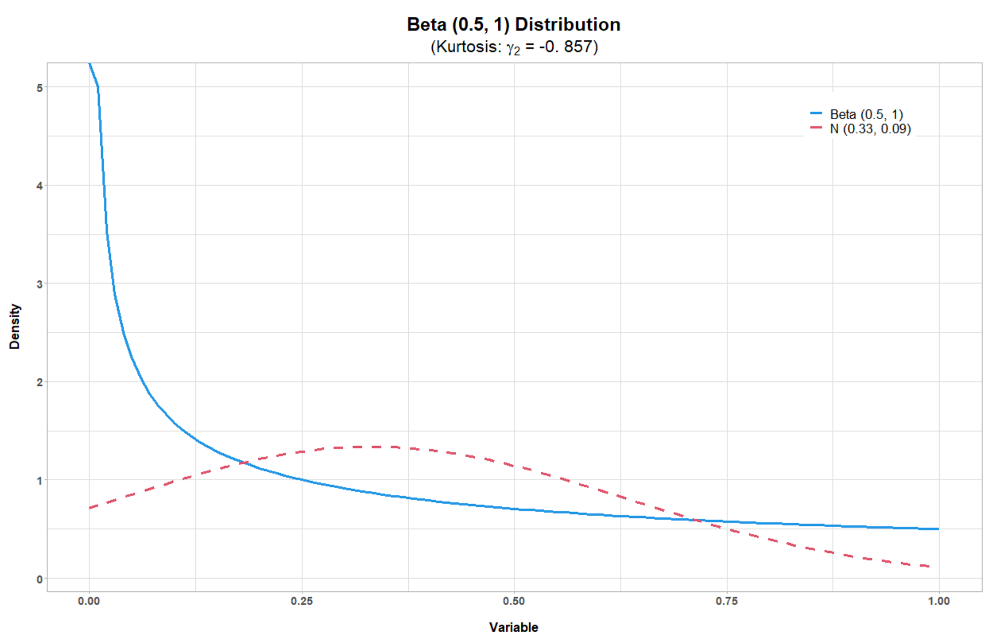


Figure 6: **PDF of Beta (0.5, 1) Distribution**

The distribution has kurtosis -0.857 (= γ_2) which is less than a normal distribution (γ_2= 0) and is therefore supposed to be *"less peaked"* than the normal but it is an infinitely peaked distribution.

- **<u>Mixture of Normal Distributions</u>**

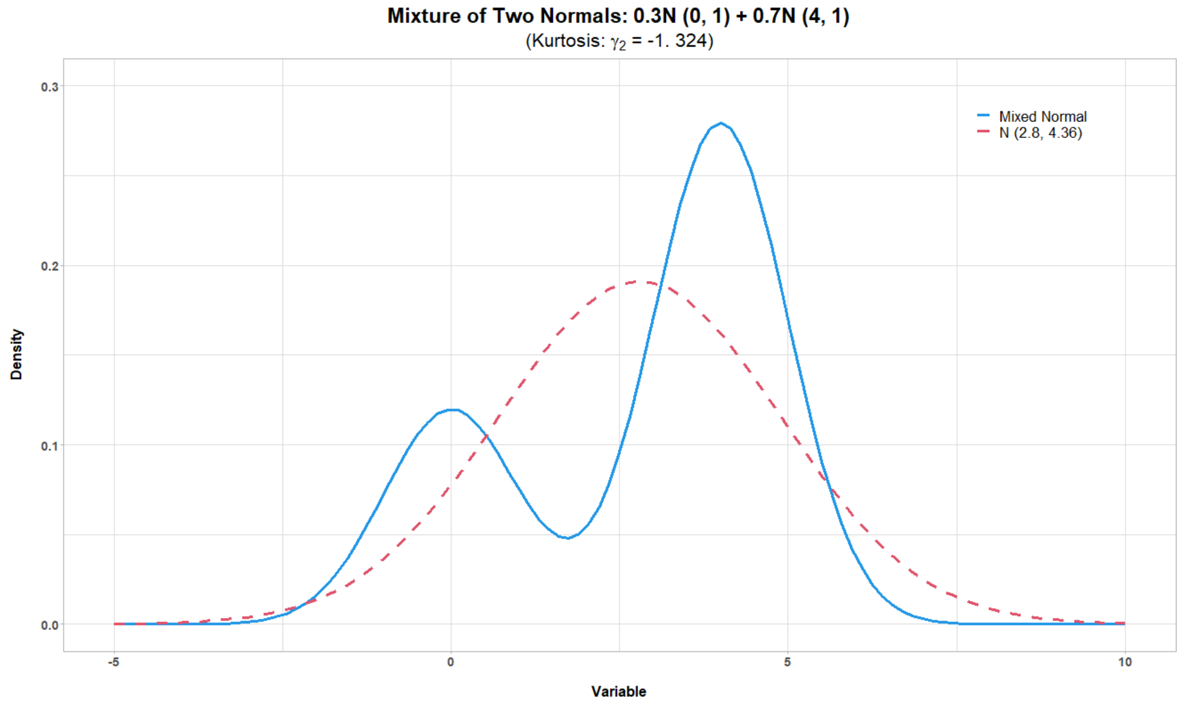


Figure 7: **PDF of Two Normal Mixtures**

The distribution has kurtosis -1.234 (= γ_2) which indicates it should be *"less peaked"* than the normal but it has relatively higher peak than normal as observed from the diagram. Again, the distribution of mixture of two normal distributions has a lighter tail than the normal.

This supports the fact that ***Kurtosis tells nothing about the 'peak' of a distribution rather it deals with the tails and outliers of a distribution***.

Does data also support the concept of "Tailedness" ?

Now, we shall deal with some numerical examples and let's see what emerges from this –

(a) Suppose the data values are - 5, 3, 3, 4, 5, 2, 5, 6, 4, 4, 1, 2, 2, 4, 3, 4, 3, 4, 6, 3.

The average of these values is 3.65 and $g_2 = -0.596$.

In this case there is no outlier and hence it gives **small kurtosis**.

(b) Now, consider this data: 42, 2, 3, 1, 1, 1, 7, 1, 1, 0, 0, 3, 6, 2, 7, 722, 1, 11, 0, 1.

The average of this values is 40.6 and $g_2 = 14.93$.

In this case there are potential outliers and hence it gives **high kurtosis**.

Conclusion

In case where there are potential outliers [as in case (b)] there will be some extremely large z_i^4 values where, $\frac{(x_i - \bar{x})}{sd(x)}$ which, when averaged with all the other z_i^4 values, gives a high kurtosis. For the values which are close to the peak (near the mean), z_i^4 values are extremely small and contribute a little to the kurtosis.

Data near the middle do not contribute much to the kurtosis statistic i.e., kurtosis does not measure the *"peakedness"* rather it is simply a **measure of detecting outlier in the sample and in case of distributions it measures the proneness of producing outliers compared to *"mesokurtic distributions"* (such as normal).**

In particular cases where the data exhibit high kurtosis, it can be said that when we draw the histogram, the peak will occupy a narrow vertical strip of the graph. The reason is there will be a very small proportion of outliers which will spread over most of the horizontal axis, which will generate such a histogram that will have sharp peak, implies a high concentration toward the mean.

Lawrence T. DeCarlo also pointed out the errors persist in many textbooks, in his article, *"...a number of textbooks, ranging from introductory to advanced graduate texts, describe positive kurtosis as indicating peakedness and light (rather than heavy) tails and negative kurtosis as indicating flatness and heavy (rather than light) tails. This is a serious error, because it leads to conclusions about the tails that are exactly the opposite of what they should be."*

A perspective of this study : Financial Risk Management

As we have clearly come to know that kurtosis of a distribution reflects the presence of outliers, it helps a lot in financial risk. If we analyze the kurtosis of the *"distribution of returns"*, we can get an idea about the risk of investing. In particular, if the kurtosis is high enough then the tail of the *"distribution of returns"* is thicker which indicates high risk for an investment since there is high chance of getting significantly high or small values of return.

If we can model the *"distribution of return"* by the distributions having lower kurtosis, then the risks become moderate. E.g. - *uniform, normal* distributions etc.

Otherwise, if we have to model it by a distribution having higher kurtosis, then the risk increases. E.g.– *Laplace* distribution, *Student's t* distribution (with lower degrees of freedom) etc.

Consider that we have a dataset of returns of ITC for the last 5 years. After standardizing it, it is modelled by a *normal* and a *t-distribution* with appropriate parameters.

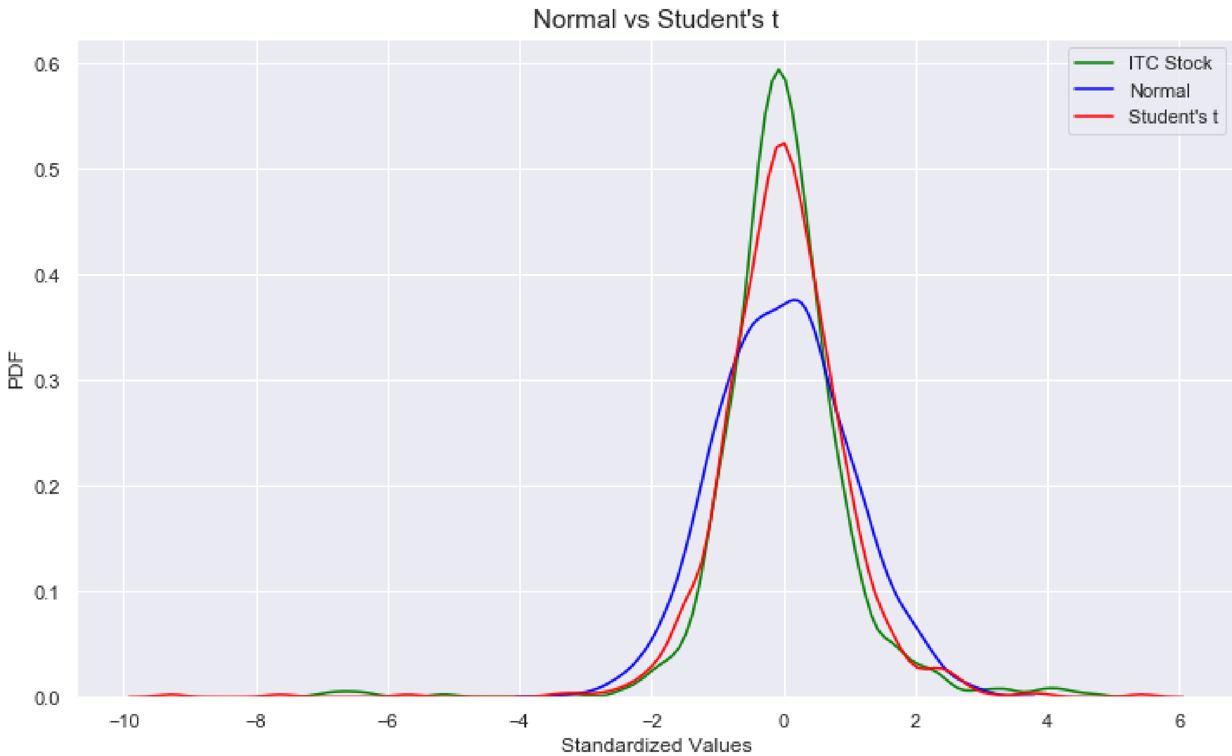
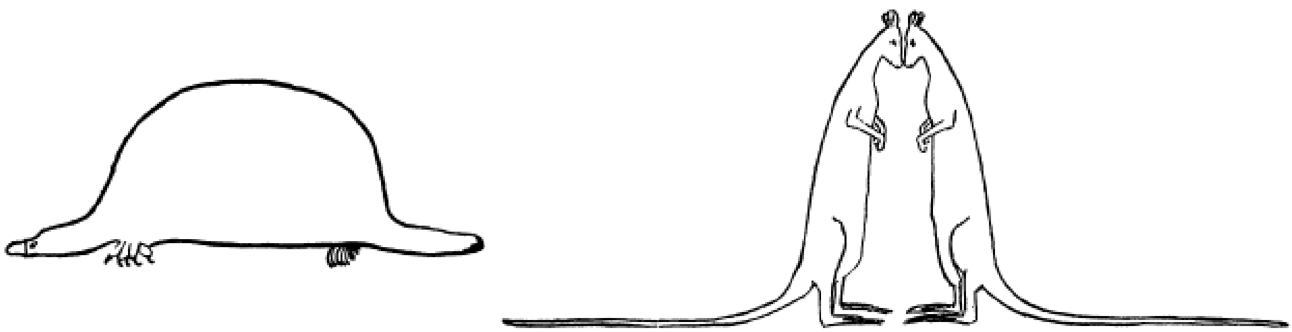


Figure 8: **ITC Stock Returns**

It is quite clear that the *Student's t-distribution* can model the data more accurately than *normal distribution*. The tail region of the distribution of return is captured well by *t-distribution*, the reason is that *Student's t-distribution* has thicker tail than that of a normal distribution for having a higher kurtosis. By modelling it by a *normal distribution*, we are basically underestimating the risk of extreme events. So, the distributions which reflects the risks more efficiently i.e., the leptokurtic ones are advantageous while calculating the risks.

* In case any of my readers may be unfamiliar with the term "kurtosis" we may define mesokurtic as "having β_2 equal to 3," while platykurtic curves have $\beta_2 < 3$ and leptokurtic > 3. The important property which follows from this is that platykurtic curves have shorter "tails" than the



normal curve of error and leptokurtic longer "tails." I myself bear in mind the meaning of the words by the above *memoria technica*, where the first figure represents platypus, and the second kangaroos, noted for "lepping," though, perhaps, with equal reason they should be hares!

Figure 9 : **Kurtosis according to William Gosset [5]**

<u>References</u>: -

[1] Wikipedia: https://en.wikipedia.org/wiki/Kurtosis

[2] "On the Meaning and Use of Kurtosis", Lawrence T. DeCarlo [1997]

[3] "Kurtosis as Peakedness, 1905 – 2014. R.I.P.", Peter H. Westfall [2014]

[4] "A Common Error concerning Kurtosis", Irving Kaplansky [1945]

[5] "Errors of Routine Analysis", William S. Gosset (a.k.a. Student) [1927]