

## Statistical Methods for Data Science (Fall 2017)

### Mini Project 3

---

#### Instructions:

- Due date: October 12, 2017.
- 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.
- 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 be 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. Suppose we would like to estimate the parameter  $\theta (> 0)$  of a Uniform  $(0, \theta)$  population based on a random sample  $X_1, \dots, X_n$  from the population. In the class, we have discussed two estimators for  $\theta$  — the maximum likelihood estimator,  $\hat{\theta}_1 = X_{(n)}$ , where  $X_{(n)}$  is the maximum of the sample, and the method of moments estimator,  $\hat{\theta}_2 = 2\bar{X}$ , where  $\bar{X}$  is the sample mean. The goal of this exercise is to compare the mean squared errors of the two estimators to determine which estimator is better. Recall that the *mean squared error* of an estimator  $\hat{\theta}$  of a parameter  $\theta$  is defined as  $E\{(\hat{\theta} - \theta)^2\}$ . For the comparison, we will focus on  $n = 1, 5, 10, 50, 100$  and  $\theta = 1, 5, 10, 50, 100$ .
  - (a) Explain how will you compute the mean squared error of an estimator using Monte Carlo simulation.
  - (b) For a given combination of  $(n, \theta)$ , compute the mean squared errors of both  $\hat{\theta}_1$  and  $\hat{\theta}_2$  using Monte Carlo simulation with  $N = 1000$  replications. *Be sure to compute both estimates from the same data.*
  - (c) Repeat (b) for the remaining combinations of  $(n, \theta)$ . Summarize your results graphically.
  - (d) Based on (c), which estimator is better? Does the answer depend on  $n$  or  $\theta$ ? Explain. Provide justification for all your conclusions.