MATH 3290 Mathematical Modeling

Chapter 5: Simulation Modeling

Kuang HUANG

February 6, 2024

Department of Mathematics
The Chinese University of Hong Kong

Course webpage

https://www.math.cuhk.edu.hk/course/2324/math3290

SCAN ME

Introduction

In empirical modeling, one needs data.

There are situations where experiments are expensive, or even impossible:

- it is harmful to inject certain drugs in body (pharmacodynamics),
- tests are impossible in the design of aircraft (computational fluid dynamics).

Therefore, one needs to simulate the situation.

We will discuss basic ideas of Monte Carlo simulation.

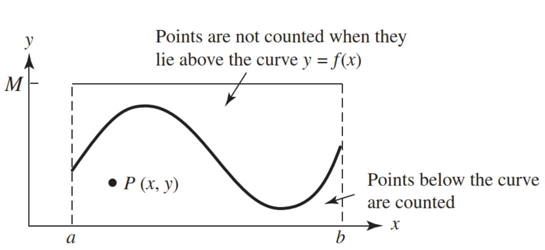
Simulating deterministic behaviors

Consider finding the area under a curve.

Let $f(x)$ defined on $a \le x \le b$ so that $0 \le f(x) \le M$.

We choose a point (x, y) from the rectangle $[a, b] \times [0, M]$ at random.

$$\frac{\text{area under curve}}{\text{area of rectangle}} \approx \frac{\text{number of points counted below curve}}{\text{total number of random points}}$$



Monte Carlo Area Algorithm

Input Total number n of random points to be generated in the simulation.

Output AREA $=$ approximate area under the specified curve $y = f(x)$ over the given interval $a \leq x \leq b$, where $0 \leq f(x) < M$.

Step 1 Initialize: COUNTER $= 0$.

Step 2 For $i = 1, 2, \ldots, n$, do Steps 3–5.

Step 3 Calculate random coordinates x_i and y_i that satisfy $a \leq x_i \leq b$ and $0 \leq y_i < M$.

Step 4 Calculate $f(x_i)$ for the random x_i coordinate.

Step 5 If $y_i \leq f(x_i)$, then increment the COUNTER by 1. Otherwise, leave COUNTER as is.

Step 6 Calculate AREA $= M(b - a)$ COUNTER$/n$.

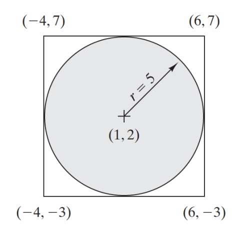
Step 7 OUTPUT (AREA)

STOP

$$\text{Area} = \int_{-\pi/2}^{\pi/2} \cos(x)\,\mathrm{d}x = 2.$$

Number of points	Approximation to area	Number of points	Approximation to area
100	2.07345	2000	1.94465
200	2.13628	3000	1.97711
300	2.01064	4000	1.99962
400	2.12058	5000	2.01429
500	2.04832	6000	2.02319
600	2.09440	8000	2.00669
700	2.02857	10000	2.00873
800	1.99491	15000	2.00978
900	1.99666	20000	2.01093
1000	1.96664	30000	2.01186

Monte Carlo approximations to the area under the curve $y = \cos(x)$ over the interval $-\pi/2 \leq x \leq \pi/2$.



Generate two random numbers R_1, R_2 in $[0,1]$. Then set

$$x = -4 + 10R_1, \quad y = -3 + 10R_2.$$

Check if (x, y) is inside the circle. Then

$$\texttt{area\_of\_circle} \approx \frac{m}{n} \texttt{area\_of\_square},$$

where n is the sample size, m is the number of points inside the circle.

```
n = 30000;  % sample size
m = 0;      % number of points inside the circle

for i = 1 : n
    % Generate two random numbers
    R1 = rand;
    R2 = rand;
    % Get (x,y) coordinates
    x = -4 + 10 * R1;
    y = -3 + 10 * R2;
    % Check if (x,y) is inside the circle
    if (x - 1)^2 + (y - 2)^2 < 25
        m = m + 1;
    end
end

area_approx = m/n * 100
area_truth = pi * 25
```

```
area_approx = 78.4600
area_truth = 78.5398
```

You could download these codes from the course webpage.

Simulating probabilistic behaviors

Use random number to simulate probabilistic behaviors.

In some cases, it may be hard to compute the analytical values.

The probability of an event is

$$\frac{\texttt{number\_of\_favourable\_events}}{\texttt{total\_number\_of\_events}}.$$

A fair coin

Let X be a random number in $[0, 1]$ that follows the uniform distribution. We want to use X to model a flip of a coin.

We let f be a function defined by

$$f(x) = \begin{cases} \text{Head}, & 0 \le x \le 0.5, \\ \text{Tail}, & 0.5 < x \le 1. \end{cases}$$

Then $f(X)$ is the result of the flipping, which is a random variable.

We can then use Monte Carlo algorithm to see the probability of getting a head.

Monte Carlo Fair Coin Algorithm

Input	Total number n of random flips of a fair coin to be generated in the simulation.
Output	Probability of getting a head when we flip a fair coin.
Step 1	Initialize: COUNTER $= 0$.
Step 2	For $i = 1, 2, \ldots, n$, do Steps 3 and 4.
Step 3	Obtain a random number x_i between 0 and 1.
Step 4	If $0 \leq x_i \leq 0.5$, then COUNTER $=$ COUNTER $+ 1$. Otherwise, leave COUNTER as is.
Step 5	Calculate $P(\text{head}) = \text{COUNTER}/n$.
Step 6	OUTPUT Probability of heads, $P(\text{head})$.
	STOP

Very close to the analytical value of 0.5.

Number of flips	Number of heads	Percent heads
100	49	0.49
200	102	0.51
500	252	0.504
1,000	492	0.492
5,000	2469	0.4930
10,000	4993	0.4993

An unfair die

A die/dice

An unfair die

Let x_i be a random number in $[0, 1]$ that follows the uniform distribution. We want to use x_i to model results of rolling an unfair die.

Roll value	P(roll)
1	0.1
2	0.1
3	0.2
4	0.3
5	0.2
6	0.1

Empirical distribution

Value of x_i	Assignment
$[0, 0.1]$	ONE
$(0.1, 0.2]$	TWO
$(0.2, 0.4]$	THREE
$(0.4, 0.7]$	FOUR
$(0.7, 0.9]$	FIVE
$(0.9, 1.0]$	SIX

Assignments

Monte Carlo Roll of an Unfair Die Algorithm

Input Total number n of random rolls of a die in the simulation.

Output The percentage or probability for rolls $\{1, 2, 3, 4, 5, 6\}$.

Step 1 Initialize COUNTER 1 through COUNTER 6 to zero.

Step 2 For $i = 1, 2, \ldots, n$, do Steps 3 and 4.

Step 3 Obtain a random number satisfying $0 \leq x_i \leq 1$.

Step 4 If x_i belongs to these intervals, then increment the appropriate COUNTER.

$$\begin{array}{rcccl l}
0 & \leq & x_i & \leq & 0.1 & \text{COUNTER 1} = \text{COUNTER 1} + 1 \\
0.1 & < & x_i & \leq & 0.2 & \text{COUNTER 2} = \text{COUNTER 2} + 1 \\
0.2 & < & x_i & \leq & 0.4 & \text{COUNTER 3} = \text{COUNTER 3} + 1 \\
0.4 & < & x_i & \leq & 0.7 & \text{COUNTER 4} = \text{COUNTER 4} + 1 \\
0.7 & < & x_i & \leq & 0.9 & \text{COUNTER 5} = \text{COUNTER 5} + 1 \\
0.9 & < & x_i & \leq & 1.0 & \text{COUNTER 6} = \text{COUNTER 6} + 1
\end{array}$$

Step 5 Calculate probability of each roll $j = \{1, 2, 3, 4, 5, 6\}$ by COUNTER$(j)/n$.

Step 6 OUTPUT probabilities.

STOP

Die value	100	1000	5000	10,000	40,000	Expected results
1	0.080	0.078	0.094	0.0948	0.0948	0.1
2	0.110	0.099	0.099	0.0992	0.0992	0.1
3	0.230	0.199	0.192	0.1962	0.1962	0.2
4	0.360	0.320	0.308	0.3082	0.3081	0.3
5	0.110	0.184	0.201	0.2012	0.2011	0.2
6	0.110	0.120	0.104	0.1044	0.1045	0.1

Remark: We always use random variables that are uniformly distributed on $[0, 1]$ to simulate any probabilistic behaviors.

Simulating probability distributions

Suppose we need a random number X that follows a target probability distribution.

Let $f(x)$ be a probability density function, and $F(x)$ be the cumulative density function. Note that

$$F(x) = \mathbb{P}\Big(\{X \leq x\}\Big), \qquad 0 \leq F(x) \leq 1.$$

It is proved that the random variable $Z = F(X)$ is uniformly distributed in $[0, 1]$.

To find a random number X follows the probability density $f(x)$:

- generate a uniformly distributed random number R in $[0, 1]$,
- compute $X = F^{-1}(R)$.

Exponential distribution

Let $f(t) = \lambda e^{-\lambda t}$ be the exponential density function. It represents the inter-arrival time at a facility with mean arrival time $1/\lambda$.

The cumulative density function is

$$F(t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}.$$

Let $R = F(t)$, we have

$$t = F^{-1}(R) = -\frac{1}{\lambda} \ln(1 - R).$$

Remark: we can replace $1 - R$ by R as they both random, that is

$$t = -\frac{1}{\lambda} \ln(R),$$

and now the random variable t follows the target distribution.

Simulation of a queueing model

Consider a barber shop operated by only one barber.

Assume that

- the inter-arrival time is exponential with mean 15 minutes;
- each hair cut takes 10 to 15 minutes, uniformly distributed.

We let p and q be random samples of inter-arrival and service times.

$$p = -15 \ln(R), \qquad q = 10 + 5R,$$

where R is a random number uniformly distributed in $[0, 1]$.

We will simulate the first 5 arrivals.

Arrival of customer 1 at $T = 0$

The simulation starts at $T = 0$ with the arrival of the first customer.

- We need to know when customer 2 arrives:

$$T = 0 + p_1 = 0 + (-15 \ln(0.0589)) = 42.48.$$

  So, customer 2 will arrive at 42.48 minutes.

- We need to know when customer 1 leaves:

$$T = 0 + q_1 = 0 + (10 + 5(0.6733)) = 13.37.$$

  So, customer 1 will leave at 13.37 minutes.

Time, T	Event
13.37	Departure of customer 1
42.48	Arrival of customer 2

Departure of customer 1 at $T = 13.37$

The queue is empty.

Note that customer 2 will arrive at 42.48 minutes.

Time, T	Event
42.48	Arrival of customer 2

Arrival of customer 2 at $T = 42.48$

- We need to know when customer 3 arrives:

$$T = 42.48 + p_2 = 42.48 + (-15\ln(0.4799)) = 53.49.$$

So, customer 3 will arrive at 53.49 minutes.

- We need to know when customer 2 leaves:

$$T = 42.48 + q_2 = 42.48 + (10 + 5(0.9486)) = 57.22.$$

So, customer 2 will leave at 57.22 minutes.

Time, T	Event
53.49	Arrival of customer 3
57.22	Departure of customer 2

Arrival of customer 3 at $T = 53.49$

- We need to know when customer 4 arrives:

$$T = 53.49 + p_4 = 53.49 + (-15\ln(0.6139)) = 60.81.$$

  So, customer 4 will arrive at 60.81 minutes.

- Customer 3 is placed in the queue since the facility is busy.

Time, T	Event
57.22	Departure of customer 2
60.81	Arrival of customer 4

Departure of customer 2 at $T = 57.22$

- Customer 3 is taken out of the queue, his waiting time is.

$$W_3 = 57.22 - 53.49 = 3.73,$$

and the departure time for customer 3 is

$$T = 57.22 + q_3 = 57.22 + (10 + 5(0.5933)) = 70.19.$$

So, customer 3 will leave at 70.19 minutes.

Time, T	Event
60.81	Arrival of customer 4
70.19	Departure of customer 3

Arrival of customer 4 at $T = 60.81$

- Customer 4 is placed in the queue.
- To find the arrival time of customer 5:

$$T = 60.81 + q_5 = 60.81 + (-15 \ln(0.9341)) = 61.83.$$

So, customer 5 will arrive at 61.83 minutes.

Time, T	Event
61.83	Arrival of customer 5
70.19	Departure of customer 3

Arrival of customer 5 at $T = 61.83$

- Customer 6 will not be generated.
- Customer 5 is place in the queue.

Time, T	Event
70.19	Departure of customer 3

Remark: both customers 4 and 5 are in the queue, customer 4 arrived at $T = 60.81$.

Departure of customer 3 at $T = 70.19$

Customer 4 is taken out from the queue, the waiting time is

$$W_4 = 70.19 - 60.81 = 9.38.$$

In addition, the departure time for customer 4 is

$$T = 70.19 + q_4 = 70.19 + (10 + 5(0.1782)) = 81.08.$$

Time, T	Event
81.08	Departure of customer 4

Remark: customer 5 is in the queue, customer 5 arrived at $T = 61.83$.

Departure of customer 4 at $T = 81.08$

Customer 5 is taken out from the queue, the waiting time is

$$W_5 = 81.08 - 61.83 = 19.25.$$

In addition, the departure time for customer 5 is

$$T = 81.08 + q_4 = 70.19 + (10 + 5(0.3473)) = 92.82.$$

Time, T	Event
92.82	Departure of customer 5

Remark: no more events.

Summary

First, we look at waiting time.

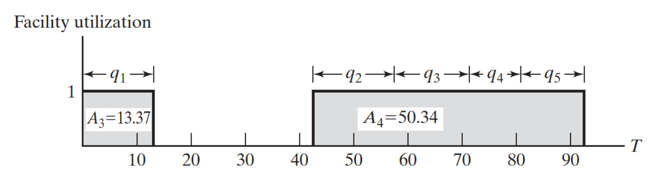
The total waiting time is

$$W_1 + W_2 + W_3 + W_4 + W_5 = 0 + 0 + 3.73 + 9.38 + 19.25 = 32.36.$$

Hence, the is average waiting time is

$$\texttt{average\_waiting\_time} = \frac{32.36}{5} = 6.47.$$

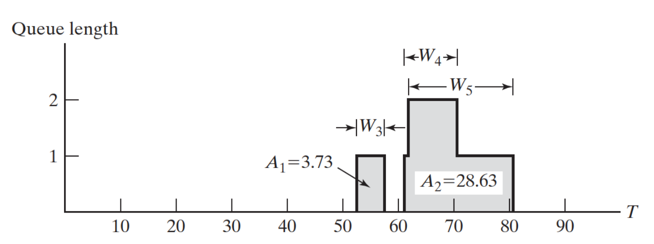
Next, we look at facility utilization.



Hence, the average utilization is

$$\texttt{average\_utilization} = \frac{A_3 + A_4}{92.82} = 0.686 \text{ } barber.$$

Finally, we look at queue length.



Hence, the average queue length is

$$\texttt{average\_queue\_length} = \frac{A_1 + A_2}{92.82} = 0.349\ customer.$$

Implement barber shop simulation

Suppose there are n customers, the sequence of inter-arrival time is

$$p_1, p_2, \ldots, p_{n-1},$$

where p_i (independently) follows the exponential distribution with λ as its mean.

Hence, the arrival time point T_i of the i-th customer has the following relation

$$T_{i+1} = T_i + p_i$$

with $T_1 = 0$. Denote by

$$L_1, L_2, \ldots, L_n$$

service time lengths for all customers. Each L_i is a random sample of the uniform distribution $U(L', L'')$.

Suppose the sequence of service starting time of customers is

$$S_1, S_2, \dots, S_n$$

with $S_1 = 0$.

We can obtain an important relation

Recursive formula of S_i

$$S_{i+1} = \max\{T_{i+1}, S_i + L_i\}.$$

Note that $S_i + L_i$ is the service ending time of the i-th customer.

This says that $\{S_i\}$ is indeed a Markov process, but it is too complicated to obtain an explicit form.

Output barber shop indexes

The waiting time of the i-th customer is $S_i - T_i$. Hence

$$\texttt{average\_waiting\_time} = \frac{\sum_{i=1}^{n}(S_i - T_i)}{n}.$$

The total time length of a simulation is $S_n + L_n$. Hence

$$\texttt{average\_utilization} = \frac{\sum_{i=1}^{n} L_i}{S_n + L_n}.$$

Similarly,

$$\texttt{average\_queue\_length} = \frac{\sum_{i=1}^{n}(S_i - T_i)}{S_n + L_n}.$$

Note that all those indexes are random variables, we need to conduct enough simulations to estimate their means, variances, etc.

Disclaimer

All figures, tables, and data appearing in the slides are only used for teaching under guidelines of Fair Use.