**Statistical Methods for Data Science**

**Mini Project 1**

**A program is divided into 3 blocks that are being compiled on 3 parallel computers. Each block takes an Exponential amount of time, 5 minutes on the average, independently of other blocks. The program is completed when all the blocks are compiled. Suppose X be the time it take the program to be compiled.**

Conduct a Monte Carlo simulation study using R to obtain the answer for E(X2). Your simulation study must involve/address the following:

* Simulate the block execution times X1, X2 and X3. Use the simulated values to simulate X2, the execution time of the whole program.
* Repeat the previous step 10,000 times. This will give you 10,000 draws from the distribution of X.
* Make a histogram of the draws of X. Superimpose the theoretical density function of X. Try using the R function `curve’ for drawing the density. Note what you see.
* Use the draws to estimate E(X2). Compare your answer with the exact answer of E(X2). Note what you see.
* Repeat the process of obtaining an estimate of E(X2) five times. Compare each estimate with the exact value. Note what you see.
* Comment on how your results would change if you use 1,000 Monte Carlo replications instead of 10,000. What if you use 100,000 replications? Justify your answers.