Simulation Exercise

Tejus

Overview :

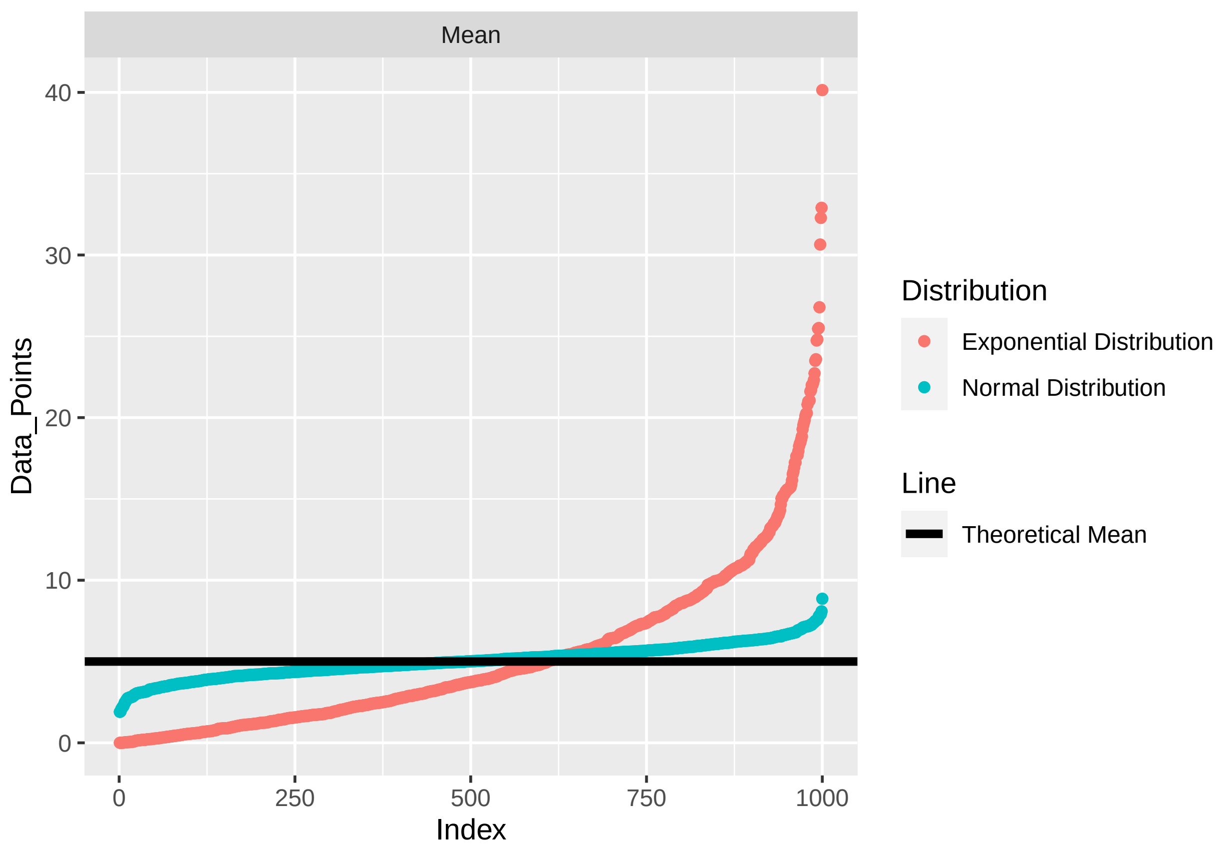
- Investigating the exponential distribution
- In R we can use the distribution using `rexp(n, lambda)`, where `lambda` is the rate parameter.
- Assuming that the population & the randomness governing our sample can be given by densities and mass functions. We work with characteristics of these densities & mass functions instead of the whole function.

Key characteristics about the distribution :

- Mean of the exponential distribution *1 / lambda* [Characterization of it's center]
- Standard deviation is also *1 / lambda* [Characterization of spread of data.]

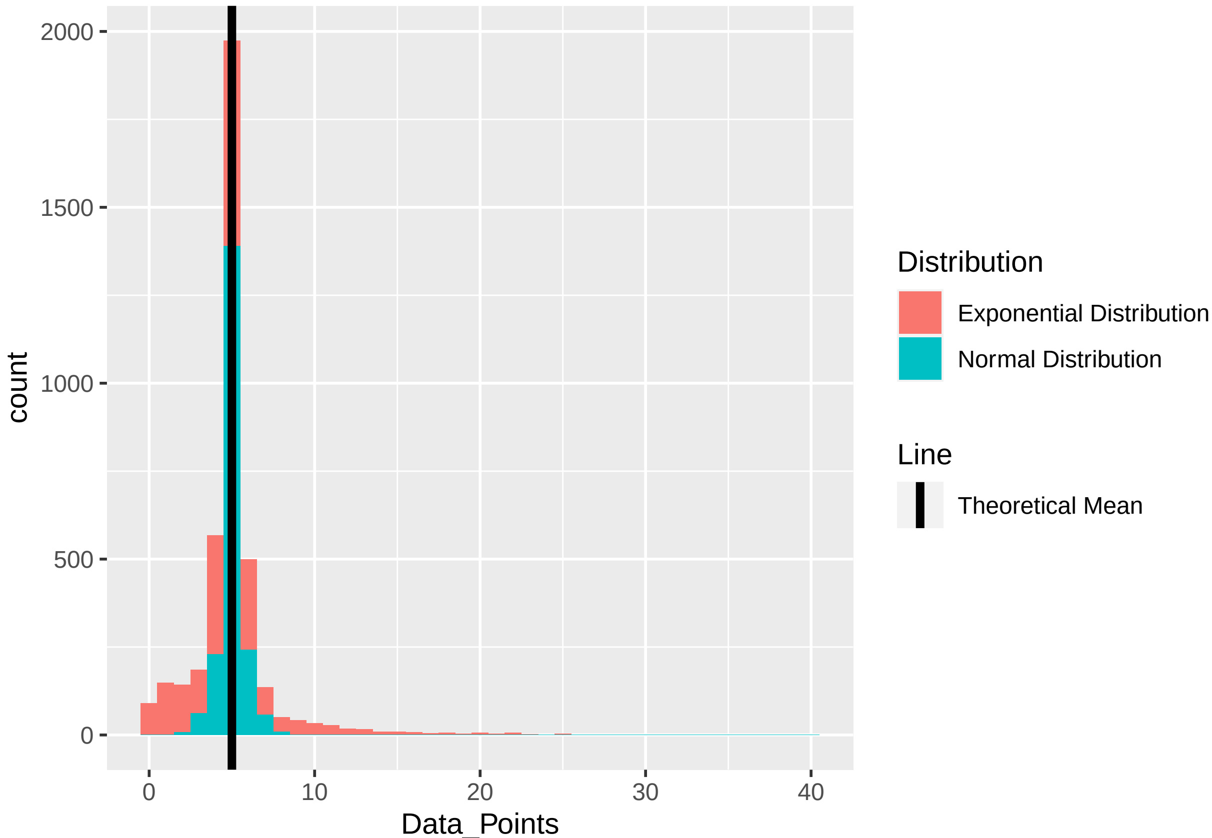
Note : Setting lambda = 2 for all the simulations & averages of 40 exponentials will be investigated by simulating 1000 times

The following plot shows the Normally distributed points along with our Exponentially distributed points both with the same mean 5. Black horizontal line is for the mean of those distributions.



- From the plot you can see that unlike Normal Distribution, data points are not symmetrical around mean. As per the name `Exponential Distribution` the data does look like it's growing exponentially.

Histogram of an typical exponential curve would look something like this



Again you can see in this histogram that the `Exponential` distribution compared to `Normal` distribution it is not symmetric around mean (black vertical line), but centered at the mean.

The sample mean and the theoretical mean of the distribution.

- The black line above is the population mean. Since we have simulated these points from the mean 5 we know the absolute value of it. However in real life experiments we want to know what that value is.

- Why ? Because as I said above that mean (true mean) of a distribution is a very good characteristic to estimate so that we get an idea about the distribution; to make inferences about the population.

Sample Mean

- Sample Mean is an estimate of **Population Mean** `(Theoretical Mean)`.

- After doing some experiments & collecting data, we take the mean of that data; that is our **Sample Mean** `the estimate for our population mean.`

Clearly the peak (sample mean) is close to the *Population mean* but not exactly the same in the first plot. It approaches the population mean as we increase the sample size (Take each value as mean of 40 such values; Inturn 4×10^4 total samples).

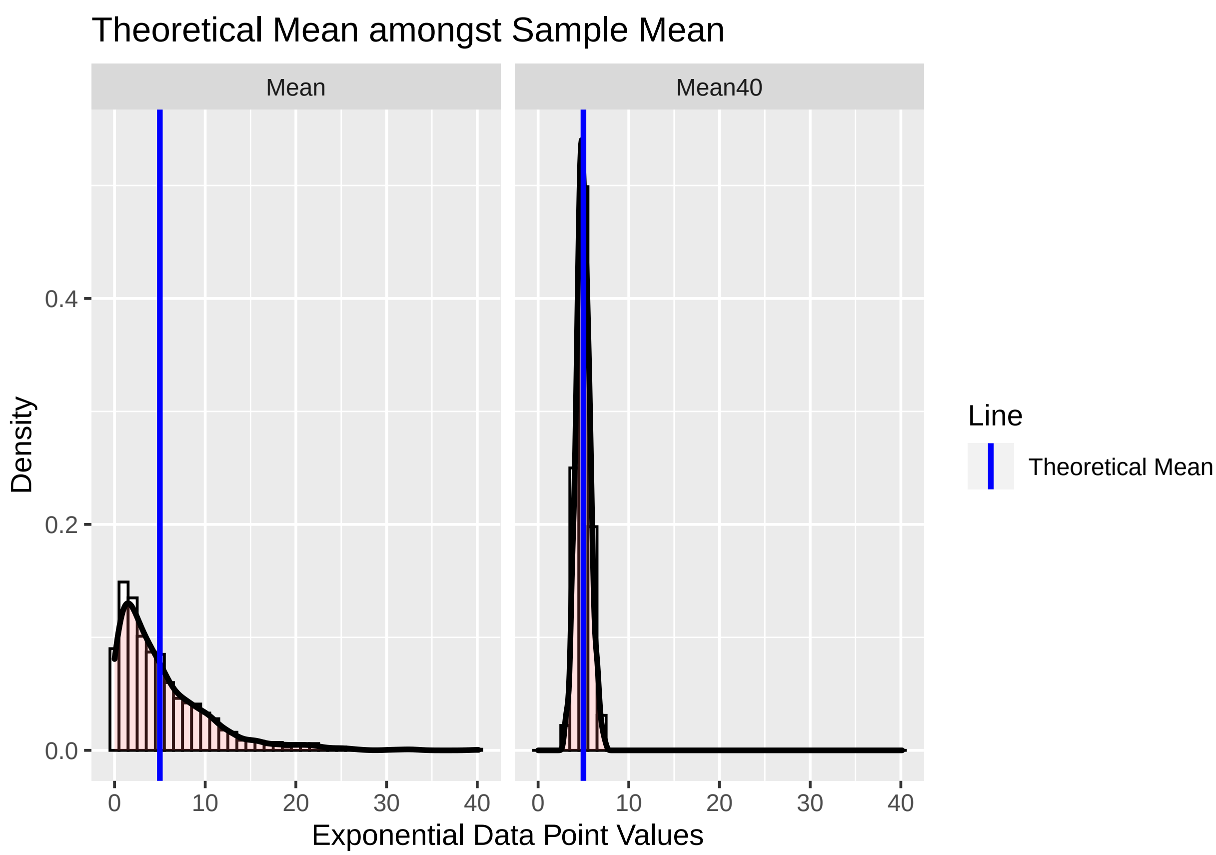


Figure 1: Distribution of 1000 simulations of 1.) Exponential Values & 2.) Mean of 40 Exponential Values

The sample variance and the theoretical variance of the distribution.

- Sample Variance is an estimate of **Population Variance**.

- After doing some experiments & collecting data the variance of that data is our **Sample variance** which is the estimate for our population variance.
- Since it's also a random variable, it must be centered around true population variance 25

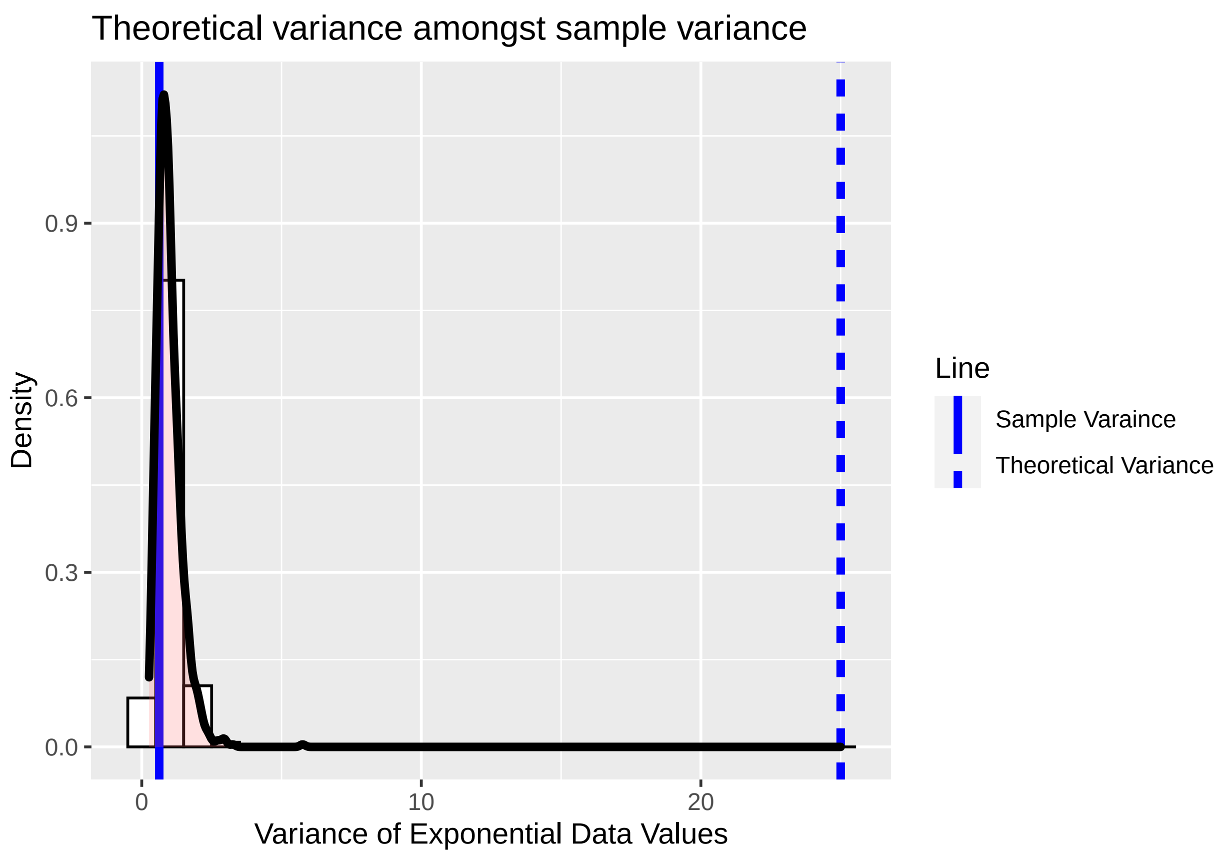


Figure 2: Distribution of variance from 1000 simulations for mean(40 values each)

Sample variance is an estimate of population variance hence as we have learned that Population variance (5) / sample size (40) gives us our **sample variance** (0.125)

It is clearly not centered at population variance because in this case it was just 40 values each. But as we increase the sample size it will approach the true **Population Variance**.

Distributions of characteristic values

In this plot we can see that, when we take the distribution of average from 40 random exponential variables we get a `Normal distribution`.

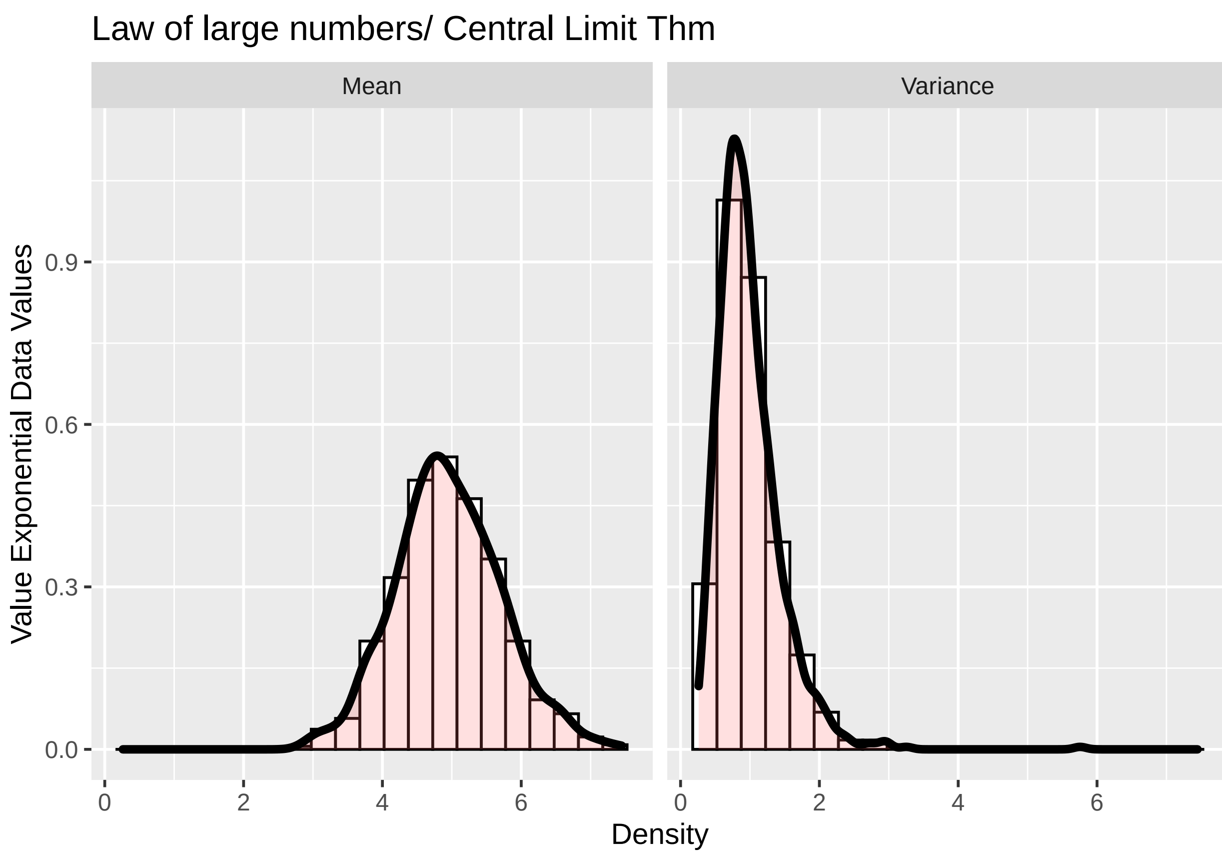


Figure 3: Distribution of mean & variance from 1000 simulations of 40 values each follows a Normal distribution centered at what it's trying to estimate (Population Mean / Population Variance)