APPLIED SYSTEMS ANALYSIS

Engineering Planning and Technology Management

Richard de Neufville
Massachusetts Institute of Technology

McGraw-Hill Publishing Company
New York St. Louis San Francisco Auckland Bogotá Caracas Hamburg
Lisbon London Madrid Mexico Milan Montreal New Delhi Oklahoma City
Paris San Juan São Paulo Singapore Sydney Tokyo Toronto

CHAPTER 15

RISK ASSESSMENT

15.1 THE ISSUE

Most of engineering planning and design assumes that we know the strength, cost, and performance of our materials, that we can determine what the loads on a system will be and how it will respond. This assumption is convenient because

- It enormously simplifies the complexity of design—we can deal with only one situation instead of the many combinations of possibilities that would occur if different parameters took on different values.
- It allows designers to bypass the mathematical difficulties of probability and statistics—never popular subjects.

Unfortunately, this convenient assumption is generally false. Nothing is really certain in this world (except death and taxes, as the saying goes, and these are even uncertain as to time and amount). The fact is that our environment is not deterministic; it is probabilistic.

Experienced designers are well aware of the inconvenient reality of risk. They know that the stated strength of materials is a crude approximation; that, for example, tests of bars of steel with a nominal strength of 40 ksi will actually

yield values mostly distributed above that value but with some below. They know that the costs of a project are extremely difficult to estimate accurately. They also know that forecasts of traffic, of growth, of demand for a product are notoriously unreliable. Numerous retrospective analyses have demonstrated the truism that "The forecast is always wrong."

Costs specifically are difficult to estimate, even in the simplest situation. Cost overruns are not a peculiarity of military spending; who has not found out that the cost of repairing a car or television set is quite different from the estimate? Figure 15.1 documents a specific case of this phenomenon. It concerns the cost of resurfacing airport runways, which is one of the very simplest projects to estimate. A resurfacing project requires one to roll asphalt over a relatively flat and smooth surface of clearly specified dimensions. It is a low-technology job, involving known quantities and no hidden elements. And yet, as Figure 15.1 shows, professional engineers have great difficulty in making correct estimates of the cost of this simple project. Weather is a factor, the performance of management or labor is variable, there may or may not be competition to lower the cost of the job. It turns out that the real costs are not only higher than the estimates on average (an understandable bias) but, most importantly, are broadly distributed!

Forecasts of future loads on a system are especially subject to large error. This is because it is people who ultimately place demands on the system—by choosing to use electricity, to call their friends, or to buy a product—and people's psychology and reasons for choice are almost beyond comprehension. This phenomenon is nicely illustrated by the analysis of the forecasts of the

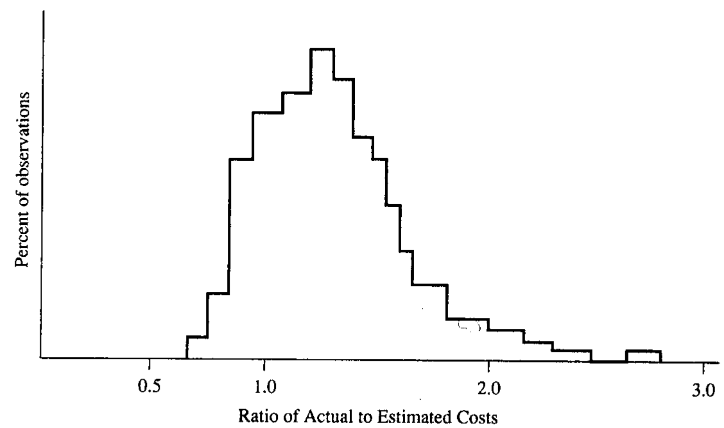

FIGURE 15.1
Probability distribution of ratio of real to estimated costs for routine airport projects (data in constant dollars for Western United States).

TABLE 15.1
Unreliability of forecasts as illustrated by large median error (source: U.S. FAA six-year forecasts)

Forecast passengers	Forecast error (median, %)
U.S. Domestic	> 15.8
U.S. International	> 20.4

U.S. Federal Aviation Administration. This agency employs the most competent professionals, using the most careful procedures, to publish annual forecasts of airline traffic. They have been doing this for about 30 years, which allows us to estimate their reliability with confidence. Table 15.1 shows the result of one of many analyses of the FAA data: their median error in forecast in just six years is about 20%! This example helps make the point: Forecasts are unreliable.

What difference does this make? One might well ask. Is it not equally likely that the actual numbers are above as below the forecast? That the errors cancel out over the long run? That we do just as well by sticking to the deterministic forecast? The answer is no and again no. The distributions do matter.

Most immediately, the distributions affect evaluation because they often involve ratios, such as for benefit-cost. The average value of such ratios is simply not the ratio of the averages of the numerator and denominators. Specifically,

$$EV\left(\frac{B}{C}\right) \neq \frac{EV(B)}{EV(C)}$$

where $EV(\bullet)$ is a standard notation to indicate expected value. Evaluations based on deterministic or average forecasts can thus easily be incorrect. (See box for concrete examples.)

Additionally, it is normal for people to feel quite differently about relative departures from the average. Typically, a catastrophic loss is much more significant to a person than a spectacular gain. In general, as developed fully in Chapters 18 to 20, people value relative gains and losses in a highly nonlinear manner. This phenomenon makes it even more inappropriate to focus on linear averages.

The bottom line is both that there is considerable uncertainty in the benefits and costs of a system, and that it matters. Neglecting the uncertainty is valid only as a first approximation, when the consequences are not especially important nor the situation especially risky. In general, however, the system designer must deal with risk, and must first of all assess it.

Average Benefit/Cost versus Ratios of Average Benefits and Costs

Consider a simple example where the benefits and costs are symmetrically distributed around their respective averages. The benefits are either 2 or 10, $EV(B) = 6$; the costs are 2 or 4, $EV(C) = 3$. The ratio of average benefits to costs is thus:

$$\frac{EV(B)}{EV(C)} = \frac{6}{3} = 2$$

Now consider two possible cases. The average benefit/cost ratio for each is significantly different from the above ratio of averages.

Case 1: Benefits and Costs Are Independent

These parameters may easily vary quite independently of each other. The benefits might fluctuate according to the whims of the public or the success of marketing, neither of which may bear any relation to the relative success in controlling costs. In this case there are four possible combinations:

Result	Benefit	Cost	*B/C*
Wild success	10	2	5
Success	10	4	2.5
Breakdown	2	2	1.0
Loss	2	4	0.5
Average			*2.25*

Case 2: Benefits and Costs Are Correlated

Both factors may also sometimes be correlated. Good management will successfully control costs and market the project and bad management will not. In this case there are only two possibilities:

Result	Benefit	Cost	*B/C*
Wild success	10	2	5
Loss	2	4	0.5
Average			*2.75*

15.2 METHODS

Estimates of probability and risk can be made by one of four basic methods. In order of increasing judgment and difficulty, these are:

- Logic
- Frequency
- Statistical Models
- Judgment

These are presented in turn.

This discussion is primarily directed toward the explicit estimate of probability distributions. It also applies to estimates of specific levels of any item of interest, such as the cost of a new facility. Indeed, once we recognize that no parameter can be known with absolute certainty, every estimate must—whether we like it or not—be considered part of a probability distribution. An estimate of the level of any parameter must realistically be considered a most likely or modal value, which could vary over some range—whether that is expressed or not.

Logic. In some cases probabilities can be deduced by logical argument. This occurs when the number of possibilities is finite and can be defined in advance, and when the mechanism that creates the outcomes is clearly specified. These cases are extremely rare in practice.

Logic is most easily applied to problems involving card games, the rolling of dice, roulette wheels, and the like. For example, we may calculate the probability of obtaining a queen from a deck of cards as $\frac{1}{13}$—assuming that the pack is a complete, standard set without any jokers and has been thoroughly shuffled.

This method permits exact calculations of the probability of many complicated situations. For example, you could use logic to estimate the probability of getting a pair of aces when your opponent has two kings, and other items of similar interest. It therefore occupies an important place in textbooks on probability. Unfortunately, real problems in systems design do not generally meet the narrow conditions necessary to make this method practical.

Frequency. Many probability distributions can easily be estimated simply by observing the frequency with which events have occurred in the past. Observing these frequencies is the same as observing the past probability distribution. This approach thus applies when we can reasonably assume that the probability of future events has not been altered by any process that has occurred between the time of observation and the moment of interest.

The frequency method is therefore routinely applied to natural phenomena such as earthquakes and patterns of rainfall. It should not be applied unquestionably to these problems, however. Just because an event is produced by nature does not mean that it is unchanged by anything we do. The number of earthquakes, for example, routinely increases when we build a major water reservoir in a region: the heavy load of the water can cause abrupt settlements. Likewise, patterns of rainfall can vary when human settlement eliminates forests and discharges warm particulate matter to the atmosphere.

This method can also be applied to many aspects of engineered systems, especially those where reason or experience indicates that the probability distributions are reasonably constant. For example, the probability distribution of the length of telephone calls, a pattern needed to design efficient telephone systems, has been traditionally estimated by the past frequency of calls of specified length. (See box for another example.)

In applying the method we must always be on the lookout for reasons why the process has changed and past frequencies are no longer good estimates of the probability distribution. This typically occurs when technology changes. The length of telephone calls in many cities, for example, has been noticeably changed by the way people use personal computers to access mainframes over the telephone network.

Probability of Failure of Dams

Until recently, the major U.S. agencies in charge of the construction of dams assumed that their probability of failure was zero. The line was that "well-built dams" (to be understood as ours) "do not fail." This is a good example of the overconfidence discussed in Section 15.5. It is also an argument difficult to maintain in the face of the failure of a major dam such as the Teton Dam in Idaho.

Colleagues and I estimated the probability of failure of major dams on behalf of the U.S. Water Resources Council. We did this by frequency analysis.

Our first step was to identify a period over which the frequency of failure could reasonably be considered stable. We took this to be the period backward toward the last major change in the technology of construction of dams, specifically the development of the methods to build large concrete dams safely.

We then turned to the catalogs listing the major dams worldwide to identify both their number in any year and the cases of failure. We could thus calculate the accumulated experience with major dams, which we measured in "dam-years," and compare it to the observed number of failures.

The probability of failure of major dams thus turns out to be

$$P(\text{Failure of Dam}) = 10^{-4} \text{ per dam-year}$$

Specifically, a major dam that operates for 100 years has an estimated probability of failure of

$$P(\text{Failure of Dam}) = 1\% \text{ over life of dam}$$

Statistical models. Statistical models for the estimation of parameters combine the main elements of both the frequency method and the judgment method, which is described below. A statistical model is essentially one or more equations describing a relationship between some parameter, y, and other quantities, x_i:

$$y = f(x_i)$$

These models are derived by statistical analysis of past data on y and x_i, as described in specialized texts.

Because these models rely on past data, they incorporate the assumption of the frequency method that the situation in the past can legitimately be extrapolated to the present or future. Additionally, these methods incorporate judgment, which is involved in specifying the exact form of the function $f(x_i)$. Even if theory suggests the general ingredients of the function, much judgment is required to specify its exact form. Economic theory indicates, for example, that the quantity of an item that will be bought depends on its price and quality. The actual form of the equation combining price and quantity, whether additive or multiplicative, and the precise way these quantities should be measured is not uniquely specified by theory. Much judgment is thus involved in using statistical models.

Statistical models tend to be deceptive. This is because they typically appear to be very technical and sophisticated. From a strictly mathematical point of view, these models may indeed be very precise. This fact does not, however, preclude the other reality that the sophisticated analysis is based on judgments that are open to question. The result is that, despite the appearance of precision, statistical models are generally about as inaccurate as methods based on judgment alone. A chain is as strong as its weakest link. Table 15.1 made the point, as does Figure 15.3 subsequently.

Judgment. Many estimates must, finally, be based on judgment. Systems designers may, for example, be required to estimate the cost or performance of a new space shuttle, computer, or material. Managers may have to estimate the public's acceptance of a new product or operation as against other alternatives. In general analysts often have to deal with situations for which there is no exact precedent. Unique situations preclude the use of frequency or statistical means to provide estimates. The analyst may, of course, use previous experience to guide the estimate but will, at the end, have to rely on judgment.

The estimates derived from judgment are known as *subjective probabilities*, subjective in that they emerge from individual feelings about a situation rather than purely from objective measures. Subjective probabilities are often highly debatable, even if they are derived from expert opinion. Individual experts are often quite positive about their estimates, an overconfidence discussed in Section 15.5, but groups of experts are quite likely to disagree.

While subjective estimates are often questionable, they will often be all the analyst has and must, therefore, be used as a start. Because these estimates are dubious it is important to revise them as soon as possible with additional information about the situation. The proper revision of estimates is a key ingredient to any risk assessment in practice.

15.3 REVISION OF ESTIMATES

A frequent problem in risk assessment is that of revising preliminary estimates of probability on the basis of new information. To appreciate the range of situations in which it is necessary to know how to revise previous estimates of probability, consider these examples:

- Exploration: Teams are sent out to prospect for desirable properties, such as geologists looking for conditions favorable to oil.
- Experimentation: Prototypes are built and tested before full-scale production is begun.
- Diagnosis: Routine tests are applied to a population to see which members warrant special attention, as for a disease.
- Market studies: New products are distributed in specific areas to see how customers will respond.

A most important feature of the problem is that, in general, the acquisition of new information does not remove all uncertainty about a situation. New information only changes our perception of the probabilities of various outcomes. When exploratory geologists find a salt dome, for example, they have not proven that oil is present; they have found a condition which makes oil more likely. Even when drillers actually tap oil, they have not removed all uncertainty about its extent or volume. Likewise, an experiment cannot prove or disprove that a full-scale process will work. The experiment may have been faulty; there may be difficulties extrapolating from the experiment to the larger reality.

There is also always uncertainty in the relation between the information acquired and the phenomenon of interest. Formally, there is always the possibility of "false positives" and "false negatives." "False positives" are the erroneous indications that a situation exists when it actually does not. For example, a person reacts positively to a tuberculosis test when not infected. A "false negative" is the opposite; it is the false indication that a situation does not exist. For example, a person passes the tuberculosis test when actually infected. (See Section 17.3 for a detailed discussion.)

These are two formal methods for revising preliminary probabilities on the basis of new information: Bayes' Theorem and Likelihood Ratios. Bayes' Theorem is the standard formula, and is best used when there is only one piece of information to be incorporated in the revision of an estimated probability. Likelihood ratios are best when there are many pieces of information.

The possibility that the new information is either incomplete or misleading means that we must be careful how we interpret it. This is especially important because, as Section 15.5 indicates, the intuitive methods people use are notoriously influenced by subjective, psychological biases.

Bayes' Theorem. Bayes' Theorem is a simple process for revising estimates of probabilities. The difficulty in understanding it lies in the elements of the formula. These are simple enough too, but generally are puzzling when first seen.

There are four elements to Bayes' Theorem. They are defined as follows.

1. The *Prior Probability*, $P(E)$, of an event E. This is the preliminary estimate of probabilities that you have before new information is acquired.
2. The *Posterior Probability*, $P(E/O)$, of the event E after some information has been acquired in the form of a specific observation O. This is the revised estimate of probability. The notation E/O is to be read "E given that (or conditional on) O having been observed." It indicates that this piece of information has been included in the estimate.
3. The *Conditional Probability*, $P(O/E)$, of the observation and the event E. This is the frequency with which an observation is associated with the existence of E, for example, that salt domes (O) are present when there is oil (E). It is important to note here that the relationship between O and E is not symmetric. For example, the probability of observing that a person is male given that the person is a king is: $P(\text{male/king}) = 1.0$, since by definition kings are men. On the other hand, since there are only a few kings on earth, the probability that any male is a king is about one in a billion: $P(\text{king/male}) \sim 10^{-9}$. In general, $P(O/E) \neq P(E/O)$.

Example for the Definition of Probabilities

Consider a factory with two kinds of staff: line workers, L, and staff, S. There are 600 line workers and 150 staff. The ratio of the sexes in each category is different: Men constitute 60% of the line workers and 10% of the staff.

Suppose that we were interested in the probability that a factory worker we meet belongs to the staff:

1. The prior probability is the frequency of staff workers. They are 150 out of a total of 750, so $P(\text{Staff}) = 0.2$.
2. The posterior probability after having made an observation, that he is male for example, is $P(\text{Staff/Male})$. This is not obvious from the data and must be calculated by Bayes' Theorem (after the observation of the worker's sex is made).
3. The conditional probabilities in this case are the frequency of male staff members: $P(\text{Male/Staff}) = 0.1$; $P(\text{Male/Line}) = 0.6$.
4. The probability of the observation of a male is their frequency among the total number of factory workers

$$P(\text{Male}) = P(\text{Male/Staff})\,P(\text{Staff}) + P(\text{Male/Line})\,P(\text{Line}) = 0.5$$

This may also be viewed as the total number of men divided by the number of factory workers.

---

4. The *Probability of the Observation*, $P(O)$. This is the probability of making the observation, O, considering all the ways it may occur. (Note, this is for a specific observation that has been made, not for a distribution over all possible values that could be made.) The observation may indeed be associated with outcome E_1, and all the other possible outcomes E_i. The probability of observing O is then

$$P(O) = \sum P(O/E_i)\,P(E_i)$$

These definitions are illustrated by the example in the preceding box.

Bayes' Theorem is a straightforward use of the above elements:

$$P(E/O) = P(E)\left\{\frac{P(O/E)}{P(O)}\right\}$$

The revised estimate of probability is simply the preliminary, prior estimate multiplied by a factor for revision, based upon an observation. Applying Bayes' Theorem is direct, once the elements have been defined (see box).

The strength of the factor of revision of the estimate, that is, the ratio of the prior and posterior estimates of the probability, depends on two considerations.

Use of Bayes' Theorem

We are at the same factory used to illustrate the definition of the different kinds of probabilities. Being just about to meet a male worker, you want to estimate the probability that he is on the staff.

You thus calculate:

$$P(\text{Staff/Male}) = P(\text{Staff})\left\{\frac{P(\text{Male/Staff})}{P(\text{Male})}\right\} = 0.2\left\{\frac{0.1}{0.5}\right\} = 0.04$$

It is thus apparent that the revision is quite strong, due to the disassociation of the observation with the event of interest: few staff members are male, $P(\text{Male/Staff}) = 0.1$. The prior probability is divided by five.

Conversely, if you wanted to use the observation that the member is a man to revise the prior estimate that he is a line worker, you would calculate:

$$P(\text{Line/Male}) = 0.8\left\{\frac{0.6}{0.5}\right\} = 0.96$$

The revision here is not particularly strong (ratio = 1.2) since the frequency of male line workers is about equal to that in the factory as a whole.

These can be seen directly from the formula for the factor: $\{P(\text{O/E})/P(\text{O})\}$. The revision is stronger when the observations are rare: the factor is greater when the denominator P(O) is small. Conversely, the revision is stronger when the observation is either closely associated with or quite disassociated from the event of concern: the factor is greater when the numerator P(O/E) is near either extreme, 0 or 1.0. It follows that strongest revisions to initial estimates of probability will be due to rare observations uniquely associated with the event of interest.

Likelihood ratios. Likelihood ratios provide a rapid means to revise prior estimates of probability when one obtains a sequence of independent observations bearing on some event. They enable us to bypass repeated applications of Bayes' Theorem. This is convenient because it is quite tedious to apply Bayes' Theorem over and over. In addition to having to use the formula once for each observation, one also has to recalculate P(O) at each iteration because it changes with each new estimation of P(E). The likelihood ratio permits us to collapse all this effort into a single formula that never requires any recalculations.

The use of the likelihood ratio involves the concept of complementary probability. The *complementary probability* of an event E is the probability that event E does not occur, that is, that some other event or events, non-E, occur instead. Since an event either occurs or not, P(E) and the complementary probability, P(non-E) sum to one:

$$P(\text{E}) + P(\text{non-E}) = 1.0$$

For example, if P(E) is the probability that a person you meet in the street is male, P(non-E) is the complementary probability that the person is female.

> **Notation:** In discussing probabilities, the use of a horizontal line over a symbol often indicates the nonexistence of that variable. This use can be confusing because the same notation is sometimes used to denote average values of a variable. To avoid difficulty, we will consistently refer to nonexistence of a variable X as "non-X."

A *likelihood ratio*, LR, is simply the ratio of the probability of event E and the probability of all complementary events, non-E:

$$\text{LR} = \frac{P(\text{E})}{P(\text{non-E})}$$

The likelihood ratio thus implicitly defines the probability of event E. Since $P(\text{non-E}) = 1 - P(\text{E})$, we can express the formula for LR in terms of P(E) only. Solving for P(E) we get

$$\text{P(E)} = \frac{\text{LR}}{(1 + \text{LR})}$$

The likelihood ratio is similar to the odds sometimes used in betting. In horse racing, for example, it is usual to give odds in the form X:Y to win. Thus, if a horse is 3:2 to win, it means that the estimate is that it has 3 chances to lose for 2 to win. The likelihood ratio on losing is $\frac{3}{2} = 1.5$. The estimated probability of losing is then

$$P(\text{Lose}) = \frac{1.5}{2.5} = 0.60$$

To explain the use of the likelihood ratio, consider first the simplest situation, in which we have one observation. We have a prior estimate of probability P(E), have made the observation, O_j, and wish to obtain the posterior probability $P(\text{E/O}_j)$. Defining LR_i as the likelihood ratio after i observations, we have by definition

$$\text{LR}_1 \equiv \frac{\{P(\text{E/O}_j)\}}{\{P(\text{non-E/O}_j)\}}$$

This can be restated by applying Bayes' Theorem to both the top and bottom of the ratio. We thus obtain

$$\text{LR}_1 = \left\{ \frac{P(\text{E})\left\{\dfrac{P(\text{O}_j/\text{E})}{P(\text{O}_j)}\right\}}{P(\text{non-E})\left\{\dfrac{P(\text{O}_j/\text{non-E})}{P(\text{O}_j)}\right\}} \right\}$$

This expression can be simplified by cancellation of the factor that is common to both top and bottom of the ratio, $P(\text{O}_j)$. This elimination explains why $P(\text{O}_j)$ does not have to be calculated when using likelihood ratios instead of Bayes' Theorem. The result is

$$\text{LR}_1 = \text{LR}_0 \left\{ \frac{P(\text{O}_j/\text{E})}{P(\text{O}_j/\text{non-E})} \right\}$$

That is, the revised likelihood ratio after one observation is the original likelihood ratio times a factor uniquely associated with the observation. The posterior probability, $P(\text{E/O}_j)$ can then be found as:

$$P(\text{E/O}_j) = \frac{\text{LR}_1}{(1 + \text{LR}_1)}$$

The likelihood ratio after some observation is conveniently restated using the concept of the conditional likelihood ratio. The *Conditional Likelihood Ratio*, CLR_j, for any observation O_j, is defined as

$$\text{CLR}_j \equiv \left\{ \frac{P(\text{O}_j/\text{E})}{P(\text{O}_j/\text{non-E})} \right\}$$

Using this concept the revised likelihood ratio after a single observation O_j is then simply

$$\text{LR}_1 = \text{LR}_0(\text{CLR}_j)$$

Note that the conditional likelihood ratio can be determined in advance regardless of the number of observations. It does not depend on the actual estimate of the probability of E. Use of this concept provides the way to define a single formula to determine, in advance, the probability of E after a specified number of observations.

The general formulation using likelihood ratios is an extension of the result for a single observation. Each time an observation of type j is observed, one

---

Use of Likelihood Ratios

Consider a bottle-making factory. Suppose that its machines can either be OK or, 10% of the time, defective:

$$P(\text{D}) = 0.1 \qquad P(\text{OK}) = 0.9$$

The bottles sometimes come out cracked due to heat stresses. The frequency of cracking depends on the state of the machine. Assume that this frequency is

$$P(\text{C/D}) = 0.2 \qquad P(\text{C/OK}) = 0.05$$

If we sample 5 bottles produced by a particular machine and observe that 2 are cracked, and 3 are uncracked, what is the probability that the machine is defective? That is, $P(\text{D/[2C, 3U]}) = ?$

To calculate this by likelihood ratios we need

$$\text{LR}_0 = \frac{P(\text{D})}{P(\text{OK})} = \frac{0.1}{0.9} = \frac{1}{9}$$

together with the conditional likelihood ratios for each type of observation:

$$\text{CLR}_C = \frac{0.2}{0.05} = 4$$

$$\text{CLR}_U = \frac{0.8}{0.95} = \frac{16}{19}$$

We can then get the likelihood ratio after the five observations:

$$\text{LR}_5 = \text{LR}_0\ (\text{CLR}_C)^2\ (\text{CLR}_U)^3 = \left(\frac{1}{9}\right)(4)^2\left(\frac{16}{19}\right)^3 = 1.062$$

Therefore:

$$P\ (\text{D/[2C, 3U]}) = \frac{\text{LR}_5}{(1 + \text{LR}_5)} = 0.515$$

---

updates the likelihood ratio by the appropriate conditional likelihood ratio. Thus

$$\text{LR}_N \equiv \text{LR}_0 \prod (\text{CLR}_j)^{N_j}$$

where N is the total number of observations, N_j is the number of observations of type j, and $N = \sum N_j$. Note carefully that the revision only depends on the number of observations of each type, not on the order in which they are presented. This general formulation is useful because it enables one to calculate directly the effect of many different observations.

In practice, the use of the general likelihood ratio is simple:

1. Calculate LR_0 and the conditional likelihood ratio CLR_j for each type of observation O_j.
2. Count the number of observations of each type j.
3. Calculate LR_N by formula.
4. Recover the revised estimate of the probability of event E.

Use of the likelihood ratio to revise estimates of probability assumes that we have conditional probabilities for any of the observations O_j that may be made. This means that we have the frequencies $P(\text{O}_j/\text{E})$ and $P(\text{O}_j/\text{non-E})$. (See box.)

15.4 CONTINUOUS PROBABILITY DISTRIBUTIONS

Conceptually, the revision of estimates of probability when the distribution is continuous is the same as when the probabilities are discrete. In practice, however, the calculations are much more complicated. This section outlines these difficulties, leaving the full treatment to specialized texts.

Dealing with continuous distributions involves a complex of related issues:

- Integrations must be used instead of summations.
- The information itself tends to come in distributions, rather than in discrete pieces of data.
- The use of Bayes' Theorem may become nearly impossible, when dissimilar distributions have to be considered jointly.

By itself, the problem of integration is the simplest concern. For example, in estimating the probability of an event, given all the ways it can occur, we simply substitute the integration for the summation to obtain

$$P(\text{O}) = \int \text{pdf (O/E)}\ d\text{E}$$

where pdf(•) denotes the probability distribution of a quantity.

The real difficulty arises when we consider the nature of the information we receive. When dealing with continuous distributions, the new information itself tends to occur as a distribution. A typical situation is that engineers have a prior estimate of the measurement of a quantity (such as the strength of a material, the speed of an aircraft on the radar, the distance of a satellite) and then obtain a second series of measurements, also as a probability distribution. The problem then becomes one of incorporating a probability distribution, instead of a single piece of data, into Bayes' Theorem.

The melding of the two probability distributions, those of the prior estimate and of the new data, is generally problematical. It is only relatively easy if the two distributions are "conjugate distributions," that is, if they have specific convenient properties. When they do, Bayes' Theorem can be applied quite directly (see box). This is not always the case, however, and the calculations can become extremely complicated.

Bayes' Theorem for Continuous Probability

Colleagues at Stanford University undertook a study to determine how the estimates of the compressibility of soil foundations were changed by the information derived from soil samples. Focusing on the soil along the San Francisco Bay, they obtained the a priori estimates of experienced soils engineers, the results of soils tests, and calculated revised estimates of the strength. All data were in the form of probability distributions.

The distributions used were those of the t-statistic, partly because it is the proper distribution for a normal distribution when both statistics are unknown, partly because t-distributions are "conjugate functions" that permit the relatively easy application of Bayes' Theorem.

The expression for the revised estimates, using these convenient assumptions, was given by

$$\text{pdf}(\text{true mean}/m, k, d) = d^{d/2}\left\{d + k(\text{true mean} - m)^2\right\}^{-(d+1)/2}(k)^{1/2}\left\{b(1/2, d/2)\right\}$$

where

m = pooled mean
k = (sample variance)$^{-1}$
d = degrees of freedom
$b(\cdot)$ = beta function

The expression makes the point: a "simple" result for continuous probability distributions is actually quite complex and tedious to calculate. Advanced texts provide the details.

15.5 BIASES IN ESTIMATION

Professionals must often estimate probabilities according to their best informed judgment. Although we might wish for objective measures, the reality must be estimated subjectively, as Section 15.2 describes.

A key difficulty here is that people are biased estimators. As repeated experiments demonstrate, both individuals and groups systematically provide skewed estimates of the probability of events. This section describes the major kinds of biases. The idea is to alert readers to their effects so that they can compensate for them in practice.

Overconfidence is arguably the root cause of the common types of biases. A general phenomenon is that all persons act as if they know much more about a situation than they actually do. Even when they know they are quite ignorant about a topic they typically endow their estimates with unwarranted precision. Psychological theory offers many other reasonable explanations of why people bias their estimates. However, the best guidance that can be offered to compensate for the biases is: restrain your confidence; be modest.

Three most obvious manifestations of bias in the estimate of probabilities are

- Overly narrow range of estimates
- Inadequate response to new information
- Hedging of estimates

These are each discussed below.

Narrow range of estimates. This is the prime case of overconfidence: people regularly will estimate a quantity very precisely, within a narrow range, even when they have little justification for such confidence. They are willing to say, in effect, that there is a very high probability that the value is what they say it is, and low probability that it is anything else. In this they are generally wrong.

In practice, this bias is manifest in two kinds of situations: the estimate of different values and the forecast of future states. It is, of course, most immediately evident when they deal with current values that can be checked. This is easily shown in a classroom or for any group by asking simple questions whose answers can be found in some reference work (the box on the following page provides an example of these "almanac questions").

The overconfidence that we can easily demonstrate using almanac questions also routinely occurs in professional practice. The only difference is that we rarely get a chance to observe it positively. We often do see that an expert's estimate turned out to be wrong. But since we rarely can see how often the expert's estimates are wrong, we cannot usually demonstrate the overconfidence.

A symposium held at MIT did demonstrate this overconfidence rather neatly, however. In preparation for a speciality conference on soil mechanics, 10 world

Length of River Nile

This is an example of the "almanac questions" that can be used to demonstrate people's overconfidence in their estimates. To conduct the experiment you need to have a willing group of participants (a class or a group of colleagues), and some specific physical facts that can be looked up in an encyclopedia or almanac. These facts could be items such as the distance to the moon, the amount of rainfall in July, or the population of Peru. The author's favorite has been the length of the River Nile.

The organizer of the experiment asks each participant to estimate the value of the fact selected, and to provide the plus or minus range for this value such that there is a 50:50 chance that it includes the true value. Note that it is easy to provide a range that must include the true value; it presumably is minus to plus infinity. This is also a uselessly broad estimate. In effect the organizer requests each member to provide a best estimate with 50% confidence limits.

If the estimates were accurate, one should find that, on average, half of the estimates actually did include the true value. This is not what happens. Typically, only 10 to 20% of the estimates include the true value. The rest have excluded it because they set their range much too narrowly: they were overconfident.

Test yourself: what do you estimate the length of the River Nile to be, with plus or minus 50% confidence limits?

When the author asks this question in class he routinely gets answers such as 500 ± 200 miles, 1800 ± 400 miles, and so on. The true value, hidden so your eye did not catch the answer before you addressed the question, is slightly more than twice the current year, in miles.

class experts were requested to estimate the strength of an embankment, with 50% confidence limits. They were given a full set of data on the soil and the state of the embankment. On one of the field trips associated with the conference, the embankment was loaded until it failed, thus creating an almost unique opportunity to demonstrate overconfidence among professionals.

The overconfidence was painfully obvious, as Figure 15.2 shows. In this case not one of the experts included the true value in their 50% confidence limits. Based on both psychological experiments and professional experience, this kind of result appears quite standard. The lesson is: do not be overconfident in your own estimates or those of others—allow generously for the possibility of being wrong.

Similarly, overconfidence in forecasting becomes evident when one compares forecasts with what actually occurs. This is most easily done when forecasters have provided high and low estimates, as they sometimes do. Figure 15.3 is a typical example of the comparison; the narrow range clearly excludes the reality, and demonstrates overconfidence. Similar comparisons can be made for all kinds of statistically based forecasts, because they normally provide confidence limits on their parameters.

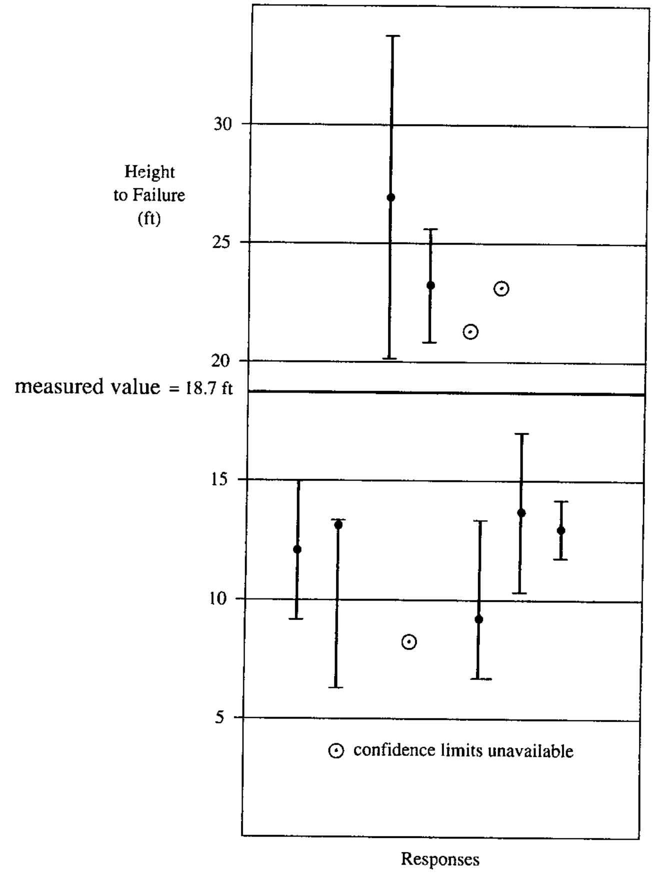

FIGURE 15.2
Demonstration of overconfidence: The true value of the strength of the embankment, as measured by the height of the load, was outside all of the experts' 50% confidence limits.

Inadequate response to new information. As another form of overconfidence, people typically fail to adjust their estimates adequately to new information. They are usually conservative, in that they tend to stick close to their initial estimates. They indicate, in effect, that they know better and do not really need to be influenced by new information.

The relative importance of the new information compared to the initial estimate becomes evident by looking at the general formula for the likelihood ratio:

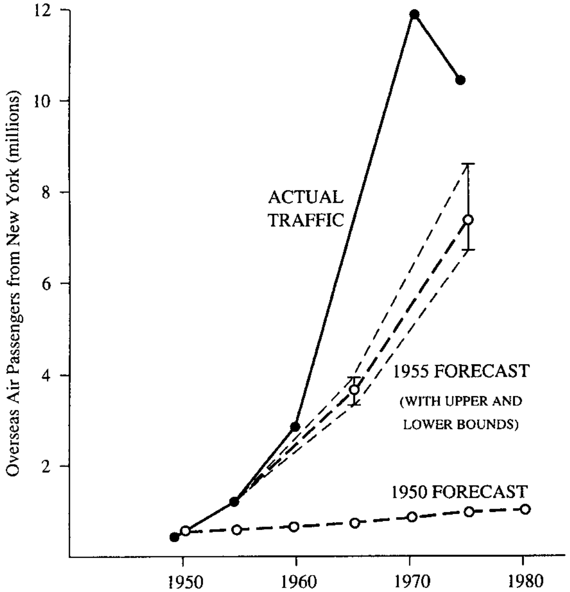

FIGURE 15.3
Overconfidence as demonstrated by the comparison of actual results and the narrow range forecast by experts.

$$LR_N = LR_0 \prod (CLR_j)^{N_j}$$

The driving factor in this equation is the multiplication of powered conditional likelihood ratios. After only a few observations the effect of LR_0 and the initial estimates have been dissipated. The previous box illustrating the use of likelihood ratios demonstrates this effect. We may thus conclude that, in general, analysts should not be confident in their initial estimates; they should rather rely on the evidence of multiple observations.

This phenomenon has been repeatedly demonstrated in carefully controlled psychological measurements. It can also be informally demonstrated with colleagues (see box).

Hedging of estimates. People are said to "hedge" when they act so as to avoid extremes and ensure that the outcomes of a situation are close to the averages. The term is generally used in connection with investments, specifically when investors buy insurance against the fluctuations of the market.

Demonstrating Inadequate Response to Information

The organizer of the demonstration prepares by setting up a simple situation and calculating possible results in advance. This person then asks a colleague or members of a group to provide estimates of a quantity according to the information provided. The comparison of the individuals' estimates with the proper estimate generated by Bayes' Theorem demonstrates the overconfidence. Typically, the individuals' estimates change slowly when they should change much more dramatically.

For example, consider an electronic assembly, with a 10% a priori probability of being faulty:

$$P(\text{Faulty}) = 0.10$$

Suppose that if the assembly is faulty it has a 50% probability of generating an error signal:

$$P(\text{Signal/Faulty}) = 0.5$$

while a good assembly can also generate error signals, but at a lower rate:

$$P(\text{Signal/Nonfaulty}) = 0.1$$

The questions to ask are then, for example, what is the probability that a part is faulty if repeated tests generate one error signal? One error and one OK? 2 error signals? And so on.

The correct answers are conveniently found by the likelihood ratios. For this case,

$$LR_0 = \frac{P(\text{Faulty})}{P(\text{OK})} = \frac{1}{9}$$

$$CLR_{\text{signal}} = 5$$

$$CLR_{OK} = \frac{5}{9}$$

so that, after n tests:

$$LR_n = \left(\frac{1}{9}\right)(5)^{\text{errors}}\left(\frac{5}{9}\right)^{\text{oks}}$$

For example, what would you estimate the probability of being faulty to be after two error signals? Write down your answer and compare it with the correct solution, calculated as

$$LR_2 = \left(\frac{1}{9}\right)(5)^2\left(\frac{5}{9}\right)^0 = \frac{25}{9}$$

so that

$$P(\text{Faulty/2 Error Signals}) \sim 70\%$$

People likewise hedge their estimates of a quantity when they provide responses that avoid extreme values. They then act as if they are unwilling to accept that the actual value of this quantity may be quite different from the average. This bias is thus similar to the overconfidence previously discussed: people avoid wide ranges.

The classic demonstration of hedging consists of asking individuals to estimate the frequency of letters in a language, for example, in English. Their answers will tend to cluster around the average value of $\frac{1}{26}$ = 4%, and they will systematically underestimate the actual high frequency of common letters such as "e," and overestimate the frequency of improbable letters such as "q" and "z." (Their usual frequency is, in fact, 13% for "e" and 0.25% each for "q" and "z.")

15.6 APPLICATIONS

This section provides a sequence of examples to illustrate the application of the methods discussed.

Frequency estimates. The probability, the risk of many kinds of events, is commonly deduced from careful examinations of the historical record. The estimate of the probability of failure of large dams discussed in Section 15.2 is a good example of the process.

The same approach can also be used to estimate probability distributions. This is commonly done for earthquakes, floods, and other natural events. It can also be applied to recurrent human situations, such as the estimate of costs discussed in Section 15.1. The method is suitable whenever the underlying causes of the event of interest have not changed significantly from the past.

Statistical models. These are most commonly found in situations thought to be well described by some theory, particularly economics. They are thus routinely used to estimate future demand for the services of a system, such as traffic on a communications network, passengers in aviation, and so on. They are equally used to estimate future prices (since price is the complement to quantity demanded in standard economics) as for oil and other commodities.

Statistical models, ultimately based on judgment for their form and thus results, nicely illustrate the pervasive problem of overconfidence. The analyses typically lead to narrow confidence limits on the values of the parameters—and thus on the results. But as the discussion of the aviation forecasts in Section 15.1, and as Figure 15.3 shows, it is quite possible to be confident and wrong.

In fields that have been extensively analyzed it is furthermore possible to obtain probability distributions on the parameters of statistical models using the frequency approach. Figure 15.4 illustrates the result. It simply shows the distribution of a key parameter of demand models, individual responsiveness to

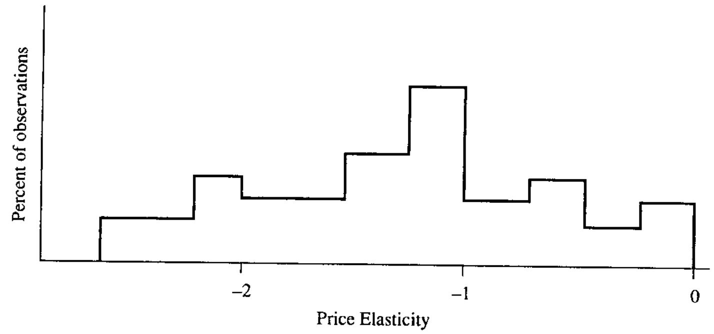

FIGURE 15.4
Probability distribution of price elasticity, for North Atlantic Business Travel by air, based on 59 statistical studies.

price, based on many independent studies of the same situation. Note that this distribution spans a broad range ($-3 <$ price elasticity < 0) and contradicts the typically overconfident claims made by individual studies that this parameter can be determined with 95% confidence limits of ± 0.1 or less.

Judgment. This is most obviously called for in situations for which there is no appropriate experience or statistics. This is common whenever one is dealing with new technology, for example. The best if not the only estimates of the performance of new aircraft, the demand for new computers, or the efficiency of robots will come from experts. Judgment is also the standard approach used to estimate risk in unique, individual situations concerning, for instance, the likelihood of being a victim of or winning a lawsuit.

Computer-based expert systems, which build on expert judgments, also naturally use Bayes' Theorem extensively. In a typical application, an expert system for finding oil (or making a medical diagnosis) will use the formula to revise prior probabilities based on new information such as the response to a probe. If the system is normally subjected to repeated measurements, as in the quality control of a production line or the warm-up testing of a computer, likelihood ratios may be used.

REFERENCES

Ascher, W., (1978). *Forecasting: An Appraisal for Policy-Makers and Planners*, Johns Hopkins University Press, Baltimore, MD.

de Neufville, R., and Marks, D. H., (1976). *Systems Planning and Design, Case Studies in Modeling Optimization and Evaluation*, Prentice-Hall, Englewood Cliffs, NJ.

Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L., and Keeney, R. L., (1981). *Acceptable Risk*, Cambridge University Press, Cambridge, England.

Folayen, J. I., Hoeg, K., and Benjamin, J. R., (1970, 1976). Chapter 23 in de Neufville and Marks (1976), also "Decision Theory Applied to Settlement Predictions" *ASCE Journal of Soil Mechanics and Foundations*, Vol. 96, July 1970, pp. 1127–1141.

PROBLEMS

15.1. *Money Bags*

You are a contestant on the "Money Bags" TV game program. Monty, the MC, has placed two bags of cash in front of you and told you that you may take one of them. One bag contains 60 $10 bills and 40 $1 bills while the other holds 20 $10 bills and 80 $1 bills. You do not know which bag has the $640, but you would very much like to choose that one.

(*a*) Monty will let you draw a bill from one of the bags before you decide which bag to choose. If you pull out a $10 bill from one of the bags, should you choose that bag?

(*b*) If Monty lets you draw three bills from one of the bags, replacing each bill before drawing the next, and you pull out one $10 bill and two $1 bills, which bag should you take?

(*c*) If Monty tells you to draw one more bill from the same bag before making your decision, and you pull out a $10 bill (total sample: 2 $10 bills and 2 $1 bills), which bag should you choose? Explain the significance of drawing this last $10 bill.

15.2. *Diskette Drives*

Suppose that you are the programming supervisor for a group developing software for a new personal computer. Among your duties, you must select the number of retries your programmers must attempt when their programs retrieve information from the diskette drives.

Within this PC, you know that during an attempt to retrieve data from the diskette drives, one of two things will occur: either the data is successfully collected or the program receives a signal that the data cannot be retrieved. Thus, you know that any program must take the following steps:

1. Request information from the diskette drive.
2. Check for the error signal.
3. If there is no signal, the data was collected successfully—continue to the next step.
4. If there is a signal, try step 1 again.
5. If the signal is detected N times in a row, abort the program.

You know that this error signal is generated whenever there is a diskette problem; it also can occur even though there is nothing wrong with the diskette. According to the device specifications, if there is a diskette problem, this signal is generated 100% of the time. Alternatively, if there is no diskette problem, this signal is generated 30% of the time. The probability that there is a diskette problem is 20%.

How many attempts to read diskette data should be made if company policy is that software should abort only if there is a 99% certainty that there is a diskette problem?

15.3. *Lie Detector*

A lie detector has a hitting rate of 80% and a false alarm rate of 50%. That is, the probability of the machine giving a positive result (indicating "LIE!") is 0.80 if the subject is lying, but 0.50 if the subject is telling the truth.

Suppose a subject is known a priori to have a 20% chance of lying on any particular test.

(*a*) What is the probability that this subject is actually lying if the machine says "LIE"?

(*b*) If the machine was improved by raising its hitting rate to 99%, how would your answer to part (*a*) be modified? Is that a significant improvement?

(*c*) Suppose you now have a subject who you think will lie four times out of five, on any trial. What is the new answer to part (*a*)? Does the lie detector improve your prior belief?

15.4. *VLSI Chips*

A manufacturing line produces VLSI chips of which 25% do not meet specifications. An automatic testing device is used to run four different independent tests on the chips. If a chip does not meet specifications, it has an 80% chance of failing any one of the tests. A chip that does meet specifications will also fail the tests 40% of the time.

(*a*) If a certain chip passes three of the tests but fails one, what is the probability that the chip meets specifications?

(*b*) If one of the tests produces independent results when repeated a number of times on a given chip, what is the minimum number of tests that must be run to achieve 90% probabilty that the chip does not meet specifications?

15.5. *Oil Drilling*

In a certain oil-rich region, there is prior probability of $\frac{2}{3}$ that any field will produce a profitable oil well. Test drillings are made to determine whether or not a well in a given field would be profitable. There is a 75% chance that the test drilling would be positive if in fact a field would support a profitable well. There is a 50% chance of a negative test if a field would not support a profitable well.

(*a*) Assuming two test drillings, one positive and one negative, are made in a field, find the revised probability of producing a profitable well in the field by (1) successive applications of Bayes' Theorem; (2) likelihood ratios.

(*b*) If five tests are made, three positive and two negative, what is the revised probability of producing a profitable well?

(*c*) When should likelihood ratios be used instead of Bayes' Theorem? What advantage do they have?

15.6. *Sonny Reyes*

Sonny Reyes, the famous photovoltaic (PV) manufacturer, is testing a new PV panel. If a panel does not meet specifications it has a 80% chance of failing the test. A panel that does meet specifications has a 20% chance of failing the test. Overall, four in five panels meet specifications.

(*a*) Define the formula for the prior likelihood ratio for this problem.

(*b*) Define the conditional likelihood ratios for this problem.

(*c*) Write the formula for the posterior likelihood ratio, if a panel first fails and then passes a second test.

(*d*) Solve for the posterior probability of a panel meeting specifications.

15.7. *SIDA Testing*
The incidence of SIDA, a deadly disease, among a certain population is 0.01%. Individuals, randomly selected from this population, are submitted to a SIDA test whose accuracy is 99% both ways. That is to say, the proportion of positive results among people known to be SIDA affected is 99%. Likewise, testing people that are not suffering from the disease yields 99% of negative results. The test gives independent results when repeated.

An individual tests positive.

(*a*) What is the probability that this person is actually affected? (Use both Bayes' Theorem and likelihood ratios.)
(*b*) Discuss the above result as regards the interpretation of the positive result.
(*c*) The test is then repeated twice. What is the probability that the person has SIDA if all three tests are positive? If the two subsequent tests are negative?

15.8. *Weather Expert*
The radio predicts a 60% chance of freezing weather. Your meteorological friend, May Vin, tells you she knows better: it is sure to freeze. From experience you know that she only gets it right 80% of the time.
(*a*) What should your estimate of freezing weather be?
(*b*) What would it be if May had predicted "no freezing weather"?

15.9. *Summer Goods*
Of the summer goods, some are bad. Two percent are defective. Visual inspection is cheap, but only correct half the time. A detailed examination, however, gives a correct diagnosis 90% of the time. Normal procedure is to look the goods over and then to examine in detail the ones that seem defective visually. Goods that fail both tests are rejected.
(*a*) What percent of the goods that pass visual inspection are in fact defective?
(*b*) What percent of defectives are not detected by the total examination process?
(*c*) If the detailed examination were applied to all goods, what percent of defectives would pass? Discuss whether you think this policy would make sense.

15.10. *Championship Playoff*
Before the infinite series, it looks as if either team A or B is equally likely to win the series. Past frequency indicates that "champions" win 70% of their games.
(*a*) If team A wins the first game, what is the probability that it is a "champion"?
(*b*) Use Bayes' Theorem to calculate how many times in a row team A should win so that the probability that it is a "champion" is greater than 90%. Then validate this by likelihood ratios.
(*c*) What is the probability that any team is a "champion" if it wins 2 out of 3 games? 3 out of 5? 4 out of 7?

CHAPTER 16

DECISION ANALYSIS

16.1 OBJECTIVE

This chapter presents the concept and methods of decision analysis, a fundamentally important method of evaluation. This is the approach that should be used whenever the outcomes of potential projects are highly uncertain. Since the planning and design of systems typically must deal with massive uncertainty about the future, as the previous chapter shows, decision analysis is a most valuable tool.

Formally, *decision analysis* is a method of evaluation that leads to three results:

1. It *structures* the problem, which otherwise appears very confusing to most people due to the complexities introduced by uncertainty (Sections 16.3 and 16.4).
2. It *defines optimal choices* for any period, based on a joint consideration of the probabilities and the nature of any outcome of a choice, specifically by calculation of an expected value (Sections 16.5 to 16.7).
3. It *identifies an optimal strategy* over many periods (Sections 16.8 and 16.9).

Decision analysis rests on the simple proposition that a planner or designer should use all the important information available about a problem, specifically the fact that the performance of any system is uncertain. This premise makes decision analysis very different from the traditional economic evaluations, which focus only on the typical or most likely outcome of a situation.