
UNIT 15 SIMULATION TECHNIQUES

Structure

- 15.1 Introduction
 - Objectives
- 15.2 Steps in Simulation
- 15.3 Monte-Carlo Method of Simulation
 - Monte-Carlo Integration
 - Simulation in Statistical Inference
 - Estimation of a Distribution of a Parameter by Monte-Carlo Simulation
 - Problem of Discrete Event Simulation
 - Monte-Carlo Simulation in Business Applications
- 15.4 Summary
- 15.5 Solutions / Answers

15.1 INTRODUCTION

Dictionary meaning of simulation is to pretend, to act like or mimic, etc. Some of the examples of simulation are: model of a car, aeroplane in a wind tunnel may act like in a real flight, etc. More precisely a simulation of a system is the operation of a model (or simulator) which is the representation of the system. The model is amenable to manipulation, which would be impossible, too expensive or impractical to perform on the entity it portrays. The operation of the model can be studied and from it, properties concerning the behaviour of the actual system (or its subsystems) can be inferred. Some examples are forest management, epidemics, traffic congestion, effect of ageing in a population, etc.

The steps involved in simulation are described in Section 15.2. The Monte-Carlo method of simulation is explained in Section 15.3. In the same section we have discussed Monte-Carlo integration, simulation in statistical inference, estimation of a distribution of a parameter by Monte-Carlo simulation, problem of discrete event simulation and Monte-Carlo simulation in business applications

Objectives

After studying this unit, you would be able to

- define the simulation;
- describe different steps in setting up simulation for solving a problem;
- describe the Monte-Carlo simulation;
- use Monte-Carlo simulation for solving some deterministic and stochastic problems; and
- describe some methods for making inference by simulation.

15.2 STEPS IN SIMULATION

The following are the principle steps in simulation:

1. Formulation of the Problem and Plan the Study

It needs clear statement of the problem and its objectives. If there are some alternatives then they should be documented. Criteria for comparing efficiencies of alternatives should be given. It should also plan for cost and time required for the study.

2. Collect Data and Define a Model

One has to collect data from the system of interest and also obtain the estimate of parameters of the model. Choice of the model is very important. As a rule one should start with a simple model and then make it more complex as the need be.

3. Validation of the Model

It is very important that the model chosen should be true representative of the system of interest. This needs verification at every stage of the operation. The adequacy of the theoretical probability distribution fitted to the observed data should be tested using goodness of fit tests.

4. Construction of a Computer Program

These days most of the simulation work is done on computers. For this, program is to be written and errors removed.

5. Make Pilot Runs of the Computer Program and Check the Output for Validity

It is important to check the program in the pilot runs by running it for some cases where correct results are known.

6. Make Production Runs

7. Analyse Output Data

8. Document and Implement Results

15.3 MONTE-CARLO METHOD OF SIMULATION

One may define simulation as a numerical technique for conducting experiments on a digital computer, which involves certain types of mathematical and logical models that describe the behaviour of the system of interest. Here we are concerned with stochastic simulation, which is also called Monte-carlo simulation. This involves sampling stochastic variables from some probability distribution. Some problems which are stochastic in nature, such as consumer demand, production, population size, number of persons waiting in a queue, etc. are some examples of this type.

Some completely deterministic problems are very difficult to solve analytically. However, by simulating a stochastic process, whose moments, density function or cumulative distribution function satisfy the solution requirements, give approximate solution of the problem. Some examples of this nature are solution of certain integral and differential equations. In this

section, we shall outline some important simulation techniques in different fields.

15.3.1 Monte-Carlo Integration

Suppose we wish to evaluate a deterministic integral

$$\theta = \int_0^1 g(x) dx$$

If this integral exists then it is merely $E[g(x)]$ where x is $U(0,1)$ variate.

Hence θ can be estimated as $\hat{\theta}$ given by

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n g(x_i)$$

where x_1, x_2, \dots, x_n is a random sample of size n from $U(0,1)$ population. $\hat{\theta}$ is an unbiased estimate of θ . As n increases, $\hat{\theta}$ approaches θ , and thus accuracy of $\hat{\theta}$ can be increased by increasing n . This is an example of solving a deterministic problem by Monte-carlo simulation.

15.3.2 Simulation in Statistical Inference

Suppose we have a sample x_1, x_2, \dots, x_n from a probability distribution function $f(x, \theta)$ where θ is not known. We wish to test the hypothesis

$$H_0: \theta = \theta_0$$

against the alternative hypothesis

$$H_1: \theta > \theta_0$$

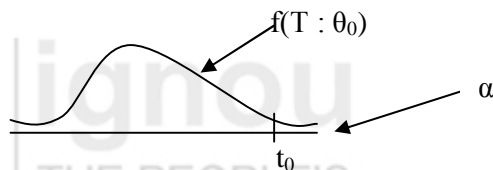
Suppose, we have a test statistic $T = g(x_1, x_2, \dots, x_n)$, where g is a known function of x_i 's. Suppose large values of T indicate the departure from the null hypothesis. Let $t = g(x_1, x_2, \dots, x_n)$ be the observed statistic from the sample.

If the theoretical distribution of T , $f(T, \theta_0)$ under $H_0: \theta = \theta_0$, is known and if

$$P(T > t_0) = \alpha,$$

then if $T > t_0$, we reject H_0 . Accept H_0 otherwise.

However, if $f(T, \theta_0)$ is not known then one can always simulate the distribution of T by taking many samples from $f(x, \theta_0)$ and calculating t for each sample and thus obtaining an empirical distribution of T . Then one can estimate critical point t_0 and test the hypothesis. Two sided hypothesis can also be tested once the empirical distribution is available.



Example 1: A random sample of size ten from a population is given as follows:

8.94, 6.31, 5.00, 3.94, 5.19, 3.80, 6.11, 5.65, 3.76, 5.03

Assuming that population can be considered as Normal with standard deviation 1, test the hypothesis H_0 that the mean of the population is 5 against the alternative hypothesis H_1 that it is greater than 5, by using statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \text{where} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Solution: In this case the exact distribution of statistic t is known, but we shall demonstrate the test by using empirical distribution of t . For this we have generated 100 random samples of size 10 from normal population with mean $\mu_0 = 5$ and $\sigma = 1$ by the methods described in Unit 14 and calculated t for all 100 samples. We then made a frequency table. Value of t calculated from the given sample is

t	frequency
-0.2-0.15	7
-0.15-.0.1	5
-0.1-.0.05	15
-0.05-0.0	25
0-0.05	16
0.05-0.10	16
0.10-0.15	9
0.15-0.2	5

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{5.373 - 5}{1/\sqrt{10}} \\ &= 1.179 \end{aligned}$$

From the frequency table, we observed that there are 14 t 's greater than observed 1.179. Hence this is a large probability i.e. $p=0.14(=14/100)$ thus we do not reject the null hypothesis of mean of the population being 5.

Hence, we used the theoretical distribution of t , which is $N(0, 1)$ and from normal table $p = 0.12 = P(Z > 1.179)$. In both cases, we accept the null hypothesis. The values of p differ because empirical distribution is not very accurate due to small number of samples (100) involved in generating it. Usually more than 1000 samples are required for estimating empirical distribution.

15.3.3 Estimation of a Distribution of a Parameter by Monte-Carlo Simulation

Suppose the purpose is to find the distribution of some of the parameters of the distribution of a random variable. The random variable, which we shall call output variable, is a known function of other random variables which have known distributions. To estimate the distribution of the output variable we draw a value for each of the input variables from their distributions and calculate the resulting output variable. Such sampling is then repeated many times and this yields an estimate of the distribution.

Example 2: Using Monte-Carlo method, estimate p where

$$p = P[g(x) < a]$$

where, $g(x) = \min(x_1, x_2)$

Here x_1 and x_2 have independent normal distributions,

$$x_1 \sim N(100, 400) \quad ; \quad x_2 \sim N(90, 100)$$

Solution: The problem of the estimation of p arises if we have a product consisting of two parts. The distribution of lives of part 1 and part 2, say x_1 and x_2 respectively, are given above. The product breaks down as soon as one of the two parts fails. We want to know the probability that the life of the product is smaller than some given value, say, a . Therefore, we want to know

p given above. For Monte-Carlo estimation of p , it is convenient to introduce a variable y defined as:

$$y = \begin{cases} 1 & \text{if } g(x) < a \\ 0 & \text{if } g(x) \geq a \end{cases}$$

The expected value of y , $E(y)$ is given by

$$\begin{aligned} E(y) &= 1 \cdot P(g(x) < a) + 0 \cdot P(g(x) \geq a) \\ &= P(g(x) < a) = p \end{aligned}$$

Now, we generate two independent random variables $x_1 \sim N(100, 400)$ and $x_2 \sim N(90, 100)$ as described in Unit 14, using independent $U(0, 1)$ variables and find y . We repeat this a large number of times say, N , and then find y_i ($i = 1, 2, \dots, N$) in each case. We estimate p as

$$\hat{p} = \bar{y} = \sum_{i=1}^N y_i / N$$

It is not necessary that x_1 and x_2 have normal distributions. One may consider any distribution and then estimate p as given above.

15.3.4 Problem of Discrete-Event Simulation

A very important example of discrete event simulation involves a queue (a waiting line). There are many practical applications of queuing problems. For example, banks, doctor's offices, supermarkets, car washers, airports and gasoline stations all involve situations in which customers may have to wait in line for service. A similar problem applies to orders entering a job shop which must wait for their turn to be processed, machine waiting to be repaired, jobs entering a computer system and so on. There are many situations in the queuing problem. Simplest situation is one in which there is only one single waiting line followed by a single server. Then there are cases where multiple servers (and multiple waiting lines) in series. Then there is a problem of multiple servers in parallel with a common waiting line. We shall discuss a simple example of this in Sub-section 16.3.1 with single line and single server.

15.3.5 Monte-Carlo Simulation in Business Applications

We shall consider a simple problem of simulation in assessing financial risk. Risk analysis is one of the most important and widely used applications of discrete event simulation. The objective is to assess the desirability of a proposed investment, based upon some financial decision criterion such as present worth. Application of the method results in a cumulative distribution being generated for the decision criterion. Hence, we can obtain not only the expected value of the decision criterion but also probabilities of obtaining much higher or lower values.

Suppose we are considering an initial outlay that will generate a series of n yearly cash flows (i.e. inflows and outflows). The present value of each cash flow can be written as

$$(PW)_j = \frac{(YCF)_j}{(1+i)^j},$$

where, $(YCF)_j$ represents the annual cash flow for the j^{th} year in the future and i is a specified annual interest rate, expressed as a decimal value. Hence the present worth of the entire proposed investment can be expressed as

$$PW = \left\{ \sum_{j=1}^n (PW)_j \right\} - I$$

$$= \sum_{j=1}^n \frac{(YCF)_j}{(1+i)^j} - I$$

where I represents the initial cash outlay.

Each of the yearly cash flows is generally comprises of several components, such as yearly sales volume, production cost, taxes, etc. These items are normally represented in terms of appropriate distribution functions. Thus, the computational procedure involves generating a random value for each cash flow component, resulting in a randomly generated value of $(PW)_j$. All the yearly cash flows are evaluated in this manner. These values are then used in above equation resulting in a single value of PW . We then repeat the entire procedure N times, which allows us to obtain, a distribution for PW . We may calculate the average PW and other quantiles of the distribution.

Example 3: Suppose an industry has started a factory with an initial capital of Rs. 50 million. Suppose $(YCF)_j$ is distributed normally with mean 10 million and variance of 1 million². For one simulation ($N=1$) obtain PW for $n=10$, taking annual interest of 10 percent ($i = 0.10$).

Solution: Suppose ten normal variables $N(0, 1)$ are given as the following:

$N(0,1)$	-0.179	0.421	0.210	-1.598	1.717	0.308	-0.421	-0.776	0.640	-0.319
$N(10,1)$ $(YCF)_j$ in million rupees	9.821	10.210	10.210	8.402	11.717	9.692	9.579	9.224	10.640	9.681
$(1+i)^j$	1.100	1.210	1.331	1.464	1.610	1.771	1.949	2.143	2.358	2.594
$(PW)_j$	8.928	8.438	7.671	5.739	7.278	5.473	4.915	4.304	4.512	3.732

$$PW = \left\{ \sum_{j=1}^n (PW)_j \right\} - I$$

$$PW = \sum_{j=1}^{10} (PW)_j - 50.000$$

$$= 60.990 - 50.000$$

$$= 10.990$$

Hence, for this simulation the worth is Rs 10.990 million. Such simulations are repeated a large number of times N say, more than 200, and distribution of PW is estimated. Every time new set of random normal variables are generated.

E1) By generating 10 uniform random variate $U(0, 1)$ estimate the integral

$$\theta = \frac{1}{\sqrt{2\pi}} \int_{-1}^2 e^{-x^2/2} dx$$

Recognizing this function as probability density function of $N(0, 1)$, compare the value of $\hat{\theta}$ with θ .

E2) Suppose it is known that height of adult male population can be approximated by a normal distribution with mean of 1.62 meters and variance of 0.14 meters².

Using random numbers generated in E1), simulate the height of ten persons from the $N(1.62, 0.14)$. Estimate the mean variance and range from this sample.

15.4 SUMMARY

In this Unit, we have discussed:

1. The nature and problems that can be solved by simulation;
2. Various steps involved in solving problems by simulation;
3. How approximate solution of purely deterministic problems can be obtained;
4. Some methods for inference based on simulation; and
5. Some examples of simulations.

15.5 SOLUTIONS/ANSWERS

E1) We have given the function

$$g(x) = (2\pi)^{-1/2} e^{-x^2/2}$$

If $x \sim U(-1, 2)$ then

$$E[g(x)] = \left(\frac{1}{2+1} \right) \frac{1}{\sqrt{2\pi}} \int_{-1}^2 e^{-x^2/2} dx$$

Then $\theta = 3 E[g(x)]$.

If we generate ten random variables from $U(-1, 2)$ then an estimator $\hat{\theta}$ is given by

$$\hat{\theta} = 3 \sum_{i=1}^{10} \frac{g(x_i)}{10}$$

Ten simulated random $U(0, 1)$ for a LCG were obtained as:

0.222, 0.198, 0.168, 0.784, 0.033, 0.932, 0.788, 0.237, 0.154, 0.587

$x \sim U(-1, 2)$ are obtained by $x = 3u - 1$,

and are given by

-0.334, -0.406, -0.496, 1.352, -0.901, 1.796, 1.364, -0.289, -0.538, 0.761

We have $\theta = 3 E [g(x)]$

$E [g(x)]$ is estimated as

$$E [g(x)] = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{10} \frac{e^{-x^2/2}}{10}$$

$$= 0.2786$$

Therefore, $\hat{\theta} = 3 \times 0.2786 = 0.8360$

So, θ from Normal distribution is given by

$\theta = 0.8185$ (from normal table)

Hence, $\hat{\theta}$ is not very good estimate of θ . Perhaps increase in sample size, which is ten here, will give a better estimate.

E2) We have $N(1.62, 0.14)$ normal variables generated by Box-Muller transformation on $U(0, 1)$ variables, given in E1) are

1.71, 1.80, 1.58, 1.49, 1.78, 1.35, 1.63, 1.70, 1.39, 1.58

Then,

Mean = 1.60m,

Standard Deviations = 0.154 m

Range = 0.45 m