Statistical Methods for Data Science (Spring 2017)
Mini Project 1

1. Suppose a random variable X has the following probability density function: f(x) equals 4*x^3 when x is between 0 and 1, and equals 0 otherwise.

(a) Compute E(X), Var(X) and P(X > 0.5) analytically, i.e., using their formulas.
(b) Explain how you would simulate a draw from the distribution of X.
(c) Approximate E(X), Var(X) and P(X > 0.5) using Monte Carlo simulation with 1,000 draws 5 times. Summarize the results in a table.
(d) Repeat (c) with 10,000 draws.
(e) Compare you results in (a), (c) and (d). Explain, with justification, what you observe.

2. IQ test scores have a population mean and standard deviation of 100 and 15, respectively. Assume that the scores follow a normal distribution.

(a) Compute the 95-th percentile of this distribution the usual way.
(b) Suppose your IQ score equals the percentile you computed in (a). What does this mean?
(c) Explain how you would simulate a draw from the distribution of the IQ scores.
(d) Approximate the 95-th percentile of the distribution using Monte Carlo simulation with 1,000 draws 5 times.
(e) Repeat (d) with 10,000 draws.
(f) Compare your results in (a), (d) and (e). Explain, with justification, what you observe.

Instructions:

- Due date: Thursday, February 2.
- Total points = 20.
- 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. 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.