# Computer Based Modelling and Simulation

## 2. Modelling and Simulation with Probability and Throwing Dice

*N K Srinivasan*

N K Srinivasan gradu-
ated from Indian
Institute of Science and
obtained his Doctorate
from Columbia Univer-
sity, New York. He has
taught in several
universities, and later
did system analysis,
wargaming and simula-
tion for defence. His
other areas of interest
are reliability engineer-
ing and software
development.

Most systems involve parameters and variables, which are
random variables due to uncertainties. Probabilistic meth-
ods are powerful in modelling such systems. In this second
part, we describe probabilistic models and Monte Carlo
simulation along with 'classical' matrix methods and differ-
ential equations as most real situations are complex (with
several variables) and involve uncertainties.

## 1. Input-output Models

Several industrial and engineering processes can be broken
down into stages, with each stage defined by input, process and
output segments. For a complex system, the output of one
subsystem may become the input for another subsystem.

Most real systems are more complex because there could be
feedback mechanisms, which connect the output to the input
(*Figure* 1). A familiar example of a feedback loop is the business
model in which part of the output or profit is fedback as input or
additional capital – for instance, a company may choose to
reinvest 10% of the profit for expansion of the business. Such
simple models, like compound-interest formula, are easily mod-
elled.

When there are several inputs and several outputs, the linear
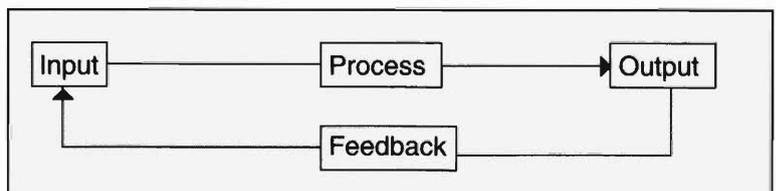model can take the form of matrix representation i.e.,

$$Y = AX,$$



**Figure 1. Process model
with feedback.**

where $Y$ is the output vector $\{y_1, y_2 \ldots, y_n\}$ and $X$ is the input vector $\{x_1, x_2 \ldots, x_m\}$ and $A$ is the $n \times m$ matrix.

A classic example of input-output model is the economic model developed by Wassily Leontief. He studied the interaction between various resources for the US economy. A typical matrix (showing only a partial list of resources) is as follows:

| Input Output →          | Agri+ Fishing | Ferrous Metals | Minerals | Construction |
|-------------------------|---------------|----------------|----------|--------------|
| Agriculture & Fishing   | 0.076         | 0.0            | 0.0      | 0.008        |
| Ferrous Metals          | 0.006         | 0.305          | 0.0      | 0.152        |
| Minerals                | 0.005         | 0.010          | 0.101    | 0.50         |
| Construction            | 0.026         | 0.004          | 0.002    | 0.0          |

The elements tell us how much of the output of one industry becomes the input for the other industry. Thus 0.6% of ferrous metals produced is used by agriculture and so on. This matrix is called the interaction matrix $[M]$.

If $[X]$ represents the production quantities in a year, called gross produce vector, then the net production available for an industry becomes:

$$[Z] = [X] - [M][X] = [I - M][X],$$

where $I$ is the unit matrix.

If an element of $[Z]$ is negative, it would indicate that the commodity should be imported from other countries.

As a simple example, consider that a nation needs only two major items: food and fuel. Let the matrix $M$ be given as follows:

$$
\begin{array}{cc}
 & \begin{array}{cc} Food & Fuel \end{array} \\
\begin{array}{c} Food \\ Fuel \end{array} & \begin{bmatrix} 0.12 & 0.07 \\ 0.15 & 0.03 \end{bmatrix}
\end{array}
$$

If required quantities are 1800 million tonnes of food and 1000 million tonnes of fuel (weighted average of coal and petroleum crude) then,

$$[I - M] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.12 & 0.07 \\ 0.15 & 0.03 \end{bmatrix} = \begin{bmatrix} 0.88 & -0.07 \\ -0.15 & 0.97 \end{bmatrix} \text{ and}$$

$$[\mathbf{X}] = [I - M]^{-1} [Z]$$

Taking the inverse of the matrix,

$$[I - M]^{-1} = \begin{bmatrix} 0.15 & 0.08 \\ 0.18 & 1.04 \end{bmatrix} \text{ and}$$

$$\mathbf{X} = \begin{bmatrix} 1.15 & 0.08 \\ 0.18 & 1.04 \end{bmatrix} \begin{bmatrix} 1800 \\ 1000 \end{bmatrix} = \begin{bmatrix} 2150 \\ 1364 \end{bmatrix}.$$

The required gross produce is 2150 million tonnes of food and 1364 million tonnes of fuel.

Such models for a nation's economy are helpful in finding the shortfalls and import/export trade needs. (*Box 1*)

## 2. Models with Differential Equations

A large category of models employ differential equations, particularly those with time-varying parameters: i.e., dynamic models. Most modelling situations involve growth, decay or saturation of a parameter such as the number of fish in a lake, radioactivity of an isotope, volume of new phases forming in a material, the number of products sold in a market zone and so on. These are modelled as follows:

Growth : $\dfrac{dx}{dt} = kx \quad x = x_0 \quad \exp{(kt)} \quad [x = x_0 \text{ at } t = 0]$

Decay : $\dfrac{dx}{dt} = -kx \quad x = x_0 \exp{(-kt)}$

Saturation : $\dfrac{dx}{dt} = k(N - x) \quad x = N(1 - \exp{(-kt)})$,

where $x$ increases from zero to $N$, the saturation value.

---

**Box 1. Matrix Methods for Nation's Economy**

The input-output model was developed by Wassily Leontief who received the Nobel Prize for Economics in 1973. His major contribution was towards the study of American economy [1919-1939] including the great depression of 1929. He used matrix methods with a $46 \times 46$ matrix!

---

The growth curves may also follow the trend in which there is exponential growth throughout but the growth rate $dx/dt$ is first positive and then becomes negative i.e., the growth rate decreases after sometime $t^*$. $t^*$ corresponds to the point of inflection: $dx/dt = 0$. Such curves are called logistic or sigmoidal curves and can be represented, in general, by the differential equation:

$$\frac{dx}{dt} = kx\,(1-x).$$

Note that $dx/dt$ reaches a maximum at $x = 0.5$. Therefore, $t$ corresponds to the time when $x = 0.5$ or the parameter is half the saturation value. As $t \rightarrow \infty$, $x$ tends to unity and $dx/dt = 0$ (*Figure* 2). The solution to the differential equation is of the form:

$$x = \frac{1}{1 + ae^{-bt}}.$$

The most common applications for logistic curves are growth of a population in a limited volume, say fishes in a pond, population of a small island nation, sales figures of a limited market and substitution of one product by another (say, replacement of analog by digital devices).

**Engineers employ models based on differential equations for electrical circuits, mechanical systems (with springs and dash-**
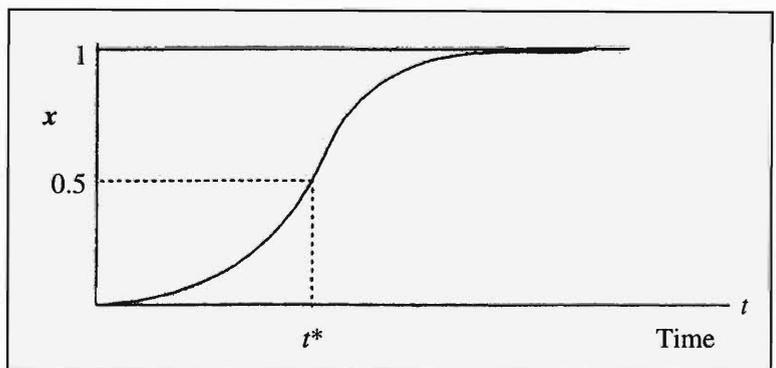


*Figure 2. Sigmoidal curve for growth and saturation.*

pots as elements) and for chemical processes. The set of linear differential equations are solved using standard techniques, such as Laplace transforms and non-linear differential equations with numerical techniques.

## 3. Discrete Event Simulation

One of the powerful techniques of modelling is to represent the modelling situation as a sequence of events. For each event, the main characteristics are often described by algebraic equations or by statistical distributions.

As an example, consider the scenario when you plan to reach railway station from your home. There are uncertainties in catching a bus and so on. You would like to estimate the likely time to reach the station. The activity can be divided into the following three events:

Event 1         packing the suitcase and taking food
Event 2         reaching bus stand and waiting for the bus
Event 3         bus travel to reach the railway station

Let us assume that the time taken for each event can be modelled with normal distribution, with the two parameters – mean and standard deviation (denoted by $N(\mu, \sigma)$). Based on past experience, we can estimate $\mu$ and $\sigma$ for each of the three events. For example:

Event 1         $N$ (40 mins, 5 mins)
Event 2         $N$ (30 mins, 10 mins)
Event 3         $N$ (70 mins, 10 mins)

Then for the complete activity, we assume that the time taken follows normal distribution, with parameters:

mean = mean 1 + mean 2 + mean 3 = 40 + 30 + 70 = 140 mins.

Standard deviation $= \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = \sqrt{25 + 100 + 100} =$

15 mins.

> One of the powerful techniques of modelling is to represent the modelling situation as a sequence of events.

Exponential distribution is widely used for 'arrival times' and 'waiting times' in queuing theory and for failure times in reliability estimation.

Therefore, the normal distribution for the entire activity becomes (140, 15). [We take the standard deviation as the root mean square for a sum of random variables, which follow normal distribution – we invoke the central limit theorem.]

We can set confidence limits to the estimated meantime of 140 minutes. With 95% confidence we can state that the time taken for reaching the station could lie between $140 - 2\sigma$ and $140 + 2\sigma$ i.e., 110 mins to 170 minutes. Since we would be interested in the safer (upper) limit, we are 95% confident that we would reach the station by 170 minutes or nearly 3 hours. Therefore, if the train departure time is 10 PM you would plan to pack at about 7 PM!

Often other distributions such as exponential and log-normal, are used. Exponential distribution is widely used for 'arrival times' and 'waiting times' in queuing theory and for failure times in reliability estimation. Lognormal distribution is used for activities such as maintenance of automobiles and aircraft, training times and manpower planning.

Discrete-event simulation is very widely used for project management, tactical training in defence, and for capacity utilization of facilities like runways in airports, piers in sea-ports, warehouses and so on.

## 4. Model Validation

Models are simplified representations of reality and are constructed based on modeller's understanding of the relevant aspects. Therefore, for the same purpose, several models can be built. A suitable or appropriate model is selected by validating the model with real-life data and examining the output of the model. Oftentimes, past data or more recent data can serve as inputs. The inputs and parameters should be taken from experts, to the extent possible. The parameters of the models are based on past experience and therefore, may require slight modifications or 'fine tuning' to provide meaningful results.

## 5. Monte Carlo Simulation

Monte Carlo, as everyone knows, is a gambling town in Europe. What has that got to do with serious work of simulation? The answer is that we use random inputs (within an interval) for simulations. This is supposed to recall gambling and hence the name Monte Carlo simulation. The procedure was developed by Stanislaw Ulam and John Van Neumann. They used the simulation method to solve partial differential equations for diffusion of neutrons! (*Box 2*).

We can illustrate the MC method by a simple example. Assume that you are running a simulation for inventory control of a factory and the demand for a hardware item, say bearings, is found to lie between 10 and 20 per month. Assume that any value between 10 and 20 is equally likely, i.e., a uniform distribution is assumed (*Figure 3*).

Since the area under the curve is unity and probability density function is constant, it is equal to 0.1 between 10 and 20 and zero elsewhere. We can generate the 'simulated demand' through the equation:

$$\text{Demand } D = 10 + (20 - 10) \times RND,$$

where *RND* is a random number from random number generator (function) which yields a value between 0 and 1. Note that if $RND = 0$, $D = 10$; if $RND = 1$, $D = 20$. Therefore for any value of *RND* between 0 and 1, we obtain a 'simulated'
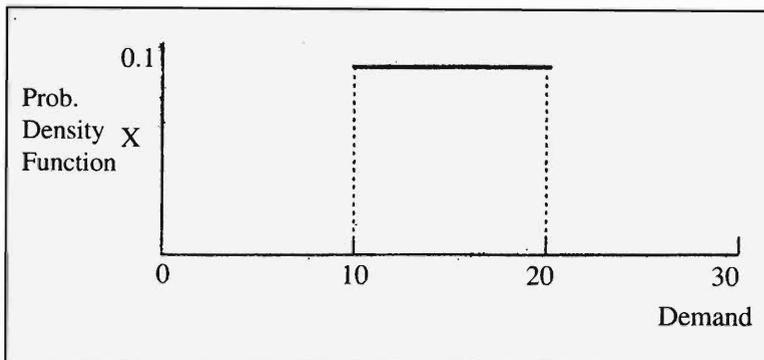
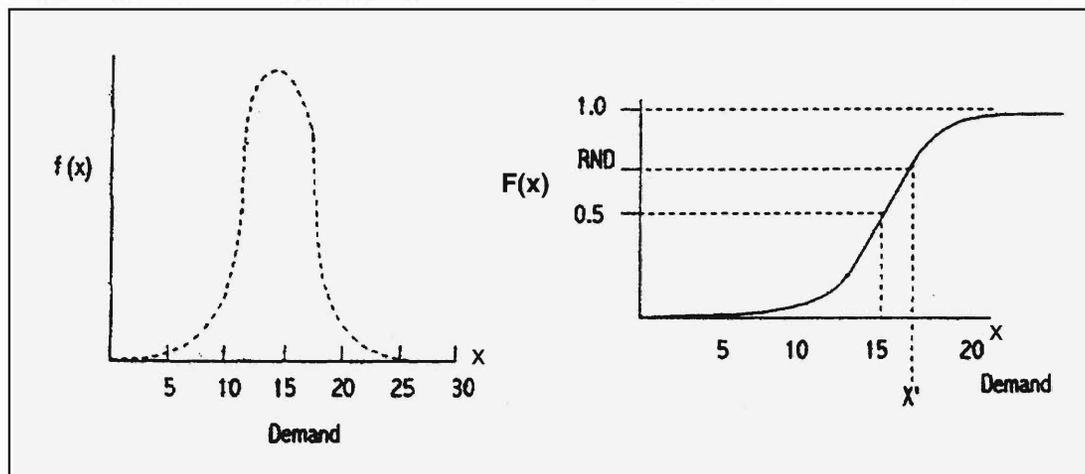*Figure 3. Probability density function for demand.*

**Figure 4. Inverse transform method for random variate.**

demand value between 10 and 20. The simulated value is called a 'random variate'.

If the parameter is a random variable following other distributions, say normal, we can derive the random variate or simulated value for input by the following procedure (Inverse transform method). (See *Figure* 4). Let $f(x)$ be the pdf and $F(x)$, the cumulative distribution function i.e.

$$F(x) \;=\; \int_{-\infty}^{x} f(x)\,\mathrm{d}x \, .$$

since $F(x)$ varies from 0 to 1 only, let us replace $F(x)$ by $RND$.

For $F(x') = RND$, find $X'$ which is $F^{-1}(x')$ as the simulated value random variate of $f(x)$.

Monte Carlo simulation runs should be repeated a large number of times (say 500-1000 times) and the results of all the runs should be analysed statistically. The simulation results fluctuate with the number of runs and with increasing number of runs, the results stabilise to a nearly constant value. The mean and standard deviation of the output (such as waiting time in a queue) can be calculated from the results. The standard deviation would decrease, as $1/n$ where $n$ is the number of runs. When the standard deviation reduces to a small value, say one tenth of

the mean, the simulation can be stopped. All these can be done effectively with simple programs that run on PC's.

## 6. Computer Based Interactive Simulation

Large-scale simulations require computer programs, not only for number crunching but also for inputting large amount of data from different sources. Consider the 'employment model' again – the data have to be collected, analysed and fed into models from different sources – authorities of school and college education, labour department and others. Computers can help in this task of interactive simulation. More importantly, decision-makers at different levels can access the results of computer based simulation. Web-based simulation models are being built – these can be used by decision-makers at different locations! In fact information technology is making a revolution in the ways in which we build models and use them.

For ease of developing simulation software, special simulation languages are available: the two most widely used languages are GPSS and SIMSCRIPT. Simulation work can also use packages for numerical methods such as MATLAB with SIMULINK and spreadsheet programs like EXCEL and LOTUS.

In the near future, large scale computer interactive simulations would help scientists, engineers and managers towards better decision-making to optimise the resources, to reduce wastage and above all, to avoid costly errors.

## Suggested Reading

[1] Web Based Simulation, *Simulation Journal*, Special Issue, Published by International Society for Computer Simulation, San Diego, July, 1999

[2] J Banks (Editor), *Handbook of Simulation*, John Wiley, 1998.

*Address for Correspondence*
N K Srinivasan
Director
Centre for Aeronautical System
Studies and Analyses
DRDO
Bangalore 560 075, India.
Email:nksrinivasan@hotmail.com

**Archimedes (Ca. 235 BC)**

Concerning levers

Give me a place to stand and I will move the earth.