MATH 252-01: Probability and Statistics II

Problem Set 2

Assigned 2016 August 30
Due 2016 September 6

Show your work on all problems! If you use a computer to assist with numerical computations, turn in your source code as well.

1 Devore Chapter 6, Problem 10

2 Devore Chapter 6, Problem 22

3 Devore Chapter 6, Problem 32

4 Computational Exercise: Bootstrapping

The *Bootstrap* technique provides a model-free way to estimate the error associated with a point estimate from a reasonable-sized sample. If the sample values are $\{x_i\} = x_1, x_2, \ldots, x_n$, we create a total of B bootstrap samples $\{x_i^{(1)}\}$, $\{x_i^{(2)}\}$, $\ldots$, $\{x_i^{(B)}\}$. Each sample has size n, and each is generated by drawing *with* replacement from $\{x_i\}$. (So in general, a given bootstrap sample will have some repeated values, and some values not represented.) To estimate the error associated with the sample mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ of the original sample, we calculate the means of the B bootstrap samples, $\overline{x}^{(1)} = \frac{1}{n}\sum_{i=1}^{n} x_i^{(1)}$, $\overline{x}^{(2)} = \frac{1}{n}\sum_{i=1}^{n} x_i^{(2)}$, $\ldots$, $\overline{x}^{(B)} = \frac{1}{n}\sum_{i=1}^{n} x_i^{(B)}$, and then take the (sample) variance of $\overline{x}^{(1)}, \overline{x}^{(2)}, ..., \overline{x}^{(B)}$,

$$s_{\overline{x}}^2 = \frac{1}{B-1}\sum_{j=1}^{B}(\overline{x}^{(j)} - \overline{\overline{x}^*})^2$$

where $\overline{\overline{x}^*} = \frac{1}{B}\sum_{j=1}^{B} \overline{x}^{(j)}$ is the average of the means of the bootstrap samples. The bootstrap estimate of the error associated with the original sample mean $\overline{x}$ is then $\sqrt{s_{\overline{x}}^2}$.

a. Download the following data set which is a sample of size $n = 5$
`http://ccrg.rit.edu/~whelan/courses/2016_3fa_MATH_252/data/ps02_prob4_small.dat`
using the username and password given in class; generate $B = 8$ bootstrap samples by randomly choosing 8 sets of 5 values each from the original dataset. Be sure to turn in a table containing these 8 bootstrap samples.

b. Calculate the mean of each of your 8 bootstrap samples.

c. Calculate the bootstrap error as the sample standard deviation of this set of 8 bootstrap means. Note that this is not a robust use of the bootstrap method, since we have a small number of samples, but it's a way to see how the calculation works explicitly.

d. Repeat the calculation using the data set
`http://ccrg.rit.edu/~whelan/courses/2016_3fa_MATH_252/data/ps01_prob4.dat`
which has size $n = 121$, with $B = 200$ bootstrap samples. (You don't need to print out the

full set of 200 × 121 values, just calculate the bootstrap error estimate and document the procedures.) If you have trouble automatically generating the bootstrap samples, you may use the datafile

`http://ccrg.rit.edu/~whelan/courses/2016_3fa_MATH_252/data/ps02_prob4_resampled.dat`

which contains 200 columns, each with a 121-row bootstrap sample.

e. The data were actually generated from a Gamma(1.5,15) distribution; use this to calculate the standard error $\sqrt{V(\overline{X})}$. (*Hint*: the sample mean of a sample of size n from a Gamma(α,β) distribution is a statistic which follows a Gamma($n\alpha$,β/n) distribution.)