Statistical Methods for Data Science (Fall 2018)

Mini Project 1

Instructions:

- Due date: Sep 6, 2018.
- Total points = 20.
- Submit a typed report.
- Justify all steps and provide all relevant explanations.
- 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.
- It is OK to discuss the project with other students in the class (even those who are not in your group), but each group must write its own code and answers. If the submitted report (including code and answer) is similar (either partially or fully) to someone else's, this will be considered evidence of academic dishonesty, and you will referred to appropriate university authorities.
- 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. Answers to the specific questions asked
  Section 2: 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.

---

1. (10 points) Consider Exercise 4.11 from the textbook. In this exercise, let X_A be the lifetime of block A, X_B be the lifetime of block B, and T be the lifetime of the satellite. The lifetimes are in years. It is given that X_A and X_B follow independent exponential distributions with mean 10 years. One can follow the solution of Exercise 4.6 to show that the probability density function of T is

$$f_T(t) = \begin{cases} 0.2\exp(-0.1t) - 0.2\exp(-0.2t), & 0 \leq t < \infty, \\ 0, & \text{otherwise,} \end{cases}$$

and $E(T) = 15$ years.

   (a) Use the above density function to analytically compute the probability that the lifetime of the satellite exceeds 15 years.

   (b) Use the following steps to take a Monte Carlo approach to compute $E(T)$ and $P(T > 15)$.

      i. Simulate one draw of the block lifetimes X_A and X_B. Use these draws to simulate one draw of the satellite lifetime T.

      ii. Repeat the previous step 10,000 times. This will give you 10,000 draws from the distribution of T. Try to avoid 'for' loop. Use 'replicate' function instead. Save these draws for reuse in later steps. [**Bonus**: 1 bonus point for not taking more than 1 line of code for steps (i) and (ii).]

iii. Make a histogram of the draws of T using 'hist' function. Superimpose the density function given above. Try using 'curve' function for drawing the density. Note what you see.

iv. Use the saved draws to estimate $E(T)$. Compare your answer with the exact answer given above.

v. Use the saved draws to estimate the probability that the satellite lasts more than 15 years. Compare with the exact answer computed in part (a).

vi. Repeat the above process of obtaining an estimate of $E(T)$ and an estimate of the probability four more times. Note what you see.

(c) Repeat part (vi) five times 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

2. (10 points) Use a Monte Carlo approach estimate the value of π based on $10,000$ replications. [**Ignorable hint**: First, get a relation between π and the probability that a randomly selected point in a unit square with coordinates — $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$ — falls in a circle with center $(0.5, 0.5)$ inscribed in the square. Then, estimate this probability, and go from there.]