ISyE 6739 – Group Activity 4

Names: Group Number:

Problem 1: Understanding Chi-squared, T, and F Distributions Through Simulations

Part a. Using R, generate 100 samples of 5 standard normal random numbers.

- Plot a histogram of all 5*100 data points.
- For each sample (row), use the definition of the Chi-squared distribution to generate one Chi-squared random number with 5 degrees of freedom. Plot the histogram for these 100 Chi-squared random numbers.
- Use R to directly generate 100 Chi-squared random numbers with 5 degrees of freedom. Plot the histogram and compare the plots.
- Find empirical $\chi^2_{0.05,5}$ from the data using `quantile` as well as the exact value of $\chi^2_{0.05,5}$ using `qchisq` function, and compare them. Note that this function works with the left tail probability.

Here are the pieces of codes you should use:

```
simdata = rnorm((1000*5), mean = 0 , sd = 1)
simdata1 = rchisq(1000, df = 5, ncp = 0)
hist(simdata1)
matrixdata = matrix(simdata, nrow = 1000, ncol = 5)
Chi2=rowSums(matrixdata^2)
par(mfrow=(c(3,1)))
hist(simdata)
hist(Chi2)
hist(simdata1)
```

Part b. Using R, generate 100 standard normal random numbers.

- Use these 100 normal random numbers and Chi-squared numbers you generated in Part a to generate 100 student's *t* random numbers.
- Use R, to directly generate 100 *t* random numbers with 5 degrees of freedom. Plot the histogram and compare the plots.
- Find empirical $t_{0.05,5}$ from the data using the `quantile` function as well as the exact value of $t_{0.05,5}$, using `qt` and compare them.

Here are the pieces of codes you should use:

```
simdata2 = rnorm(100, mean = 0 , sd = 1)
simdata3 = rt(100, df = 5, ncp = 0)
```

Part c. Using R, generate two independent series of 100 Chi-squared random numbers with 1 and 5, using `rchisq`.

- Use these 2 sets of 100 Chi-squared numbers to generate 100 *F*(1,5) random numbers.
- Use R, to directly generate 100 *F* random numbers with 1 and 5 degrees of freedom using rf. Plot the histogram and compare the plots.
- Find empirical $F_{0.05,1,5}$ from the data as well as the exact value of $F_{0.05,1,5}$ using `qf` and compare them.

Problem 2: Understanding CLT Through Simulations

Part a. Using R, generate 100 samples of *n* exponential random variables with the rate of 0.2.

- Plot a histogram of all *n**100 data points.
- Compute the average of each sample (you should have 100 average values) and plot a histogram for Xbar values.
- Repeat this procedure for various *n* values, namely (5, 10, 30, 50, 100), and find the minimum *n* whose corresponding Xbar histogram looks normal.
- In your final answer, you need to report histogram of all *n**100 data points, the minimum *n*, and its corresponding histogram.

Part b. Using R, generate 100 samples of *n* gamma random variables with the shape parameter of 10 and the scale parameter of 5.

- Plot a histogram of all *n**100 data points.
- Compute the average of each sample (you should have 100 average values) and plot a histogram for Xbar values.
- Repeat this procedure for various *n* values, namely (5, 10, 30, 50, 100), and find the minimum *n* whose corresponding Xbar histogram looks normal.
- In your final answer you need to report histogram of all *n**100 data points, the minimum *n*, and its corresponding histogram.

Compare the results of Parts a and b. What is the conclusion you can draw?

Here are the pieces of codes you should use:

```
simdata = rexp( (100*n), rate = .2)
simdata = rgamma((100*n), shape = 5, scale = 5)
hist(simdata)
matrixdata =  matrix(simdata, nrow = 100, ncol = n)
rowSums(matrixdata)
```