

## Statistics 624: Statistical Computing

### Homework # 9

Due Tuesday, 16 Oct., 8am

In this assignment, please complete, organize, and turn in **electronically** to `stat624@stat.byu.edu`.

This assignment demonstrates library usage in C, random number generation in C and usage of the sweep operator.

1. Estimate the sampling distribution of the mean ( $\bar{X}$ ) of the Landau distribution using C. Present the sampling distribution for samples of size ( $n_1 = 10, n_2 = 50, n_3 = 100, n_4 = 500$ , and,  $n_5 = 1000$ ). Note: to approximate the sampling distribution, use 10000 samples of size  $n_i$ .
  - Include in your output, the sampling distribution mean, standard deviation, median, first quartile and third quartile.
  - Include a plot of the density estimator (in R) for the sampling distribution of the sample mean.
  - Plot the standard deviation of the sampling distribution of  $\bar{X}$  as a function of  $n$  and include that plot in your output.

The sweep operator on the  $k$ th diagonal defined for a  $n \times n$  matrix  $\mathbf{A}$  is

$$b_{ij} = \begin{cases} 1/a_{kk} & \text{if } i = j = k \\ a_{kj}/a_{kk} & \text{if } i = k \\ -a_{ik}/a_{kk} & \text{if } j = k \text{ but } i \neq k \\ a_{ij} - a_{ik}a_{kj}/a_{kk} & \text{if } i \neq k \text{ and } j \neq k \end{cases}$$

2. Write a C program with the following features:
  - read an  $m \times m$  matrix from a file
  - function that takes argument  $\mathbf{a}$  (a matrix) and  $\mathbf{k}$  (an integer) that performs the sweep operation on the  $\mathbf{k}$ th diagonal
  - write the resulting matrix from sweeping on  $k_1, k_2, \dots, k_\ell$  on  $\mathbf{A}$  to a file
3. Test your code on the data in `xtxtest.txt`. The columns in this file represent features of homes in Iron County, UT. In order: sales price, lot size, number of floors, construction quality (1=poor, 4=excellent), roof condition (1=poor, 4=excellent), home condition (1=poor, 4=excellent), square feet, year built, age, number of baths, garage (0=no, 1=yes), basement (0=no, 1=yes).

- (a) Use R to construct the matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Y} \\ \mathbf{Y}'\mathbf{X} & \mathbf{Y}'\mathbf{Y} \end{bmatrix}$$

where  $\mathbf{Y}$  is the  $n \times 1$  vector containing sales price and  $\mathbf{X}$  is the  $n \times p$  matrix containing all the remaining columns of `xtxtest.txt` augmented with a column of ones. Create a file containing  $\mathbf{A}$  to read into your C program.

- (b) Compute  $\hat{\beta}$ ,  $s^2$ , and  $\mathbf{X}'\mathbf{X}^{-1}$  for the regression model containing the  $Y$ -intercept, lot size, square feet, and number of baths using the sweep operator. How do the results compare to the same computation in R?