Assignment 2 - Report

Team Members:

Sathyanarayanan R - S20180010154
Darshan G - S20180010046
Jaswanth K - S20180010068
Ejurothu Pavan Sai Santhosh - S20180010053

Part-1 Simulating Chi-Square Distribution

The Chi-Squared Distribution (χ^2-distribution) is the distribution of the sum of squares of *n* standard normal random variables, where n is the degrees of freedom of the chi-square random variable (df = n).

Mean of the distribution = df
Variance of the distribution = 2 * df

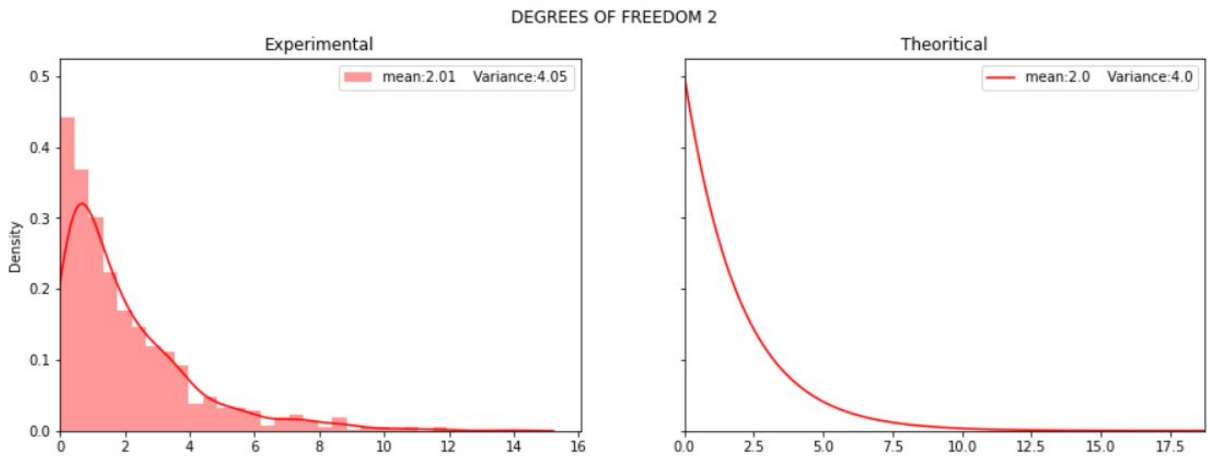
Values of Mean and Variance found by simulation and theoretical:

Degrees of Freedom, (df)	Simulated		Theoretical	
	Mean	Variance	Mean	Variance
1	1.0	1.89	1	2
2	2.01	4.05	2	4
3	2.98	5.98	3	6
4	4.07	8.76	4	8
5	5.06	10.85	5	10
6	6.12	12.69	6	12
7	7.05	14.32	7	14
8	8.13	16.6	8	16
9	9.11	18.04	9	18
10	10.12	20.27	10	20

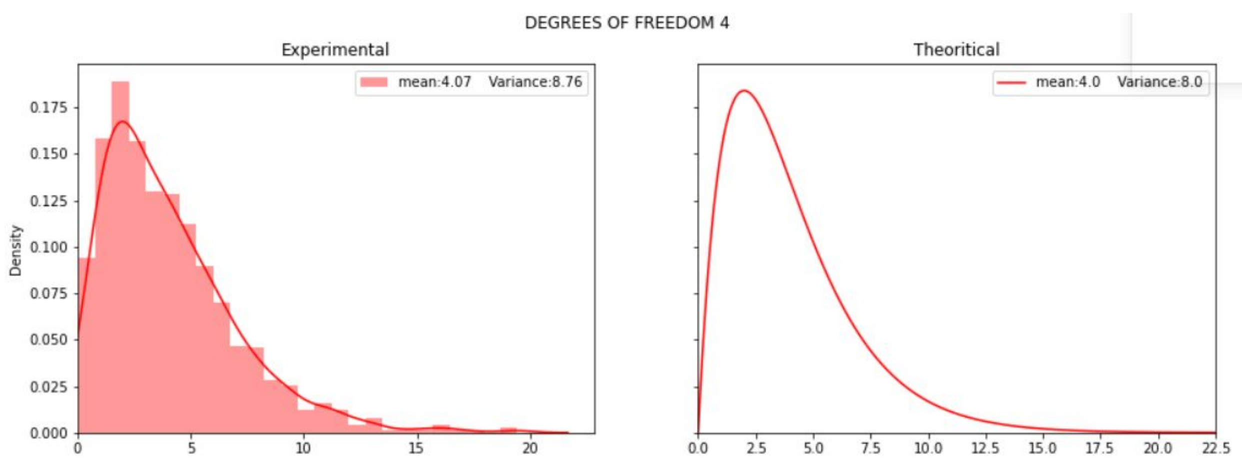
Chi-Square Distribution for the above tabulation:

Random variables were generated for the above parameters and the pdf formula f(x) was applied to each of those points to get the corresponding distribution

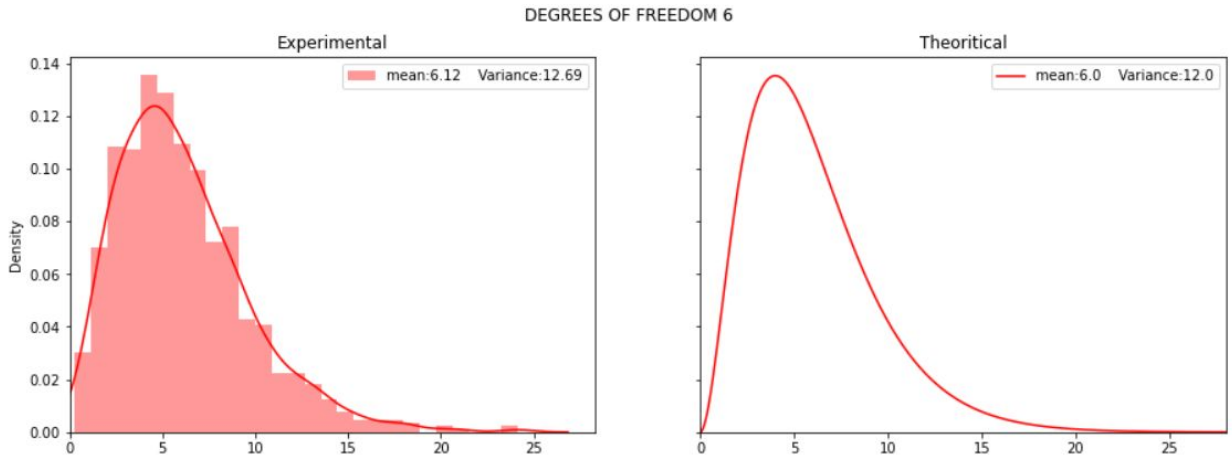
DF=2



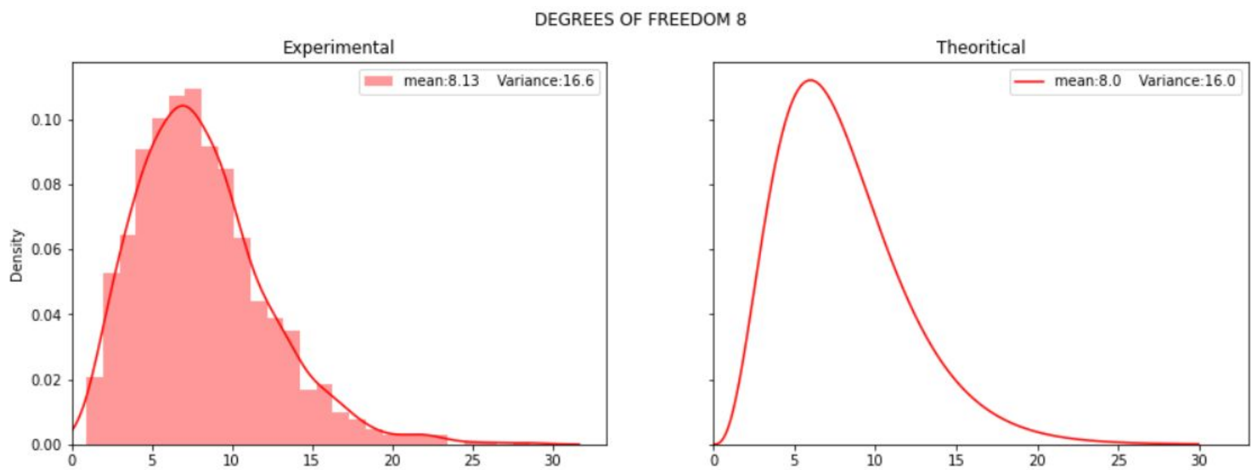
DF=4



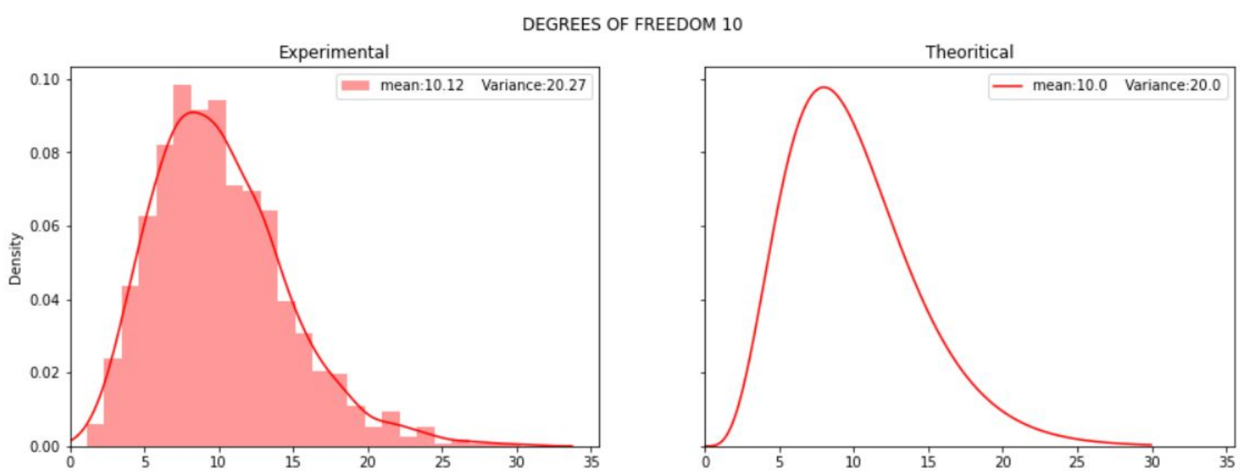
DF=6

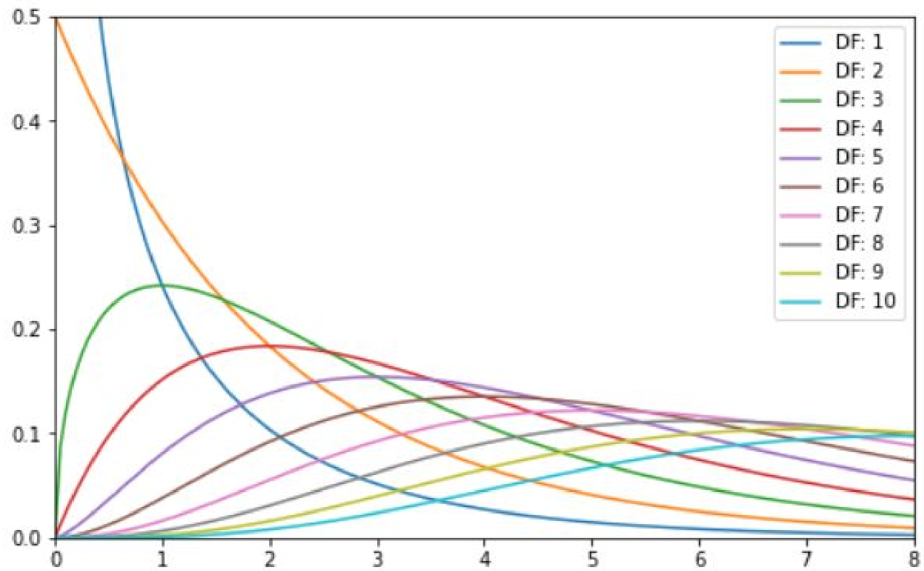

DEGREES OF FREEDOM 6
Experimental
mean:6.12 Variance:12.69
Density
Theoritical
mean:6.0 Variance:12.0

DF=8


DEGREES OF FREEDOM 8
Experimental
mean:8.13 Variance:16.6
Density
Theoritical
mean:8.0 Variance:16.0

DF=10


DEGREES OF FREEDOM 10
Experimental
mean:10.12 Variance:20.27
Density
Theoritical
mean:10.0 Variance:20.0



Conclusion:

The Chi-Square Distribution was successfully simulated by taking random samples and the parameters are found to be approximately equal to the theoretical values.

The mean of a Chi-Square distribution is equal to the DF, the variance is equal to two times DF.

The Chi-Square PDF looks more and more normal with an increase in degrees of freedom.

Part 2 - Simulating t-distributions

A t-distribution can be expressed as $\frac{z}{\sqrt{\chi^2/n}}$

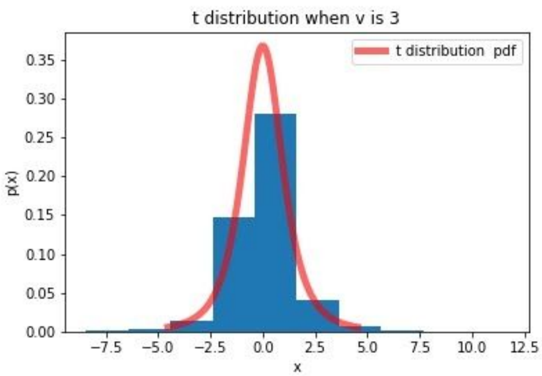
Where Z is a standard normal distribution random variable and X^2 is a chi-squared distribution with 'n' degrees of freedom.

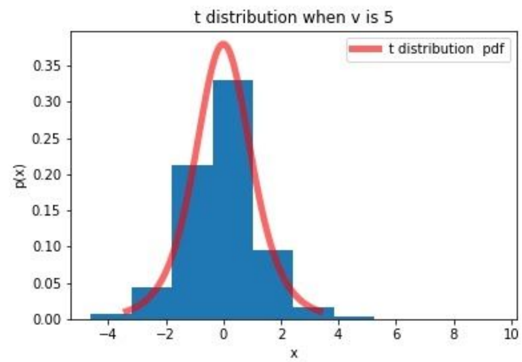
Mean of a t-distribution is zero and variance is n/(n-2)

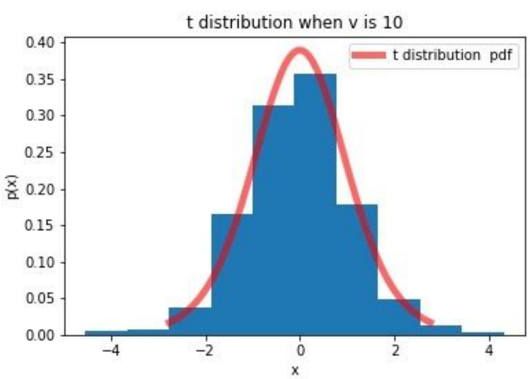
Method used :

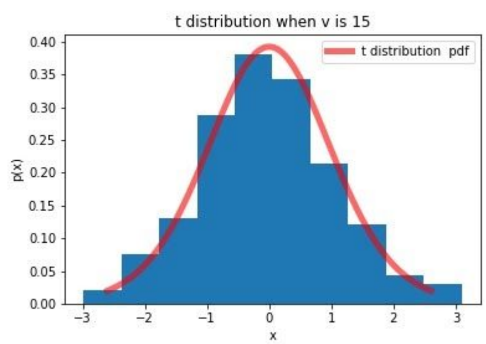
Random variables generated for Z-distribution and Chi-squared were used for each degrees of freedom and with the help of the above formula t-distribution was simulated (v - Degrees of freedom)

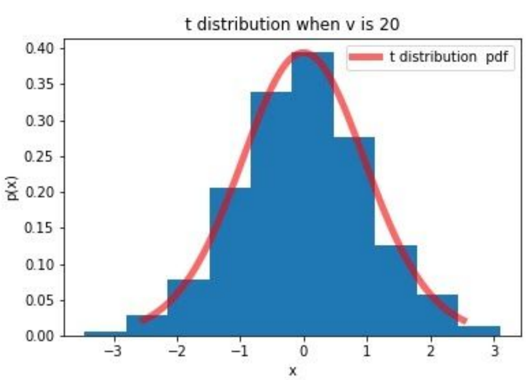
Plots showing t-distribution with different degrees of freedom

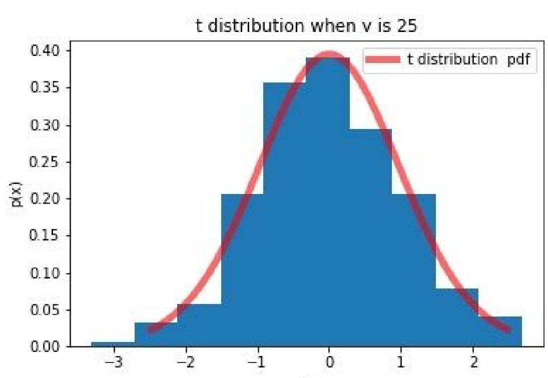




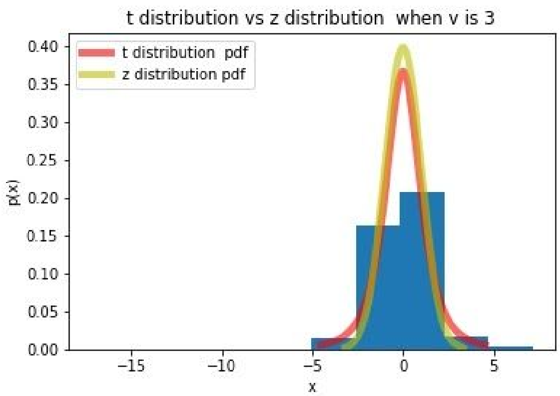


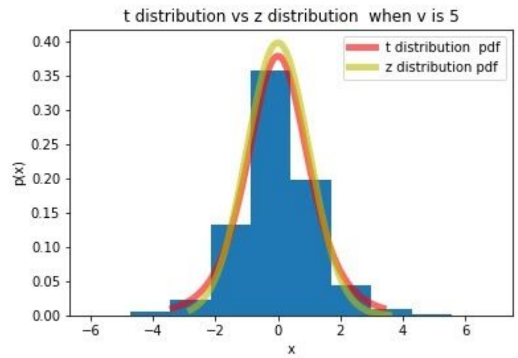


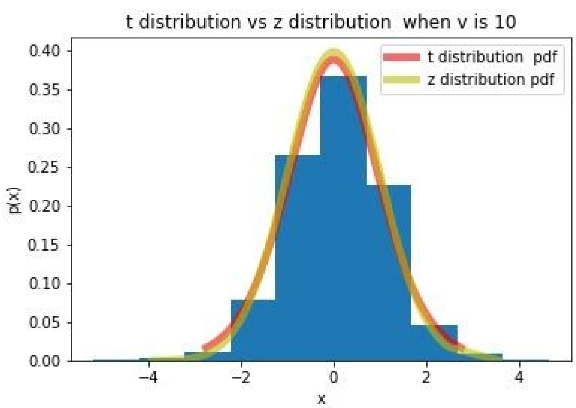


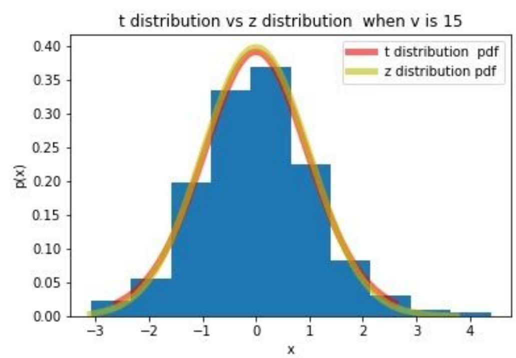


Plots comparing Standard Normal Distribution with t-distribution









Conclusion

In terms of mean both Standard normal distribution and t-distribution are centered around zero. Also, t-distribution has fatter tails than normal distribution. This means that the probability of getting values far away from the mean is larger with a t-distribution than a normal distribution. Based on the observations in the above plots above, as the degrees of freedom increase t-distribution looks similar to the normal distribution

Part-3 Simulating F - Distributions using X^2 :

The F distribution is the ratio of two chi-square distributions with degrees of freedom n_1 and n_2, respectively, where each chi-square has first been divided by its degrees of freedom.

F(n1,n2) = (U1/n1)/(U2/n2)

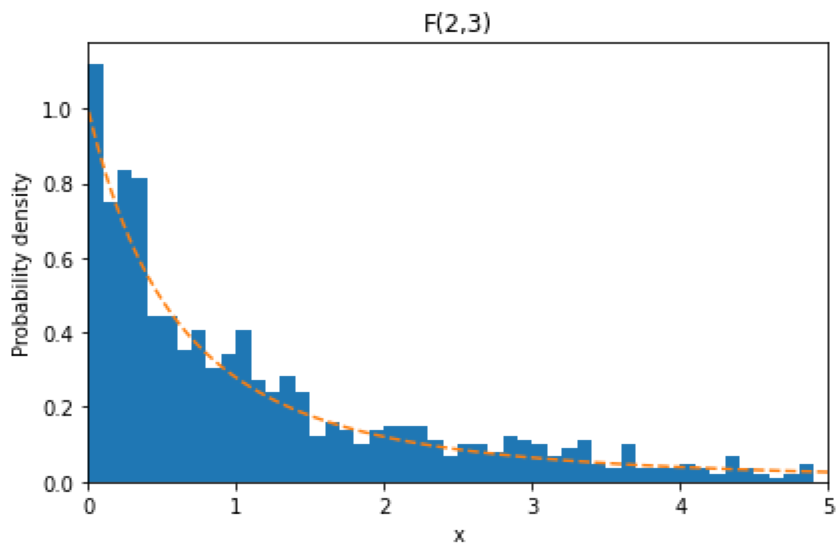
Here, U1 has X^2 distribution with n1 degrees of freedom. U2 has X^2 distribution with n2 degrees of freedom.

Method:

F(2,3) is simulated using $X^2(2)$ and $X^2(3)$ taken from part-1.

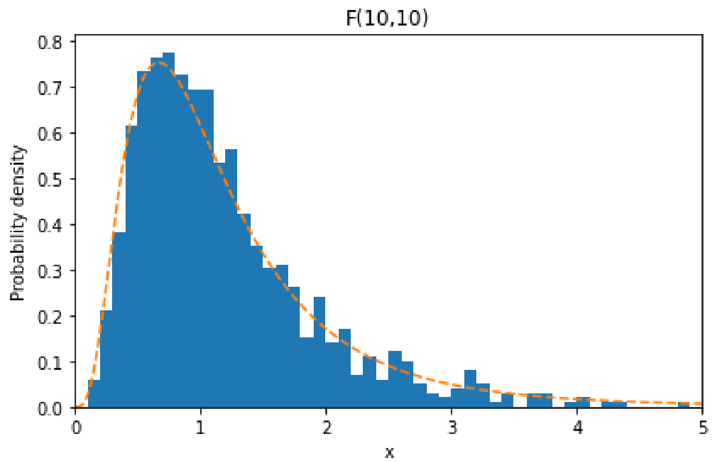
We divide chisq2 and chisq3 by their degrees of freedom and then the ratio is taken to find F(2,3).

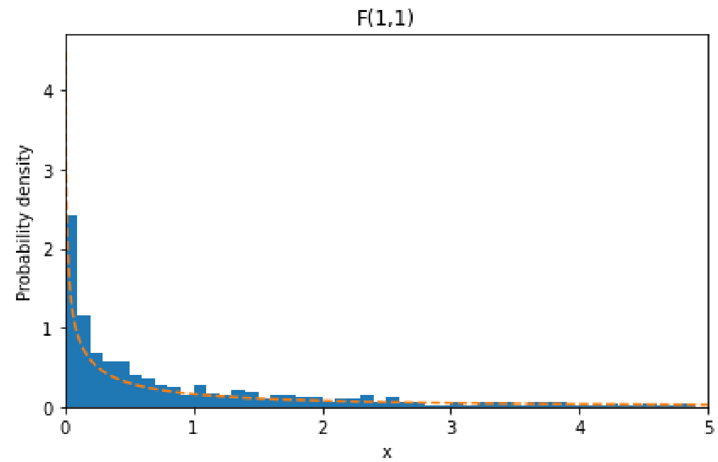
```
dfn = 2
dfd = 3
limit = 5
chn=chisq2/dfn
chd= chisq3/dfd
np.random.shuffle(chd)

x=np.divide(chn,chd)
```

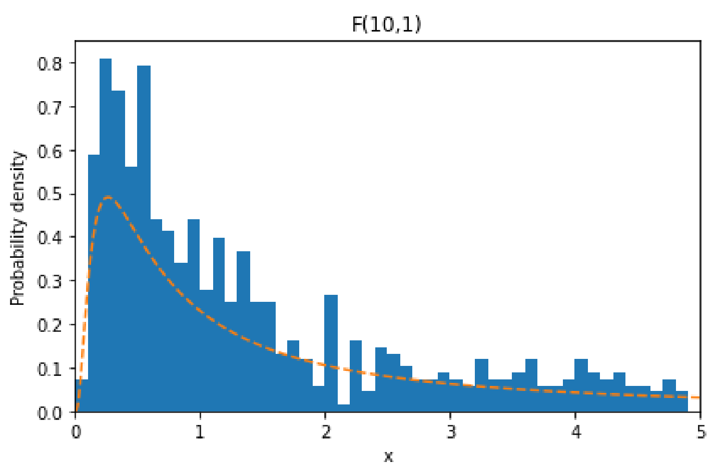


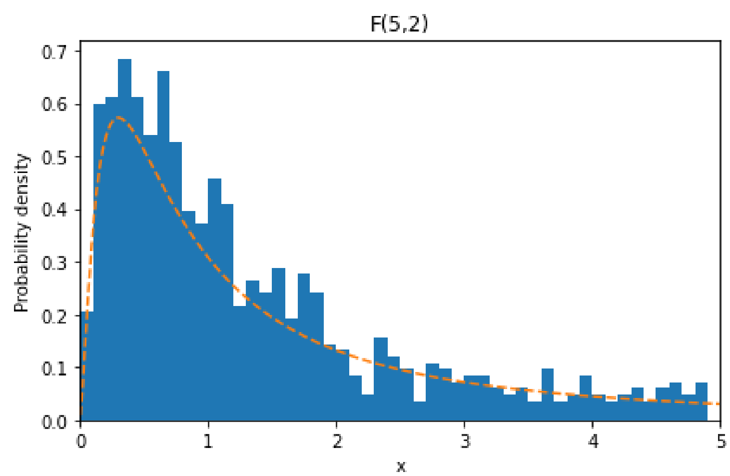
Similarly, the following F- distributions have been simulated and plotted using respective X^2 - distributions:

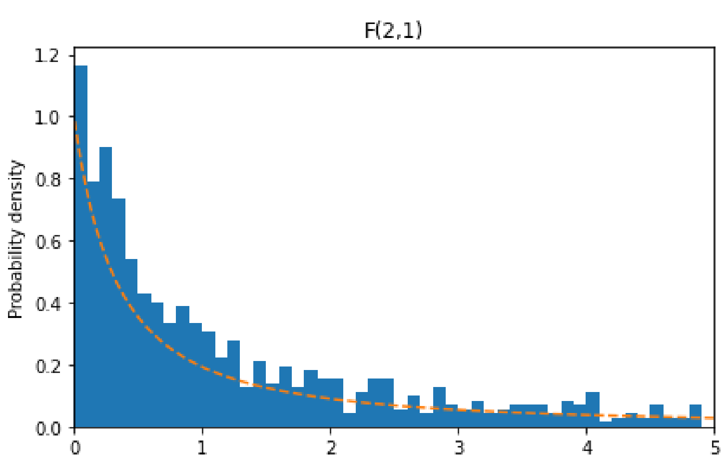
*Orange lines(--) denote theoretical pdf

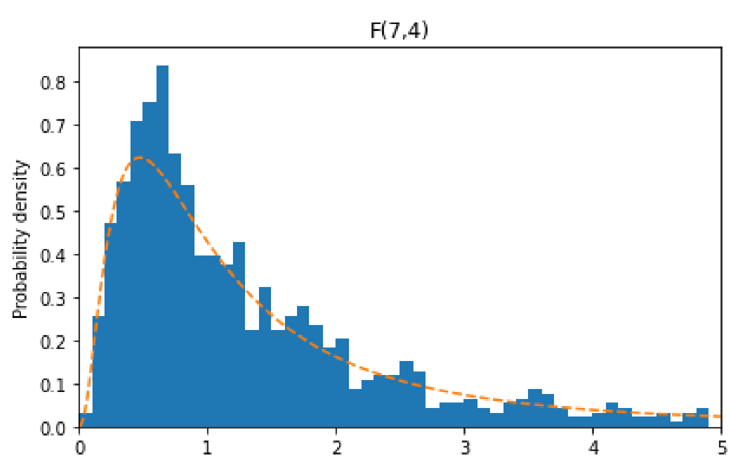












Conclusion:
Since theoretical pdf matches with simulated ones, it can be concluded that F-distribution is successfully simulated.

Part-4 Check Normality:

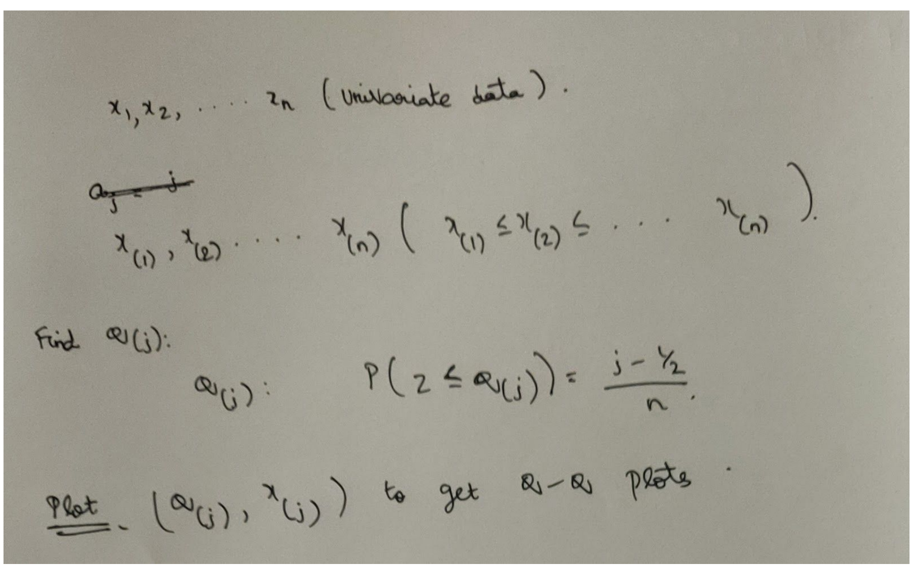
A multivariate normal distribution was simulated with 3 variables for 100 observations

Mean used to generate random variables = [5,3,7]
Covariance matrix used to generate random variables= [[4,-1,0],[-1,4,2],[0,2,9]]

A) Univariate Normality Check using Q-Q Plots:

Q-Q plots are used to check univariate normality

Method to plot Q-Q plots


$x_1, x_2, \ldots x_n$ (univariate data).

$\cancel{Q_j = j}$

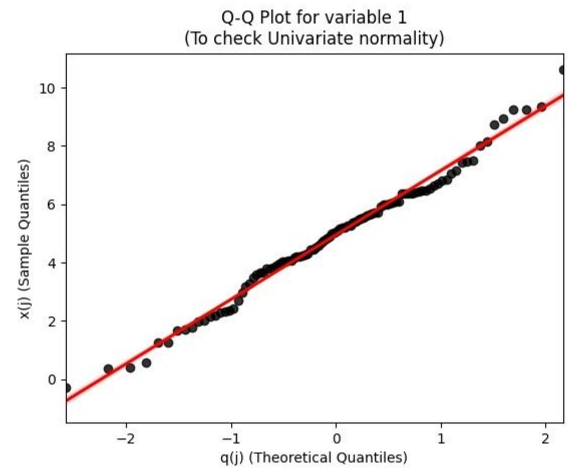
$x_{(1)}, x_{(2)} \ldots x_{(n)}$ $(x_{(1)} \le x_{(2)} \le \ldots x_{(n)})$.

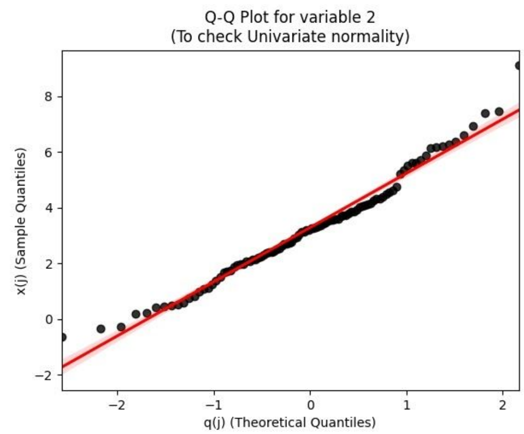
Find $Q(j)$:

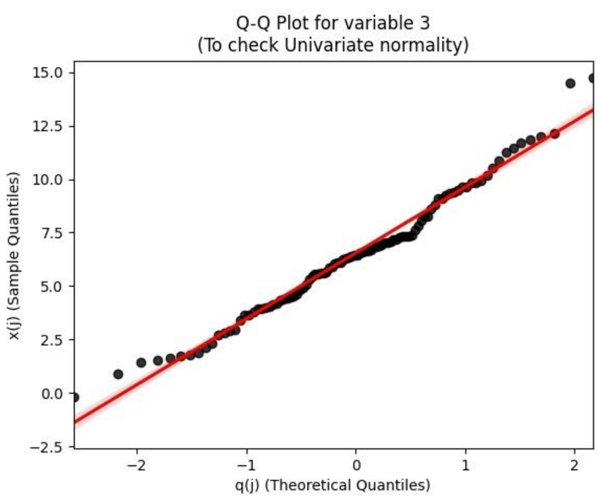
$Q_{(j)}$: $P(z \le Q_{(j)}) = \frac{j - 1/2}{n}$.

Plot $(Q_{(j)}, x_{(j)})$ to get Q-Q plots.

Figure Q-Q plots for each variable:







Observation:

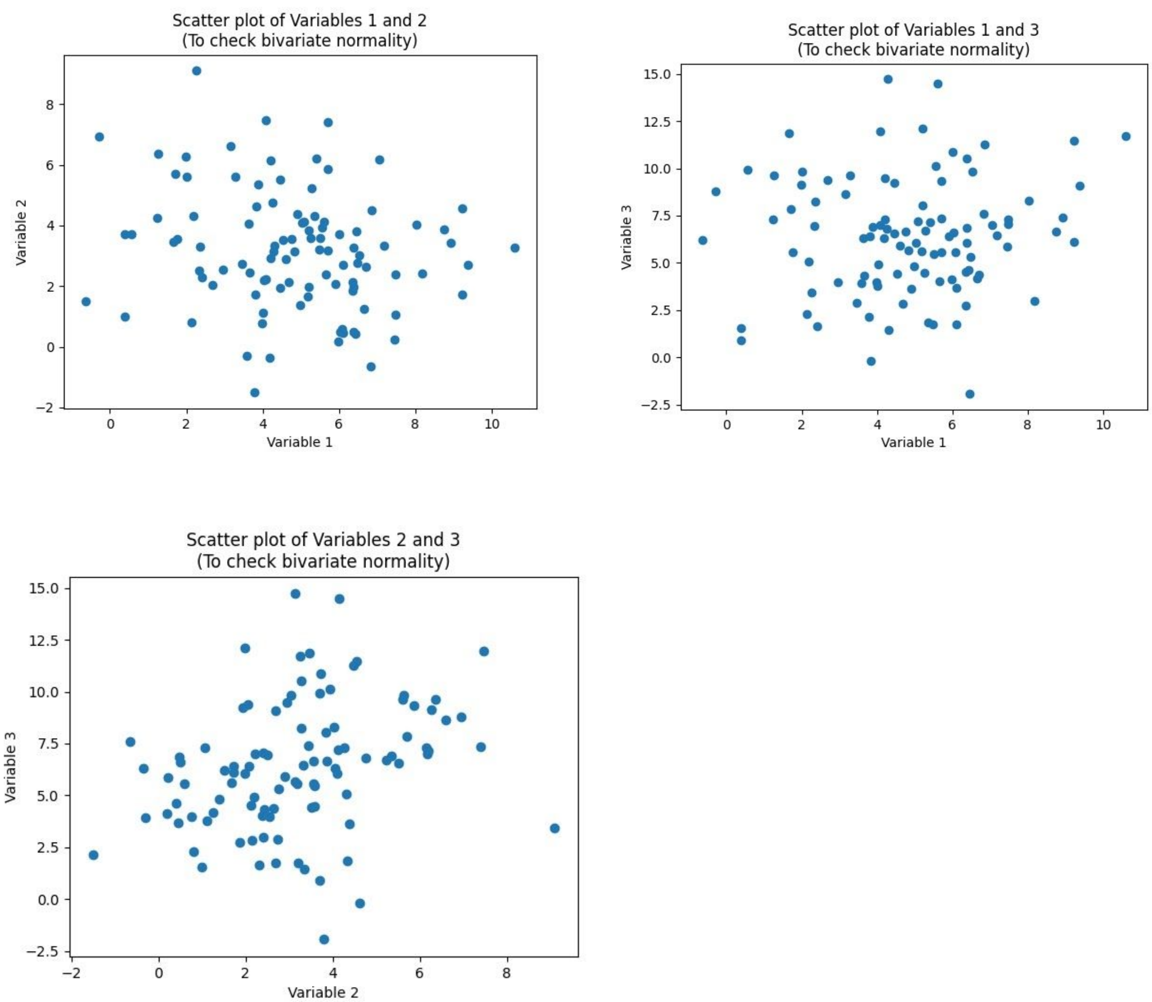
As observed all Q-Q plots have points which can be fitted with a straight line, so this shows that the given univariate data is normal

B)

Bivariate Normality Test:

To do bivariate normality test, two variables are taken at a time their scatter plots are plotted

Figure: Scatter plots of two variables at a time



Observation:

The scatter plots form an elliptical shape which confirms that two variables taken a time are bivariate normals

Multivariate Normality Test:

Method used:

$x_1, x_2, \ldots\ldots x_n$ (Multivariate data).

$$d_j^2 = (x_j - \bar{x})^{-1} S^{-1} (x_j - \bar{x}) \quad [\text{This follows chi } \chi^2 \text{ distribution}]$$

$$d_1^2, d_2^2, \ldots\ldots, d_n^2$$

Sort it,

$$d_{(1)}^2, d_{(2)}^2, \ldots d_{(n)}^2 \quad [d_{(1)}^2 \le d_{(2)}^2 \le \ldots d_{(n)}^2].$$

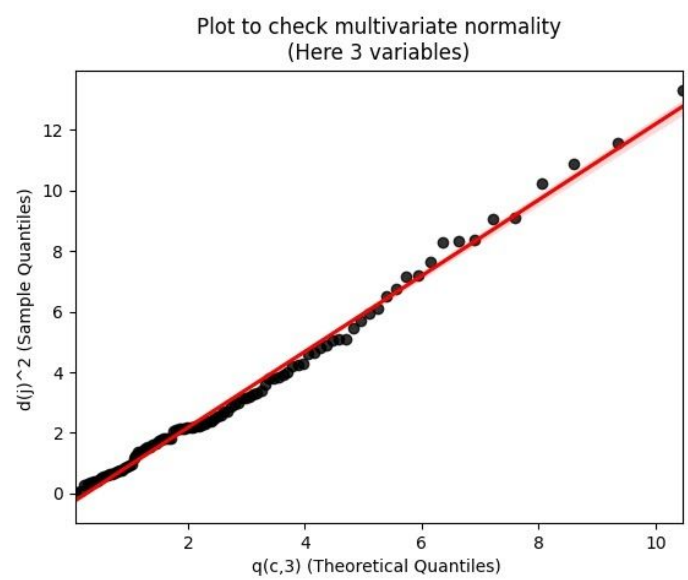
$$q_{c,p}\left(\left(j - \tfrac{1}{2}\right)/n\right) = \chi_p^2\left(\left(n - j + \tfrac{1}{2}\right)/n\right)$$

Plot $\left(q_{c,p}\left(\left(j-\tfrac{1}{2}\right)/n\right), d_{(j)}^2\right)$

Each d_j^2 follows chi-squared distribution and using chi-squared table q(c,p) is found which is plotted along with d_j^2 values

Figure:

Plot to check Multivariate Normality:



Observation:

The scatter plot forms a straight line which shows that the simulated Multivariate data (here Trivariate) is normal