STAT 440, Take home final

due 5.5.2015 at 6 pm

Note that all work must be done only by you. You cannot discuss this take home final with any of your classmates or anyone else. First,load the data. This can be done by using the command `load("final.Rdata")`. You should now have variable `x` with a single set of 40 data points. What you need to turn in for this final is a word document with any explanations and plots and R code in a .R file. Both of these should be your last name with the appropriate extension.

1. Using on the `sample()` command (nothing from bootstrap packages), do the following with the enclosed data set.

   (a) Use non-parametric bootstrap to give an approximate histogram for the sampling distribution of the median.

   (b) Use non-parametric bootstrap to approximate the MSE for the median of the data. Report your result in your word file and give your code in your .R file.

   (c) Use non-parametric bootstrap to approximate the bias for the median of the data. Report your result in your word file and give your code in your .R file.

2. Suppose that your data consists of 40 observations from a beta distribution

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

   where $0 < x < 1$ with two parameters, α and β. Write R code with `optim()` to find the maximum likelihood estimates for α and β. (Be sure to initialize `optim` with values of the parameters larger than one. Submit your estimated values along with the approximate variance-covariance matrix in the word file. Submit the code that you use in the .R file.

3. With only 40 observations, the variance-covariance matrix may not accurately reflect the variation in your estimates. Use `\rbeta()` to simulate 1000 new samples (of size 40 each). Then answer the following questions with comments in the word file and by providing corresponding code for each in your .R file

   (a) Are the MLE biased?

   (b) Does the variance-covariance matrix from the previous problem reflect the true variance-covariance in the sampling distributions.

   (c) The maximum likelihood theory implies that the sampling distribution of the MLE are normal when the data is large enough. Are the sampling distributions normal with these 40 observations? Provide plots to support your conclusion.