Probability Models.S4
Simulating Random Variables

In the fashion of the last several sections, we will often create probability models of certain aspects of systems. For simple cases, we compute probabilities of events using probability theory. In the remainder of this book, we will find other situations that are affected by more than one random variable. Because of their complexity, however, many problems will not yield to easy analysis. Rather, it is necessary to *simulate* the several random variables that impact the situation and observe and analyze their effects through statistical analysis. The approach is widely used in operations research. We begin the discussion of simulation in this section and carry it throughout the rest of the book. In addition to its power as a tool for analyzing complex systems, simulation provides a medium for illustrating the ideas of variability, uncertainty and complexity. We construct a simulated reality and perform experiments on it.

A Production Process

Ex. 18. Consider an assembly process with three stations. Each station produces a part, and the three parts are assembled to produce a finished product. The stations have been balanced in terms of work load so that each has an average daily production equal to ten units. The production of each station is variable, however, and we have established that the daily production in a station has a discrete uniform distribution ranging from 8 to 12. Work in process cannot be kept from one day to the next, so the production of the line is equal to the smallest station production. Let X_1, X_2, and X_3 be random variables that are the production amounts of the three stations. The daily production of the line is a random variable, Y, such that: $Y = \text{Min}\{X_1, X_2, X_3\}$. Our interest is learning about the distribution of the random variable Y.

The dedicated student may be able to derive the p.d.f. of Y given this information, however, we will take the conceptually easier approach of simulation. We observe in Table 15 ten days of simulated operation. The production in each station is simulated[1] using the methods described later in this chapter. With these observations we compute the simulated production of the line as

$$y = \text{Min}\{x_1, x_2, x_3\}.$$

[1]The observations were simulated by: $x_i = \text{ROUND}(7.5 + 5r)$. This provides a discrete uniform distribution with integer values 8 through 12.

Table 15. Simulated Observations

x_1	11	10	8	11	9	12	8	11	12	9
x_2	12	10	10	12	8	9	8	9	10	9
x_3	9	8	8	9	9	12	12	9	10	10
y	9	8	8	9	8	9	8	9	10	9

The simulation provides information about the system. The statistical mean and standard deviation for the system production is

$$\bar{y} = \; = 8.7, \; s^2 = 0.4556, \; s = 0.6749.$$

We conclude that the average daily production of the system is well below the average of the individual stations. The cause of this reduction in capacity is variability in station production.

Since the statistics depend on the specific simulated observations, the analyst usually runs replications of the simulation to learn about the variability of the simulated results. The table below shows six replications of this experiment. Although the results do vary, they stay within a narrow range. Our conclusion about the reduction in the system capacity is certainly justified.

Replication	1	2	3	4	5	6
$\bar{y}$	8.7	8.6	9.1	9	8.8	8.8

Sample Statistics

The *mean:*

$$\bar{y} = \sum_{i=1}^{n} y_i \Big/ n$$

The *sample variance*:

$$s^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 \Big/ (n-1)$$

The *standard deviation*:

$$s = \sqrt{s^2}$$

The grand average and the standard deviation of the six observations is 8.833 and 0.186, respectively. Since each of the replication averages are determined by the sum of ten numbers, we can conclude from the Central Limit Theorem that the average observations are approximately Normally distributed. This allows a number of additional statistical tests on the results of the simulation replications.

Simulation is a very powerful tool for the analysis of complex systems involving probabilities and random variables. Virtually any system can be simulated to obtain numerical estimates of its parameters when its logic is understood and the distributions of its random variables are known. The approach will be used extensively in the following chapters to illustrate the theoretical concepts, as well as to provide solutions when the theoretical results are not available or too difficult to apply.

Simulating with the Reverse Transformation Method

The conceptually simplest way to simulate a random variable, X, with density function, $f_X(x)$, is called the reverse transformation method. The cumulative distribution function is $F_X(x)$.

Let R be a random variable taken from a uniform distribution ranging from 0 to 1, and let r be a specific observation in that range. Because R is from a uniform distribution

$$P\{R \leq r\} = r.$$

Let the simulated value x_S be that value for which the c.d.f., $F(x_S)$, equals r.

$$r = F(x_S) \text{ or } x_S = F^{-1}(r).$$

The random observation is the reverse transformation of the c.d.f., thus providing the name of the method. The process is illustrated for the example in Fig. 14 where we show the c.d.f for the production in a station. Say we select the random number 0.65 from a continuous uniform distribution. We locate 0.65 on the $F(x)$ axis, project to the cumulative distribution function (the dotted line), and find the value of the random variable x_s for which $F(x_s) = 0.65$. For this instance $x_s = 11$.

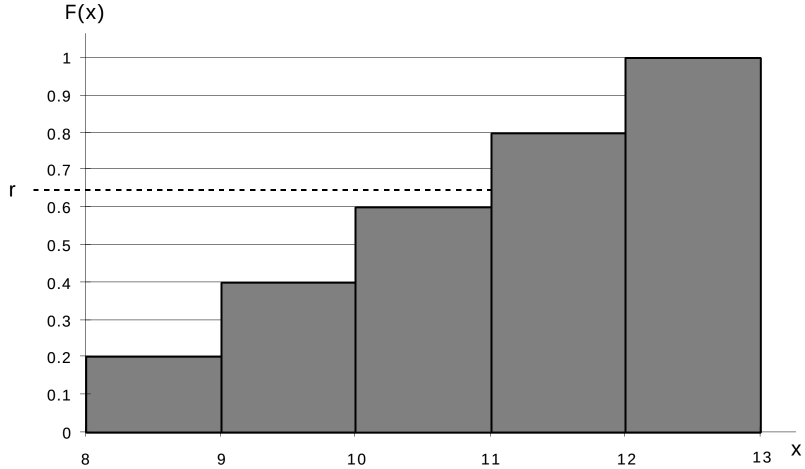

Figure 14. The reverse transformation method for a discrete distribution.

Table 16 shows the random numbers that provided the observations of Table 15.

Table 16. Random Numbers

	1	2	3	4	5	6	7	8	9	10
1	0.6151	0.5569	0.18	0.7291	0.3046	0.869	0.1611	0.6633	0.9535	0.3298
2	0.861	0.5541	0.4883	0.983	0.1391	0.251	0.1389	0.3261	0.5123	0.3793
3	0.3108	0.0923	0.141	0.3124	0.2558	0.9785	0.9638	0.3351	0.5778	0.4323

Discrete Random Variables

To illustrate the reverse transformation method for a discrete random variable, consider the binomial distribution with $n = 3$ and $p = 0.9$. The p.d.f. is :

$$P_X(k) = \binom{3}{k}(0.9)^{k}(0.1)^{3-k} \text{ for } k = 0,1,2,3.$$

Table 17 shows the probabilities associated with this distribution.

We next define intervals on the real number line from 0 to 1 for the possible values of the random variable. The intervals are computed using the cumulative distribution. In general, when r is a real number and x is a finite discrete random variable, the intervals are

$$\boldsymbol{I}(k) = \{r \mid F_X(k-1) < r < F_X(k)\} \quad k = 0,1,2, \ldots n.$$

For the example case the intervals are shown in Table 17.

Table 17. Simulation Intervals for Binomial Distribution with $n = 3$, $p = 0.9$.

k	0	1	2	3
$P(k)$	0.001	0.027	0.243	0.729
$F(k)$	0.001	0.028	0.271	1.000
$\boldsymbol{I}(k)$	0 - 0.001	0.001 - 0.028	0.028 - 0.271	0.271 - 1.0

To simulate an observation from the distribution, we select a random number and determine the simulation interval in which the random number falls. The corresponding value of k is the simulated observation. For instance, using the random number 0.0969, we find that 0.0969 is in the range $\boldsymbol{I}(2)$. The simulated observation is then 2. In general, let r_i be a random number drawn from a uniform distribution in the range (0, 1). The simulated observation is

$$x_i = \{k \mid F_X(k - 1) < r_i < F_X(k)\}.$$

Six simulated values are shown in Table 18.

Table 18. Random Numbers and Simulated Values

Random Number	0.0969	0.2052	0.0013	0.2637	0.6032	0.5552
Simulated Observation	2	2	1	2	3	3

Based on a sample of 30 observations, the sample mean, variance and standard deviation are computed and shown in Table 19. We also estimate the probabilities of each value of the random variable from the proportion of times that value appears in the data. The sample results are compared to the population values in the table.

Table 19. Distribution Information Estimated from the Simulation

	Estimated from Sample	Population Value
Mean	2.667	2.7
Variance	0.2989	0.27
Standard Deviation	0.5467	0.5196
$P(0)$	0	0.001
$P(1)$	0.0333	0.027
$P(2)$	0.2667	0.243
$P(3)$	0.7	0.729

The statistics obtained from the sample are estimates of the population parameters. If we simulate again using a new set of random numbers we will surely obtain different estimates. Comparing the statistics to the population values we note that the statistical mean and variance are reasonably close to the population parameters that they estimate. We did not observe any values of 0 from the sample. This is not surprising as the probability of such an occurrence is one in a thousand, and there were only 30 observations.

Continuous Random Variables

Ex. 19. A particular job consists of three tasks. Tasks A and B are to be done simultaneously. Task C can begin only when both tasks A and B are complete. The times required for the tasks are T_A, T_B, and T_C respectively, and all times are random variables. T_A has an exponential distribution with a mean of 10 hours, T_B has a uniform distribution that ranges between 6 and 14 hours, and T_C has a Normal distribution with a mean of 10 hours and a standard deviation of 3 hours. The time to complete the project, Y, is a random variable that depends on the task times as

$$Y = \text{Max.}\{ T_A, T_B \} + T_C .$$

What is the probability that the promised completion time of 20 hours is met.

An analytical solution to this problem would be difficult because the completion time is a complex combination of random variables. Simulation provides statistical estimates concerning the completion time.

Table 20 shows the results for 10 simulations. The row labeled r provides the random numbers used to simulate the random variables and the rows labeled t_A, t_B and t_c show the associated observations. The row labeled y is computed using the system equation and the observations in each column. The process of producing the simulated observations are described in the following.

Table 20. Simulated results for the example problem

	1	2	3	4	5	6	7	8	9	10
r	0.663	0.979	0.256	0.312	0.141	0.092	0.311	0.251	0.139	0.983
t_A	10.89	38.41	2.954	3.745	1.52	0.968	3.722	2.89	1.497	40.76
r	0.953	0.33	0.139	0.326	0.512	0.379	0.964	0.335	0.578	0.432
t_B	13.63	8.638	7.111	8.609	10.1	9.034	13.71	8.681	10.62	9.459
r	0.731	0.615	0.557	0.18	0.729	0.305	0.869	0.861	0.554	0.488
t_C	11.85	10.88	10.43	7.253	11.83	8.466	13.36	13.25	10.41	9.912
y	25.48	49.28	17.54	15.86	21.93	17.5	27.07	21.94	21.03	50.67

It is clear from the ten completion times in Table 20, that there is quite a bit of variability in an estimate of the time to complete the project. There are 3 numbers below 20, so an initial estimate of the probability of a completion time less than 20 would be 0.3. A much larger sample would be necessary for an accurate determination.

Closed Form Simulation

We continue to use the reverse transformation method to simulate continuous random variables. For continuous variables, however, probabilities do not occur at discrete values. Rather we have the c.d.f. that gives probabilities

$$P\{X < x\} = F(x)$$

The process for generating random follows.

- Select a random number r from the uniform distribution.
- Find the value of x for which $r = F(x)$ or $x = F^{-1}(r)$.
- The resulting x is the simulated value.

For some named distributions we have a closed form functional representation of the c.d.f., while for others we must use a numerical procedure to find the inverse probability function. When we have a closed form expression for the c.d.f. that can be solved for x given r, the simulation procedure is easily performed. For example consider the exponential distribution for task A. The mean of 10 hours yields the parameter $\lambda = 0.1$. The general form of the p.d.f. is

$$f(x) = \lambda exp(-\lambda x) \text{ and } F(x) = 1 - exp(-\lambda x) \text{ for } x \geq 0.$$

Setting the cumulative distribution equal to r and solving we obtain

$$r = 1 - exp(-\lambda x) \text{ or } x = -(1/\lambda)\ln(1 - r).$$

The first line labeled r in Table 20 is used to simulate T_A, shown in the following line. For example when

$$r = 0.663,\ t_A = -10\ln(1 - 0.663) = 10.89.$$

The uniform distribution associated with T_B is also simulated with a closed form expression. The c.d.f. for a uniform distribution ranging from a to b is

$$F(x) = (x - a)/(b - a) \text{ for } a \leq x \leq b.$$

Setting r equal to the c.d.f. and solving form x yields the simulated value

$$x = a + r(b - a).$$

To illustrate, we compute the first simulated value of T_B using the random number $r = 0.953$.

$$t_B = 8 + 0.953(14 - 8) = 13.63.$$

The Weibull distribution also allows a closed form simulation. We set the expression for the c.d.f. equal to the random number

$$r = F(x) = 1 - exp\{-\alpha x^{\beta}\} \text{ for } x \geq 0.$$

Solving for x, we obtain the expression for the simulated random variable

$$x = \sqrt[\beta]{\frac{-ln[1 - r]}{\alpha}}.$$

We recognize that when r is a random number from the range 0 to 1, $(1 - r)$ is also a random number from the same range. This simplifies the expression for the simulated observation to

$$x = \sqrt[\beta]{\frac{-ln(r)}{\alpha}}\ .$$

Simulation using Inverse Probability Functions

When a closed form expression is not available for the cumulative distribution, a computer program may be used to numerically compute the inverse probability function. For the example we used the inverse Normal distribution function available with the Excel spreadsheet program to simulate the observations from T_C. To simulate we must find

$$t_C = F^{-1}(r)$$

where $F^{-1}(r)$ is the inverse probability function for the Normal distribution with mean 10 and standard deviation 3 evaluated for the value r. Fig. 15 illustrates the reverse transformation in the case where a random number equal to 0.731 yields t_C equal to 11.85.

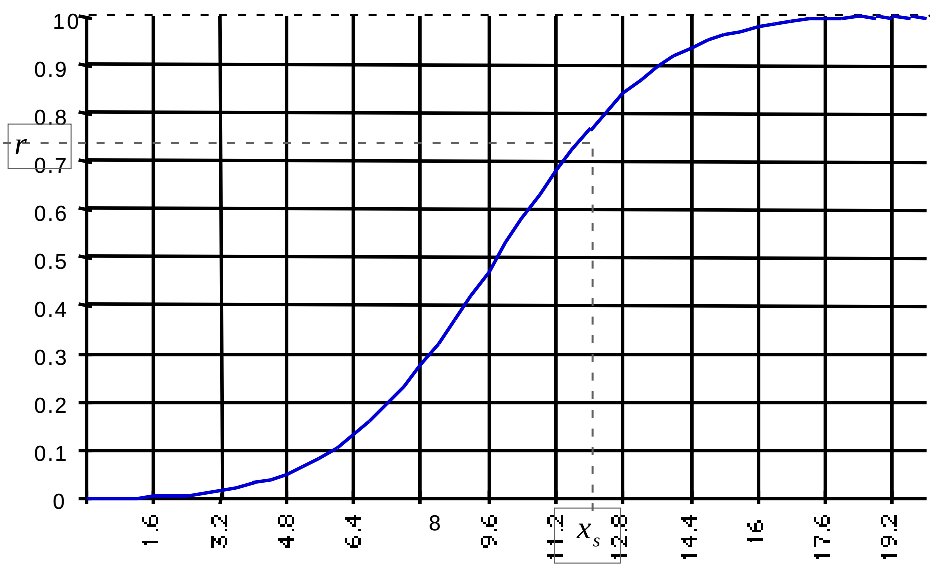

Figure 15 The reverse transformation method for a Normal distribution

We complete the first column of the example by computing y with the equation defining the interaction between the random variables.

$$y = \text{Max}(10.89, 13.63) + 11.85 = 25.48.$$

Using Simulation to Solve Problems

The main purpose for simulation is to analyze problems involving combinations of random variables that are too complex to be conveniently analyzed by probability theory.

Simulation is a very powerful tool for the analysis of complex systems involving probabilities and random variables. Virtually any system can be simulated to obtain numerical estimates of its parameters when its logic is understood and the distributions of its random variables are known. The approach is used extensively in the practice of operations research.

Simulation does have its disadvantages however. Except for very simple systems, simulation analysis is computationally expensive. Inverse probabilities functions that do not have a closed form may be difficult to compute and complex systems with many random variables will use computer time. Simulation results are statistical. One selection of random variables yields a single observation. Since most situations involve variability, it is necessary to make a large number of observations and use statistical techniques for drawing conclusions. Simulation results are sometimes difficult to interpret. This is partly due to the complexity of many simulated situations and partly due to their inherent variability. Effects may be observed, but causes are difficult to assign.