CHAPTER 4 MONTE CARLO METHODS AND SIMULATION

In this Chapter, you will learn:

- the random numbers and pseudo-random numbers,
- the estimation of areas and volumes by Monte Carlo methods,
- the simulation using random numbers.

1. RANDOM NUMBERS

Introduction

Good random numbers play a central part in Monte Carlo simulations. Usually they are generated using a deterministic algorithm; the numbers generated in this way are called ***pseudorandom numbers***. Typically the generators produce a random integer (with a definite number of bits), which is converted to a floating point number X $\in$ [0,1) or [0, 1] by multiplying with a suitable constant. The generators are initialized once before use with a *seed number*, typically an integer value or values. This sets the initial state of the generator. The essential properties of a good random number generator are:

a) Random pattern: passes statistical tests of randomness.

b) Long period: goes as long as possible before repeating.

c) Efficiency: executes rapidly and requires little storage.

d) Repeatability: produces same sequence if started with same initial conditions.

e) Portability: runs on different kinds of computers and is capable of producing same sequence on each.

➢ **Linear Congruential Generators (LCG)**

Mixed linear congruential generators have the form

$$X_n = (aX_{n-1} + c)(\mathrm{mod}\, m)$$

where m is the modulus, c is the increment, a is the multiplier and X_0, is the seed (starting value). X_n is the positive remainder when we divide $aX_{n-1} + c$ by m, thus $0 \le X_n \le m-1$.

The LCG $X_n = (aX_{n-1} + c)(\text{mod}\, m)$ has full period (m) if and only if all two of the following conditions hold:

(i) c and m are relatively prime (i.e., the only positive integer that divides both c and m is 1) (ii) If q is any prime number that divides m, then q also divides a-1.

Advantages of full period mixed LCG are:

(i) Since large-scale simulation projects can use hundreds of thousands of random numbers, it is desirable to have LCG `s with long periods.

(ii) If the LCG `s is with full period, it may contribute towards the uniformity of the random numbers.

Multiplicative LCG's and *Additive LCG's*:

In LCG $X_n = (aX_{n-1} + c)(\text{mod}\, m)$

(i) if $c = 0$, the LCG form is called as the *multiplicative LCG*. (ii) if $a = 1$, the LCG form is called as the *additive LCG*.

***But we cannot have full period in these two cases.*

<u>*Example 1*</u> : Consider $m = 8$, $a = 5$, $c = 7$ and $X_0 = 4$, generate a series of 8 random numbers.

➢ **Random Variate Generator for Exponential Distribution**

Let X be a r.v which is exponentially distributed whose probability density function (pdf) is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$$

and cumulative density function, cdf $F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$

The parameter λ is the mean number of occurrence per unit time. For example, $\lambda =$ average numbers of customers per minute.

From the cdf, random variate generator can be derived:

Step 1: Let $F(X) = R$, where $R = [0,1]$

So $1 - e^{-\lambda x} = R$

Step 2: Solve for *X* in terms of *R*, as a result we have:

$$X = (-1/\lambda)\ln(1 - R)$$

This is the random variate generator for the exponential distribution. Then generate random numbers $R_1, R_2, \ldots$ and compute the series of random variates by

$$X_i = (-1/\lambda)\ln(1 - R_i)$$

Example 2: Given a sequence of random numbers, $R_1 = 0.1306$, $R_2 = 0.0422$, $R_3 = 0.6597$, $R_4 = 0.7965$ and $R_5 = 0.7696$, generate five exponential random variates using the generator with $\lambda = 1$.

➢ Random Variate Generator for Uniform Distribution

Let *X* be uniformly distributed on the interval $[a,b]$, the probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases}$$

and cumulative density function, cdf $F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$

From the cdf, random variate generator can be derived:

Step 1: Let $F(X) = R$, where $R = [0,1]$

So, $\frac{x-a}{b-a}$ = R

Step 2: Solve for X in terms of R, as a result we have:

$$X = a + (b-a)R$$

This is the random variate generator for the uniform distribution. Then generate random numbers R_1, R_2, …. and compute the series of random variates by

$$X_i = a + (b-a)R_i$$

Example 3: Given a sequence of random numbers, $R_1 = 0.9271$, $R_2 = 0.4590$, $R_3 = 0.4210$, $R_4 = 0.4111$ and $R_5 = 0.8239$, within interval $[-5,5]$ generate five uniform random variates.

2. MONTE CARLO METHODS

Introduction

Monte Carlo methods can be thought of as statistical simulation methods that utilize a sequences of random numbers to perform the simulation. The name "Monte Carlo" was coined by Nicholas Constantine Metropolis (1915-1999) and inspired by Stanslaw Ulam (1909-1986), because of the similarity of statistical simulation to games of chance, and because Monte Carlo is a center for gambling and games of chance.

Monte Carlo Method for π

We start the familiar example of finding the area of a circle. Figure 1 below shows a circle with radius $r = 1$ inscribed within a square. The area of the circle is πr^2, and the area of the square $(2r)^2$. The ratio of the area of the circle to the area of the square is

$$p = \text{area of circle / area of square} = \pi/4$$

So we have, $\pi = p * 4$

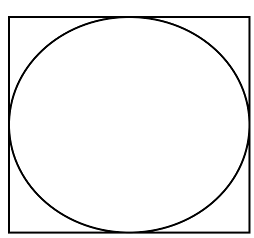


Figure 1

Then we can use a Monte Carlo simulation to approximate the value of π :

i) Randomly selecting *n* points in the unit square.

ii) Determine the ratio $p = m/n$, where *m* is number of points in the unit circle which satisfying $x^2 + y^2 \le 1$.

iii) Estimate the value of π

$$\pi \approx p*4$$

In a typical simulation of sample size *n* = 1000 , there were 787 points satisfying $x^2 + y^2 \le 1$. Using this data,

$$\pi \approx p*4 \approx 0.787*4 \approx 3.148$$

Estimation of Area and Volume: Monte Carlo Integration

Monte Carlo simulation can be used to approximate the area under a curve $y = f(x)$ for $a \le x \le b$ and $f(x) \ge 0$:

Area under the curve = $\int_a^b f(x)dx$

The steps for approximation the integral are:

(i) Pick *n* randomly distributed points $x_1, x_2, ..., x_n$ in the interval $[a,b]$.

(ii) Determine the average value of the function

$$\bar{f} = \frac{\sum_{i=1}^{n} f(x_i)}{n}$$

(iii) Compute the approximation to the integral

$$\int_a^b f(x)dx \approx (b-a)\frac{\sum_{i=1}^{n} f(x_i)}{n}$$

Likewise, approximating of integrals by the Monte Carlo method to estimate volumes is given by:

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2}\int_{a_3}^{b_3} f(x,y,z)dx \approx \frac{(b_1-a_1)(b_2-a_2)(b_3-a_3)}{n}\sum_{i=1}^{n} f(x_i, y_i, z_i)$$

3. SIMULATION

Simulation using random numbers is a technique for estimating probabilities.

- **Two Dice Problem**

Consider the experiment of tossing two unbiased dice. What is the probability that the sum of the numbers on two dice showing less than or equal to 4? The theoretical solution is 6/36=0.16667.

However, Monte Carlo simulation can be used to model this problem. The advantage is the results of Monte Carlo simulation can be compared with theoretical solution. For example, we can simulate *n* tosses of the dice. The number showing on each die is an integer uniformly distributed between 1 and 6. The flows of simulation are as follows:

i) Generate random numbers between 1 and 6 for 2 x *n* array (2 dice).

ii) Calculate the sums of the two arrays (numbers on two dice).

iii) Count how many of these sums less than or equal to 4, *m*.

iv) Then probability can be obtained by:

$$P(X \le 4) \approx \frac{m}{n}$$

- **Birthday Problem**

Suppose that in a room of *n* persons, each of these 365 days of the year is equally likely to be someone's birthday. From theory of probability, after someone is asked his or her birthday, the chances that next person asked will not have the same birthday are 364/365. The chances that the third person's birthday will not match first two persons are 363/365. Therefore, the probability that the *n*th person asked will have birthday different from that of anyone already asked is

$$\left(\frac{(365)(364)(363)...(365-(n-1)}{365^n} \right)$$

Then the probability that the *n*-th person asked will match is:

$$1-\left(\frac{(365)(364)(363)...(365-(n-1)}{365^n} \right)$$

Table 1 below shows the theoretical probabilities for different *n* persons generated by using formula above and by simulation:

n	Theoretical	Simulation
5	0.027	0.028
10	0.117	0.110
15	0.253	0.255
20	0.411	0.412
25	0.569	0.553

Table 1

The simulation to calculate the probability of repeated birthdays can be carried out as follows:

i) Generate *n* random birthdays in the range 1 to 365 (ignore leap years).

ii) Check if two coincide

iii) Repeat for different trials

iv) Probabilities = numbers of coincidences/numbers of trials.

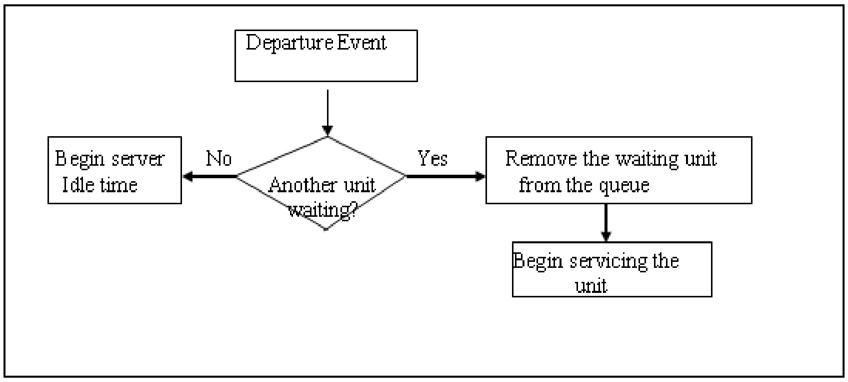
4. QUEUING SIMULATION

A queuing system is described by its calling population, the nature of the arrivals and services, the system capacity, and the queuing discipline.

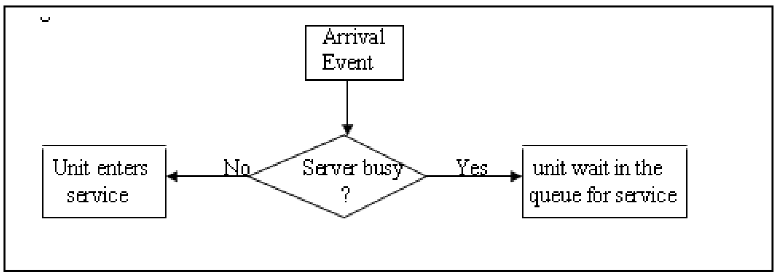
The state of the system is the number of units in the system and the status of the server (busy or idle).

In a single channel queuing system there are only two possible events: arrival event and the departure event.

If the service has just been completed, the simulation proceeds in the manner shown in the following diagram:



The second event occurs when a unit enters the system. The flow diagram is shown in the figure:



Simulation of queuing systems generally requires the maintenance of an event list for determining what happens next. The event list indicates the times at which the different types of events occur for each unit in the queuing system.

The times are kept on a "clock" which marks the occurrences of events in time.

The randomness needed to imitate the real life is made possible through the "random numbers".

- **Simulation of single server**

A grocery store has only one check out counter. Customers arrive at this counter at random and inter-arrival times are uniformly distributed on [1, 8].

The service times vary from 1 to 6 minutes with the following probabilities:

Service Time	1	2	3	4	5	6
Probability	0.1	0.2	0.3	0.25	0.1	0.05

Assume we have generated the following random digits

For inter-arrival time: 931, 949, 648, 258, 420, 396, 187, 689, 154

For service time: 70, 66, 79, 90, 85, 34, 5, 57, 31, 82

Simulate the arrival and service of 10 customers, find the

i) Average waiting time of a customer in queue
ii) Probability that a customer has to wait in the queue
iii) Average service time
iv) Average interarrival time
v) Average time a customer spend in the system

Solution:

Service Time	Probability	CDF, F(x)	Random No Range
1	0.10	0.10	1-10
2	0.20	0.30	11-30
3	0.30	0.60	31-60
4	0.20	0.85	61-85
5	0.1	0.95	86-95
6	0.05	1	96-100

Interarrival Time	Probability	CDF, F(x)	Random No Range
1	0.125	0.125	1-125
2	0.125	0.250	126-250
3	0.125	0.375	251-375
4	0.125	0.500	376-500
5	0.125	0.625	501-625
6	0.125	0.750	626-750
7	0.125	0.875	751-875
8	0.125	1.000	876-1000

cust no	RN arrivals	Inter arrivals time	Arrival time	RN service	time service begins	Service time	Time service ends	Time spend in system	Time in queue
1		0	0	70	0	4	4	4	0
2	931	8	8	66	8	4	12	4	0
3	949	8	16	79	16	4	20	4	0
4	645	6	22	90	22	5	27	5	0
5	258	3	25	85	27	4	31	6	2
6	420	4	29	34	31	3	34	5	2
7	396	4	33	5	34	1	35	2	1
8	187	2	35	57	35	3	38	3	0
9	689	6	41	31	41	3	44	3	0
10	154	2	43	82	44	4	48	5	1

i) Average waiting time of a customer in queue=6/10
ii) Probability that a customer has to wait in the queue=4/10
iii) Average service time=35/10
iv) Average interarrival time=43/9
v) Average time a customer spend in the system=41/10

- **Simulation of a two-server queue**

Consider a drive in restaurant where carhops take orders and bring food to the car. Cars arrive in the manner shown in the following table:

Time between arrivals (minutes)	1	2	3	4
Probability	0.25	0.40	0.20	0.15

There are two carhops - Ali and Ahmad. Ali is better to do the job, and works somewhat faster than Ahmad. The distribution of service times is shown in the following table:

Service time distribution of Ali:

Service time (minutes)	2	3	4	5
Probability	0.30	0.28	0.25	0.17

Service time distribution of Ahmad:

Service time (minutes) :	3	4	5	6
Probability :	0.35	0.25	0.20	0.20

Assume we have generated the following random digits

For inter-arrival time: 26,98,90,26,42,74,80,68,22,48,34

For service time: 95,21,51,92,89,38,13,61,50,49,39,53

Simulate this system for 12 customers, find

i) Percentage of time Ali was busy.
ii) Percentage of time Ahmad was busy
iii) Probability that a customer has to wait in the queue.
iv) Average waiting time in the queue.

Solution:

Interarrival Time	Probability	CDF, F(x)	Random No Range
1	0.25	0.25	1-25
2	0.40	0.65	26-65
3	0.20	0.85	66-85
4	0.15	1.00	86-100

Ali

Service Time	Probability	CDF, F(x)	Random No Range
2	0.30	0.30	1-30
3	0.28	0.58	31-58
4	0.25	0.83	59-83
5	0.17	1.00	84-100

Service Time	Probability	CDF, F(x)	Random No Range
3	0.35	0.35	1-35
4	0.25	0.60	36-60
5	0.20	0.80	61-80
6	0.20	1.00	81-100

					Ali			Ahmad			
cust no	RN arrivals	Inter arrivals time	Arrival time	RN service	time service begins	Service time	Time service ends	time service begins	Service time	Time service ends	Time in queue
1			0	95	0	5	5				0
2	26	2	2	21				2	3	5	0
3	98	4	6	51	6	3	9				0
4	90	4	10	92	10	5	15				0
5	26	2	12	89				12	6	18	0
6	42	2	14	38	15	3	18				1
7	74	3	17	13	18	2	20				1
8	80	3	20	61	20	4	24				0
9	68	3	23	50				23	4	27	0
10	22	1	24	49	24	3	27				0
11	48	2	26	39	27	3	30				1
12	34	2	28	53				28	4	32	0
						28			17		

i) Percentage of time Ali was busy = (28/32) *100%=87.5%
ii) Percentage of time Ahmad was busy = (17/32)*100%=53125%
iii) Probability that a customer has to wait in the queue = 3/12
iv) Average waiting time in the queue =3 /12