I M.Sc - Mathematics

Semester : 02 , Operations Research

Unit : II : Simulation , Decision Analysis.

Simulation

Page No

(i) Monte Carlo Simulation : 605 / 16.1

(ii) Types of Simulation : 610 / 16.2

(iii) Elements of Discrete - Event Simulation
611 / 16.3

(iv) Generation of Random Numbers
622 / 16.4

(v) Mechanics of Discrete Simulation -
Manual Simulation of single server model
624 / 16.5, 16.5.1

Decision Analysis.

(i) Decision making under certainty - Analytic Hierarchy Process (AHP)
Page No : 490 / 13.1

(ii) Decision Making under Risk.
500 / 13.2 ; 13.2.1, 13.2.2.

(iii) Decision under Uncertainty
515 / 13.3.

CHAPTER 16

Simulation Modeling

Chapter Guide. Simulation is the next best thing to observing a real system. It deals with a computerized imitation of the random behavior of a system for the purpose of estimating its measures of performance. Basically, simulation views an operational situation as a waiting line in a service facility. By literally following the movements of customers in the facility, pertinent statistics (e.g., waiting time and queue length) can be collected. The task of using simulation starts with the development of the logic of the computer model in a manner that will allow collecting needed data. A number of computer languages are available to facilitate these tedious computations.

A common misuse of simulation is to run the model for an arbitrary time period, and then view the results as the "true gospel." In fact, simulation output changes (sometimes drastically) with the length of the run. For this reason, simulation modeling deals with a statistical experiment whose output must be interpreted by appropriate statistical tests. As you study the material in this chapter, pay special attention to the peculiarities of the simulation experiment, including (1) the important role of (0,1) random numbers in sampling from probability distributions, and (2) the special methods used to collect observations to satisfy the underlying assumption of a true statistical experiment.

The prerequisite for this chapter is a basic knowledge of probability and statistics. A background in queuing theory is helpful.

This chapter includes 10 solved examples, 2 Excel templates, and 44 end-of-section problems. The AMPL/Excel/Solver/TORA programs are in folder ch16Files.

16.1 MONTE CARLO SIMULATION

A forerunner to present-day simulation is the Monte Carlo technique, a modeling scheme that estimates stochastic or deterministic parameters based on random sampling. Examples of Monte Carlo applications include evaluation of multiple integrals, estimation of the constant π ($\cong$ 3.14159), and matrix inversion.

This section uses an example to demonstrate the Monte Carlo technique. The objective of the example is to emphasize the statistical nature of the simulation experiment.

Example 16.1-1

We will use Monte Carlo sampling to estimate the area of a circle defined as

$$(x-1)^2+(y-2)^2=25$$

The radius of the circle is $r = 5$ cm, and its center is $(x, y) = (1, 2)$.

The procedure for estimating the area requires enclosing the circle tightly in a square whose side equals the diameter of the circle, as shown in Figure 16.1. The corner points are determined from the geometry of the square.

The estimation of the area of the circle is based on the assumption that all the points in the square are equally likely to occur. Taking a random sample of n points in the square, if m of these points fall within the circle, then

$$\begin{pmatrix}\text{Estimate of the}\\ \text{area of the circle}\end{pmatrix} \cong \frac{m}{n}\begin{pmatrix}\text{Area of}\\ \text{the square}\end{pmatrix} = \frac{m}{n}(10 \times 10)$$

To ensure that all the points in the square occur with equal probabilities, we represent the coordinates x and y of a point in the square by the following *uniform* distributions:

$$f_1(x) = \frac{1}{10}, \; -4 \le x \le 6$$

$$f_2(y) = \frac{1}{10}, \; -3 \le y \le 7$$

A sampled point (x, y) based on the distribution $f_1(x)$ and $f_2(y)$ guarantees that all points in the square are equally likely to be selected.

The determination of a sample (x, y) is based on the use of independent and uniformly distributed random numbers in the range (0, 1). Table 16.1 provides a small list of such numbers which we will use in the example computations. For the purpose of general simulation, special arithmetic operations are used to generate the 0-1 random numbers, as will be shown in Section 16.4.

For a pair of 0-1 random numbers, R_1 and R_2, a random point (x, y) in the square is determined by mapping them on the x and y axes of Figure 16.1 using the following formulas:

$$x = -4 + [6 - (-4)]R_1 = -4 + 10R_1$$

$$y = -3 + [7 - (-3)]R_2 = -3 + 10R_2$$

FIGURE 16.1

Monte Carlo estimation of the area of a circle

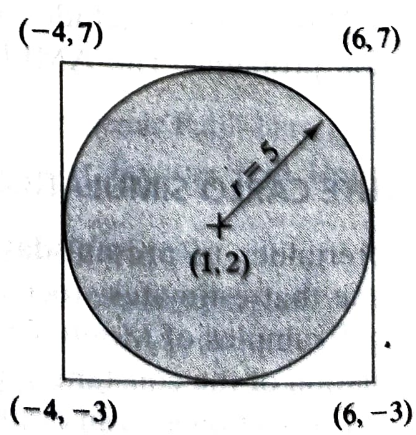


TABLE 16.1 A Short List of 0-1 Random Numbers

.0589	.3529	.5869	.3455	.7900	.6307
.6733	.3646	.1281	.4871	.7698	.2346
.4799	.7676	.2867	.8111	.2871	.4220
.9486	.8931	.8216	.8912	.9534	.6991
.6139	.3919	.8261	.4291	.1394	.9745
.5933	.7876	.3866	.2302	.9025	.3428
.9341	.5199	.7125	.5954	.1605	.6037
.1782	.6358	.2108	.5423	.3567	.2569
.3473	.7472	.3575	.4208	.3070	.0546
.5644	.8954	.2926	.6975	.5513	.0305

To demonstrate the application of the procedure, consider $R_1 = .0589$ and $R_2 = .6733$. Then

$$x = -4 + 10R_1 = -4 + 10 \times .0589 = -3.411$$

$$y = -3 + 10R_2 = -3 + 10 \times .6733 = 3.733$$

This point falls inside the circle because

$$(-3.411 - 1)^2 + (3.733 - 2)^2 = 22.46 < 25$$

The procedure is repeated n times, keeping track of the number of points m that fall within the circle. The estimate of the area is then computed as $100\frac{m}{n}$.

Remarks. To increase the reliability of estimating the area of the circle, we use the same procedures employed in ordinary statistical experiments:

1. Increase the sample size.
2. Use replications.

The discussion in Example 16.1-1 poses two questions regarding the simulation experiment:

1. How large should the sample size, n, be?
2. How many replications, N, are needed?

There are some formulas in statistical theory for determining n and N, and they depend on the nature of the simulation experiment as well as the desired confidence level. However, as in any statistical experiment, the golden rule is that higher values of n and N mean more reliable simulation results. In the end, the sample size will depend on the cost associated with conducting the simulation experiment. Generally speaking, however, a selected sample size is considered "adequate" if it produces a relatively "small" standard deviation.

Because of the random variation in the output of the experiment, it is necessary to express the results as a confidence interval. Letting $\overline{A}$ and s be the mean and variance of N replications, then, for a confidence level α, the confidence interval for the true area A is

$$\bar{A} - \frac{s}{\sqrt{N}} t_{\frac{\alpha}{2},N-1} \le A \le \bar{A} + \frac{s}{\sqrt{N}} t_{\frac{\alpha}{2},N-1}$$

The parameter $t_{\frac{\alpha}{2},N-1}$ is determined from the t-distribution tables given a confidence level α and $N - 1$ degrees of freedom (see the t-table in Appendix B or use excelStatTables.xls). Note that N equals the number of replications, which is distinct from n, the sample size.

Excel Moment

Because the computations associated with each sample in Example 16.1-1 are voluminous, Excel template excelCircle.xls (with VBA macros) is used to test the effect of sample size and number of replications on the accuracy of the area estimate. The input data include the circle radius, r, and its center, (cx, cy), sample size, n, and number of replications, N. The entry *Steps* in cell D4 allows executing several sample sizes in the same run. For example, if $n = 30{,}000$ and Steps $= 3$, the template will automatically produce output for $n = 30{,}000, 60{,}000, 90{,}000$. Each time the command button Press to Execute Monte Carlo is pressed, new estimates are realized, because Excel refreshes the random number generator to a different sequence.

Figure 16.2 summarizes the results for 5 replications and sample sizes of 30,000, 60,000, and 90,000. The exact area is 78.54 cm^2, and the Monte Carlo results show that

FIGURE 16.2

Excel output of Monte Carlo estimation of the area of a circle (file excelCircle.xls)

	B	C	D	E
1	**Monte Carlo Estimation of the Area of a Circle**			
2		**Input data**		
3	Nbr. Replications, N =	5		
4	Sample size, n =	30,000	Steps =	3
5	Radius, r =	5		
6	Center, cx =	1		
7	Center, cy =	2		
8		**Output results**		
9	Exact area =	78.540		
10	Press to Execute Monte Carlo			
11	Monte Carlo Calculations:			
12		n=30000	n=60000	n=90000
13	Replication 1	78.207	78.555	78.483
14	Replication 2	78.673	78.752	78.581
15	Replication 3	78.300	78.288	78.281
16	Replication 4	78.503	78.347	78.343
17	Replication 5	78.983	78.775	78.760
18				
19	Mean =	78.533	78.543	78.490
20	Std. Deviation =	0.308	0.225	0.191
21				
22	95% lower conf. limit =	78.151	78.263	78.253
23	95% upper conf. limit =	78.915	78.823	78.727

the mean estimated area for the three sample sizes varies from $\overline{A} = 78.533$ to $\overline{A} = 78.490 \text{ cm}^2$. We note also that the standard deviation decreases from $s = .308$ for $n = 30{,}000$ to $s = .191$ for $n = 90{,}000$, an indication that accuracy increases with the increase in the sample size.

In terms of the present experiment, we are interested in establishing the confidence interval based on the largest sample size (i.e., $n = 90{,}000$). Given $N = 5$, $\overline{A} = 78.490 \text{ cm}^2$, and $s = .191 \text{ cm}^2$, $t_{.025,4} = 2.776$, and the resulting 95% confidence interval is $78.25 \le A \le 78.73$. In general, the value of N should be at least 5 to realize reasonable accuracy in the estimation of the confidence interval.

PROBLEM SET 16.1A

1. In Example 16.2-1, estimate the area of the circle using the first two columns of the (0, 1) random numbers in Table 16.1. (For convenience, go down each column, selecting R_1 first and then R_2.) How does this estimate compare with the ones given in Figure 16.2?
2. Suppose that the equation of a circle is

$$(x - 3)^2 + (y + 2)^2 = 16$$

   (a) Define the corresponding distributions $f(x)$ and $f(y)$, and then show how a sample point (x, y) is determined using the (0, 1) random pair (R_1, R_2).

   (b) Use excelCircle.xls to estimate the area and the associated 95% confidence interval given $n = 100{,}000$ and $N = 10$.

3. Use Monte Carlo sampling to estimate the area of the lake shown in Figure 16.3. Base the estimate on the first two columns of (0, 1) random numbers in Table 16.1.
4. Consider the game in which two players, Jan and Jim, take turns in tossing a fair coin. If the outcome is heads, Jim gets $10 from Jim. Otherwise, Jan gets $10 from Jan.

   ***(a)** How is the game simulated as a Monte Carlo experiment?

   (b) Run the experiment for 5 replications of 10 tosses each. Use the first five columns of the (0, 1) random numbers in Table 16.1, with each column corresponding to one replication.

   (c) Establish a 95% confidence interval on Jan's winnings.

   (d) Compare the confidence interval in (c) with Jan's expected theoretical winnings.

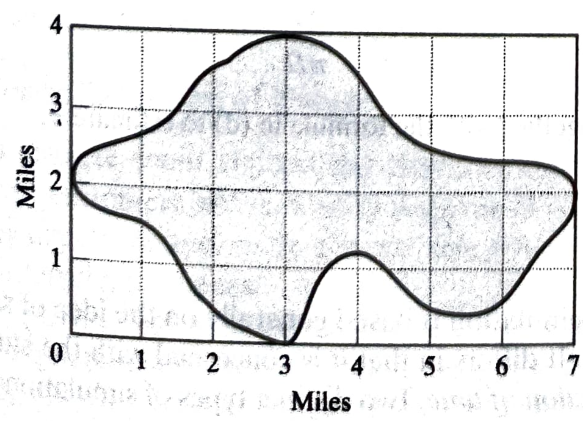


FIGURE 16.3
Lake map for Problem 3, Set 16.1a

5. Consider the following definite integral:

$$\int_0^1 x^2\,dx$$

(a) Develop the Monte Carlo experiment to estimate the integral.
(b) Use the first four columns in Table 16.1 to evaluate the integral based on 4 replications of size 5 each. Compute a 95% confidence interval, and compare it with the exact value of the integral.

6. Simulate five wins or losses of the following game of craps: The player rolls two fair dice. If the outcome sum is 7 or 11, the player wins \$10. Otherwise, the player records the resulting sum (called *point*) and keeps on rolling the dice until the outcome sum matches the recorded *point*, in which case the player wins \$10. If a 7 is obtained prior to matching the *point*, the player loses \$10.

*7. The lead time for receiving an order can be 1 or 2 days, with equal probabilities. The demand *per day* assumes the values 0, 1, and 2 with the respective probabilities of .2, .7, and .1. Use the random numbers in Table 16.1 (starting with column 1) to estimate the joint distribution of the demand and lead time. From the joint distribution, estimate the pdf of demand during lead time. (*Hint*: The demand during lead time assumes discrete values from 0 to 4.)

8. Consider the Buffon needle experiment. A horizontal plane is ruled with parallel lines spaced D cm apart. A needle of length d cm ($d < D$) is dropped randomly on the plane. The objective of the experiment is to determine the probability that either end of the needle touches or crosses one of the lines. Define

h = Perpendicular distance from the needle center to a (parallel) line

θ = Inclination angle of the needle with a line

(a) Show that the needle will touch or cross a line only if

$$h \le \frac{d}{2}\sin\theta,\ 0 \le h \le \frac{D}{2},\ 0 \le \theta \le \pi$$

(b) Design the Monte Carlo experiment, and provide an estimate of the desired probability.
(c) Use Excel to obtain 4 replications of size 10 each of the desired probability. Determine a 95% confidence interval for the estimate. Assume $D = 20$ cm and $d = 10$ cm.
(d) Prove that the theoretical probability is given by the formula

$$p = \frac{2d}{\pi D}$$

(e) Use the result in (c) together with the formula in (d) to estimate π.

16.2 TYPES OF SIMULATION

The execution of present-day simulation is based generally on the idea of sampling used with the Monte Carlo method. It differs in that it is concerned with the study of the behavior of real systems *as a function of time*. Two distinct types of simulation models exist.

1. Continuous models deal with systems whose behavior changes *continuously* with time. These models usually use difference-differential equations to describe the interactions among the different elements of the system. A typical example deals with the study of world population dynamics.

2. Discrete models deal primarily with the study of waiting lines, with the objective of determining such measures as the average waiting time and the length of the queue. These measures change only when a customer enters or leaves the system. The instants at which changes take place occur at specific discrete points in time (arrivals and departure events), giving rise to the name **discrete event simulation.**

This chapter presents the basics of discrete event simulation, including a description of the components of a simulation model, collection of simulation statistics, and the statistical aspect of the simulation experiment. The chapter also emphasizes the role of the computer and simulation languages in the execution of simulation models.

PROBLEM SET 16.2A

1. Categorize the following situations as either discrete or continuous (or a combination of both). In each case, specify the objective of developing the simulation model.
   - ***(a)** Orders for an item arrive randomly at a warehouse. An order that cannot be filled immediately from available stock must await the arrival of new shipments.
   - **(b)** World population is affected by the availability of natural resources, food production, environmental conditions, educational level, health care, and capital investments.
   - **(c)** Goods arrive on pallets at a receiving bay of an automated warehouse. The pallets are loaded on a lower conveyor belt and lifted through an up-elevator to an upper conveyor that moves the pallets to corridors. The corridors are served by cranes that pick up the pallets from the conveyor and place them in storage bins.
2. Explain why you would agree or disagree with the following statement: "Most discrete event simulation models can be viewed in some form or another as queuing systems consisting of *sources* from which customers are generated, *queues* where customers may wait, and *facilities* where customers are served."

16.3 ELEMENTS OF DISCRETE-EVENT SIMULATION

This section introduces the concept of events in simulation and shows how the statistics of the simulated system are collected.

16.3.1 Generic Definition of Events

All discrete-event simulations describe, directly or indirectly, queuing situations in which customers arrive, wait in a queue if necessary, and then receive service before they depart the system. In general, any discrete-event model is composed of a network of interrelated queues.

Given that a discrete-event model is in reality a composite of queues, collection of simulation statistics (e.g., queue length and status of the service facility) take place only when a customer arrives or leaves the facility. This means that two principal events

control the simulation model: arrivals and departures. These are the only two instants at which we need to examine the system. At all other instants, no changes affecting the statistics of the system take place.

Example 16.3-1

Metalco Jobshop receives two types of jobs: regular and rush. All jobs are processed on two consecutive machines with ample buffer areas. Rush jobs always assume nonpreemptive priority over regular jobs. Identify the events of the situation.

This situation consists of two tandem queues corresponding to the two machines. At first, one may be inclined to identify the events of the situation as follows:

$A11$: A regular job arrives at machine 1.
$A21$: A rush job arrives at machine 1.
$D11$: A regular job departs machine 1.
$D21$: A rush job departs machine 1.
$A12$: A regular job arrives at machine 2.
$A22$: A rush job arrives at machine 2.
$D12$: A regular job departs machine 2.
$D22$: A rush job departs machine 2.

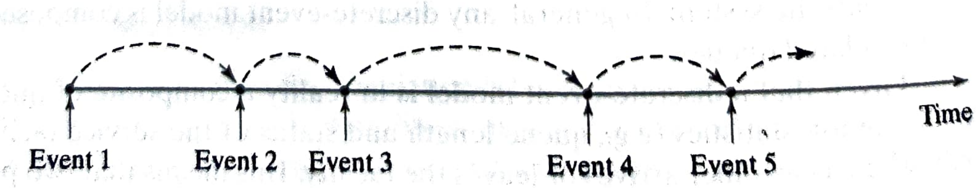
In reality, we have only two events: an arrival of a (new) job at the shop and a departure of a (completed) job from a machine. First notice that events $D11$ and $A12$ are actually one and the same. The same applies to $D21$ and $A22$. Next, in discrete simulation we can use one event (arrival or departure) for both types of jobs and simply "tag" the event with an **attribute** that identifies the job type as either regular or rush. (We can think of the attribute in this case as a *personal identification number*, and, indeed, it is.) Given this reasoning, the events of the model reduce to (1) an arrival A (at the shop) and (2) a departure D (from a machine). The actions associated with the departure event will depend on the machine at which they occur.

Having defined the basic events of a simulation model, we show how the model is executed. Figure 16.4 gives a schematic representation of typical occurrences of events on the simulation time scale. After all the actions associated with a current event have been performed, the simulation advances by "jumping" to the next chronological event. In essence, the execution of the simulation occurs at the instants at which the events occur.

How does the simulation determine the occurrence time of the events? The arrival events are separated by the interarrival time (the interval between successive arrivals), and the departure events are a function of the service time in the facility. These times may be deterministic (e.g., a train arriving at a station every 5 minutes) or probabilistic (e.g., the random arrival of customers at a bank). If the time between events is deterministic, the determination of their occurrence

FIGURE 16.4

Example of the occurrence of simulation events on the time scale



times is straightforward. If it is probabilistic, we use a special procedure to sample from the corresponding probability distribution. This point is discussed in the next section.

PROBLEM SET 16.3A

1. Identify the discrete events needed to simulate the following situation: Two types of jobs arrive from two different sources. Both types are processed on a single machine, with priority given to jobs from the first source.
2. Jobs arrive at a constant rate at a carousel conveyor system. Three service stations are spaced equally around the carousel. If the server is idle when a job arrives at the station, the job is removed from the conveyor for processing. Otherwise, the job continues to rotate on the carousel until a server becomes available. A processed job is stored in an adjacent shipping area. Identify the discrete events needed to simulate this situation.
3. Cars arrive at a two-lane, drive-in bank, where each lane can house a maximum of four cars. If the two lanes are full, arriving cars seek service elsewhere. If at any time one lane is at least two cars longer than the other, the last car in the longer lane will jockey to the last position in the shorter lane. The bank operates the drive-in facility from 8:00 A.M. to 3:00 P.M. each work day. Define the discrete events for the situation.

*4. The cafeteria at Elmdale Elementary provides a single-tray, fixed-menu lunch to all its pupils. Kids arrive at the dispensing window every 30 seconds. It takes 18 seconds to receive the lunch tray. Map the arrival-departure events on the time scale for the first five pupils.

16.3.2 Sampling from Probability Distributions

Randomness in simulation arises when the interval, t, between successive events is probabilistic. This section presents three methods for generating successive random samples ($t = t_1, t_2, \ldots$) from a probability distribution $f(t)$:

1. Inverse method.
2. Convolution method.
3. Acceptance-rejection method.

The inverse method is particularly suited for analytically tractable probability density functions, such as the exponential and the uniform. The remaining two methods deal with more complex cases, such as the normal and the Poisson. All three methods are rooted in the use of independent and identically distributed uniform (0, 1) random numbers.

Inverse Method. Suppose that it is desired to obtain a random sample x from the (continuous or discrete) probability density function $f(x)$. The inverse method first determines a closed-form expression of the cumulative density function $F(x) = P\{y \leq x\}$, where $0 \leq F(x) \leq 1$, for all defined values of y. Given that R is a random value obtained from a uniform (0, 1) distribution, and assuming that F^{-1} is the inverse of F, the steps of the method are as follows:

Step 1. Generate the (0, 1) random number, R.
Step 2. Compute the desired sample, $x = F^{-1}(R)$.

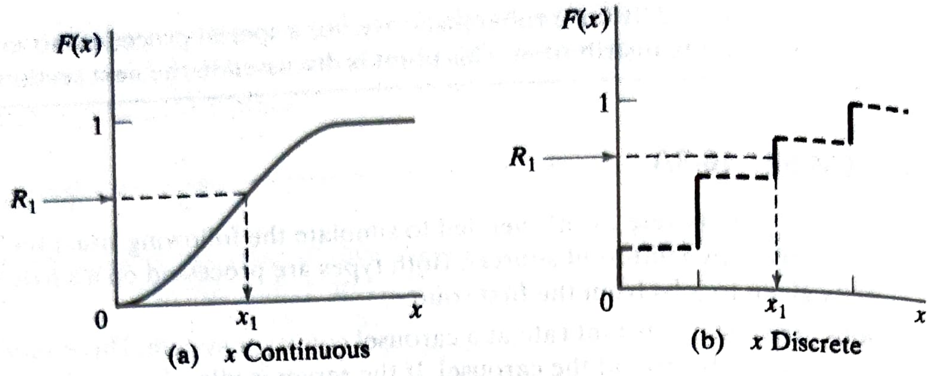


FIGURE 16.5
Sampling from a probability distribution by the inverse method

Figure 16.5 illustrates the procedures for both a continuous and a discrete random distribution. The uniform (0, 1) random value R_1 is projected from the vertical $F(x)$-scale to yield the desired sample value x_1 on the horizontal scale.

The validity of the proposed procedure rests on showing that the random variable $z = F(x)$ is uniformly distributed in the interval $0 \le z \le 1$, as the following theorem proves.

Theorem 16.3-1. *Given the cumulative density function $F(x)$ of the random variable x, $-\infty < x < \infty$, the random variable $z = F(x)$, $0 \le z \le 1$, has the following uniform 0-1 density function:*

$$f(z) = 1, 0 \le z \le 1$$

Proof. The random variable is uniformly distributed if, and only if,

$$P\{z \le Z\} = Z, 0 \le Z \le 1$$

This result applies to $F(x)$ because

$$P\{z \le Z\} = P\{F(x) \le Z\} = P\{x \le F^{-1}(Z)\} = F[F^{-1}(Z)] = Z$$

Additionally, $0 \le Z \le 1$ because $0 \le P\{z \le Z\} \le 1$.

Example 16.3-2 (Exponential Distribution)

The exponential probability density function

$$f(t) = \lambda e^{-\lambda t}, t > 0$$

represents the interarrival time t of customers at a facility with a mean value of $\frac{1}{\lambda}$. Determine a random sample t from $f(t)$.

The cumulative density function is determined as

$$F(t) = \int_0^t \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda t}, t > 0$$

Setting $R = F(t)$, we can solve for t, which yields

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

Because $1 - R$ is the complement of R, $\ln(1 - R)$ may be replaced with $\ln(R)$.

In terms of simulation, the result means that arrivals are spaced t time units apart. For example, for $\lambda = 4$ customers per hour and $R = .9$, the time period until the next arrival occurs is computed as

$$t_1 = -\left(\frac{1}{4}\right)\ln(1 - .9) = .577 \text{ hour} = 34.5 \text{ minutes}$$

The values of R used to obtain successive samples must be selected *randomly* from a uniform (0, 1) distribution. We will show later in Section 16.4 how these (0, 1) random values are generated during the course of the simulation.

PROBLEM SET 16.3B

*1. In Example 16.3-2, suppose that the first customer arrives at time 0. Use the first three random numbers in column 1 of Table 16.1 to generate the arrival times of the next 3 customers and graph the resulting events on the time scale.

*2. *Uniform Distribution.* Suppose that the time needed to manufacture a part on a machine is described by the following uniform distribution:

$$f(t) = \frac{1}{b - a}, a \leq t \leq b$$

Determine an expression for the sample t given the random number R.

3. Jobs are received randomly at a one-machine shop. The time between arrivals is exponential with mean 2 hours. The time needed to manufacture a job is uniform between 1.1 and 2 hours. Assuming that the first job arrives at time 0, determine the arrival and departure time for the first five jobs using the (0, 1) random numbers in column 1 of Table 16.1.

4. The demand for an expensive spare part of a passenger jet is 0, 1, 2, or 3 units per month with probabilities .2, .3, .4, and .1, respectively. The airline maintenance shop starts operation with a stock of 5 units, and will bring the stock level back to 5 units immediately after it drops below 2 units.

   *(a) Devise the procedure for sampling demand.

   (b) How many months will elapse until the first replenishment occurs? Use successive values of R from the first column in Table 16.1.

5. In a simulation situation, TV units are inspected for possible defects. There is an 80% chance that a unit will pass inspection, in which case it is sent to packaging. Otherwise, the unit is repaired. We can represent the situation symbolically in one of two ways.

   goto REPAIR/.2, PACKAGE/.8
   goto PACKAGE/.8, REPAIR/.2

   These two representations appear equivalent. Yet, when a given sequence of (0, 1) random numbers is applied to the two representations, different decisions (REPAIR or PACKAGE) may result. Explain why.

6. A player tosses a fair coin repeatedly until a head occurs. The associated payoff is 2^n, where n is the number of tosses until a head comes up.
   (a) Devise the sampling procedure of the game.
   (b) Use the random numbers in column 1 of Table 16.1 to determine the cumulative payoff after two heads occur.
7. *Triangular Distribution.* In simulation, the lack of data may make it impossible to determine the probability distribution associated with a simulation activity. In most of these situations, it may be easy to describe the desired variable by estimating its smallest, most likely, and largest values. These three values are sufficient to define a triangular distribution, which can then be used as a "rough cut" estimation of the real distribution.
   (a) Develop the formula for sampling from the following triangular distribution, whose respective parameters are a, b, and c:

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)}, & a \le x \le b \\ \dfrac{2(c-x)}{(c-b)(c-a)}, & b \le x \le c \end{cases}$$

   (b) Generate three samples from a triangular distribution with parameters (1,3,7) using the first three random numbers in column 1 of Table 16.1.
8. Consider a probability distribution that consists of a rectangle flanked on the left and right sides by two symmetrical right triangles. The respective ranges for the triangle on the left, the rectangle, and the triangle on the right are $[a, b]$, $[b, c]$, and $[c, d]$, $a < b < c < d$. Both triangles have the same height as the rectangle.
   (a) Develop a sampling procedure
   (b) Determine five samples with $(a, b, c, d) = (1, 2, 4, 6)$ using the first five random numbers in column 1 of Table 16.1.

*9. *Geometric distribution*. Show how a random sample can be obtained from the following geometric distribution:

$$f(x) = p(1-p)^x, \quad x = 0, 1, 2, \ldots$$

The parameter x is the number of (Bernoulli) failures until a success occurs, and p is the probability of a success, $0 < p < 1$. Generate five samples for $p = .6$ using the first five random numbers in column 1 of Table 16.1.

10. *Weibull distribution.* Show how a random sample can be obtained from the Weibull distribution with the following probability density function:

$$f(x) = \alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^{\alpha}}, x > 0$$

where $\alpha > 0$ is the shape parameter, and $\beta > 0$ is the scale parameter.

Convolution Method. The basic idea of the convolution method is to express the desired sample as the statistical sum of other easy-to-sample random variables. Typical among these distributions are the Erlang and the Poisson whose samples can be obtained from the exponential distribution samples.

Example 16.3-3 (Erlang Distribution)

The m-Erlang random variable is defined as the statistical sum (convolutions) of m independent and identically distributed exponential random variables. Let y represent the m-Erlang random variable; then

$$y = y_1 + y_2 + \cdots + y_m$$

where y_i, $i = 1, 2, \ldots, m$, are independent and identically distributed exponential random variables whose probability density function is defined as

$$f(y_i) = \lambda e^{-\lambda y_i}, \; y_i > 0, i = 1, 2, \ldots, m$$

From Example 16.3-2, a sample from the ith exponential distribution is computed as

$$y_i = -\left(\frac{1}{\lambda}\right) \ln(R_i), i = 1, 2, \ldots, m$$

Thus, the m-Erlang sample is computed as

$$\begin{aligned} y &= -\left(\frac{1}{\lambda}\right)\{\ln(R_1) + \ln(R_2) + \cdots + \ln(R_m)\} \\ &= -\left(\frac{1}{\lambda}\right) \ln\left(\prod_{i=1}^{m} R_i\right) \end{aligned}$$

To illustrate the use of the formula, suppose that $m = 3$, and $\lambda = 4$ events per hour. The first 3 random numbers in column 1 of Table 16.1 yield $R_1R_2R_3 = (.0589)(.6733)(.4799) = .0190$, which yields

$$y = -\left(\tfrac{1}{4}\right) \ln(.019) = .991 \text{ hour}$$

Example 16.3-4 (Poisson Distributions)

Section 15.3.1 shows that if the distribution of the time between the occurrence of successive events is exponential, then the distribution of the number of events per unit time must be Poisson, and vice versa. We use this relationship to sample the Poisson distribution.

Assume that the Poisson distribution has a mean value of λ events per unit time. Then the time between events is exponential with mean $\frac{1}{\lambda}$ time units. This means that a Poisson sample, n, will occur during t time units if, and only if,

$$\text{Period till event } n \text{ occurs} \leq t < \text{Period till event } n + 1 \text{ occurs}$$

This condition translates to

$$\begin{aligned} t_1 + t_2 + \cdots + t_n \leq t < t_1 + t_2 + \cdots + t_{n+1}, \; n > 0 \\ 0 \leq t < t_1, \; n = 0 \end{aligned}$$

where t_i, $i = 1, 2 \ldots, n + 1$, is a sample from the exponential distribution with mean $\frac{1}{\lambda}$. From the result in Example 16.3-3, we have

$$-\left(\frac{1}{\lambda}\right)\ln\left(\prod_{i=1}^{n} R_i\right) \leq t < -\left(\frac{1}{\lambda}\right)\ln\left(\prod_{i=1}^{n+1} R_i\right),\ n > 0$$

$$0 \leq t < -\left(\frac{1}{\lambda}\right)\ln(R_1),\ n = 0$$

which reduces to

$$\prod_{i=1}^{n} R_i \geq e^{-\lambda t} > \prod_{i=1}^{n+1} R_i,\ n > 0$$

$$1 \geq e^{-\lambda t} > R_1,\ n = 0$$

To illustrate the implementation of the sampling process, suppose that $\lambda = 4$ events per hour and that we wish to obtain a sample for a period $t = .5$ hour. This gives $e^{-\lambda t} = .1353$. Using the random numbers in column 1 of Table 16.1, we note that $R_1 = .0589$ is less than $e^{-\lambda t} = .1353$. Hence, the corresponding sample is $n = 0$.

Example 16.3-5 (Normal Distribution)

The central limit theorem (see Section 12.4.4) states that the sum (convolution) of n independent and identically distributed random variables becomes asymptotically normal as n becomes sufficiently large. We use this result to generate samples from normal distribution with mean μ and standard deviation σ.

Define

$$x = R_1 + R_2 + \cdots + R_n$$

The random variable is asymptotically normal by the central limit theorem. Given that the uniform (0,1) random number R has a mean of $\frac{1}{2}$ and a variance of $\frac{1}{12}$, it follows that x has a mean of $\frac{n}{2}$ and a variance of $\frac{n}{12}$. Thus, a random sample, y, from a normal distribution with mean μ and standard deviation σ, $N(\mu, \sigma)$, can be computed from x as

$$y = \mu + \sigma\left(\frac{x - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right)$$

In practice, we take $n = 12$ for convenience, which reduces the formula to

$$y = \mu + \sigma(x - 6)$$

To illustrate the use of this method, suppose that we wish to generate a sample from $N(10, 2)$ (mean $\mu = 10$ and standard deviation $\sigma = 2$). Taking the sum of the first 12 random numbers in columns 1 and 2 of Table 16.1, we get $x = 6.1094$. Thus, $y = 10 + 2(6.1094 - 6) = 10.2188$.

The disadvantage of this procedure is that it requires generating 12 random numbers for each normal sample, which is computationally inefficient. A more efficient procedure calls for using the transformation

$$x = \cos(2\pi R_2)\sqrt{-2\ln(R_1)}$$

Box and Muller (1958) prove that x is a standard $N(0, 1)$. Thus, $y = \mu + \sigma x$ will produce a sample from $N(\mu, \sigma)$. The new procedure is more efficient because it requires two (0, 1) random numbers only. Actually, this method is even more efficient than stated, because Box and Miller prove that the preceding formula will produce another $N(0, 1)$ sample if $\sin(2\pi R_2)$ replaces $\cos(2\pi R_2)$.

To illustrate the implementation of the Box-Muller procedure to the normal distribution $N(10, 2)$, the first two random numbers in column 1 of Table 16.1 yield the following $N(0, 1)$ samples:

$$x_1 = \cos(2\pi \times .6733)\sqrt{-2\ln(.0589)} \approx -1.103$$

$$x_2 = \sin(2\pi \times .6733)\sqrt{-2\ln(.0589)} \approx -2.109$$

Thus, the corresponding $N(10, 2)$ samples are

$$y_1 = 10 + 2(-1.103) = 7.794$$

$$y_2 = 10 + 2(-2.109) = 5.782$$

PROBLEM SET 16.3C[1]

*1. In Example 16.3-3, compute an Erlang sample, given $m = 4$ and $\lambda = 5$ events per hour.

2. In Example 16.3-4, generate three Poisson samples during a 2-hour period, given that the mean of the Poisson is 5 events per hour.

3. In Example 16.4-5, generate two samples from $N(8, 1)$ by using both the convolution method and the Box-Muller method.

4. Jobs arrive at Metalco jobshop according to a Poisson distribution, with a mean of six jobs per day. Received jobs are assigned to the five machining centers of the shop on a strict rotational basis. Determine one sample of the interval between the arrival of jobs at the first machine center.

5. The ACT scores for the 1994 senior class at Springdale High are normal, with a mean of 27 points and a standard deviation of 3 points. Suppose that we draw a random sample of six seniors from that class. Use the Box-Muller method to determine the mean and standard deviation of the sample.

*6. Psychology professor Yataha is conducting a learning experiment in which mice are trained to find their way around a maze. The base of the maze is square. A mouse enters the maze at one of the four corners and must find its way through the maze to exit at the same point where it entered. The design of the maze is such that the mouse must pass by each of the remaining three corner points exactly once before it exits. The multi-paths of the maze connect the four corners in a strict clockwise order. Professor Yataha estimates that the time the mouse takes to reach one corner point from another is uniformly

[1]For all the problems of this set, use the random numbers in Table 16.1 starting with column 1.

distributed between 10 and 20 seconds, depending on the path it takes. Develop a sampling procedure for the time a mouse spends in the maze.

7. In Problem 6, suppose that once a mouse makes an exit from the maze, another mouse instantly enters. Develop a sampling procedure for the number of mice that exit the maze in 5 minutes.

8. *Negative Binomial.* Show how a random sample can be determined from the negative binomial whose distribution is given as

$$f(x) = C_x^{r+x-1} p^r (1-p)^x, x = 0, 1, 2, K$$

where x is the number of failures until the rth success occurs in a sequence of independent Bernoulli trials and p is the probability of success, $0 < p < 1$. (*Hint:* The negative binomial is the convolution of r independent geometric samples. See Problem 9, Set 16.3b.)

Acceptance-Rejection Method. The acceptance-rejection method is designed for complex pdfs that cannot be handled by the preceding methods. The general idea of the method is to replace the complex pdf $f(x)$ with a more analytically manageable "proxy" pdf $h(x)$. Samples from $h(x)$ can then be used to sample the original pdf $f(x)$.

Define the **majorizing function** $g(x)$ such that it dominates $f(x)$ in its entire range—that is,

$$g(x) \geq f(x), -\infty < x < \infty$$

Next, define the proxy pdf, $h(x)$, by normalizing $g(x)$ as

$$h(x) = \frac{g(x)}{\int_{-\infty}^{\infty} g(y)d(y)}, -\infty < x < \infty$$

The steps of the acceptance-rejection method are thus given as

Step 1. Obtain a sample $x = x_1$ from $h(x)$ using the inverse or the convolution method.

Step 2. Obtain a (0, 1) random number R.

Step 3. If $R \leq \frac{f(x_1)}{g(x_1)}$, accept x_1 as a sample from $f(x)$. Otherwise, discard x_1 and return to step 1.

The validity of the method is based on the following equality:

$$P\{x \leq a | x = x_1 \text{ is accepted}, -\infty < x_1 < \infty\} = \int_{-\infty}^{a} f(y)\, dy, -\infty < a < \infty$$

This probability statement states that the sample $x = x_1$ that satisfies the condition of step 3 in reality is a sample from the original pdf $f(x)$, as desired.

The efficiency of the proposed method is enhanced by the decrease in the rejection probability of step 3. This probability depends on the specific choice of the majorizing function $g(x)$ and should decrease with the selection of a $g(x)$ that "majorizes" $f(x)$ more "snugly."

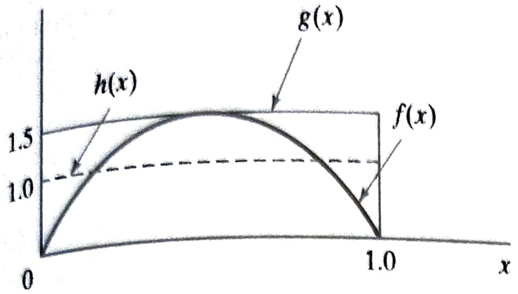


FIGURE 16.6
Majorizing function, $g(x)$, for the beta distribution, $f(x)$

Example 16.3-6 (Beta Distribution)

Apply the acceptance-rejection to the following beta distribution:

$$f(x) = 6x(1 - x), 0 \le x \le 1$$

Figure 16.6 depicts $f(x)$ and a majorizing function $g(x)$.

The height of the majorizing function $g(x)$ equals the maximum of $f(x)$, which occurs at $x = .5$. Thus, the height of the rectangle is $f(.5) = 1.5$. This means that

$$g(x) = 1.5, 0 \le x \le 1$$

The proxy pdf $h(x)$, also shown in Figure 16.6, is computed as

$$h(x) = \frac{g(x)}{\text{Area under } g(x)} = \frac{1.5}{1 \times 1.5} = 1, 0 \le x \le 1$$

The following steps demonstrate the procedure using the $(0, 1)$ random sequence in Table 16.1.

Step 1. $R = .0589$ gives the sample $x = .0589$ from $h(x)$.

Step 2. $R = .6733$.

Step 3. Because $\frac{f(.0589)}{g(.0589)} = \frac{.3326}{1.5} = .2217$ is less than $R = .6733$, we accept the sample $x_1 = .0589$.

To obtain a second sample, we continue as follows:

Step 1. Using $R = .4799$, we get $x = .4799$ from $h(x)$.

Step 2. $R = .9486$.

Step 3. Because $\frac{f(.4799)}{g(.4799)} = .9984$ is larger than $R = .9486$, we reject $x = .4799$ as a valid beta sample. This means that the steps must be repeated again with "fresh" random numbers until the condition of step 3 is satisfied.

Remarks. The efficiency of the acceptance-rejection method is enhanced by selecting a majorizing function $g(x)$ that "jackets" $f(x)$ as tightly as possible while yielding an analytically tractable proxy $h(x)$. For example, the method will be more efficient if the rectangular majorizing function $g(x)$ in Figure 16.6 is replaced with a step-pyramid function (see Problem 2, Set 16.3d, for an illustration). The larger the number of steps, the more tightly will $g(x)$ majorize $f(x)$, and hence the higher is the probability of accepting a sample. However, a "tight" majorizing function generally entails additional computations which, if excessive, may offset the savings resulting from increasing the probability of acceptance.

PROBLEM SET 16.3D

1. In Example 16.3-5, continue the steps of the procedure until a valid sample is obtained. Use the (0, 1) random numbers in Table 16.1 in the same order in which they are used in the example.
2. Consider the beta pdf of Example 16.3-6. Determine a two-step pyramid majorizing function $g(x)$ with two equal jumps of height $\frac{1.5}{2} = .75$ each. Obtain one beta sample based on the new majorizing function using the same (0, 1) random sequence in Table 16.1 that was employed in Example 16.3-6. The conclusion, in general, is that a tighter majorizing function will increase the probability of acceptance. Observe, however, that the amount of the computations associated with the new function is larger.
3. Determine the functions $g(x)$ and $h(x)$ for applying the acceptance-rejection method to the following function:

$$f(x) = \frac{\sin(x) + \cos(x)}{2}, 0 \le x \le \frac{\pi}{2}.$$

Use the (0, 1) random numbers from column 1 in Table 16.1 to generate two samples from $f(x)$. [*Hint*: For convenience, use a rectangular $g(x)$ over the defined range of $f(x)$.]

4. The interarrival time of customers at HairKare is described by the following distribution:

$$f_1(t) = \frac{k_1}{t}, 12 \le t \le 20$$

The time to get a haircut is represented by the following distribution:

$$f_2(t) = \frac{k_2}{t^2}, 18 \le t \le 22$$

The constant k_1 and k_2 are determined such that $f_1(t)$ and $f_2(t)$ are probability density functions. Use the acceptance-rejection method (and the random numbers in Table 16.1) to determine when the first customer will leave HairKare and when the next customer will arrive. Assume that the first customer arrives at $T = 0$.

16.4 GENERATION OF RANDOM NUMBERS

Uniform (0,1) random numbers play a key role in sampling from distributions. True (0, 1) random numbers can only be generated by electronic devices. However, because simulation models are executed on the computer, the use of electronic devices to generate random numbers is much too slow for that purpose. Additionally, electronic devices are activated by laws of chance, and hence it will be impossible to duplicate the same sequence of random numbers at will. This point is important because debugging, verification, and validation of the simulation model often require duplicating the same sequence of random numbers.

The only plausible way for generating (0, 1) random numbers for use in simulation is based on arithmetic operations. Such numbers are not truly random because they can be generated in advance. It is thus more appropriate to refer to them as **pseudo-random numbers.**

The most common arithmetic operation for generating (0, 1) random numbers is the **multiplicative congruential method**. Given the parameters u_0, b, c, and m, a pseudo-random number R_n can be generated from the formulas:

$$u_n = (bu_{n-1} + c) \bmod (m), n = 1, 2, \ldots$$

$$R_n = \frac{u_n}{m}, n = 1, 2, \ldots$$

The initial value u_0 is usually referred to as the **seed** of the generator.

Variations of the multiplicative congruential method that improve the quality of the generator can be found in Law and Kelton (1991).

Example 16.4-1

Generate three random numbers based on the multiplicative congruential method using $b = 9$, $c = 5$, and $m = 12$. The seed is $u_0 = 11$.

$$u_1 = (9 \times 11 + 5) \bmod 12 = 8, R_1 = \frac{8}{12} = .6667$$

$$u_2 = (9 \times 8 + 5) \bmod 12 = 5, R_2 = \frac{5}{12} = .4167$$

$$u_3 = (9 \times 5 + 5) \bmod 12 = 2, R_3 = \frac{2}{12} = .1667$$

Excel Moment

Excel template excelRN.xls is designed to carry out the multiplicative congruential calculations. Figure 16.7 generates the sequence associated with the parameters of Example 16.4-1. Observe carefully that the cycle length is exactly 4, after which the sequence repeats itself. The conclusion here is that the choice of u_0, b, c, and m is critical in determining the (statistical) quality of the generator and its cycle length. Thus, "casual" implementation of the congruential formula is not advisable. Instead, one must use a reliable and tested generator. Practically all commercial computer programs are equipped with dependable random number generators.

PROBLEM SET 16.4A

*1. Use excelRN.xls with the following sets of parameters and compare the results with those in Example 16.4-1:

$$b = 17, c = 111, m = 103, \text{seed} = 7$$

2. Find a random number generator on your computer, and use it to generate 500 zero-one random numbers. Histogram the resulting values (using the Microsoft histogram tool, see Section 12.5) and visually convince yourself that the obtained numbers reasonably follow

	A	B
1	Multiplicative Congruential Method	
2	Input data(B7<=1000)	
3	b =	9
4	c =	5
5	u0 =	11
6	m =	12
7	How many numbers?	10
8	Output results	
9	Press to Generate Sequence	
10	Generated random numbers:	
11	1	0.66667
12	2	0.41667
13	3	0.16667
14	4	0.91667
15	5	0.66667
16	6	0.41667
17	7	0.16667
18	8	0.91667
19	9	0.66667
20	10	0.41667

FIGURE 16.7

Excel random numbers output for the data of Example 16.4-1 (file excelRN.xls)

the (0, 1) uniform distribution. Actually, to test the sequence properly, you would need to apply the following tests: chi-square goodness of fit (see Section 12.6), runs test for independence, and correlation test (see Law and Kelton [1991] for details).

16.5 MECHANICS OF DISCRETE SIMULATION

This section details how typical statistics are collected in a simulation model. The vehicle of explanation is a single-queue model. Section 16.5.1 uses a numeric example to detail the actions and computations that take place in a single-server queuing simulation model. Because of the tedious computations that typify the execution of a simulation model, Section 16.5.2 shows how the single-server model is modeled and executed using Excel spreadsheet.

16.5.1 Manual Simulation of a Single-Server Model

The interarrival time of customers at HairKare Barbershop is exponential with mean 15 minutes. The shop is operated by only one barber and it takes between 10 and 15 minutes, uniformly distributed, to do a haircut. Customers are served on a first-in, first-out (FIFO) basis. The objective of the simulation is to compute the following measures of performance:

1. The average utilization of the shop.
2. The average number of waiting customers.
3. The average time a customer waits in queue.

The logic of the simulation model can be described in terms of the actions associated with the arrival and departure events of the model.

Arrival Event

1. Generate and store chronologically the occurrence time of the next arrival event (= current simulation time + interarrival time).
2. If the facility (barber) is idle
   - **(a)** Start service and declare the facility busy. Update the facility utilization statistics.
   - **(b)** Generate and store chronologically the time of the departure event for the customer (= current simulation time + service time).
3. If the facility is busy, place the customer in the queue and update the queue statistics.

Departure Event

1. If the queue is empty, declare the facility idle. Update the facility utilization statistics.
2. If the queue is not empty
   - **(a)** Select a customer from the queue, and place it in the facility. Update the queue and facility utilization statistics.
   - **(b)** Generate and store chronologically the occurrence time of the departure event for the customer (= current simulation time + service time).

From the data of the problem, the interarrival time is exponential with mean 15 minutes, and the service time is uniform between 10 and 15 minutes. Letting p and q represent random samples of interarrival and service times, then, as explained in Section 16.3.2, we get

$$p = -15 \ln(R) \text{ minutes}, \quad 0 \le R \le 1$$

$$q = 10 + 5R \text{ minutes}, \quad 0 \le R \le 1$$

For the purpose of this example, we use R from Table 16.1, starting with column 1. We also use the symbol T to represent the simulation clock time. We further assume that the first customer arrives at $T = 0$ and that the facility starts empty.

Because the simulation computations are typically voluminous, the simulation is limited to the first 5 arrivals only. The example is designed to cover all possible situations that could arise in the course of the simulation. Later in the section we introduce the excelSingleServer.xls template that allows you to experiment with the model without the need to carry out the computations manually.

Arrival of Customer 1 at $T = 0$. Generate the arrival of customer 2 at

$$T = 0 + p_1 = 0 + [-15 \ln(.0589)] = 42.48 \text{ minutes}$$

Because the facility is idle at $T = 0$, customer 1 starts service immediately. The departure time is thus computed as

$$T = 0 + q_1 = 0 + (10 + 5 \times .6733) = 13.37 \text{ minutes}$$

The *chronological* list of future events is thus given as:

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

Departure of Customer 1 at T = 13.37. Because the queue is empty, the facility is declared idle. At the same time, we record that the facility has been busy between $T = 0$ and $T = 13.37$ minutes. The updated list of future events becomes

Time, T	Event
42.48	Arrival of customer 2

Arrival of Customer 2 at T = 42.48. Customer 3 will arrive at

$$T = 42.48 + [-15 \ln(.4799)] = 53.49 \text{ minutes}$$

Because the facility is idle, customer 2 starts service and the facility is declared busy. The departure time is

$$T = 42.48 + (10 + 5 \times .9486) = 57.22 \text{ minutes}$$

The list of future events is updated as

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

Arrival of Customer 3 at T = 53.49. Customer 4 will arrive at

$$T = 53.49 + [-15 \ln(.6139)] = 60.81 \text{ minutes}$$

Because the facility is currently busy (until $T = 57.22$), customer 3 is placed in queue at $T = 53.49$. The updated list of future events is

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 to start service. The waiting time is

$$W_3 = 57.22 - 53.49 = 3.73 \text{ minutes}$$

The departure time is

$$T = 57.22 + (10 + 5 \times .5933) = 70.19 \text{ minutes}$$

The updated list of future events is

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

Arrival of Customer 4 at T = 60.81. Customer 5 will arrive at

$$T = 60.81 + [-15 \ln(.9341)] = 61.83 \text{ minutes}$$

Because the facility is busy until $T = 70.19$, customer 4 is placed in the queue. The updated list of future events is

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

Arrival of Customer 5 at T = 61.83. The simulation is limited to 5 arrivals only, hence customer 6 arrival is not generated. The facility is still busy, hence the customer is placed in queue at $T = 61.83$. The updated list of events is

Time, T	Event
70.19	Departure of customer 3

Departure of Customer 3 at T = 70.19. Customer 4 is taken out of the queue to start service. The waiting time is

$$W_4 = 70.19 - 60.81 = 9.38 \text{ minutes}$$

The departure time is

$$T = 70.19 + [10 + 5 \times .1782] = 81.08 \text{ minutes}$$

The updated list of future events is

Time, T	Event
81.08	Departure of customer 4

Departure of Customer 4 at T = 81.08. Customer 5 is taken out of the queue to start service. The waiting time is

$$W_5 = 81.08 - 61.83 = 19.25 \text{ minutes}$$

The departure time is

$$T = 81.08 + (10 + 5 \times .3473) = 92.82 \text{ minutes}$$

The updated list of future events is

Time, T	Event
92.82	Departure of customer 5

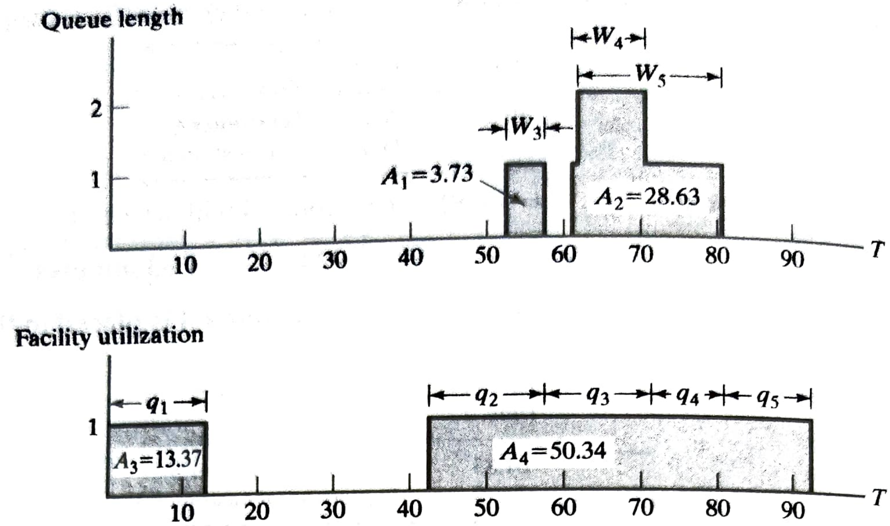


FIGURE 16.8
Changes in queue length and facility utilization as a function of simulation time, T

Departure of Customer 5 at T = 92.82. There are no more customers in the system (queue and facility) and the simulation ends.

Figure 16.8 summarizes the changes in the length of the queue and the utilization of the facility as a function of the simulation time.

The queue length and the facility utilization are known as **time-based** variables because their variation is a function of time. As result, their average values are computed as

$$\begin{pmatrix}\text{Average value of a}\\ \text{time-based variable}\end{pmatrix} = \frac{\text{Area under curve}}{\text{Simulated period}}$$

Implementing this formula for the data in Figure 16.8, we get

$$\begin{pmatrix}\text{Average queue}\\ \text{length}\end{pmatrix} = \frac{A_1 + A_2}{92.82} = \frac{32.36}{92.82} = .349 \text{ customer}$$

$$\begin{pmatrix}\text{Average facility}\\ \text{utilization}\end{pmatrix} = \frac{A_3 + A_4}{92.82} = \frac{63.71}{92.82} = .686 \text{ barber}$$

The average waiting time in the queue is an **observation-based** variable whose value is computed as

$$\begin{pmatrix}\text{Average value of an}\\ \text{observation-based variable}\end{pmatrix} = \frac{\text{Sum of observations}}{\text{Number of observations}}$$

Examination of Figure 16.8 reveals that the area under the queue-length curve actually equals the sum of the waiting time for the three customers who joined the queue; namely,

$$W_1 + W_2 + W_3 + W_4 + W_5 = 0 + 0 + 3.73 + 9.38 + 19.25 = 32.36 \text{ minutes}$$

The average waiting time in the queue for all customers is thus computed as

$$\overline{W}_q = \frac{32.36}{5} = 6.47 \text{ minutes}$$

PROBLEM SET 16.5A

1. Suppose that the barbershop of Section 16.5.1 is operated by two barbers, and customers are served on a FCFS basis. Suppose further that the time to get a haircut is uniformly distributed between 15 and 30 minutes. The interarrival time of customers is exponential, with a mean of 10 minutes. Simulate the system manually for 75 time units. From the results of the simulation, determine the average time a customer waits in queue, the average number of customers waiting, and the average utilization of the barbers. Use the random numbers in Table 16.1.
2. Classify the following variables as either *observation based* or *time based*:
   *(a) Time-to-failure of an electronic component.
   *(b) Inventory level of an item.
   (c) Order quantity of an inventory item.
   (d) Number of defective items in a lot.
   (e) Time needed to grade test papers.
   (f) Number of cars in the parking lot of a car-rental agency.

*3. The following table represents the variation in the number of waiting customers in a queue as a function of the simulation time.

Simulation time, T (hr)	No. of waiting customers
$0 \le T \le 3$	0
$3 < T \le 4$	1
$4 < T \le 6$	2
$6 < T \le 7$	1
$7 < T \le 10$	0
$10 < T \le 12$	2
$12 < T \le 18$	3
$18 < T \le 20$	2
$20 < T \le 25$	1

Compute the following measures of performance:

(a) The average length of the queue.

(b) The average waiting time in the queue for those who must wait.

CHAPTER 13

Decision Analysis and Games

Chapter Guide. Decision problems involving a finite number of alternatives arise frequently in practice. The tools used to solve these problems depend largely on the type of data available (deterministic, probabilistic, or uncertain). The analytic hierarchy process (AHP) is a prominent tool for dealing with decisions under certainty, where subjective judgment is quantified in a logical manner and then used as a basis for reaching a decision. For probabilistic data, decision trees comparing the expected cost (or profit) for the different alternatives are the basis for reaching a decision. Decisions under uncertainty use criteria reflecting the decision maker's attitude toward risk, ranging from optimism to pessimism. Another tool of decision under uncertainty is game theory, where two opponents with conflicting goals aim to achieve the best out of the worst conditions available to each. To demonstrate the importance of these tools in practice, four case analyses in Chapter 24 on the CD deal with using AHP to determine the layout of a CIM laboratory, using decision-tree analysis to determine booking limits in hotel reservations, applying Bayes probabilities to evaluate the results of a medical test, and using game theory to rank golfers in Ryder Cup matches. To assist you in understanding the details of the different tools, the chapter provides 4 spreadsheets. You will also find TORA useful in carrying out the graphical and algebraic solution of games. A basic knowledge of probability and statistics is needed for this chapter.

This chapter includes summaries of 4 real-life applications, 10 solved examples, 4 spreadsheets, 63 end-of-section problems, and 5 cases. The cases are in Appendix E on the CD. The AMPL/Excel/Solver/TORA programs are in folder ch13Files.

Real-Life Application—Layout Planning of a Computer Integrated Manufacturing (CIM) Facility

The engineering college in an academic institution wants to establish a CIM laboratory in a vacated building. The new lab will serve as a teaching and research facility and a center of technical excellence for industry. Recommendations are solicited from the faculty regarding a layout plan for the new laboratory, from which the ideal and absolute minimum square footage for each unit are compiled. The study uses both AHP (analytic hierarchy process) and goal programming to reach a satisfactory compromise solution that meets the needs for teaching, research, and service to industry. The details of the study are given in Case 9, Chapter 24 on the CD.

13.1 DECISION MAKING UNDER CERTAINTY—ANALYTIC HIERARCHY PROCESS (AHP)

The LP models presented in Chapters 2 through 9 are examples of decision making under certainty in which all the functions are well defined. AHP is designed for situations in which ideas, feelings, and emotions affecting the decision process are quantified to provide a numeric scale for prioritizing the alternatives.

Example 13.1-1 (Overall Idea of AHP)

Martin Hans, a bright high school senior, has received full academic scholarships from three institutions: U of A, U of B, and U of C. To select a university, Martin specifies two main criteria: location and academic reputation. Being the excellent student he is, he judges academic reputation to be five times as important as location, giving a weight of approximately 17% to location and 83% to reputation. He then uses a systematic analysis (which will be detailed later) to rank the three universities from the standpoint of location and reputation. The following table ranks the two criteria for the three universities:

	Percent weight estimates for		
Criterion	*U of A*	*U of B*	*U of C*
Location	12.9	27.7	59.4
Reputation	54.5	27.3	18.2

The structure of the decision problem is summarized in Figure 13.1. The problem involves a single hierarchy (level) with two criteria (location and reputation) and three decision alternatives (U of A, U of B, and U of C).

The ranking of each university is based on computing the following *composite* weights:

$$\text{U of A} = .17 \times .129 + .83 \times .545 = \mathbf{.4743}$$

$$\text{U of B} = .17 \times .277 + .83 \times .273 = .2737$$

$$\text{U of C} = .17 \times .594 + .83 \times .182 = .2520$$

Based on these calculations, U of A has the highest composite weight, and hence represents the best choice for Martin.

Remarks. The general structure of AHP may include several hierarchies of criteria. Suppose in Example 13.1-1 that Martin's twin sister, Jane, was also accepted with full scholarship to the three universities. Their parents stipulate that they both must attend the same university so they can share one car. Figure 13.2 summarizes the decision

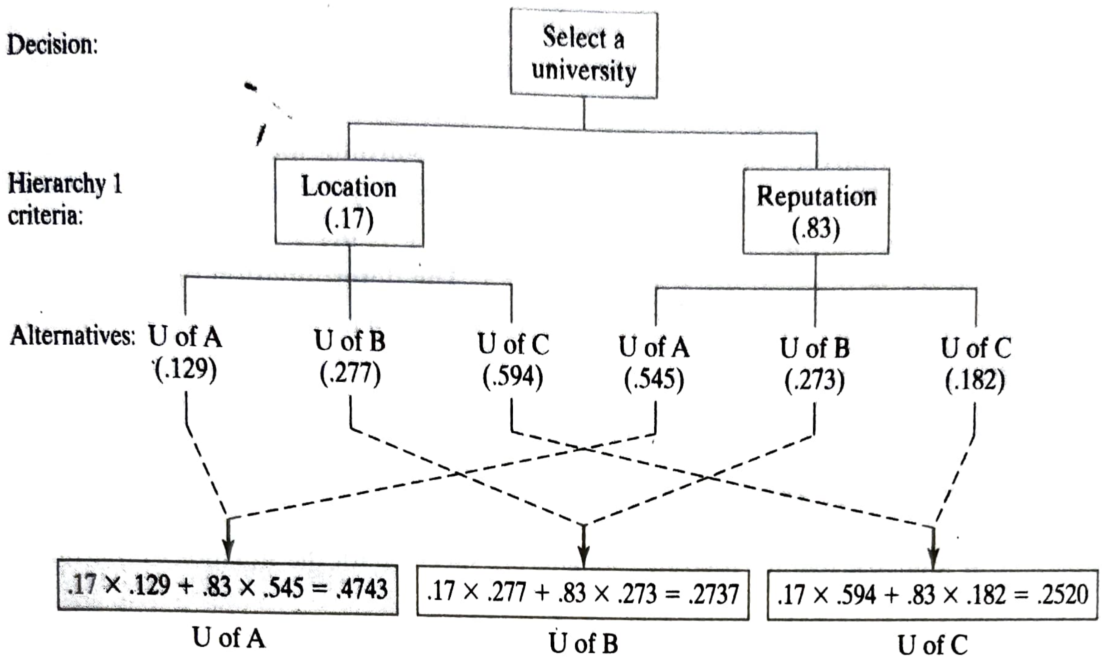
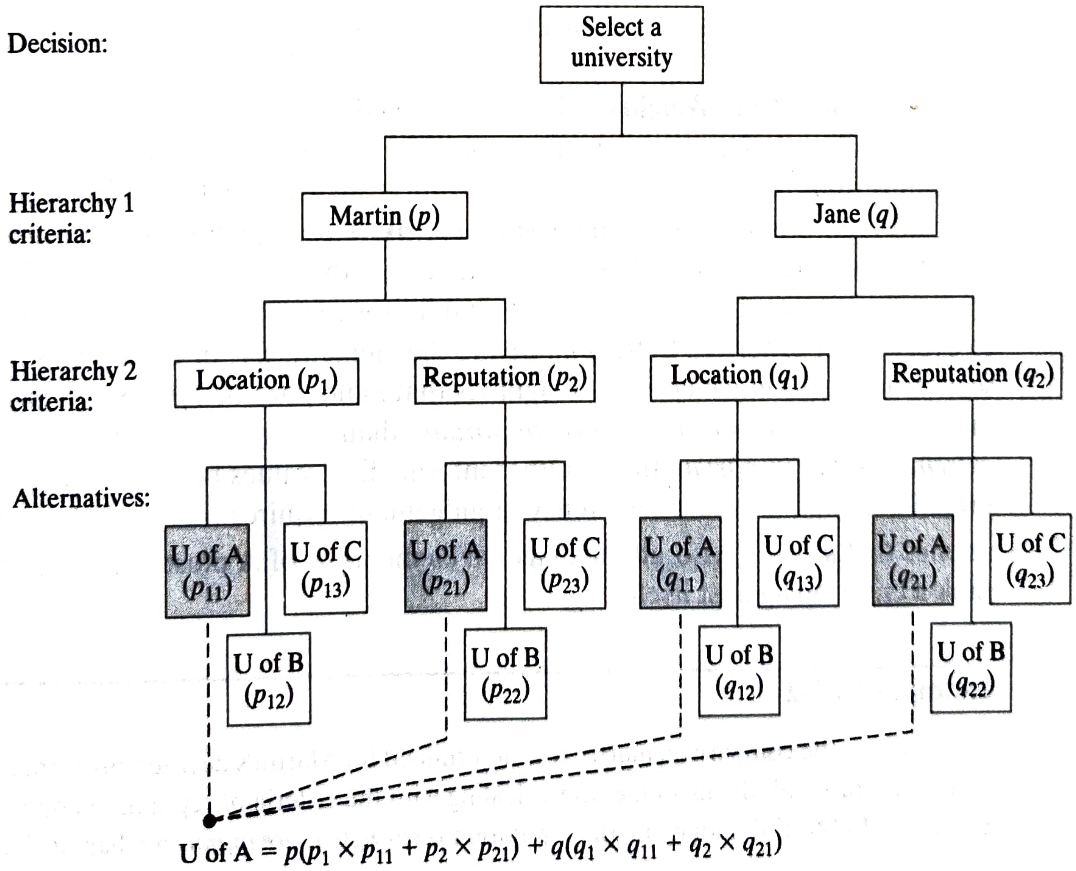


FIGURE 13.1

Summary of AHP calculations for Example 13.1-1

FIGURE 13.2

Embellishment of the decision problem of Example 13.1-1



problem, which now involves two hierarchies. The values p and q (presumably equal) at the first hierarchy represent the relative weights given to Martin's and Jane's opinions about the selection process. The second hierarchy uses the weights (p_1, p_2) and (q_1, q_2) to reflect Martin's and Jane's preferences regarding location and reputation of each university. The remainder of the decision-making chart can be interpreted similarly. Note that $p + q = 1$, $p_1 + p_2 = 1$, $q_1 + q_2 = 1$, $p_{11} + p_{12} + p_{13} = 1$, $p_{21} + p_{22} + p_{23} = 1$, $q_{11} + q_{12} + q_{13} = 1$, $q_{21} + q_{22} + q_{23} = 1$. The determination of the U of A composite weight, shown in Figure 13.2, demonstrates the manner in which the computations are carried out.

PROBLEM SET 13.1A

*1. Suppose that the following weights are specified for the situation of Martin and Jane:

$$p = .5, q = .5$$

$$p_1 = .17, p_2 = .83$$

$$p_{11} = .129, p_{12} = .277, p_{13} = .594$$

$$p_{21} = .545, p_{22} = .273, p_{23} = .182$$

$$q_1 = .3, q_2 = .7$$

$$q_{11} = .2, q_{12} = .3, q_{13} = .5$$

$$q_{21} = .5, q_{22} = .2, q_{23} = .3$$

Based on this information, rank the three universities.

Determination of the Weights. The crux of AHP is the determination of the relative weights (such as those used in Example 13.1-1) to rank the decision alternatives. Assuming that we are dealing with n criteria at a given hierarchy, the procedure establishes an $n \times n$ *pairwise* **comparison matrix, A**, that quantifies the decision maker's judgment regarding the relative importance of the different criteria. The pairwise comparison is made such that the criterion in row i $(i = 1, 2, \ldots, n)$ is ranked relative to every other criterion. Letting a_{ij} define the element (i, j) of **A**, AHP uses a discrete scale from 1 to 9 in which $a_{ij} = 1$ signifies that i and j are of *equal importance*, $a_{ij} = 5$ indicates that i is *strongly more important* than j, and $a_{ij} = 9$ indicates that i is *extremely more important* than j. Other intermediate values between 1 and 9 are interpreted correspondingly. Consistency in judgement requires that $a_{ij} = k$ automatically implies that $a_{ji} = \frac{1}{k}$. Also, all the diagonal elements a_{ii} of **A** must equal 1, because they rank a criterion against itself.

Example 13.1-2

To show how the comparison matrix **A** is determined for Martin's decision problem of Example 13.1-1, we start with the main hierarchy dealing with the criteria of reputation and location of a university. In Martin's judgment, the reputation is *strongly more important* than the location, and hence $a_{12} = 5$. This assignment automatically implies that $a_{21} = \frac{1}{5}$. Using the symbols R and L to represent reputation and location, the associated comparison matrix is given as

$$A = \begin{matrix} & L & R \\ L & 1 & \frac{1}{5} \\ R & 5 & 1 \end{matrix}$$

The relative weights of R and L can be determined from **A** by normalizing it into a new matrix **N**. The process requires dividing the elements of each column by the sum of the elements of the same column. Thus, to compute **N**, we divide the elements of columns 1 by $(5 + 1 = 6)$ and those of column 2 by $\left(1 + \frac{1}{5} = 1.2\right)$. The desired relative weights, w_R and w_L, are then computed as the row average:

$$N = \begin{matrix} & L & R \\ L & .17 & .17 \\ R & .83 & .83 \end{matrix} \qquad \begin{matrix} \text{Row average} \\ w_L = \frac{.17 + .17}{2} = .17 \\ w_R = \frac{.83 + .83}{2} = .83 \end{matrix}$$

The computations yield $w_L = .17$ and $w_R = .83$, the weight used in Figure 13.1. The columns of **N** are identical, a characteristic that occurs only when the decision maker exhibits perfect *consistency* in specifying the entries of the comparison matrix **A**. This point is discussed further later in this section.

The relative weights of the alternatives U of A, U of B, and U of C are determined within each of the L and R criteria using the following two comparison matrices, whose elements are based on Martin's judgment regarding the relative importance of the three universities.

$$\mathbf{A}_L = \begin{matrix} & A & B & C \\ A & 1 & \frac{1}{2} & \frac{1}{5} \\ B & 2 & 1 & \frac{1}{2} \\ C & 5 & 2 & 1 \end{matrix}, \quad \mathbf{A}_R = \begin{matrix} & A & B & C \\ A & 1 & 2 & 3 \\ B & \frac{1}{2} & 1 & \frac{3}{2} \\ C & \frac{1}{3} & \frac{2}{3} & 1 \end{matrix}$$

Summing the columns, we get

$$\mathbf{A}_L\text{-column sum} = (8, 3.5, 1.7)$$
$$\mathbf{A}_R\text{-column sum} = (1.83, 3.67, 5.5)$$

The following normalized matrices are determined by dividing all the entries by the respective column-sums:

$$\mathbf{N}_L = \begin{matrix} & A & B & C \\ A & .125 & .143 & .118 \\ B & .250 & .286 & .294 \\ C & .625 & .571 & .588 \end{matrix} \qquad \begin{matrix} \text{Row averages} \\ w_{LA} = \frac{.125 + .143 + .118}{3} = .129 \\ w_{LB} = \frac{.250 + .286 + .294}{3} = .277 \\ w_{LC} = \frac{.625 + .571 + .588}{3} = .594 \end{matrix}$$

$$\mathbf{N}_R = \begin{matrix} & A & B & C \\ A & .545 & .545 & .545 \\ B & .273 & .273 & .273 \\ C & .182 & .182 & .182 \end{matrix} \qquad \begin{matrix} \text{Row averages} \\ w_{RA} = \frac{.545 + .545 + .545}{3} = .545 \\ w_{RB} = \frac{.273 + .273 + .273}{3} = .273 \\ w_{RC} = \frac{.182 + .182 + .182}{3} = .182 \end{matrix}$$

The values $(w_{LA}, w_{LB}, w_{LC}) = (.129, .277, .594)$ provide the respective location weights for U of A, U of B, and U of C. Similarly, $(w_{RA}, w_{RB}, w_{RC}) = (.545, .273, .182)$ give the relative weights regarding academic reputation.

Consistency of the Comparison Matrix. In Example 13.1-2, all the columns of the normalized matrices $\mathbf{N}$ and $\mathbf{N}_R$ are identical, and those of $\mathbf{N}_L$ are not. As such, the original comparison matrices $\mathbf{A}$ and $\mathbf{A}_R$ are said to be *consistent*, whereas $\mathbf{A}_L$ is not.

Consistency implies coherent judgment on the part of the decision maker regarding the pairwise comparisons. Mathematically, we say that a comparison matrix $\mathbf{A}$ is consistent if

$$a_{ij}a_{jk} = a_{ik}, \text{ for all } i, j, \text{ and } k$$

For example, in matrix $\mathbf{A}_R$ of Example 13.1-2, $a_{13} = 3$ and $a_{12}a_{23} = 2 \times \frac{3}{2} = 3$. This property requires all the columns (and rows) of $\mathbf{A}$ to be linearly dependent. In particular, the columns of any 2×2 comparison matrix are by definition dependent, and hence a 2×2 matrix is always consistent.

It is unusual for all comparison matrices to be consistent. Indeed, given that human judgment is the basis for the construction of these matrices, some "reasonable" degree of inconsistency is expected and tolerated.

To determine whether or not a level of consistency is "reasonable," we need to develop a quantifiable measure for the comparison matrix $\mathbf{A}$. We have seen in Example 13.1-2 that a perfectly consistent $\mathbf{A}$ produces a normalized matrix $\mathbf{N}$ in which all the columns are identical—that is,

$$\mathbf{N} = \begin{pmatrix} w_1 & w_1 & \cdots & w_1 \\ w_2 & w_2 & \cdots & w_2 \\ \vdots & \vdots & \vdots & \vdots \\ w_n & w_n & \cdots & w_n \end{pmatrix}$$

It then follows that the original comparison matrix $\mathbf{A}$ can be determined from $\mathbf{N}$ by dividing the elements of column i by w_i (which is the reverse of the process of determining $\mathbf{N}$ from $\mathbf{A}$). We thus have

$$\mathbf{A} = \begin{pmatrix} 1 & \frac{w_1}{w_2} & \cdots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & 1 & \cdots & \frac{w_2}{w_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdots & 1 \end{pmatrix}$$

From the given definition of $\mathbf{A}$, we have

$$\begin{pmatrix} 1 & \frac{w_1}{w_2} & \cdots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & 1 & \cdots & \frac{w_2}{w_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} nw_1 \\ nw_2 \\ \vdots \\ nw_n \end{pmatrix} = n \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

More compactly, given that $\mathbf{w}$ is the column vector of the relative weights w_i, $i = 1, 2, \ldots, n$, $\mathbf{A}$ is consistent if,

$$\mathbf{A}\mathbf{w} = n\mathbf{w}$$

For the case where $\mathbf{A}$ is not consistent, the relative weight, w_i, is approximated by the average of the n elements of row i in the normalized matrix $\mathbf{N}$ (see Example 13.1-2). Letting $\overline{\mathbf{w}}$ be the computed average vector, it can be shown that

$$\mathbf{A}\overline{\mathbf{w}} = n_{\max}\overline{\mathbf{w}},\ n_{\max} \geq n$$

In this case, the closer $n_{\max}$ is to n, the more consistent is the comparison matrix $\mathbf{A}$. Based on this observation, AHP computes the **consistency ratio** as

$$CR = \frac{CI}{RI}$$

where

$$CI = \text{Consistency index of } \mathbf{A} = \frac{n_{\max} - n}{n - 1}$$

$$RI = \text{Random consistency of } \mathbf{A} = \frac{1.98(n-2)}{n}$$

The random consistency index, RI, was determined empirically as the average CI of a large sample of randomly generated comparison matrices, $\mathbf{A}$.

If $CR \leq .1$, the level of inconsistency is acceptable. Otherwise, the inconsistency is high and the decision maker may need to reestimate the elements a_{ij} of $\mathbf{A}$ to realize better consistency.

We compute the value of $n_{\max}$ from $\mathbf{A}\overline{\mathbf{w}} = n_{\max}\overline{\mathbf{w}}$ by noting that the ith equation is

$$\sum_{j=1}^{n} a_{ij}\overline{w}_j = n_{\max}\overline{w}_i,\ i = 1, 2, \ldots, n$$

Given $\sum_{i=1}^{n}\overline{w}_i = 1$, we get

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}\overline{w}_j\right) = n_{\max}\sum_{i=1}^{n}\overline{w}_i = n_{\max}$$

This means that the value of $n_{\max}$ can be determined by first computing the column vector $\mathbf{A}\overline{\mathbf{w}}$ and then summing its elements. / 3

Example 13.1-3

In Example 13.1-2, the matrix $\mathbf{A}_L$ is inconsistent because the columns of its $\mathbf{N}_L$ are not identical. Test the degree of consistency of $\mathbf{N}_L$.

We start by computing n_{max}. From Example 13.1-2, we have

$$\bar{w}_1 = .129, \bar{w}_2 = .277, \bar{w}_3 = .594$$

Thus,

$$\mathbf{A}_L\bar{\mathbf{w}} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{5} \\ 2 & 1 & \frac{1}{2} \\ 5 & 2 & 1 \end{pmatrix}\begin{pmatrix} .129 \\ .277 \\ .594 \end{pmatrix} = \begin{pmatrix} 0.3863 \\ 0.8320 \\ 1.7930 \end{pmatrix}$$

This yields

$$n_{max} = .3863 + .8320 + 1.7930 = 3.0113$$

Hence, for $n = 3$,

$$CI = \frac{n_{max} - n}{n - 1} = \frac{3.0113 - 3}{3 - 1} = .00565$$

$$RI = \frac{1.98(n - 2)}{n} = \frac{1.98 \times 1}{3} = .66$$

$$CR = \frac{CI}{RI} = \frac{.00565}{.66} = .00856$$

Because CR < .1, the level of inconsistency in $\mathbf{A}_L$ is acceptable.

Excel Moment

Template excelAHP.xls is designed to handle comparison matrices with sizes up to 8 × 8. As in the Excel models in Chapters 10 and 11, user input drives the model. Figure 13.3 demonstrates the application of the model to Example 13.1-2.[1] The comparison matrices of the problem are entered *one at a time* in the (top) input data section of the spreadsheet. The order in which the comparison matrices are entered is unimportant, though it makes more sense to consider them in their natural hierarchal order. Upon entering the data for a comparison matrix, the output (bottom) section of the spreadsheet will provide the associated normalized matrix together with its consistency ratio, *CR*. The user must copy the weights, *w*, from column J and paste them into the solution summary area (the right section of the spreadsheet). Remember to use **Paste Special** ⇒ **Values** when performing this step to guarantee a permanent record. The process is repeated until all the comparison matrices have been stored in columns K:R.

In Figure 13.3, the final ranking is given in cells (K20:K22). The formula in cell K20 is

=$L$4*$L7+$L$5*$N7

This formula provides the final evaluation of alternative UA, and is copied in cells K21 and K22 to evaluate alternatives UB and UC. Note how the formula in K20 is constructed: Cell reference to the alternative UA must be column-fixed (namely, $L7 and $N7), whereas *all* other references must be row-and-column-fixed (namely, $L$4 and

---

[1]The more accurate results of the spreadsheet differ from those in Example 13.1-2 and 13.1-3 because of manual roundoff approximation. Columns F:I and rows 11:13 are suppressed to conserve space.

	A	B	C	D	E	J	K	L	M	N
1			AHP-Analytic Hierarchy Process							
2		Input: Comparison matrix						Solution summary		
3	Matrix name:	AL						A		
4	Matrix size=	3	<<Maximum 8				R	0.83333		
5	Matrix data:	UA	UB	UC			L	0.16667		
6	UA	1	0.5	0.2				AR		AL
7	UB	2	1	0.5			UA	0.54545	UA	0.1285
8	UC	5	2	1			UB	0.27273	UB	0.27661
9							UC	0.18182	UC	0.59489
13										
14	Col sum	8	3.5	1.7						
15		Output: Normalized martix								
16		nMax=	3.00746	CR=	0.0056					
17		UA	UB	UC		Weight				
18	UA	0.12500	0.14286	0.11765		0.12850				
19	UB	0.25000	0.28571	0.29412		0.27661		Final ranking		
20	UC	0.62500	0.57143	0.58824		0.59489		UA= 0.47596		
21								UB= 0.27337		
22								UC= 0.25066		
26										

FIGURE 13.3

Excel solution of Example 13.1-2 (file excelAHP.xls)

\$L\$5). The validity of the copied formulas requires that the (column-fixed) *alternative* weights of each matrix appear in the same column with no intervening empty cells. For example, in Figure 13.3, the AR-weights in column L cannot be broken between two columns. The same applies to the AL-weights in column N. There are no restrictions on the placement of the A-weights because they are row- and column-fixed in the formula.

You can embellish the formula to capture the names of the alternatives directly. Here is how the formula for alternative UA should be entered:

=\$K7&"="&TEXT(\$L\$4*\$L7+\$L\$5*\$N7,"####0.00000")

The procedure for evaluating alternatives can be extended readily to any number of hierarchy levels. Once you develop the formula correctly for the first alternative, the same formula applies to the remaining alternatives simply by copying it into (same column) succeeding rows. Remember that *all* cell references in the formula must be row-and-column-fixed, except for references to the alternatives, which must be column-fixed only. Problem 1, Set 13.1b, asks you to develop the formula for a 3-level problem.

PROBLEM SET 13.1B[2]

1. Consider the data of Problem 1, Set 13.1a. Copy the weights in a logical order into the solution summary section of the spreadsheet excelAHP.xls, then develop the formula for evaluating the first alternative, UA, and copy it to evaluate the remaining two alternatives.

[2]Spreadsheet excelAHP.xls should be helpful in verifying your calculations.

7. An individual is in the process of buying a car and has narrowed the choices to three models, *M1*, *M2*, and *M3*. The deciding factors include purchase price (PP), maintenance cost (MC), cost of city driving (CD), and cost of rural driving (RD). The following table provides the relevant data for 3 years of operation:

Car model	PP($)	MC($)	CD($)	RD($)
M1	6,000	1800	4500	1500
M2	8,000	1200	2250	750
M3	10,000	600	1125	600

Use the cost data to develop the comparison matrices. Assess the consistency of the matrices, and determine the choice of model.

13.2 DECISION MAKING UNDER RISK

Under conditions of risk, the payoffs associated with each decision alternative are described by probability distributions. For this reason, decision making under risk can be based on the *expected value criterion*, in which decision alternatives are compared based on the maximization of expected profit or the minimization of expected cost. However, because the approach has limitations, the expected value criterion can be modified to encompass other situations.

Real-Life Application—Booking Limits in Hotel Reservations

Hotel La Posada has a total of 300 guest rooms. Its clientele includes both business and leisure travelers. Rooms can be sold in advance (usually to leisure travelers) at a discount price. Business travelers, who usually are late in booking their rooms, pay full price. La Posada must thus set a *booking limit* on the number of discount rooms sold to leisure travelers in order to take advantage of the full-price business customers. Decision-tree analysis is used in Case 10, Chapter 24 on the CD to determine the booking limit.

13.2.1 Decision Tree-Based Expected Value Criterion

The expected value criterion seeks the maximization of expected (average) profit or the minimization of expected cost. The data of the problem assumes that the payoff (or cost) associated with each decision alternative is probabilistic.

Decision Tree Analysis. The following example considers simple decision situations with a finite number of decision alternatives and explicit payoff matrices.

Example 13.2-1

Suppose that you want to invest $10,000 in the stock market by buying shares in one of two companies: *A* and *B*. Shares in Company *A*, though risky, could yield a 50% return on investment during the next year. If the stock market conditions are not favorable (i.e., "bear" market), the stock may lose 20% of its value. Company *B* provides safe investments with 15% return in a "bull" market and only 5% in a "bear" market. All the publications you have consulted (and there is always a flood of them at the end of the year!) are predicting a 60% chance for a "bull" market and 40% for a "bear" market. Where should you invest your money?

The decision problem can be summarized as follows:

	1-year return on $10,000 investment	
Decision alternative	*"Bull" market ($)*	*"Bear" market ($)*
Company *A* stock	5000	−2000
Company *B* stock	1500	500
Probability of occurrence	.6	.4

The problem can also be represented as a **decision tree** as shown in Figure 13.4. Two types of nodes are used in the tree: a square (□) represents a *decision point* and a circle (○) represents a *chance event*. Thus, two branches emanate from decision point 1 to represent the two alternatives of investing in stock A or stock B. Next, the two branches emanating from chance events 2 and 3 represent the "bull" and the "bear" markets with their respective probabilities and payoffs.

From Figure 13.4, the expected 1-year returns for the two alternatives are

$$\text{For stock } A = \$5000 \times .6 + (-2000) \times .4 = \$2200$$

$$\text{For stock } B = \$1500 \times .6 + \$500 \times .4 = \$1100$$

Based on these computations, your decision is to invest in stock A.

Remarks. In the terminology of decision theory, the "bull" and the "bear" markets in the preceding example are referred to as **states of nature**, whose chances of occurrence are probabilistic (.6 versus .4). In general, a decision problem may include n states of nature and m alternatives. If $p_j\ (> 0)$ is the probability of occurrence for state of nature j and a_{ij} is the payoff of alternative i, given state of nature $j (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$, then the expected payoff for alternative i is computed as

$$EV_i = a_{i1}p_1 + a_{i2}p_2 + \cdots + a_{in}p_n, i = 1, 2, \ldots, n$$

By definition, $p_1 + p_2 + \cdots + p_n = 1$.

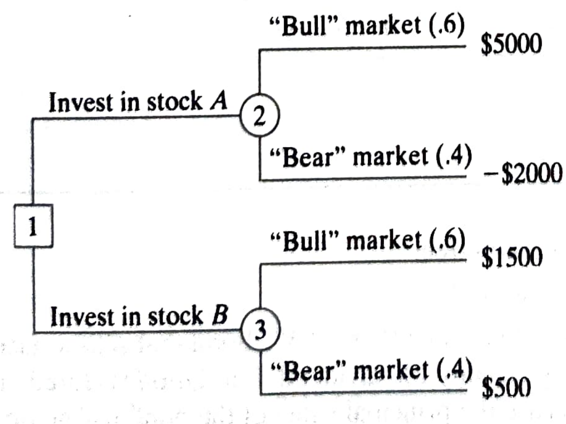


FIGURE 13.4

Decision-tree representation of the stock market problem

The best alternative is the one associated with $EV^* = \max_i\{EV_i\}$ or $EV^* = \min_i\{EV_i\}$, depending, respectively, on whether the payoff of the problem represents profit (income) or loss (expense).

PROBLEM SET 13.2A

1. You have been invited to play the Fortune Wheel game on television. The wheel operates electronically with two buttons that produce hard (H) or soft (S) spin of the wheel. The wheel itself is divided into white (W) and red (R) half-circle regions. You have been told that the wheel is designed to stop with a probability of .3 in the white region and .7 in the red region. The payoff you get for the game is

	W	R
H	\$800	\$200
S	−\$2500	\$1000

Draw the associated decision tree, and specify a course of action.

*2. Farmer McCoy can plant either corn or soybeans. The probabilities that the next harvest prices of these commodities will go up, stay the same, or go down are .25, .30, and .45, respectively. If the prices go up, the corn crop will net \$30,000 and the soybeans will net \$10,000. If the prices remain unchanged, McCoy will (barely) break even. But if the prices go down, the corn and soybeans crops will sustain losses of \$35,000 and \$5000, respectively.
   (a) Represent McCoy's problem as a decision tree.
   (b) Which crop should McCoy plant?

3. You have the chance to invest in three mutual funds: utility, aggressive growth, and global. The value of your investment will change depending on the market conditions. There is a 10% chance the market will go down, 50% chance it will remain moderate, and 40% chance it will perform well. The following table provides the percentage change in the investment value under the three conditions:

	Percent return on investment		
Alternative	*Down market (%)*	*Moderate market (%)*	*Up market (%)*
Utility	+5	+7	+8
Aggressive growth	−10	+5	+30
Global	+2	+7	+20

   (a) Represent the problem as a decision tree.
   (b) Which mutual fund should you select?

4. You have the chance to invest your money in either a 7.5% bond that sells at face value or an aggressive growth stock that pays only 1% dividend. If inflation is feared, the interest rate will go up to 8%, in which case the principal value of the bond will go down by

	Price increase (O1)	No price increase (O2)
Original mix (A1)	\$400,000	\$295,500
New mix (A2)	\$372,000	\$350,000

(a) Develop the associated decision tree and determine which action should be adopted.

(b) The manufacturer can invest \$1000 to obtain additional information about whether or not the price will increase. This information says that there is a 58% chance that the probability of price increase will be .9 and a 42% chance that the probability of price increase will be .3. Would you recommend the additional investment?

***19.** *Aspiration Level Criterion.* Acme Manufacturing uses an industrial chemical in one of its processes. The shelf life of the chemical is 1 month, following which any amount left is destroyed. The use of the chemical by Acme (in gallons) occurs randomly according to the following distribution:

$$f(x) = \begin{cases} \dfrac{200}{x^2}, & 100 \leq x \leq 200 \\ 0, & \text{otherwise} \end{cases}$$

The actual consumption of the chemical occurs instantaneously at the start of the month. Acme wants to determine the level of the chemical that satisfies two conflicting criteria (or aspiration levels): The average excess quantity for the month does not exceed 20 gallons and the average shortage quantity for the month does not exceed 40 gallons.

13.2.2 Variations of the Expected Value Criterion

This section addresses three issues relating to the expected value criterion. The first issue deals with the determination of *posterior probabilities* based on experimentation, and the second deals with the *utility* versus the actual value of money.

Posterior (Bayes') Probabilities. The probabilities used in the expected value criterion are usually determined from historical data (see Section 12.5). In some cases, these probabilities can be adjusted using additional information based on sampling or experimentation. The resulting probabilities are referred to as **posterior (or Bayes') probabilities**, as opposed to the **prior probabilities** determined from raw data.

Real-Life Application—Casey's Problem: Interpreting and Evaluating a New Test

A screening test of a newborn baby, named Casey, indicated a C14:1 enzyme deficiency. The enzyme is required to digest a particular form of long-chain fats, and its absence could lead to severe illness or mysterious death (broadly categorized under sudden infant death syndrome or SIDS). The test had been administered previously to approximately 13,000 newborns and Casey was the first to test positive. Though the screening test does not in itself constitute a definitive diagnosis, the extreme rarity of the condition led her doctors to conclude that that there was an 80–90% chance that she was suffering from this deficiency. Given that Casey tested positive, Bayes' posterior probability is used to assess whether or not the child has the C14:1 deficiency. The analysis of this situation is detailed in Case 11, Chapter 24 on the CD.

Example 13.2-2

This example demonstrates how the expected-value criterion is modified to take advantage of the posterior probabilities. In Example 13.2-1, the (prior) probabilities of .6 and .4 of a "bull" and a "bear" market are determined from available financial publications. Suppose that rather than relying solely on these publications, you have decided to conduct a more "personal" investigation by consulting a friend who has done well in the stock market. The friend offers the general opinion of "for" or "against" investment quantified in the following manner: If it is a "bull" market, there is a 90% chance the vote will be "for." If it is a "bear" market, the chance of a "for" vote is lowered to 50%. How do you make use of this additional information?

The statement made by the friend provides conditional probabilities of "for/against," given that the states of nature are "bull" and "bear" markets. To simplify the presentation, let us use the following symbols:

$$v_1 = \text{"For" vote}$$
$$v_2 = \text{"Against" vote}$$
$$m_1 = \text{"Bull" market}$$
$$m_2 = \text{"Bear" market}$$

The friend's statement may be written in the form of probability statements as

$$P\{v_1|m_1\} = .9,\ P\{v_2|m_1\} = .1$$
$$P\{v_1|m_2\} = .5,\ P\{v_2|m_2\} = .5$$

With this additional information, the decision problem can be summarized as follows:

1. If the friend's recommendation is "for," would you invest in stock A or in stock B?
2. If the friend's recommendation is "against," would you invest in stock A or in stock B?

The problem can be summarized in the form of a decision tree as shown in Figure 13.5. Node 1 is a chance event representing the "for" and "against" prossibilities. Nodes 2 and 3 are decision points for choosing between stocks A and B, given that the friend's votes are "for" and "against," respectively. Finally, nodes 4 to 7 are chance events representing the "bull" and "bear" markets.

To evaluate the different alternatives in Figure 13.5, it is necessary to compute the *posterior* probabilities $P\{m_i|v_j\}$ shown on the m_1- and m_2-branches of chance nodes 4, 5, 6, and 7. These posterior probabilities take into account the additional information provided by the friend's "for/against" recommendation and are computed according to the following general steps:

Step 1. The conditional probabilities $P\{v_j|m_i\}$ of the problem can be summarized as

	v_1	v_2
m_1	.9	.1
m_2	.5	.5