

Statistical Methods for Data Science (Fall 2016)

Mini Project 1

The goal of this exercise is to consider the setting of Exercise 4.6 in Homework 1, where we had to do a theoretical calculation to compute an expected value, and compute this expected value by Monte Carlo simulation using R. We will also compute a probability. Your simulation study must involve/address the following:

1. Simulate the block execution times X_1 , X_2 and X_3 . Use the simulated values to simulate X , the execution time of the whole program.
2. Repeat the previous step 10,000 times. This will give you 10,000 draws from the distribution of X . Try to avoid 'for' loop. Use 'replicate' function instead. Save these draws for reuse in later steps. [**Bonus:** 2 bonus points for not taking more than 1 line of code for the first two steps.]
3. Make a histogram of the draws of X . Use 'hist' function in R. Superimpose the density function obtained in Exercise 4.6. Try using the R function 'curve' for drawing the density. Note what you see.
4. Use the draws in Step 2 to estimate $E(X)$. Compare your answer with the exact answer obtained in Exercise 4.6. Note what you see.
5. Use the draws in Step 2 to estimate the probability that the entire program takes more than 20 minutes to compile. We can compute the exact value of this probability but it is not needed here.
6. Repeat the process of obtaining an estimate of $E(X)$ in Step 4 and an estimate of the probability in Step 5 five times. Note what you see.
7. Repeat Step 6 using 1,000 and 100,000 Monte Carlo replications instead of 10,000. Make a table of results. Comment on what you see, and provide an explanation.

Instructions:

- Due date: Thursday, September 15.
- Total points = 10
- Submit a typed report.
- You can work on the project either individually or in a group of no more than two students. In case of the latter, submit only one report for the group, and include a description of the contribution of each member.
- Do a good job.
- You must use the following template for your report:

Mini Project #

Name

Names of group members (if applicable)

Contribution of each group member

Section 1: Give an overview of how you are going to simulate the experiment. Be sure to provide a stepwise description the algorithm in **simple** language.

Section 2: Provide answers to the specific questions asked in the exercise.

Section 3: Provide the R code. Your code must be annotated. No points may be given if a brief look at the code does not tell us what it is doing.