CPE 213 Data Model – Lab

29 March 2019

Instruction: This lab aims to allow students to gain more understanding in statistical modeling. Students should work on each task consecutively. Write down a technical report on your work in this lab. The report should contain the explanation, code, result (whether numerical and graphical) and discussion.

Due: Please submit your work as a PDF file to LEB2 before March 4th, 2019, 23:59.

Monte Carlo Simulation

[Ref: Probabilistic Engineering Design]

Monte Carlo simulation is named after the city of Monte Carlo in Monaco, which is famous for gambling such as roulette, dice, and slot machines. Since the simulation process involves generating chance variables and exhibits random behaviors, it has been called Monte Carlo simulation. Monte Carlo simulation is a powerful statistical analysis tool and widely used in both non-engineering fields and engineering fields. It was initially used to solve neutron diffusion problems in atomic bomb work at Alamos Scientific Laboratory in 1944. Monte Carlo simulation has been applied to diverse problems ranging from the simulation of complex physical phenomena such as atom collisions to the simulation of traffic flow and Dow Jones forecasting. Monte Carlo is also suitable for solving complex engineering problems because it can deal with a large number of random variables, various distribution types, and highly nonlinear engineering models.

Different from a physical experiment, Monte Carlo simulation performs random sampling and conducts a large number of experiments on computer. Then the statistical characteristics of the experiments (model outputs) are observed, and conclusions on the model outputs are drawn based on the statistical experiments. In each experiment, the possible values of the input random variables $X = (X_1 , X_2 ,\ldots, X_n)$ are sampled (generated) according to their distributions. Then the values of the output variable Y are calculated through the performance function $Y = g(X)$ at the samples of input random variables. With a number of experiments carried out in this manner, a set of samples of output variable Y are available for the statistical analysis, which estimates the characteristics of the output variable Y.

The outline of Monte Carlo simulation is depicted in Fig. 1. Three steps are required in the simulation process: Step 1 – sampling on random input variables X, Step 2 – evaluating model output Y, and Step 3 – statistical analysis on model output.

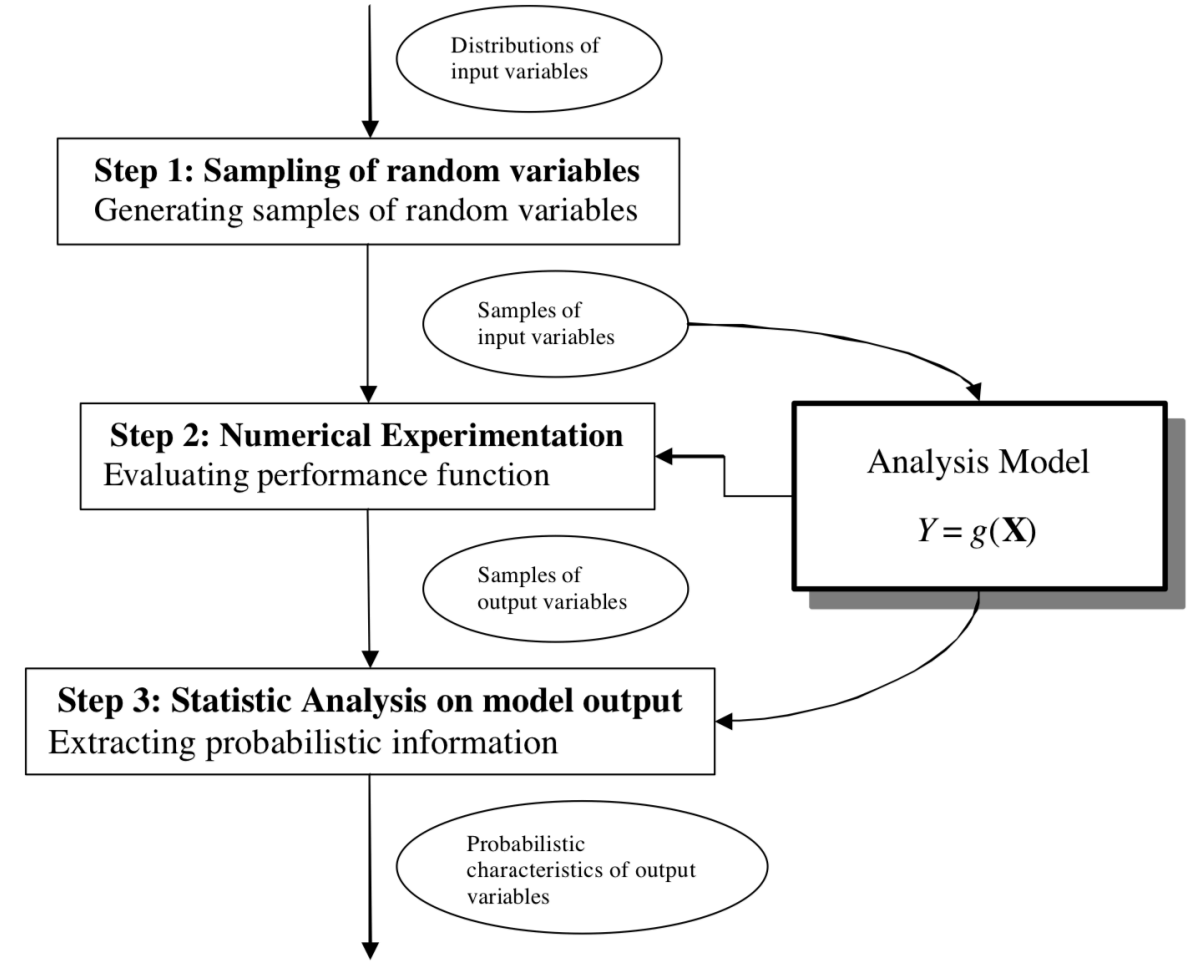

Figure 1. Monte Carlo Simulation

Example of Monte Carlo Simulation: Coin Tossing

Task 1: Pi Estimation

The goal of this problem is to illustrate how π = 3.1415... can be computed by random sampling of the unit disk. Starting from the following pseudocode, write R program to calculate Pi.

```
algorithm simple_pi
    initialize n_hits      0
    initialize m_trials    10000
    initialize counter     0

    while(counter < m_trials) do
        x = rand(0,1)
        y = rand(0,1)
        if(x**2 + y**2 < 1)
            n_hits++
        fi
        counter++
    done

return pi = 4*n_hits/m_trials
```

In your simulation run the code multiple times for $N=10^i$, i = 1, 2, 3, ... random numbers. See how the estimate for π improves with increasing N and compute the deviation from the exact result: error = $| \pi - \pi_{estimate} |$.

Perform a log-log plot of the error as a function of N and show that the data can be fit to a straight line of slope -1/2.

Task 2: Sales Simulation

Jimmy prints a neighborhood newspaper with 10 subscribers. He also sells it to whoever comes by from his front lawn on Friday afternoons. His mother has kept track of his demand (including requests made after he had sold out) for the past 100 weeks, and observed the pattern shown below.

Papers Demanded	Number of weeks	Probability	Cumulative Probability
13	1		
14	2		
15	4		
16	9		
17	10		
18	15		
19	16		
20	15		
21	12		
22	9		
23	4		
24	2		
25	1		
26	0		
Total	**100**		

The papers cost 30 cents to print and Jimmy sells them for 50 cents. Assume that he prints 20 copies a week. Mom makes him throw away unsold copies. Simulate his sales for the next 12 weeks and determine his earnings.

Week	Random Number	Number Demanded	Number Sold	Number Thrown	Revenue	Cost	Profit
1	0.4175						
2	0.8434						
3	0.5227						
4	0.1624						
5	0.0149						
6	0.0067						
7	0.0957						
8	0.6233						
9	0.9990						
10	0.0391						
11	0.2901						
12	0.0779						

Should Jimmy increase his paper copies to 30? Would it increase the earning? Why?

Task 3: Simulation from Data

Using the Superstore data,

1. Analyze the distribution of daily demand of each order.
2. Analyze the time between order of different product category.
3. Analyze the distribution of order processing time of each order.
4. Simulate the superstore order and processing time.
5. Is there relationship between order size and average processing time?
6. What would happen, if every Friday, the number of orders is boosted by 50%?