# SDA

November 8, 2020

# 1 Statistical Data Analysis - Assignment 2

## 1.1 Group Members:


1. Ashutosh Chauhan (S20180010017)
2. Pradum Singh (S20180010136)
3. Utkarsh Ajay Aditya (S20180010182)
4. Vipul Rawat (S20180010192)


# Contents



Approach: The entire code is object oriented.

## 1.2 Base Distribution Class

A base distribution class is defined which depicts that a distribution will have parameters, a mean function, a variance function and graph plotting functions. The distributions would be derived from this base class and will have their own unique structure. The graph plotting function, the sample class and the simulation class is common for all the distributions, which has been written only once and it is being utilised by all the derived distribution classes.

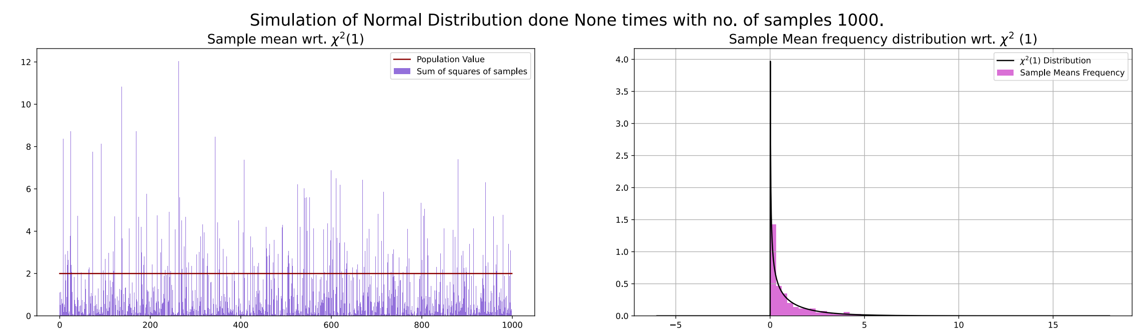
## 1.3 Sample Class

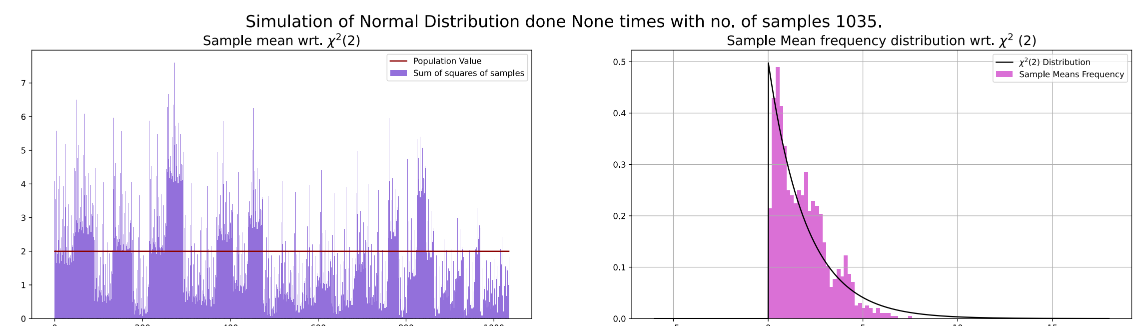
The Sample class is used for representing samples and it defines the size of sample, the distribution from which the sample is drawn, the population parameters of that distribution, a generate function to create a sample of the given distribution, a sample mean function and a sample variance function.

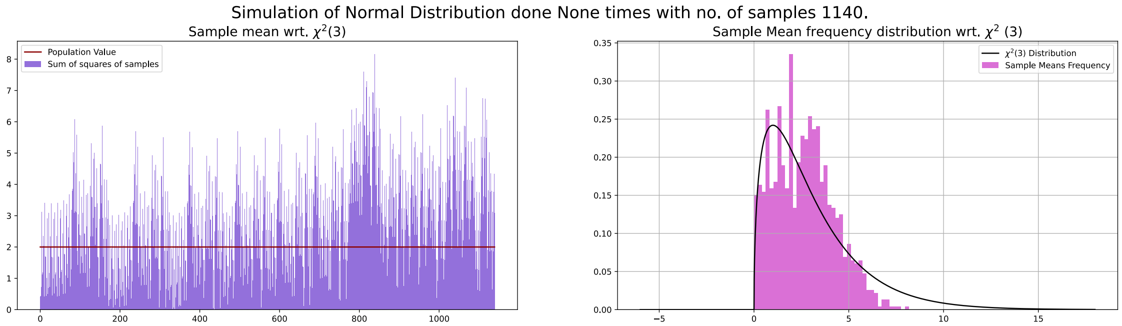
## 1.4 Simulation 1

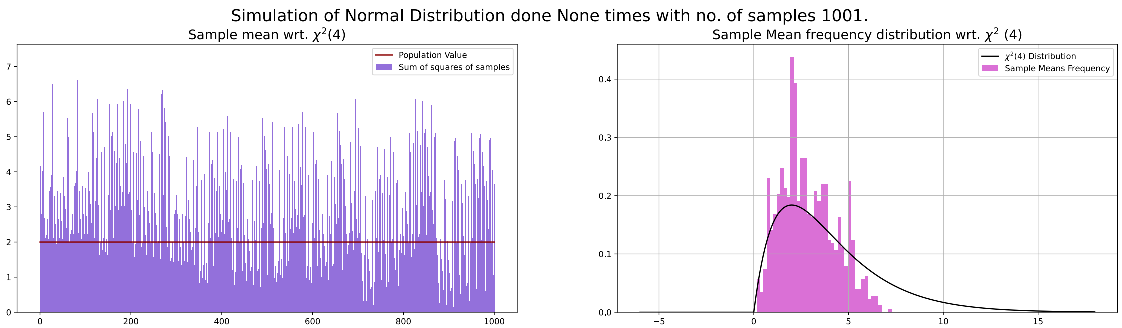
Pick 1000 samples several times from standard normal distribution $N(0, 1)$: name them St1, St2, St3, St4 etc.
Compute St12+St22 and all possible combinations with Sts. Plot the histogram and compare with the $\chi^2(2)$ pdf. Compute (St1)2 +(St2)2 + (St3)2 +(St4)2.
Compare the plot with $\chi^2(4)$ pdf. Similarly do the computation for others and plot 2 histograms for several degrees of freedom. Write the conclusion.

### 1.4.1 Chi Squared Simulation for d.o.f. 2,3,4,5






### 1.4.2 Observation from Simulation of $\chi^2$ distribution using random Normal variables

$$\chi^2(k) \approx \sum_{i=1}^{k} X_i^2, \forall X_i \sim Normal(\mu = 0, \sigma = 1)$$

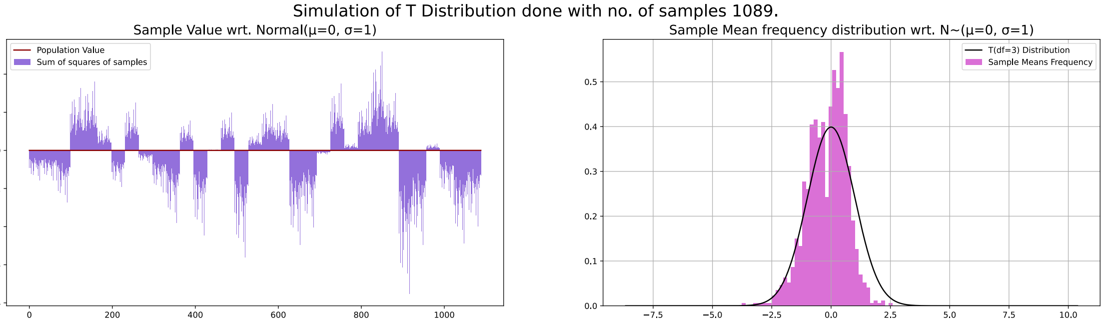
The $\chi^2(k)$ Distribution is approximately similar to sum of squares of $p$ random Normal variables.

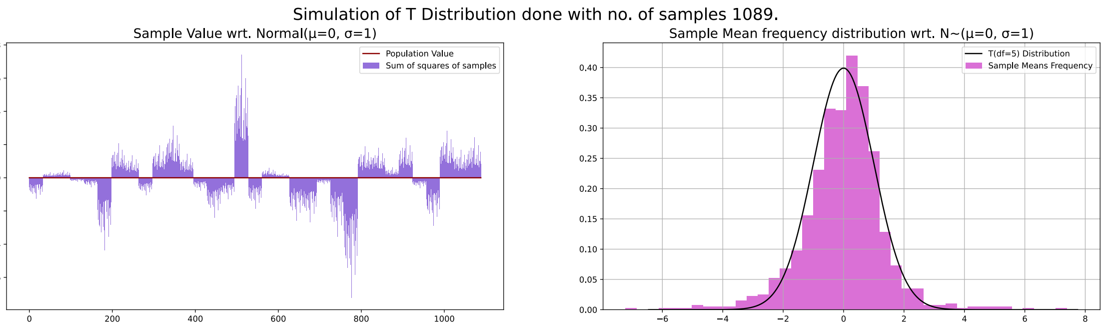
It is evident from the graphs that for a large sample size $\chi^2(k)$ is good approximate for $\sum\limits_{i=1}^{k} X_i^2$.

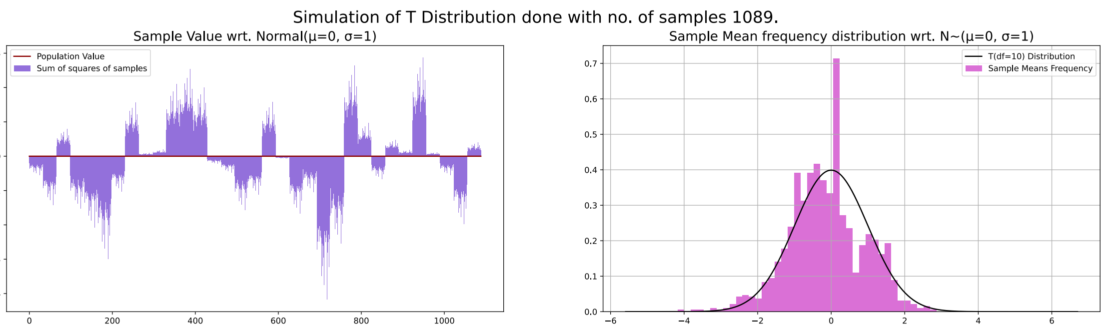
## 1.5 Simulation 2

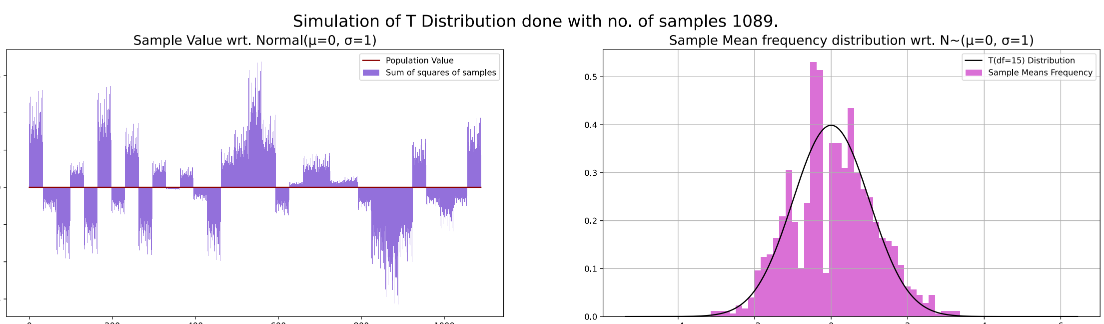
Suppose $Z \sim N(0,1)$, $V \sim \chi^2(v)$, : Degrees of freedom, Let $t = \frac{Z}{\sqrt{(V/v)}}$; Simulate $t$ taking $\chi^2(v)$ as simulated from the previous question. (Take $v$=3, 5, 10, 15 etc). Plot $t$ histograms for $v$=(3, 5, 10, 15). In the same plot show the standard normal distribution $Z$ pdf. Try to infer about the relationship between $t$ distribution and standard normal distribution $Z$.

### 1.5.1 T Distribution Simulation for df = (3, 5, 10, 15)



### 1.5.2 Observation from Simulation of T distribution using Normal and $\chi^2$ Random Variables

$$T(v) = \frac{Z}{\sqrt{(V/v)}}, \text{ where } Z \sim Normal(\mu = 0, \sigma = 1) \text{ and } V \sim \chi^2(v)$$

The $T(v)$ Distribution is approximately similar to the division of random Normal Distribution $Z$ by $\sqrt{\frac{\chi^2(v)}{v}}$.

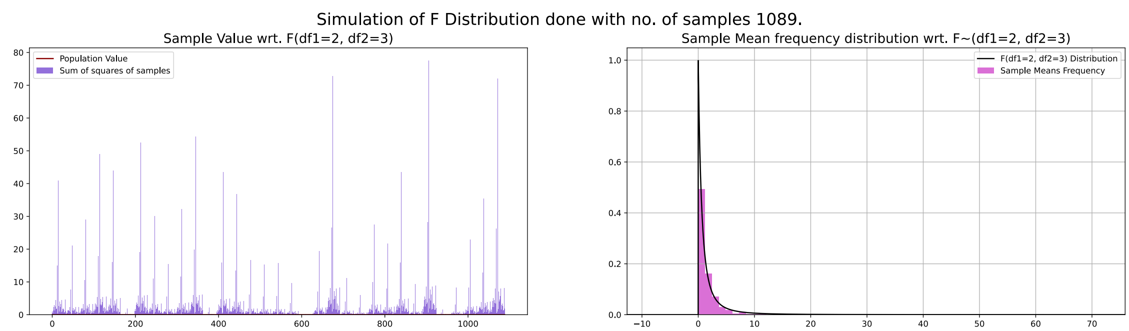
It is evident from the graphs that for a large sample size, the distribution of $T(v)$ is good approximate for $\frac{Z}{\sqrt{(V/v)}}$.

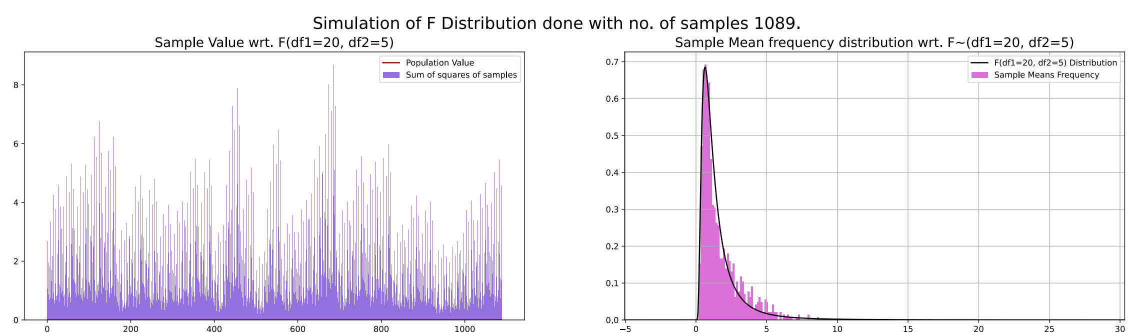
## 1.6 Simulation 3

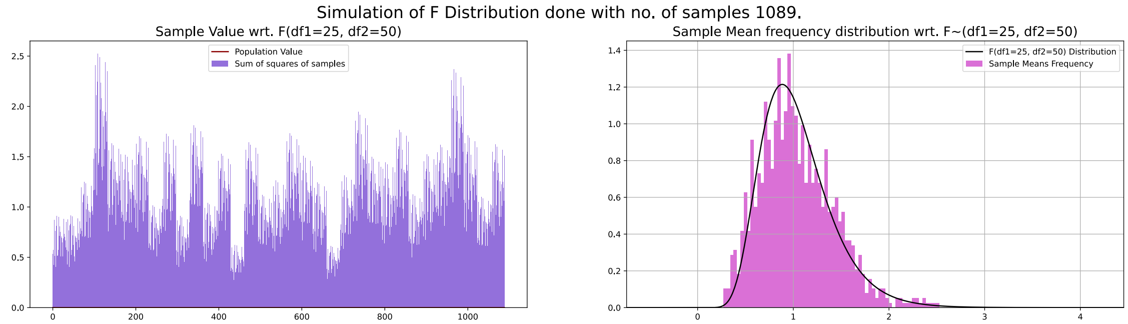
$F(n_1, n_2) = (U_1/n_1)/(U_2/n_2)$;
$U_1$ has $\chi^2$ distribution with n1 degrees of freedom.
$U_2$ has $\chi^2$ distribution with $n_2$ degrees of freedom.

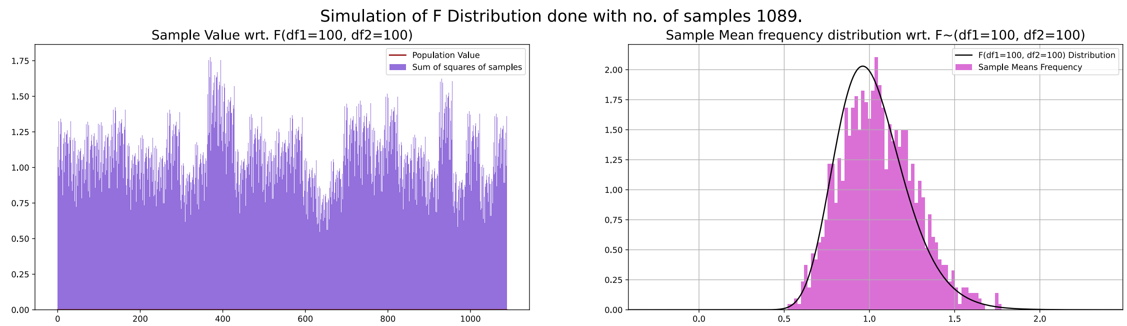
Get the corresponding simulations for $\chi^2$ from question no. 1 and plot $F(2, 3)$. Plot other $F$ histograms using the other $\chi^2$ data obtained in Q1 with different degrees of freedom.

### 1.6.1 F Distribution Simulation for df's (2,3) (20,5) (25,50) (100,100)






### 1.6.2 Observation from Simulation of F distribution using two $\chi^2$ Random Variables

$$F(u,v) = \frac{(U/u)}{(V/v)}, \text{ where } U \sim \chi^2(u) \text{ and } V \sim \chi^2(v)$$

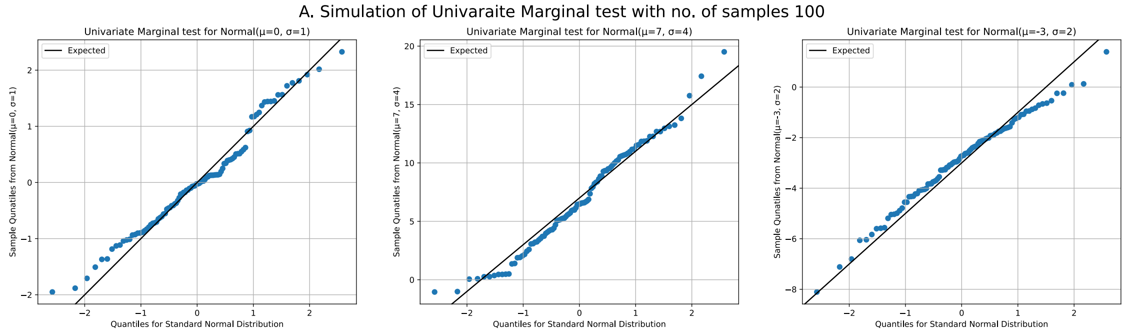
The $F(u,v)$ Distribution is approximately similar to the ratio of two $\chi^2$ distribution with degree of freedom $u$ and $v$.

It is evident from the graphs that for a large sample size, the distribution of $F(u,v)$ is good approximate for $\frac{(U/u)}{(V/v)}$.

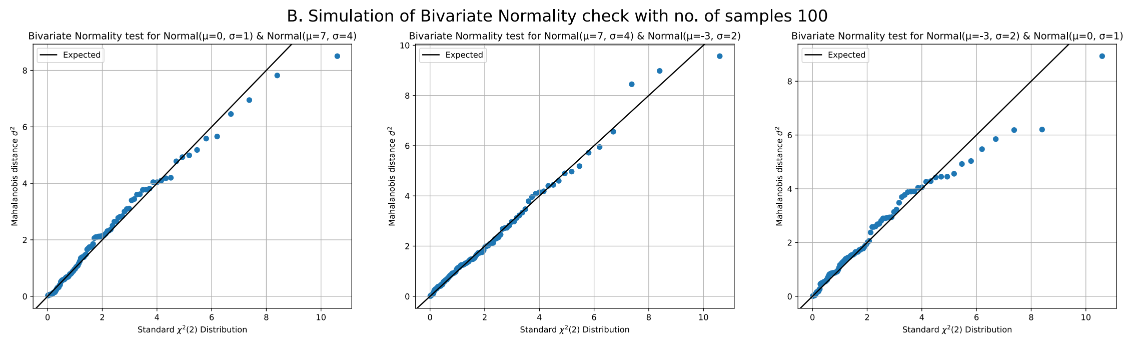
## 1.7 Simulation 4

Simulate multivariate normal data for 3 variables and 100 observations.
A. Check for univariate marginal tests by Q-Q plot.
B. Check for bivariate and multivariate normality.
The three Normal Distributions used are as follows:
* Normal(0,1) * Normal(7,4) * Normal(-3,2)

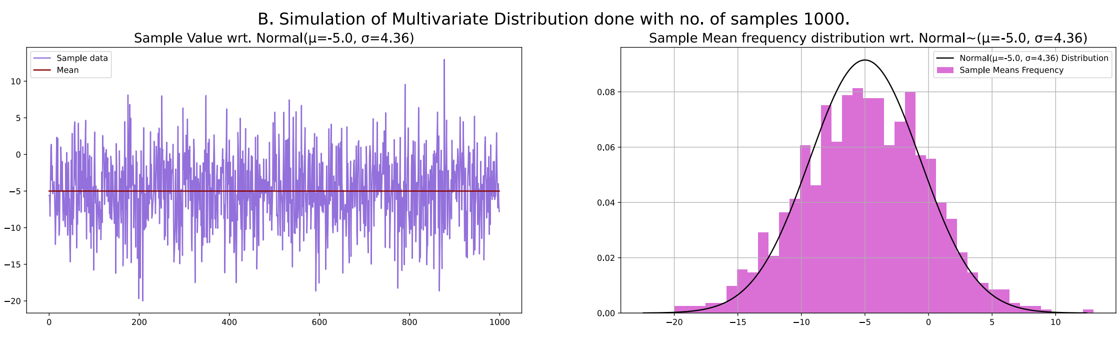
### 1.7.1 Check for univariate marginal tests by Q-Q plot.




### 1.7.2 Bivariate Normality Check




### 1.7.3 Multivariate Normality Check




### 1.7.4 Observation from Simulation Multivariate Normal Distributions with 3 Normal Random Variables

A. The Univariate Normal test produced good approximation to standard Normal Distribution using the Q-Q Plot.

B. The Bivariate test showed good approximation of Mahalanobis distance with $\chi^2(k)$.
The Multivariate test showed good approximation of linear combination of parameters with Normal Variables.