Stat 580 (Spring 2015): Homework 4

Due, in class, March 12, 2015

Please submit your homework with codes (hard copy) in class and upload the corresponding codes to the Blackboard. Problems marked with * will be graded in detail and they are worth 50% of the total score. Remaining problems, worth the remaining 50% of the total score, will be given full mark if reasonable amount of work is shown.

For this homework, use R for programming parts unless otherwise specified.

1. * Write a C program that performs a linear regression and returns the regression coefficients. This program should read in the data specified by the user via command line arguments. It should have only two arguments:

```
Arguments: data intercept
       data: data file
  intercept: 1=intercept, 0=no intercept
```

   A sample data file ("reg.dat") is posted in the Blackboard. Here are the first 5 lines of "reg.dat":

```
5.1 3.5 1.4
4.9 3 1.4
4.7 3.2 1.3
4.6 3.1 1.5
5 3.6 1.4
```

   Each line corresponds to the measurements from the same subject. For each line, the first element is the response while the remaining ones are the covariates. Take the first line as an example, the response is 5.1 and there are two covariates being 3.5 and 1.4 respectively. Assume that the data file is well formatted (i.e. you don't have to perform any checking on the data file), number of elements are the same in each line and the covariates are all continous variables. Note that the sample size and number of predictors depend on the data file. The output of the sample data file should be: (the first line is the input)

```
> ./a.out reg.dat 1
Sample size and number of predictors are 150 and 2 respectively.
The regression coefficients: 2.249140 0.595525 0.471920
```

   Note that you may have to check the C standard library for some useful functions (about reading data from files).

2. * Use Monte Carlo integration to evaluate each of the following integrals:
   (a) $\int_0^1 x^2 dx$
   (b) $\int_0^1 \int_{-2}^2 x^2 \cos(xy) dx dy$
   (c) $\int_0^\infty \frac{3}{4} x^4 e^{-x^3/4} dx$

3. Let
$$I = \frac{1}{\sqrt{2\pi}} \int_1^2 e^{-x^2/2} dx.$$
   Estimate I using importance sampling. Take g to be $N(1.5, \nu^2)$ with $\nu = 0.1, 1$ and 10. Plot a histogram of the values you are averaging to see if there are any extreme values.

4. We will approximate the following integral using both the simple Monte Carlo integration and the control variate method:
$$I = \int_0^1 \frac{1}{1+x} dx.$$

(a) Set $h(x) = \frac{1}{1+x}$ and let $U_1, \ldots, U_n$ be iid Unif$[0, 1]$. Estimate I using $\hat{I}_{\mathrm{MC}} = \frac{1}{n}\sum_i h(U_i)$ with $n = 1500$. The exact value for I is $\ln 2$.

(b) Next introduce $c(x) = 1 + x$ as a control variate, and estimate I with

$$\hat{I}_{\mathrm{CV}} = \frac{1}{n}\sum_{i=1}^{n} h(U_i) - b\left[\frac{1}{n}\sum_{i=1}^{n} c(U_i) - E\{c(U)\}\right].$$

Note: you will need to analytically calculate $E\{c(U)\}$, and estimate the optimal value for b. Also use $n = 1500$.

(c) Estimate and compare the variances of $\hat{I}_{\mathrm{MC}}$ and $\hat{I}_{\mathrm{CV}}$.

(d) Can you design a new estimator for I that has a smaller variance than $\hat{I}_{\mathrm{CV}}$?