CHAPTER 7
Simulation

Introduction

The previous chapters were concerned with formulating and solving mathematical models for system reliability. The solution may be either in the form of an explicit expression or the results may be obtained by numerical methods. This type of approach portrays the cause and effect relationships in the physical system and enhances the insight. Methods have been described both for Markovian and non-Markovian systems. It is obvious, however, that problems involving non-exponential distributions can become very complicated. The analytical approach is efficient and should always be employed when it is possible to develop a model which is a reasonable representation of the physical system and also when such a model is amenable to solution. Some problems are, however, too complex to be solved in this manner and simulation techniques have to be used.

In simulation, the system is divided into elements whose behaviour can be predicted either deterministically or by probability distributions. These elements are then combined to determine the system reliability. Simulation, therefore, also employs a mathematical model but it proceeds by performing sampling experiments on this mathematical model. Simulation experiments are virtually the same as ordinary statistical experiments except that they are performed on the mathematical model rather than on the actual system.

Simulation is an imprecise technique by virtue of its very statistical nature. Mathematical methods discussed in the previous chapters generally give exact results under the assumptions made. Simulation techniques, however, provide only the estimates of the exact results. Moreover they provide only a numerical value and to obtain another numerical value for a different set of parameters, the whole simulation experiment may have to be repeated. Sensitivity analysis, using a simulation approach is therefore quite expensive. It is, however, a very flexible approach and for many problems may be the only answer.

It should be noted this book deals only with digital simulation. It is called digital because most often it is executed on the digital computer but there is no inherent relationship between the two. It is a vast field and a separate book is needed to do justice to it. This book discusses the basics on which the reader can build further simulation programs.

Basic Procedure

It has been previously noted that simulation experiments are similar to ordinary statistical experiments except that they use a mathematical model of the system rather than the physical system itself. This is illustrated with the help of an example of two independent components in parallel. The system is failed when both the components are failed. The system could be constructed and operated for a long time. The history of operation and failure of the system could be recorded and the different reliability measures obtained using statistical methods. Such a method would be very expensive, especially where costly equipment is involved and would require a long time before any statement could be made about its reliability. The simulation of this system is conducted by making a mathematical model where the behaviour of the components are represented by probability distributions. Assume that component 1 is in the up state at the beginning of the experiment. Using a random number and the probability distribution of the up time of component 1, the time at which this component will fail is determined. Methods for doing this are explained later in the chapter. In a similar manner a possible duration of the repair time is generated. A history of the component generated in this manner is one possible realization of the stochastic process. The realization of component 2 is also constructed and the overlapping outage durations represent the durations of the system failure. A number of realizations of the system history can be constructed in this manner and the reliability measures obtained from these realizations using statistical methods.

In essence, simulation consists of constructing realizations of the stochastic process underlying the system and then extracting the required system performance parameters from these realizations. Most of the refinements in the theory of simulation are concerned either with developing more efficient methods of constructing realizations or extracting the information from the least possible number of realizations.

Random Number Generation

Random numbers are needed to generate random observations from the probability distributions. Tables of random numbers have been generated using mechanical or electronic devices. The basic requirement for the numbers to be random is that in a sequence each number should have equal probability of taking on any one of the possible values and it must be statistically independent of the other numbers in the sequence. While executing simulation on a digital computer, the table of random numbers can be provided externally. It is, however, more common to have the computer generate its own random numbers. There are several good methods available and only one is described here. It is a multiplicative congruential method and obtains the $(n+1)$th random number R_{n+1} from the nth random number R_n by using the following recurrence relation due to Lehmer

$$R_{n+1} = (aR_n) (\text{modulo } m)$$

where a and m are positive integers, $a < m$. The above notation signifies that R_{n+1} is the remainder when $(a R_n)$ is divided by m. The first random number R_0 is assumed and the subsequent random numbers can be generated by this recurrence relation. The sequence of random numbers so generated is periodic. Great care has to be exercised in the selection of a combination of R_0, a and m. The sequence cycle should be larger than the number of random numbers required. One combination which is satisfactory is

a = 455 470 314
m = 2^{31} - 1 = 2 147 483 647
R_0 = Any integer between 1 and 2 147 483 646.

Now if random numbers between say 0 and 999 are required, then the computer can be instructed to take the last three digits of the random number so generated. It can be seen that the sequence of random numbers so produced is predictable and reproducible and is not therefore strictly a sequence of random numbers. For this reason these random numbers are called pseudo-random numbers. They can, however, satisfactorily play the role of random numbers in digital simulation. In fact in many applications where alternative design configurations are being evaluated, the use of the same sequence of random numbers may be desirable.

Simulation Model

A simulation model representing the system to be simulated is required. The analyst should become thoroughly familiar with the system as in the case of mathematical modelling. In complex systems, failure modes and effects analysis is quite useful in gaining an insight into the system behaviour. The system is broken into elements whose behaviour can be predicted either in a deterministic manner or in the form of probability distributions. In reliability evaluation, continuous probability distributions are most often used. When historical data is available, either the frequency distribution of these data or the probability distribution which best fits these data may be used. The latter alternative is, however, more satisfactory as it comes closer to predicting the expected future performance rather than repeating the idiosyncrasies of the recorded data.

The operating rules which define the effect of the elements on each other and on the system should be specified. These rules may be either probability distributions, tables or some set of rules. It may be preferable to draw logical

flow diagrams of the system specifying the rules and logical linkages. The tendency to be over-realistic at the expense of simplicity should be guarded against.

Timing Controls

Simulation studies deal with the passage of time. There is no connection between the simulated time which represents the passage of time in the actual world and the computational time. There are two methods of representing time in computer simulation programs:

i. fixed time interval method
ii. next event method

A brief description of each is given below.

Fixed Time Interval Method

This is also called the synchronous timing method. This is a two step method. The basic time interval is Δt which may be microseconds, minutes or days. The interval length Δt will be chosen depending upon the operating characteristics of the system. Starting in the initial state, time is advanced by Δt and the program then looks to see if an event has occurred. The system is then up dated by determining the resulting state of system. If no event has occurred then the system stays in the same state. These two steps may be repeated as many times as desired.

Next Event Method

This is also called the asynchronous timing method. Simulated time, in this method, is advanced by a variable amount rather than a fixed amount each time. The computer proceeds by keeping a record of the next few simulated events scheduled to occur. The most imminent event is assumed to have occurred and the simulated time is advanced to the point of occurrence of the event. The cycle is repeated as many times as desired.

In essence, in the synchronous timing method, the time is advanced by definite amounts and every time the system is updated by determining the event that occurred during this interval and in the asynchronous timing method, the next event is determined and the time is advanced to the occurrence of this event. The occurrence of an event during an interval or the time to the occurrence of an event is determined using the following sampling techniques.

Random Sampling

When all the elements operate and interact in a deterministic manner, the event occurring during an interval or the time till the next event, is easily determined. In systems with stochastic elements, these random observations are obtained from the probability distributions using random numbers and methods of generating random observations from probability distributions are required.

Discrete Distributions

Two methods for modelling discrete distributions are described below.

1 *Proportionate Allocation Technique*

It consists in allocating the possible values of the random number to the various values of the random variable underlying the distribution in direct proportion to their respective probabilities. A random number is selected and the corresponding value of the random variable is the random observation. The method is quite useful in simulating discrete time Markov chains when the transition probability matrix is specified. This method is illustrated by the example of man who if he does his exercises one day, is 70% sure not to do them next day. On the other hand, if he does not do his exercises one day, is 60% sure not to do them the next day. Denoting the doing and not doing of the exercises by 0 and 1 respectively, the transition probability matrix is

	Final state	
Initial state	0	1
0	0·3	0·7
1	0·4	0·6

The realizations are constructed using a table of random digits. If the man is in state 0, select a single random digit and the next state is determined as follows

digit	*event*
0 – 2	stay in 0
3 – 9	transit to 1

Similarly if the man is in state 1

digit	*event*
0 – 3	transit to 0
4 – 9	stay in 1

The construction of a realization for ten days is shown in Table 1. It is assumed that the man does his exercises on the first day. It should be noted that this is only one possible realization of the stochastic process. The stochastic nature is evident from the sequence of states occupied. For deriving probabilites,

a number of such realizations have to be constructed. The methods for deriving measures from these realizations is described later in the chapter.

Table 7.1 *Constructing a Realization*

Day	*Random number*	*State*
1		0
2	4	1
3	3	0
4	4	1
5	1	0
6	2	0
7	2	0
8	2	0
9	2	0
10	0	0

2 *Inverse of the Probability Distribution Method*

This method is practically the same as method 1 but is a little more involved and proceeds in the following steps:

1. Construct the distribution function of the random variable X, i.e. $F(x) = P(X \leqslant x)$. The distribution function has the property that is monotonically increasing. The probability mass function of an arbitrary random variable X and its distribution function are shown in Fig. 7.1.

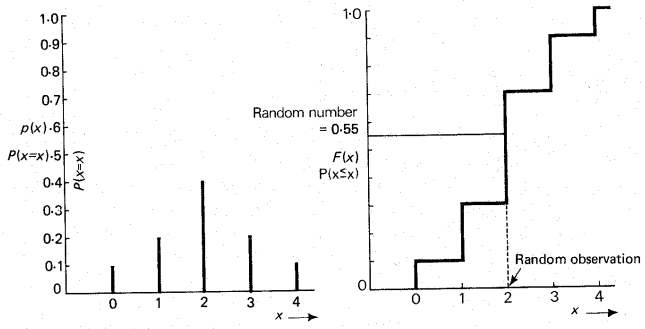

Fig. 7.1 Probability mass function and probability distribution function

2. Generate a random decimal number between 0 and 1. This is achieved by obtaining a random integer with the desired digits and then placing a decimal point before it.
3. Set $F(x)$ equal to the random number and select the value of x corresponding to $F(x)$. This value of x is the desired random observation from the probability distribution.

It should be noticed that $F(x_i) - F(x_{i-1})$ is equal to $P(X = x_i)$, and if the random number falls in the interval $(F(x_i), F(x_{i-1}))$, the value of $X = x_i$ will be selected. The procedure therefore essentially allocates the random numbers to the random variables in the proportion of their probabilities of occurrence. The basic procedure is therefore the same as that of method 1.

Continuous Distributions

The above procedure can also be used for continuous distributions. Continuous distributions are approximated by discrete distributions whose irregularly spaced points have equal probabilities. The accuracy can be increased by increasing the number of intervals into which (0,1) is divided. This requires additional data in the form of tables. Although the method is quite general, its disadvantages are the great amount of work required to develop tables and possible computer storage problems. The following analytic inversion approach is simpler.

Let z be a random number in the range 0 to 1 with either a uniform probability density function or a triangular distribution function, i.e.

$$f(z) = \begin{cases} 0 & Z < 0 \\ 1 & 0 \leqslant Z \leqslant 1 \\ 0 & Z > 1 \end{cases}$$

Similarly

$$F(z) = \begin{cases} 0 & Z < 0 \\ z & 0 \leqslant Z \leqslant 1 \\ 1 & Z \geqslant 1 \end{cases}$$

Let $F(x)$ be the distribution function from which the random observations are to be generated. Let

$$z = F(x)$$

Solving the equation for x gives a random observation of X. That the observations so generated do have $F(x)$ as the probability distribution can be shown as follows.

Let ϕ be the inverse of F, then

$$x = \phi(z)$$

Now x is the random observation generated. To determine its probability distribution

$$P(x \leqslant X) = P(F(x) \leqslant F(X)) = P(z \leqslant F(X)) = F(X)$$

Therefore the distribution function of x is $F(x)$ as required. In the case of several important distributions, special techniques have been developed for efficient random sampling. A few cases are described in the following section. The reader can refer to Reference 5 for a more detailed treatment.

Exponential Distribution

The exponential distribution has the following probability distribution

$$P(X \leqslant x) = 1 - e^{-\rho x}$$

where $1/\rho$ is the mean of the random variable X. Setting this function equal to a random decimal number between 0 and 1

$$z = 1 - e^{-\rho x}$$

Since the complement of such a random number is also a random number the above equation can be written as

$$z = e^{-\rho x}$$

Taking the natural logarithm of both sides and simplifying

$$x = \frac{\ln(z)}{-\rho}$$

which is the desired random observation from the exponential distribution having $1/\rho$ as the mean.

Erlang Distribution

The above procedure can be readily extended to generate random observations from an Erlang distribution. It has been shown in Chapter 6 that the sum of 'a' independent exponentially distributed random variables each with mean $\frac{1}{\rho}$ has the Erlang distribution with mean $\frac{a}{\rho}$ and 'a' as the shape parameter. Therefore if we have a sequence of 'a' random decimal numbers in the interval (0,1), denoted by $z_1, z_2, \ldots, z_a$, the random observation from Erlang distribution is

$$x = \sum_{i=1}^{a} \frac{\ln(z_i)}{-\rho}$$

$$= -\frac{1}{\rho}\left(\ln \prod_{i=1}^{a} z_i\right)$$

Normal Distribution

The following describes a technique developed by Box and Muller. The method proceeds by generating pairs of normal deviates. The joint distribution of two independent standardized normal deviates is given by

$$f_{XY}(x,y) = \frac{1}{2\pi} \exp\{-\tfrac{1}{2}(x^2 + y^2)\} \tag{7.1}$$

Consider the polar transformation

$$x = u \operatorname{Cos} v$$

$$y = u \operatorname{Sin} v$$

The inverse transformation is

$$u = (x^2 + y^2)^{\frac{1}{2}}$$

and

$$v = \arctan \frac{y}{x}$$

The above functions can be written in a general form

$$U = U(X, Y) \qquad V = V(X, Y)$$

and

$$X = X(U, V) \qquad Y = Y(U, V)$$

When X is near x and Y is near y, U and V must be near u and v. Therefore

$$\begin{aligned} P(x < X \leqslant x + dx, y < Y \leqslant y + dy) &= f_{XY}(x,y)\, dx\, dy \\ &= P(u < U \leqslant u + du, v < V \leqslant v + dv) \\ &= f_{UV}(u,v)\, du\, dv \end{aligned} \tag{7.2}$$

Therefore

$$f_{UV}(u,v) = f_{XY}(x,y)\begin{vmatrix} \frac{dx}{du} & \frac{dy}{dv} \end{vmatrix}$$

Absolute values are used so that the expression is applicable for both non-increasing and non-decreasing functions.

$$dx\,dy = u\,du\,dv$$

Therefore

$$f_{UV}(u, v)\,du\,dv = \frac{1}{2\pi} e^{-u^2} u\,du\,dv \tag{7.4}$$

$$= e^{-u^2} d(\tfrac{1}{2}u^2) \frac{1}{2\pi} dv$$

This expression can be interpreted in this manner: u^2 is exponentially distributed and v has a uniform distribution in the interval $(0, 2\pi)$. Now if there are two random decimal numbers z_1, z_2 in $(0, 1)$

$$u^2 = -ln\, z_1$$

i.e.

$$u = \sqrt{ln\, 1/z_1}$$

and

$$v = 2\pi z_2$$

Hence

$$X = \sqrt{ln \frac{1}{z_1}} \text{Cos}\, 2\pi z_2$$

and

$$Y = \sqrt{ln \frac{1}{z_1}} \text{Sin}\, 2\pi z_2$$

are exact independent normal deviates.

This book covers only a few cases of analytic inversion. Many other methods exist for analytic inversion of particular probability distributions including inversion by graphical or tabular means.

Estimating Reliability Measures

Reliability measures can be calculated from the realizations using statistical methods. It is possible to construct probability distributions for the various residence times but usually the mean values are the main parameters.

Time Specific Probability of X^+

If N observations of the state of system are made at time t, and n_+ of the times the system is found in X^+, the estimate of the probability of X^+ is found by

$$\hat{p}_+(t) = \frac{n_+}{N} \tag{7.5}$$

It is well known from the frequency concept of probability that

$$p_+(t) = \hat{p}_+(t) = \frac{n_+}{N}$$

$$N \to \infty$$

Interval Frequency

If in N realizations of the system state, the system visited X^+ n_c times, the estimate of $F(0, t)$ is

$$\hat{F}_+(0, t) = \frac{n_c}{N} \tag{7.6}$$

Fractional Duration

The fractional duration in X^+ is estimated by

$$\hat{D}_+(0, t) = \frac{\sum_{i=1}^{N} u_i}{tN} \tag{7.7}$$

where u_i is the time spent in X^+ in the ith realization.

Steady State Probability of being in X^+

This can be calculated either from a number of realizations, allowing sufficient time for letting each realization to reach equilibrium state or it can be calculated as the fraction of time spent in X^+ in a very long realization, i.e.

$$\hat{P}_+ = \frac{t_+}{T}$$

where t_+ is the time spent in X^+ in interval $(0,T)$ when T is large. This latter approach is less time consuming as in the former approach considerable time is spent in reaching the equilibrium condition.

Mean Cycle Time

After the simulated system has reached an equilibrium, the mean cycle time can be estimated as follows

$$\hat{T}_+ = \frac{\sum_{i=1}^{n} t_i}{n}$$

where t_i is the time interval between the $(i-1)$th and the ith encounter of X^+.

Equilibrium Conditions and Sample Size

The most important reliability measures in repairable systems are obtained from the steady state or equilibrium conditions. In this case the system state probabilities are independent of the initial conditions and the time elapsed since the start of system. The system reaches a steady state condition when the system state probability distribution reaches a limiting equilibrium distribution. It should be remembered that steady state condition can only be approached but never exactly attained.

In determining the mean cycle time, the data obtained during the initial period of simulated system operation should be excluded. It is difficult to know how long the system should be operated before taking observations. It could, however, be roughly estimated by having a few trial runs and estimating the probability distributions at various points in time. It should be noted that even when steady state measures are the basis of evaluating alternative designs, the same initial conditions should preferably be used. Simulation or Monte Carlo techniques are used to obtain a numerical estimate of the inherent system reliability parameters. As the sample size increases the estimated value approaches the estimand. The question which now arises is how big should the sample size be? It is not adequate to simulate the system for an arbitrary long time and then simply assume that the results are sufficiently precise.

Reliability measures obtained from each simulated sample run are generally different and one object is to determine the mean value of the measure. A simple case is when these observations of the measure are statistically independent and have a common normal distribution. This case is considered and then extended to more general situations. The confidence interval for the mean m of the normal distribution having variance σ^2 can be obtained as follows. Let $\bar{X}$ be the sample mean obtained from a random sample of size n. $\bar{X}$ is then also a random variable and it can be shown that

$$Z = \frac{(\bar{X} - m)\sqrt{n}}{v}$$

has a t distribution with $(n-1)$ degrees of freedom. The value of v is

$$v^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$$

where X_i is the ith observation. Now

$$P\left\{-t_{\alpha/2}^{n-1} \leqslant \frac{(\bar{X} - m)\sqrt{n}}{v} \leqslant t_{\alpha/2}^{n-1}\right\} = 1 - \alpha$$

where $t_{\alpha/2}^{n-1}$ is the 100 $\alpha/2$ percent point of the distribution with $(n-1)$ degrees of freedom and can be found from Tables of the t distribution. The above

expression can be rearranged as

$$P\left\{\bar{X} - t_{\alpha/2}^{n-1} \frac{v}{\sqrt{n}} \leqslant m \leqslant \bar{X} + t_{\alpha/2}^{n-1} \frac{v}{\sqrt{n}}\right\} = 1 - \alpha$$

Therefore with confidence $1 - \alpha$, the upper and lower bounds of m are

$$\text{upper bound} = \bar{X} + t_{\alpha/2}^{n-1} \frac{v}{\sqrt{n}}$$

$$\text{lower bound} = \bar{X} - t_{\alpha/2}^{n-1} \frac{v}{\sqrt{n}}$$

The required sample size can be predicted by obtaining an estimate of the standard deviation of the observations either from pilot runs or from initial observations. As can be seen from above, the interval between the upper and lower bounds can be made as narrow as desired by making the sample size sufficiently large.

Two assumptions made above are that the observations of the measure are statistically independent and have a common normal distribution. This, however, is not true in general and the following methods may be employed to realize these assumptions in practice.

The first method is to have a number of independent simulated runs. The mean measure obtained from each run can be used as an independent observation. These can be assumed to be normally distributed in accordance with the Central Limit Theorem. The confidence interval therefore can be found by the procedure described.

When steady state measures are being estimated, the above procedure can be quite wasteful since in each simulation run, the initial period is unproductive. The alternative is to use a single simulated run and to divide the steady state period into equal long intervals. The value derived from each interval can be used as an observation. It should be noted that these observations are not completely independent but by making the intervals sufficiently long, the correlation can be decreased.

Variance Reducing Techniques

It has been pointed out that the precision of sample estimates can be increased by making the sample size large enough. Increasing the precision is equivalent to decreasing the variance of the sample estimates. The simple method of repeated runs (or making a single run very large and dividing it into equal intervals), treating the measures obtained from each sample as independent sample values until the variance has been reduced to the desired level is usually quite time-consuming. Special techniques for reducing variance have

been devised. These techniques, extract as much and as precise information as possible from the amount of simulation that can be economically executed. These are three generally used techniques:

1. stratification
2. control variates
3. antithetic variables

The reader is referred to References 4–5 for discussion of these methods.

References

1. D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes*, Methuen, London (1965).
2. C.D. Flagle, W.H. Huggins and R.H. Roy, *Operations Research and Systems Engineering*, Johns Hopkins Press, Baltimore (1960).
3. F.S. Hillier and G.J. Lieberman, *Introduction To Operations Research*, Holden-Day (1970).
4. J.M. Hammersley and D.C. Handscomb, *Monte Carlo Methods*, Methuen, London (1964).
5. K.D. Tocher, *The Art of Simulation*, English Universities Press, London (1963).
6. C.F. DeSieno and L.L. Stine, A Probability Method for Determining the Reliability of Electric Power Systems, *Trans. AIEE* (February 1964).