ISQS-7339: Homework 2

Note:

The MS-DS students: Solve problems 1-5. Your grade will be: X = (your score*100)/85. If you solve one of problems 6 or 7, you receive maximum 5 points as a bonus (your grade will be X+5).

The Ph.D. students: Solve problems 1-7. Your grade will be: X = (your score*100)/110.

Problem 1

(15 points) The time for an automated storage and retrieval system in a warehouse to locate a part consists of three movements. Let X be the time to travel to the correct aisle. Let Y be the time to travel to the correct location along the aisle. And let Z be the time to travel up to the correct location on the shelves. Assume that the distributions of X, Y, and Z are as follows:

$X \sim$ normal with mean 25 and standard deviation 4 seconds

$Y \sim$ triangular with minimum 10, maximum 22 seconds, and most likely travel time of 15 seconds.

Z $\sim$ truncated-normal with minimum 4, mean 6, and standard deviation 1 seconds.

a) Develop a Monte Carlo simulation model based on 100,000 random observations that estimate the mean and standard deviation of the total time it takes to locate a part.
b) Create a histogram for the total time that it takes to locate a part.
c) Estimate the probability that the time to locate a part exceeds 60 seconds. Graphically explain the probability by adding a vertical line at 60 seconds to the histogram of part b.

Problem 2

(25 points) Miller Pharmaceuticals needs to decide whether to conduct clinical trials and seek FDA approval for a newly developed drug.

Suppose that analysts have made the following assumptions:

- R&D costs: Triangular(min= \$450, max = \$800, most likely = \$700) <u>in millions</u> of dollars
- Clinical trials costs: Uniform(min= \$135, max = \$170) <u>in millions</u> of dollars
- Market size: Normal(mean = 2,200,000, sd = 250,000)
- Market share in year 1: Uniform(min = 5%, max = 10%)
- Discount Rate = 0.10

All other data are considered constant.

- Market size growth = 3% per year
- Market share growth = 15% per year
- Monthly revenue/prescription = $130
- Monthly variable cost/prescription = $40

Develop and run a Monte Carlo simulation model to estimate

a) (15 points) The expected (mean) net present value (NPV) over 5 years and its 95% confidence interval. Explain your result.
b) (10 points) What is the probability of negative NPV? Draw a histogram and explain the probability by adding a vertical line at zero.

Problem 3

(20 points) Use the Newsvendor Model to set up and run a Monte Carlo simulation assuming that demand is Truncated-Normal with a mean of 45, the standard deviation of 3, the minimum value of 40, and the maximum value of 50.

a) Suggest the optimal purchase quantity when the cost per unit (C) is $12, selling price (R) is $18, and the salvage value (S) is $9.
b) Report a 95% confidence interval for the optimal profit.

Problem 4

(15 points) The observations available in the wind turbine file represent

a) "FailCount": the count of the number of failures on a windmill turbine farm per year

b) "RepairTime": the time that it takes to repair a windmill turbine on each occurrence in minutes.

c) "DriveTime": the time that it takes to notice the failure and drive from the control station to the windmill turbine in minutes

Using the techniques discussed in the input-analysis lecture, recommend an input distribution model for the "FailCount," "RepairTime," and "DriveTime" variables. Please pay attention to the fact that which variable is naturally discrete or continuous. You can find the data on Blackboard.

Problem 5

(10 points) Estimate

a) the true median and 95% confidence interval for the median of "income" variable in the "Prestige" dataset by using bootstrapping.
b) the true standard deviation and 95% confidence interval for the standard deviation of the "income" variable in the "Prestige" dataset using bootstrapping.

```
library("car")
data("Prestige")
income <- Prestige$income
```

Problem 6 (only for Ph.D. students) (10 points)

Suppose there is a data from an iid sampling from the Poisson[1] distribution: `data = c(1, 0, 2, 0, 0, 4, 1, 0, 0, 0)`.

Suppose you use the prior $p(\lambda) = 0.01e^{-0.01\lambda}$.

The posterior distribution for λ, based on Bayes' theorem without finding the normalizing constant, is as follows.

$$p(\lambda \mid data) \propto p(data \mid \lambda)p(\lambda)$$

Since the pdf $p(data \mid \lambda)$ and the likelihood function $L(\lambda \mid data)$ are identical, you can rewrite Bayes' theorem as follows.

$$p(\lambda \mid data) \propto L(\lambda \mid data)\, p(\lambda)$$

The likelihood function $L(\lambda \mid data) = \frac{\lambda^1 e^{-\lambda}}{1!} \times \frac{\lambda^0 e^{-\lambda}}{0!} \times \frac{\lambda^2 e^{-\lambda}}{2!} \times \ldots \times \frac{\lambda^0 e^{-\lambda}}{0!} = \frac{\lambda^8 e^{-10\lambda}}{48}$

Accordingly, the posterior distribution $p(\lambda \mid data) \propto \frac{\lambda^8 e^{-10\lambda}}{48} \times 0.01e^{-0.01\lambda} = \frac{0.01\lambda^8 e^{-10.01\lambda}}{48}$

You can ignore all constants, so the posterior distribution for λ is

$$p(\lambda \mid data) \propto \lambda^8 e^{-10.01\lambda}.$$

Now please do the following activities:

(a) Sample values for λ based on the posterior distribution using the random-walk MH algorithm given using your choice of proposal function (you may need to use a truncated normal distribution with a lower bound of zero because λ (the mean of Poisson) is always positive). Pick an appropriate value for standard deviation to reach an acceptance-rate between 30-40%. Report the mean and standard deviation of λ.

(b) Discard the first 100 MH samples, then present the remaining by a histogram. Report a 95% credible interval for λ.

---

[1] The expression for Poison distribution is $p(X = x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$ For $x = 0, 1, \ldots, \infty$.

Problem 7 (only for PhD students) (15 points)

Use the firm data set posted on blackboard.

The dataset contains information on n = 1,642 firms in a given year. The variables of interest are given somewhat as follows:

```
##    y = A measure of firm performance- something like ROA. Negative
values are possible and do exist in the data set.
##   x = Extent of adoption of a particular strategy.Something like
number of quality management initiatives undertaken.

# install.packages("MCMCpack")
library(MCMCpack)
```

Our goal is to estimate β_0 , β_1, and σ for the classical regression model y = $\beta_0 + \beta_1 x + \epsilon$, and the error term $\epsilon \sim N(0, \sigma^2)$ (normally distributed as a common assumption in linear regression), by Markov Chain Monte-Carlo Sampling.

a) Simulate plausible values for β_0 and β_1 using MCMCregress (in the MCMCpack library).

b) Draw the scatterplot of the plausible pairs (β_0, β_1) that MCMCregress produced. Based on the scatterplot, identify five plausible functions $\beta_0 + \beta_1 x$.

c) Identify 90% credible intervals for β_0, β_1, and σ using the quantile function.