STAT 740 – Spring 2004 – Homework 3

Due: Friday, February 20^{th}

1) Consider the `newrt` function from in class for generating *t* random variables with the accept-reject method. Modify the function to output not only the random variable, but also the number of rejections needed to construct it.

   Just by thinking about the shape of the $g(x)$ function and the properties of the t-distribution would you expect more rejections to be needed for low degrees of freedom or higher? Why?

   Perform a brief simulation study where you generate the random variables using df=2, 8, and 30 and examine the number of rejections needed for each. Use 500 replications for each df and perform an appropriate statistical analysis to see if your hypothesis was correct at the α=0.05 level. Was it?

2) Many multivariate distributions are very difficult to generate, and the parameters often do not match the ones we are used to talking about. The multivariate normal is an exception however; it is described completely by its mean vector and covariance matrix. The most common method of doing this is to generate a vector of independent standard normal random variables and then apply a linear function. This is carried out in R using the function `mvrnorm` in the MASS library (use `library(MASS)` to load the library).

   Unfortunately, while this function is set up to use the estimated covariance matrix, we often relate to the distribution better in terms of the correlation matrix. Create a function `mvrnorm2` that will generate normal random variables with the specified mean vector, variance vector, and correlation matrix. (It can call `mvrnorm` to do much of the work).

3) (Continuing Problem 4 on Homework 2) Choose a distribution, combination of distributions, or Fleishman's power method (including all of the appropriate parameters) that seems to approximate the distribution of your chosen data set. Justify your choice, for example using plots, comparing the estimated moments, or a test such as Kolmogorov-Smirnov. Finally, write code (preferably in R) that will simulate data following your chosen distribution.