DSSC 221 – Probability and Statistics
Lab 7. Distributions arising from normal distribution

Data file "timeBOAK.csv" is in the blackboard.

1. We wish to inspect the fit of various distributions to data collected at Oakland International Airport. In "timeBOAK.csv", the last column filled in represents the time between arrivals (min) for aircraft arrivals at Oakland International Airport (OAK). The data set includes arrivals on 9/12/07 between 7AM and 12 noon.
   a. Plot the histogram of the time between arrivals. Compute its sample mean and standard deviation.
   b. **Calculate** and **plot** the theoretical probability density function having the mean in part (a) for the exponential distribution. (Calculate means report your pdf values, plot means display your plot).
   c. **Calculate** and **plot** the theoretical probability density function having the mean in part (a) for the uniform distribution function.
   d. **Challenge Question (Extra pts.) -** Compare your histogram with the PDFs you generated. Which PDF is more similar to your histogram? What does this tell you about how your data is more likely distributed: exponential or uniform?

2. Chi square distribution

Random variable following Chi square distribution with degree of freedom of n is the sum of n standard normal random variables squared. In other words, if $Y \sim \chi^2(n)$, then $Y = \sum_1^n X_i^2$, where $X_i \sim N(0,1)$ for all i.

a) Generate two vectors Z1 and Z2, such that each vector contains 10000 random numbers drawn from standard normal distribution
b) Create a vector $myK2 = Z1^2 + Z2^2$
c) Generate a vector K2 containing 10000 random numbers drawn from $\chi^2(2)$
d) **Challenge Question (Extra pts.) -** Plot the pdfs of myK2 and K2. Do they coincide with each other?

3. Sum of two random variables following Chi square distribution

Sum of two Chi square random variables is a chi square random variable.

a) Generate vector K3 with 10000 random numbers from $\chi^2(3)$
b) Generate vector K5 with 10000 random numbers from $\chi^2(5)$
c) Generate vector K8 with 10000 random numbers from $\chi^2(8)$

d) **Challenge Question (Extra pts.) -** Plot the pdfs of K3+K5 and K8. Compare the results.

4. Student t distribution

a. Generate vector Z with 10000 random numbers from Normal(0,1)
b. Generate vector $myT3 = \frac{Z}{\sqrt{\frac{K3}{3}}}$
c. Generate vector T3 with 10000 random numbers from T distribution with degree of freedom of 3
d. Plot the pdfs of myT3 and T3 and compare.
e. Generate vector T6 with 10000 random numbers from T distribution with degree of freedom of 6
f. Generate vector T12 with 10000 random numbers from T distribution with degree of freedom of 12
g. Generate vector T25 with 10000 random numbers from T distribution with degree of freedom of 25
h. **Challenge Question (Extra pts.) -** Plot the pdfs of T3, T6, T12, T25 and Z together. Can you see that as the higher degree of freedom gets the closer it looks like Z?