8.0 COST RISK AND UNCERTAINTY

8.1 Introduction

Risk and uncertainty exist in cost estimating because a cost estimate is a prediction of the future. There is a chance that the estimated cost may differ from the actual cost. Moreover, the lack of knowledge about the future is only one possible reason for such a difference. Another equally important cause is errors resulting from historical data inconsistencies, cost estimating equations, and factors that typically are used in an estimate. For instance, the standard error of the estimate (discussed in Section 9.5.5 of Chapter 9) and the limitations of historical data (covered in Section 5.2.2 of Chapter 5) are both examples of error sources. Thus, when viewed in its totality, a cost estimate can include a substantial amount of error. Once this is recognized, the question becomes one of dealing with those errors, which is what the subject of risk and uncertainty is about.

The rest of this section discusses the purpose of both risk and uncertainty analysis and provides useful definitions. Section 8.2 provides general background information and a discussion on the nature of risk and uncertainty. Section 8.3 reviews the typical approaches that have been proposed and applied in dealing with the uncertainty problem. Section 8.4 provides a summary of models and methods currently being employed. The main purpose of this section is not to provide a detailed description of each model or method, but rather to show how cost estimation approaches in the previous section were implemented by field practitioners. Qualitative indices are covered in Section 8.5.

8.1.1 Purpose of Risk and Uncertainty Analysis

In general, people associate one number with an estimate. The use of one number rather than a range of numbers probably has its origins in the need to put one value in a budget request. The budget request quickly becomes a very practical document, with organizational obligations made based on approved budgets. Obligations involve payment to individuals or entities; and payment is made in discrete dollars and cents, not ranges of estimated amounts. However, estimates are prepared long before actual obligations are incurred; and by the time the estimate turns into a payment for services rendered, it often has grown from the original amount. Significant cost growth generally is viewed as an indicator of poor planning, mismanagement, underestimation of cost, and/or incomplete/inadequate identification of requirements. After all, during the investment analysis phase, the choice of the best alternative is made based on a comparison of cost estimates. If cost growth on one alternative is significantly higher than it might have been on another, the original choice could be called into question. Therefore, decision makers require a way of measuring the inherent risk and uncertainty in an estimate.

8.1.2 Common Terms and Definitions

A prerequisite to discussing risk and uncertainty analysis requires that certain terms and definitions be provided. Table 8.1 lists these terms and definitions.

Table 8.1 Risk and Uncertainty Analysis Definitions

TERM	DEFINITION
Risk	A situation in which the outcome is subject to an uncontrollable, random event stemming from a known probability distribution.
Uncertainty	Occurs in a situation in which the outcome is subject to an uncontrollable, random event stemming from an unknown probability distribution.
Engineering Change Orders (ECO)	That amount of money in a program specifically set-aside for uncertainty. ECO generally is referred to as the money set-aside for "known-unknowns."
Management Reserve (MR)	This term represents a value within the negotiated contract target cost that the contractor has withheld at the management level for uncertainties. The contractor is required to track and report to the government the application of MR. Generally, MR is referred to as the money set-aside for "unknown-unknowns."
Monte Carlo Analysis	Simulation technique, which varies all relevant input parameters to arrive at the potential range of outcomes expressed in terms of probability distributions.
Sensitivity Analysis	Estimating technique in which a relevant non-cost input parameter is varied to determine the probable cost.
Most Likely Cost	The most likely or most probable estimate of the cost that ultimately will be realized for a program, project, or task.
Standard Error of the Estimate	Represents a measure of the variation around the fitted line of regression, measured in units of the dependent variable.
Budgeting to Most Likely Cost	The most likely or most probable estimate of the cost that ultimately will be realized for a program, project, or task. Inherent in the estimate should be all funding necessary to ensure that the program can be managed properly in an environment of undefined technical complexity, schedule uncertainty, and the associated cost risk.
ECO Funding	ECO is the best estimate for contract changes, based on historical precedence (e.g., safety of flight, correction of deficiencies, and value engineering). ECO is a reserve for known-unknown contract changes and does not include reserves for requirements creep. ECO is an identifiable and traceable element of cost. ECO applies to both development and production and varies by both program and fiscal year.

8.2 Classical Treatments

This section covers several aspects of risk and uncertainty that set the stage for the later sections that deal more with approaches and actual practice. As a result, the focus here is to examine the nature of risk and uncertainty.

8.2.1 Risk versus Uncertainty

The terms risk and uncertainty often are used interchangeably. However, in the more strict definitions of statistics they have distinct meanings. Reviewing these definitions helps clarify the problem confronting the cost estimator. Three reports were consulted for the following definitions. They were Frank Husic's *Cost Uncertainty Analysis,* Paul Dieneman's *Estimating*

Cost Uncertainty Using Monte Carlo Techniques, and Gene Fisher's *Cost Considerations in Systems Analysis.*

The traditional view of risk is a situation in which the outcome is subject to an uncontrollable random event stemming from a known probability distribution, e.g., drawing an ace of spades. There is only one chance in 52. In drawing one card from the deck, the outcome is not known, but the probability associated with each outcome is known. The probability of drawing an ace of spades with replacement is 1/52, and the probability of drawing a spade with replacement is 13/52, etc.

Uncertainty is a situation in which the outcome is subject to an uncontrollable, random event stemming from an unknown probability distribution. That is, there is insufficient information available to form an objective view of the outcomes and their associated probabilities.

In most cost estimating situations, it is impossible to collect enough data to generate anything like a frequency distribution; in many cases five or six data points is a bonanza. The general conclusion is that cost estimating is much more in the realm of uncertainty than risk. Therefore, in the interest of both clarity and simplicity, the remainder of this chapter will use only the term uncertainty.

8.2.2 Elements of Uncertainty

The term cost growth seems to represent an inherent aspect of almost any government acquisition. It usually is measured by comparing the estimated cost of an item with its final actual cost. In this respect, cost growth is a monetary realization of the uncertainty that existed at the time the estimate was made. If all the events and circumstances that occurred between estimate and final cost were known at the time of the estimate and the source data/estimating techniques were sound, there would have been no uncertainty and, hence, no cost growth. The time element in the comparison of estimated to actual cost is critical; the earlier an estimate is made, the less is known about the item and the more opportunity there is for change. In terms of the acquisition cycle, the system cost estimate developed during the investment analysis phase has more uncertainty than the system estimate developed during the solution implementation phase. The reason for this is that at the beginning of a program there are several aspects for which only general statements can be made. As the program progresses, these aspects become clearer and more refined; as a result, uncertainty is reduced. That is, the known-unknowns are becoming known with the passage of time and experience.

This relationship of uncertainty and the acquisition cycle is shown in Figure 8.1. In the figure, the vertical lines represent the most likely cost (the point estimate). Panel A depicts the situation just described; all variables affecting the system are known but their magnitudes are originally uncertain. Again, as the program progresses, the measure or value of these variables becomes known, the uncertainty is reduced, and the probability of the point estimate increases. Panel B illustrates, basically, the same situation as A, except that the point estimate is increasing with time. The reason for the increase is due to factors and changes that could not be anticipated. These are the so-called unknown-unknowns. Past experience indicates that most programs resemble more panel B than panel A situations, which means the uncertainty surrounding an estimate is a composite of known-unknowns and unknown-unknowns.

Figure 8.1 System Cost Uncertainty and the Acquisition Cycle

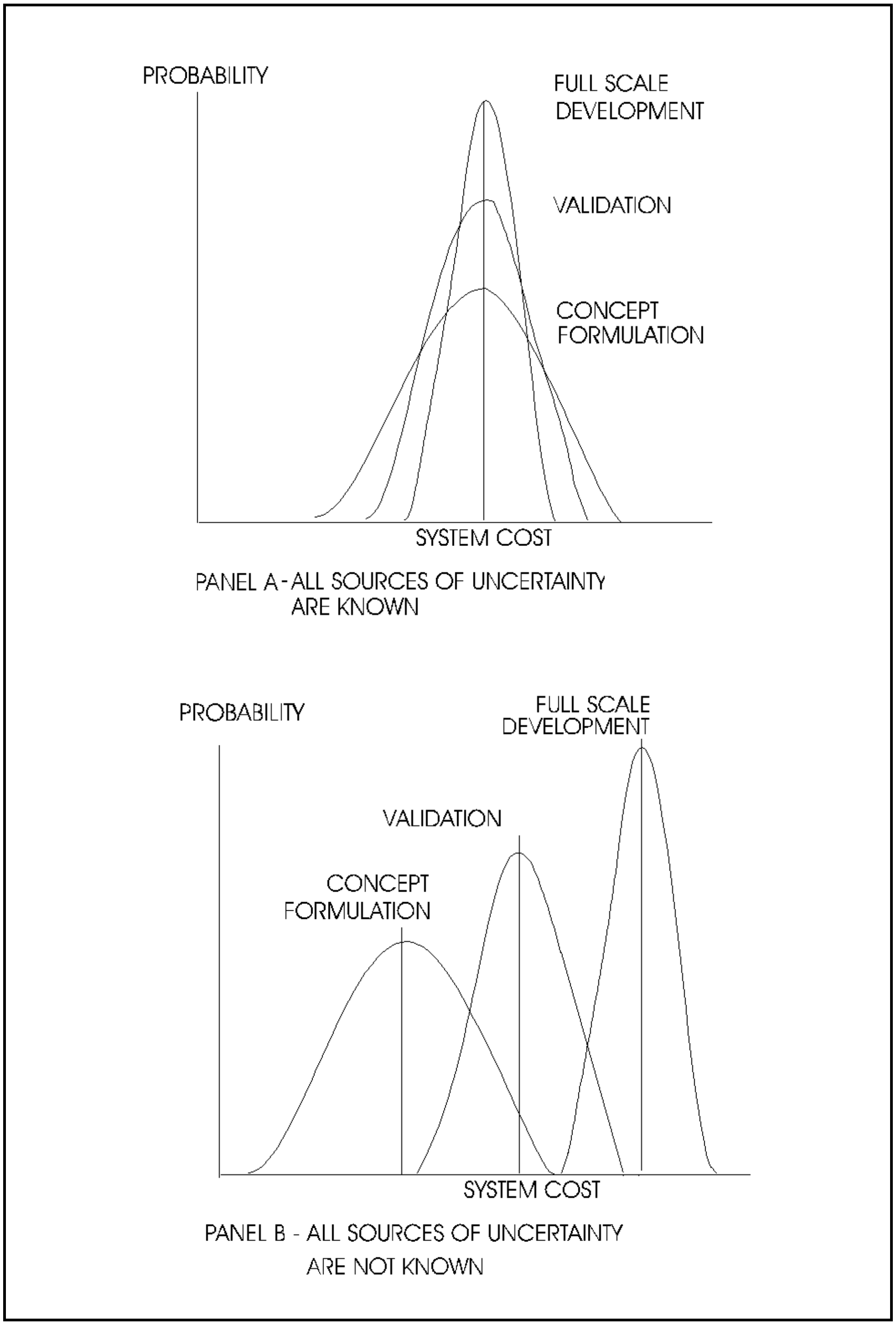


Table 8.2 provides a listing of one researcher's (John D. Hwang) findings about the economic, technical, and program factors causing these uncertainties. Generally, these sources of uncertainty are categorized as requirements uncertainty and cost estimating uncertainty.

Table 8-2. Sample List of Factors Causing Uncertainty

• Current/Future State of Technology	• Present Defense Systems Capabilities
• Defined Threat or Proposed Change/Innovation	• Production Facilities and Factory Test Equipment
• Desired Date for Operational Capability	• Production Hardware Including Necessary Spares
• End Item Interfaces Defined	• Quality Assurance and Test Requirements
• Equipment Schedules Delivery Dates	• Recommended Changes to System Design
• Estimated Production Rates/Quantities/Deliveries	• Reliability, Maintainability, Evaluation Criteria
• Expected Operational Environment	• Required Training Equipment and Facilities
• Field Requirements for Trained Personnel	• Subsystem Specifications
• Fiscal Information/Available Resources	• Support Facilities/Equipment on Hand
• Identified/Approved Engineering Design Changes	• System Operational/Functional Requirements
• Maintenance and Logistics Plans	• System Performance Demonstration Plans
• Material Sources and Market Prices	• System Performance/Design Requirements
• Mission Objectives and Priorities	• Test and Evaluation Concepts
• Mission Responsibility Assignment	• Test Facility, Support Equipment, Instrumentation
• National Objectives and Strategies	• Test Measurements, Data, Variables, Parameters
• Necessary Technology Advance and Risk Assessment	• Test Objectives, Environment, Expected Results
• Operational Plans Instructions and Manuals	• Tooling Design Jigs and Fixtures
• Performance Envelopes/Design Constraints	• Training and Personnel Requirements
• Personnel Subsystem Evaluation Plans	• Training Course Materials

Requirements uncertainty refers to the variation in cost estimates caused by changes in the general configuration or nature of an end item. This would include deviations or changes to specifications, hardware characteristics, program schedule, operational/deployment concepts, and support concepts.

Cost estimating uncertainty refers to variations in cost estimates when the configuration of an end item remains constant. The source of this uncertainty results from errors in historical data, cost estimating relationships, input parameter specification, analogies, extrapolation, or differences among analysts.

This form of categorization has been employed in the study of weapon system cost growth. According to Leroy Baseman's article in the *Journal of Cost Analysis*, in the 1960s and later in the early 1970s, requirements uncertainty accounted for about 75 percent of cost growth with the remaining 25 percent attributed to cost estimating uncertainty. By 1983, the percent attributed to cost estimating uncertainty had dropped to around five percent, and current information indicates the percentage will be even smaller in the future. Thus, cost growth today is not so much a matter of cost estimating error. Instead, it is a matter of how the end item originally estimated is different from the item finally produced due to changes in technology, national strategy, deployment concepts, operations procedures, or other end systems.

8.2.3 Point Estimates versus Interval Estimates

Development of a cost estimate usually involves the application of a variety of techniques to produce estimates of the individual elements of the item. The summation of these individual estimates becomes the singular, best (and most likely) estimate of the total system and is referred to as a point estimate. In and of itself, the point estimate provides no information about uncertainty other than it is the value judged more likely to occur than any other value. A confidence interval, on the other hand, provides a range within which the actual cost should fall, given the confidence level specified.

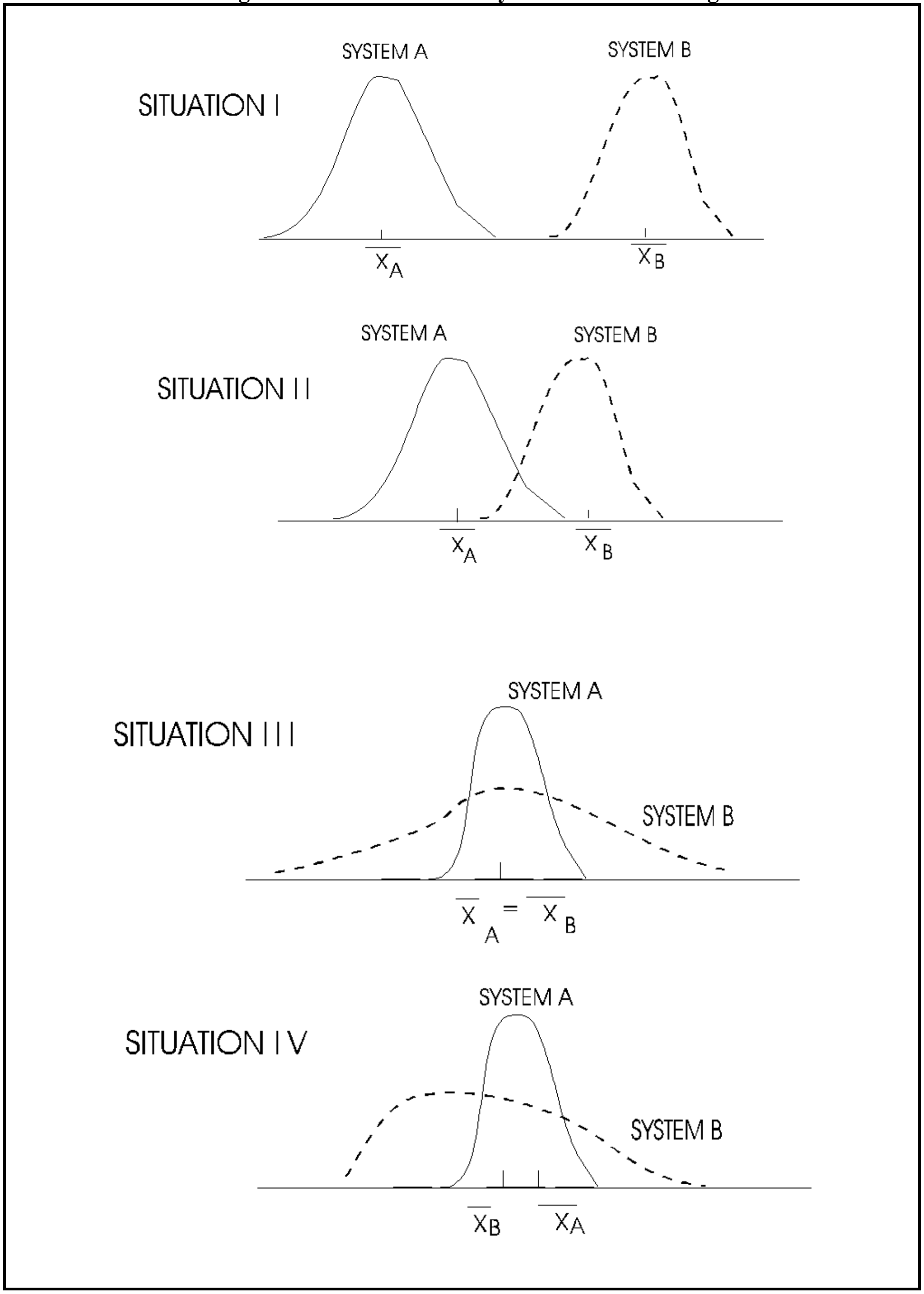
For example, suppose an estimating team has provided a point estimate for a system of $10M. Also, because of the way the estimate was built, the standard deviation has been estimated at $2.5M, and the distribution of cost is assumed to be normal. The interval estimate for the hypothetical system would be $5M to $15M, at the 95 percent level of confidence. This tells the manager that there is a 95 percent probability that the actual cost of the system will be between $5M and $15M, but the exact amount is unknown.

8.2.4 Uncertainty in Decision Making

The point estimate provides a best single value, but with no consideration of uncertainty. In contrast, the interval estimate provides significant information about the uncertainty but little about the single value itself. However, when both measures are taken together, they provide valuable information to the decision maker.

An example of the value of this information is in situations involving choice among alternatives, as in the case of source selection or systems analysis studies. For instance, suppose systems A and B are being evaluated; and because of equal technical merit, the choice will be made on the basis of estimated cost. According to Paul Dieneman, in his report *Estimating Cost Uncertainty Using Monte Carlo Techniques*, if the choice is made solely on the basis of the most probable cost, the decision may be a poor one (depending upon which of the four situations in Figure 8.2 applies.)

Figure 8.2 Cost Uncertainty In Decision Making



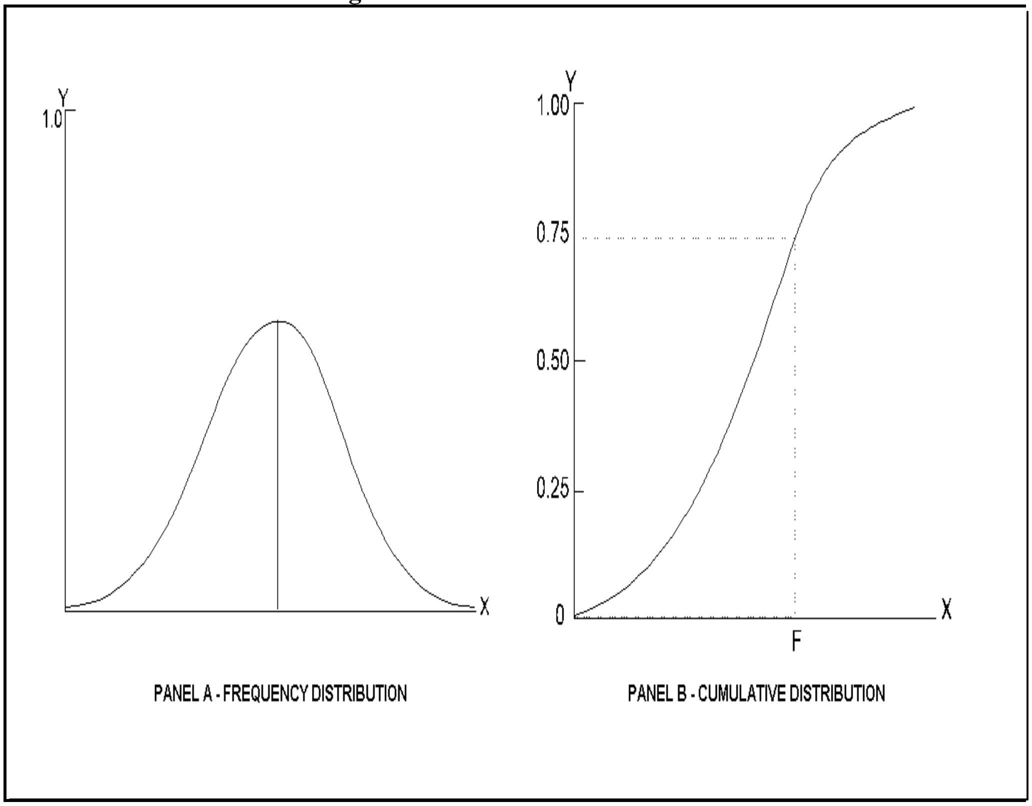
In situation I, there is no problem in the choice, since all possible costs for A are lower than B. A's most probable cost is the obvious choice. Situation II is not quite so clear because there is some chance of A's costs being higher than B's. If this chance is low, A's most probable cost is still the best choice. However, if the overlap is great, then the most probable cost is no longer a valid criterion. In situation III, both estimates are the same, but the uncertainty ranges are different. At this point, it is the decision maker's disposition toward risk that decides. If the preference is a willingness to risk possible high cost for the chance of obtaining a low cost system, then B is the choice. If the preference is to minimize risk, then A is the appropriate choice. Finally, situation IV poses a more complicated problem, since the most probable cost of B is lower but with much less certainty than A. If the manager uses only the point estimates in this case, the most probable choice would be the less desirable alternative. In the preceding situations, uncertainty information was a method used to select between alternatives. A quite different use of uncertainty information is when a point estimate must be adjusted for uncertainty, as in the case of establishing a budget.

8.2.5 Budget Realities

Establishing the funding level for a program or system is one of the primary purposes of developing an estimate. Unfortunately, the budgeting process is not designed to accommodate an interval estimate, which means that a single monetary value must be chosen. The program manager will, in most cases, not select the point estimate as the budget since it does not reflect any adjustments for uncertainty or circumstances beyond the realm of the cost estimate (such as affordability, availability of funds, the cost and relative priority of other systems/items competing for funds, and the manager's disposition toward taking a chance). Since it is likely that the choice will be somewhere between the point estimate and the upper level of a conservative interval estimate, the selection of a value suitable to external constraints and the cost uncertainty of the estimate becomes an obvious concern. Such a selection must be made by the manager, but the estimator can assist in the decision by providing uncertainty information for various budget values.

One particularly effective method of portraying the uncertainty implications of alternative choices is to depict the estimate and its related uncertainty in the form of a cumulative probability distribution, as shown in Figure 8.3. The utility of this approach is the easy-to-understand, convenient manner in which the information is presented to the decision maker. In the figure, panel A shows the cost estimate as it might normally be depicted with the most likely value (point estimate at the center); panel B shows the same information in the form of a cumulative curve. It is easy to see, for instance, that the selection of the funding level, F, is at the 75 point, which means that there is only a 25 percent chance of actual cost exceeding this funding level. The manager can see the implications of a particular choice immediately.

Figure 8.3 Distribution Forms



This completes the discussion of the nature and makeup of uncertainty. Before proceeding on to the next section, which covers the methods of dealing with uncertainty, there is a point that needs to be made. Rarely are there ever enough data available to generate a useable frequency distribution that could be employed like those in the examples used in this section. However, estimators do try to approximate such distributions through the use of some of the techniques discussed in the next section.

8.3 Dealing with Uncertainty

When actually treating uncertainty in an estimate, several approaches are available, ranging from very subjective judgment calls to rather complex statistical approaches. This section is not intended to be an exhaustive discussion of every possible approach or variation of an approach, but rather to provide an insight into the more fundamental and traditional techniques that form the basis for current field use. The order of presentation of these techniques is intentional, because it tends to portray the evolution that has taken place in terms of the tools used to handle uncertainty.

Before beginning actual discussions of the uncertainty approaches, there are a few points for an estimator to keep in mind. First, to the extent that actual historical cost information has been used in developing the point estimate, data already include the realities of both requirements and cost uncertainty. This leads to a natural question of why there is any need to treat uncertainty separately. The need appears to come from the view that a point estimate includes an inherent amount for expected uncertainty. However, there is a bias toward caution by adding an amount to the point estimate to cover uncertainties over and above what might be expected. Other than lacking the specific precision of statistics, this is not any different from adding some number of standard deviations to the mean to arrive at a higher specified level of confidence.

A second point to keep in mind is whether cost uncertainty or requirements uncertainty, or both, are to be treated. Several of the approaches discussed here require the estimator to provide a highest and lowest possible value. The point becomes one of knowing whether these values presume a fixed baseline and, therefore, only reflect cost uncertainty, or whether they reflect possible variations of the baseline itself. Whatever the case, it must be communicated clearly so that the decision-maker knows exactly what is included in, or excluded from, the estimate.

8.3.1 Subjective Estimator Judgment

This is perhaps one of the oldest methods of accounting for uncertainty and, in some respects, is the basis for most other approaches. Under this approach the estimator merely reflects upon the assumptions and judgments that were made during the development of the estimate. After evaluating all the "ingredients," a final adjustment is made to the estimate, usually as a percentage increase. This yields a revised total cost, which explicitly recognizes the existence of uncertainty. The logic to support this approach is that the estimator is more aware of the uncertainty in the estimate than anyone else, especially if the estimator is a veteran and has experience in systems or items similar to the one being estimated. One method for assisting estimators is to use a questionnaire, which provides a yardstick of their uncertainty beliefs when arriving at their judgment. The following questions, drawn from John D. Hwang's *Analysis of Risk for the Material Acquisition Process Part I: Fundamentals*, provide some examples:

- What cost is as likely to be greater than or less than the actual cost (this gives the median or 50 percent probability level)?

- What is the greatest imaginable cost of the project (this gives the 100 percent probability level)?

- What cost is just as likely to be above median as it is to be below the greatest amount (this gives the 75 percent probability level)?

- What cost is just as likely to be above the cost from the preceding statement as it is to be below the greatest amount (this generates the 87.5 percent probability level)?

This questionnaire device is equally applicable to a single cost estimator or team of estimators. Regardless of how subjective judgment is determined, there comes a time where the complexity and sophistication of the item is beyond the estimator's subjective assessment abilities. One method to overcome this is to use the expert judgment/executive jury technique discussed in the next paragraphs.

8.3.2 Expert Judgment/Executive Jury

A variant of estimator subjective judgment is a technique wherein an independent jury of experts is gathered to review, understand, and discuss the system and its costs. The specific objective from their collective deliberation is some measure of uncertainty that can be quantified into dollars and used to adjust the point estimate cost. The strengths of such an approach are related directly to the diversity, experience, and availability of the group members.

The use of such panels or juries requires careful planning, guidance, and control to ensure that the product of the group is objective and reflects the best, unmitigated efforts of each member. Approaches have been designed to contend with the group dynamics of such panels. One classical approach is the Delphi technique, which originally was suggested by RAND. With this technique, a panel of experts is drawn together to evaluate some particular subject and submit their answers anonymously. Next, a composite feedback of all answers is communicated to each panelist, and a second round begins. This process may be repeated a number of times, and ideally, convergence toward a single best solution takes place. By keeping the identities anonymous rather than in a committee session, the panelists can change their minds more easily after each round and provide better assessments, rather than defending their initial evaluation. The principle drawback of Delphi is that it is cumbersome, and the time elapsed in processing input may present some difficulty to respondents as to their reasons for the ratings. However, it is possible to automate the process with online computer terminals for automatic processing and immediate feedback. Other group dynamics schemes have been proposed as alternatives to Delphi; but, as with Delphi, there has been no definitive analysis of how well they work.

8.3.3 Sensitivity Analysis

Another common approach is to measure how sensitive the system cost is to variations in non-cost system parameters. For instance, if system weight is a critical issue, then weight would be varied over its relevant range, and the influence on cost could be observed. Analysis of this type helps to identify major sources of uncertainty. It also provides valuable information to the system designer in terms of highlighting elements that are cost sensitive, areas in which design research is needed to overcome cost obstacles to achieve better program performance, and areas in which system performance can be upgraded without increasing program cost substantially. The traditional criticism of this procedure is that it does not reveal the extent to which the estimated system cost might differ from the actual cost. That is, it tends to address uncertainty of requirements more than cost uncertainty.

8.3.4 High/Low Analysis

Another approach, which has been used to express cost uncertainty, requires the estimator to specify the lowest and highest possible values for each system element cost, in addition to its most likely value. These sets of input values are then summed to total system cost estimates. The most likely values establish the central tendency of the system cost, while the sums of the lowest possible values and highest possible values determine the uncertainty range for the cost estimate.

Although this approach has a logical appeal, it tends to greatly exaggerate the uncertainty of system cost estimates because it is unlikely that all system element costs will be at the lowest (or highest) values at the same time. While the high/low approach is plausible, its shortcoming is that it restricts measurement to three points, without consideration to intermediate values or their likelihood. The approaches described in the next paragraph provide solutions to this shortcoming.

8.3.5 Mathematical Approaches

If the individual cost elements can be regarded as random variables and their distributions can be determined, then the system cost can also be expressed as a probability distribution around an expected value. This is the basis for the approaches covered in this section. What these approaches do is to overlay the high/low approach with probability distributions for each cost element. Doing so requires the solution of two distinct problems. The first is how to determine the probability distribution for each cost element. The beta and triangular distributions are both described as solutions to this problem. The second is how to combine the individual cost elements and their measures of uncertainty into a total estimate of cost and uncertainty. The summation of moments and Monte Carlo simulation are described as solutions to this problem.

<u>The Beta Distribution</u>

This distribution is particularly useful in describing cost risk because it is finite, continuous, can easily accommodate a unimodal shape requirement ($\alpha > 0$, $\beta > 0$), and allows virtually any degree of kurtosis and skewness. Kurtosis characterizes the relative peakedness or flatness of a distribution as compared to the normal distribution. Skewness characterizes the degree of asymmetry of a distribution around its mean. S. Sobel, in *A Computerized Technique to Express Uncertainty in Advanced System Cost Estimates*, described a few of the many shapes of the Beta as shown in Figure 8.4. Per H. W. Darrwachter et al. and Gerald R. McNichols, the Generalized Beta Family of Distributions is defined over an interval (a, a+b) as in Equation 8.1.

Equation 8.1

$$f(x;\alpha,\beta,a,b) = \left(\frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)b} \right) \bullet \left(\frac{(x-a)}{b}^{\alpha} \right) \bullet \left(1 - \left(\frac{(x-a)}{b}^{\beta} \right) \right)$$

$$a \le x \le a+b$$

Where:

$a \le x \le a+b$ *defines an interval*

α, β *are the shape parameters of the Beta Distribution (values follow)*

Γ *is the Gamma Distribution (see Appendix 8-A for values to use)*

The following transformation is frequently used as in Equation 8.2.

Equation 8.2

$$f(x) = K \bullet \left(\frac{(H-L)^{\alpha} \quad (H-X)^{\beta}}{(H-L)^{(\alpha+\beta+1)}} \right)$$

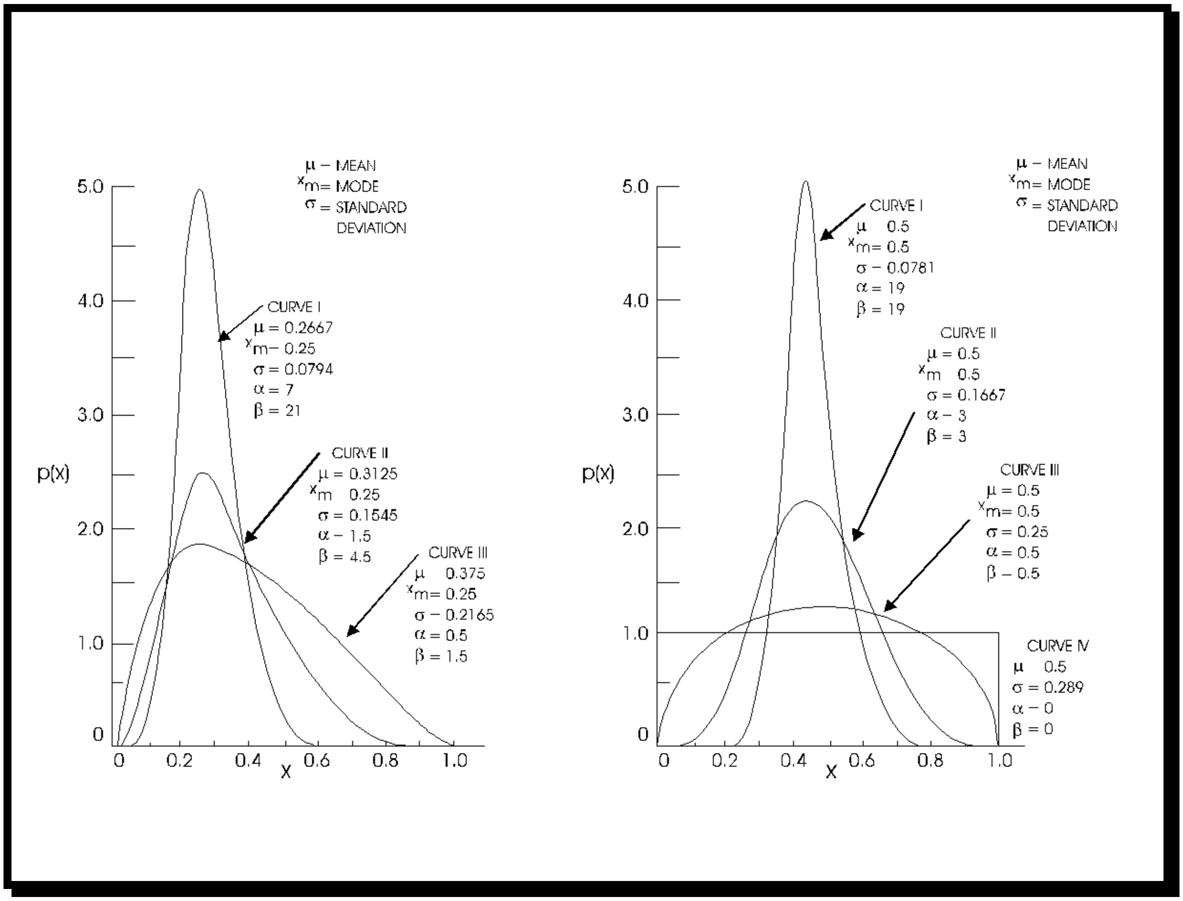
Where:

L = *Lowest value*

H = *Highest value*

$$K = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1) \bullet \Gamma(\beta+1)}$$

Figure 8.4 Beta Distribution Shape Examples

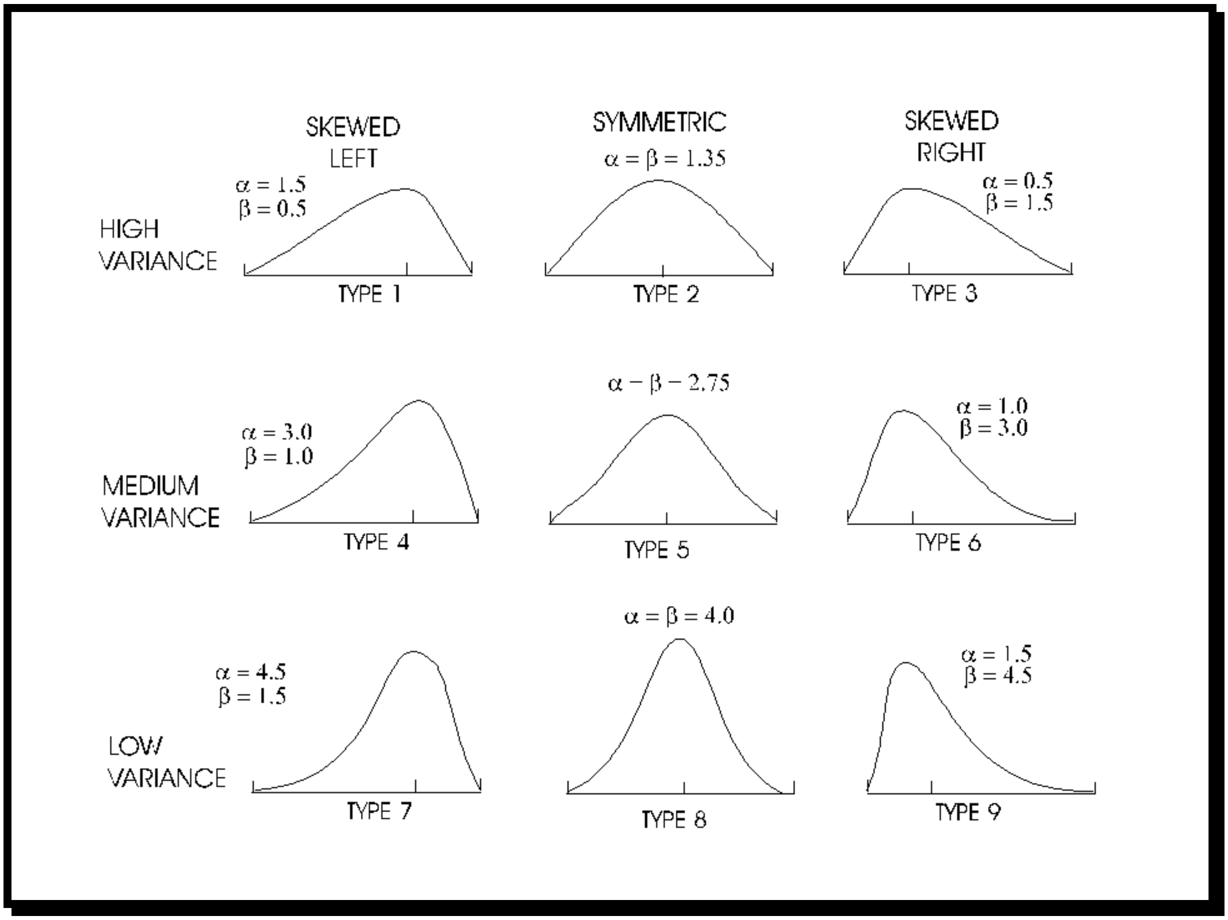


The values of α and β are the shape parameters, and each combination produces a unique shape. However, the process of deriving the appropriate values for a particular shape can be quite involved. Fortunately, a few observations about α and β lead to a rather useful approach in approximating the appropriate values. In the case of skewness, when α and β are equal the distribution is symmetric; when α>β the distribution is negatively skewed; and when α<β the distribution is positively skewed. Similarly, variance (kurtosis) can be categorized as high, medium, or low, based upon the magnitude of α and β. When these notions of skewness and kurtosis are combined, the result is nine combinations as shown in Table 8.3. These nine types tend to be fairly descriptive of most situations an estimator might confront. For that reason, Paul F. Dieneman translated them into the specific beta distributions shown in Figure 8.5.

Table 8.3 Beta Shape Combinations

Combination Type	Skewness	Kurtosis
1	Negative	High
2	Symmetric	High
3	Positive	High
4	Negative	Medium
5	Symmetric	Medium
6	Positive	Medium
7	Negative	Low
8	Symmetric	Low
9	Positive	Low

Figure 8.5 Beta Probability Distributions for Uncertainty Analysis



The advantage of the figure is that estimators can choose the distribution which best approximates their subjective view of the cost element uncertainty without having to derive α or β. Although the nine distributions do represent a rather restrictive set of options, the selection generally is considered sufficient from the standpoint that an estimator probably cannot distinguish among more variations accurately. These nine shapes have been adopted as a kind of standard by several researchers and practitioners. Also, it should be noted that these nine limit the location of the mode to the first, second, or third quartiles of the distribution range. The estimator should be conscious of these locations when specifying the high and low values relative to the most likely (point estimate) value. If these conditions are unsatisfactory, others can be developed by varying α and β (the ratio of α to β locates the mode within the range of the distribution). The value of this approach is that the estimator uses the point estimate as the most likely value and specifies a lowest possible value and highest possible value consistent with the distribution shape, based upon subjective judgment of variability.

At this point, the cost element can be described by its expected value and variance as shown in Equations 8.3 and 8.4.

Equation 8.3

$$\overline{X} = L + \frac{(H-L)(\alpha+1)}{(\alpha+\beta+2)}$$

In the case where the estimator specifies only the lowest and highest value and the chosen distribution, the most likely value (MO) can be calculated as in Equation 8.5:

Equation 8.1

$$S^2 = \frac{(H-L)^2(\alpha+1)(\beta+1)}{(\alpha+\beta+3)(\alpha+\beta+2)^2}$$

Equation 8.2

$$MO = \frac{\alpha(H) + \beta(L)}{(\alpha+\beta)}$$

Triangular Distribution

An alternative approach to assigning a distribution shape to a cost element is the triangular distribution. Like the Beta, it can take on virtually any combination of skewness and kurtosis, but the distribution represented by a triangle rather than the smoother curve of Beta, as shown in Figure 8.6. Albin D. Kazanowski wrote in *A Quantitative Methodology for Estimating Total System Cost Risk*, about the triangular distribution. The triangular distribution is specified by the lowest, most likely (usually the point estimate), and the highest value. Any point within the range of the distribution can be chosen to locate the mode and the relationship among the three values specifies the amount of kurtosis. Given the selection of the values and the triangular shape inherent to those values, both the mean and the variance can be calculated as in Equations 8.6 and 8.7.

Equation 8.3

$$\overline{X} = \frac{1}{3}(L + ML + H)$$

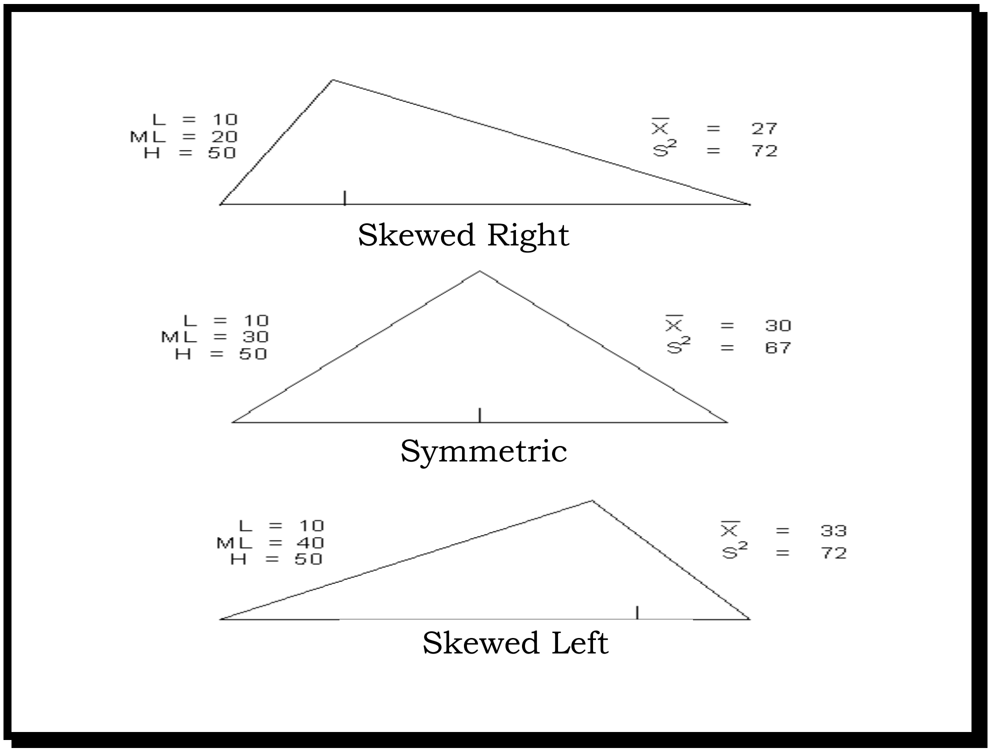
Equation 8.4

$$S2 = \frac{1}{18}[\,(H\text{-}L)2 + (ML\text{-}L)\ (ML\text{-}H)\,]$$

Where:

L = Lowest likely value
ML = Most likely value
H = Highest likely value

Figure 8.6 Triangular Distribution Examples



In contrast to Beta, the triangular distribution is much easier to use and produces equally satisfactory results. For this reason, the triangular generally is preferred over the more common Beta distribution.

Once distribution shapes have been identified for each cost element (or grouping of elements), the next step is to find the expected value and measure of uncertainty for the total system cost.

The Summation of Moments

This method takes the approach of measuring or describing a distribution through the use of moment statistics. According to Paul G. Hoel in *Introduction to Mathematical Statistics*, the first moment is the mean (x) and the second, third, and fourth moments (about the mean) take the form of Equation 8.8.

Equation 8.5

$$M_r = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^r$$

Where r is the rth moment around the mean.

As can be seen, the second moment is the variance. The third and fourth moments do not have any particular name, but they are used to calculate two measures that provide additional insight into the shape of a particular distribution. Those measures are: 1) the coefficient of skewness, which provides a measure of symmetry, and 2) the coefficient of kurtosis, which measures the peakedness or height of a distribution.

Coefficient of skewness = M3/(M2) 3/2

Coefficient of kurtosis = M/(M2) 2

The relevance of moment statistics to the development of a measure of total system cost uncertainty hinges upon one fact. That is, the moment measures for each cost element can be summed to produce the moment measures for the total system (or item) cost, when the variables (cost elements) are independent. (If, for some reason, independence among variables does not exist, then the covariance of the interdependent variables must be incorporated in estimating the moment of the sum.) For instance, the system mean is the sum of the individual element means; the variance (second moment) of the sum of independent variables is equal to the sum of the variances, etc. In fact, some authors use only the first and second moments to arrive at a measure of uncertainty. That is, with both the mean and variance of the total system cost determined through the summation process, the standard deviation is computed directly and the total cost portrayed as either a normal probability distribution or cumulative density distribution, as shown in Figure 8.3.

According to E. H. Yates, et. al. and Edward L. Murphy, Jr., the critical assumption in this approach is that even though the individual cost element distributions may not be normal, the total cost distribution will be. The basis for this normality assumption is both the central limit theorem and a sufficiently large number of cost elements (a minimum of thirty). This particular approach is shown in Figure 8.7. However, it is possible that if the variance of the distribution for an individual cost element is an order of magnitude greater than others, it may dominate the resulting aggregate distribution, which then may take on any of the non-normal characteristics of the dominant cost element. When this, or any other condition occurs which might jeopardize the central limit assumption, the approaches described in the next paragraphs offer possible solutions.

A more specific approach, advocated by several researchers and authors, is to take advantage of all four moments at the total system cost level by computing the mean, variance, and coefficients of skewness and kurtosis. These four measures can be analyzed then to determine the approximate distribution shape, without being limited to the central limit theorem and the normal assumption. One such method is to compare the characteristics of the estimated total system cost distribution with those of known distributions, such as shown in Table 8.4. According to J. J. Wilder, in *An Analytic Method for Cost Risk Analysis*, "If the correspondence is close enough (we leave that to the judgment of the analyst), we can conclude that the matching distribution is a good model of the unknown distribution, and use the appropriate density function for our calculations." There are other approaches to identifying the proper distribution, based upon moment statistics. However, they are beyond the scope of this chapter.

Figure 8.1 Summation of Moments--Central Limit Theorem Assumption

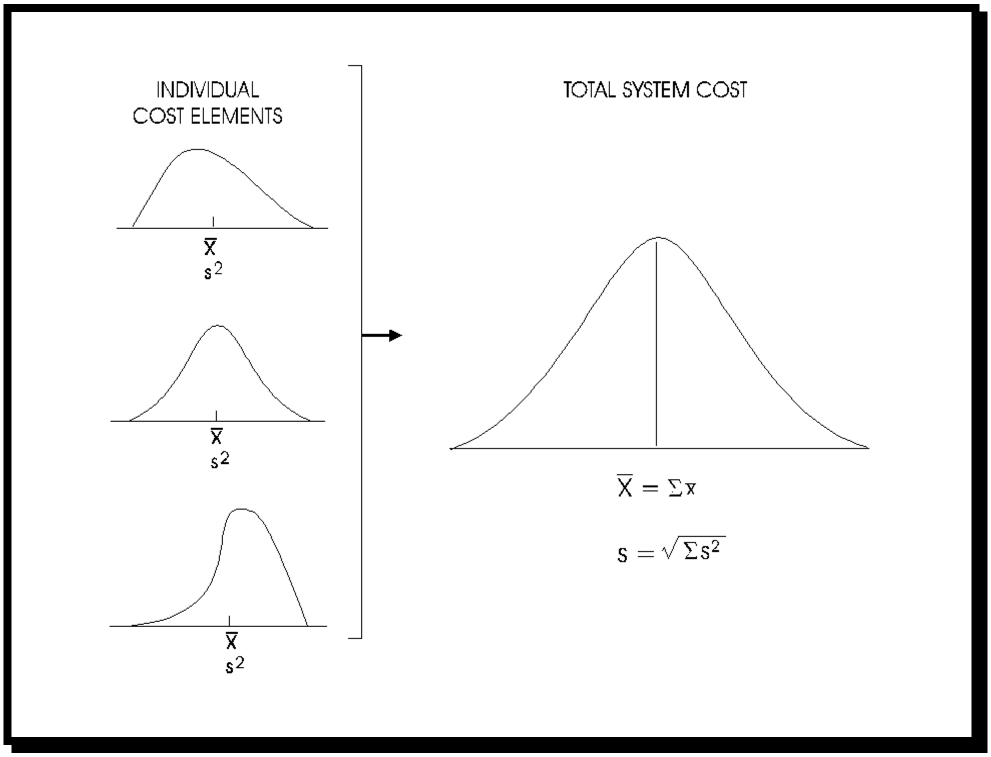


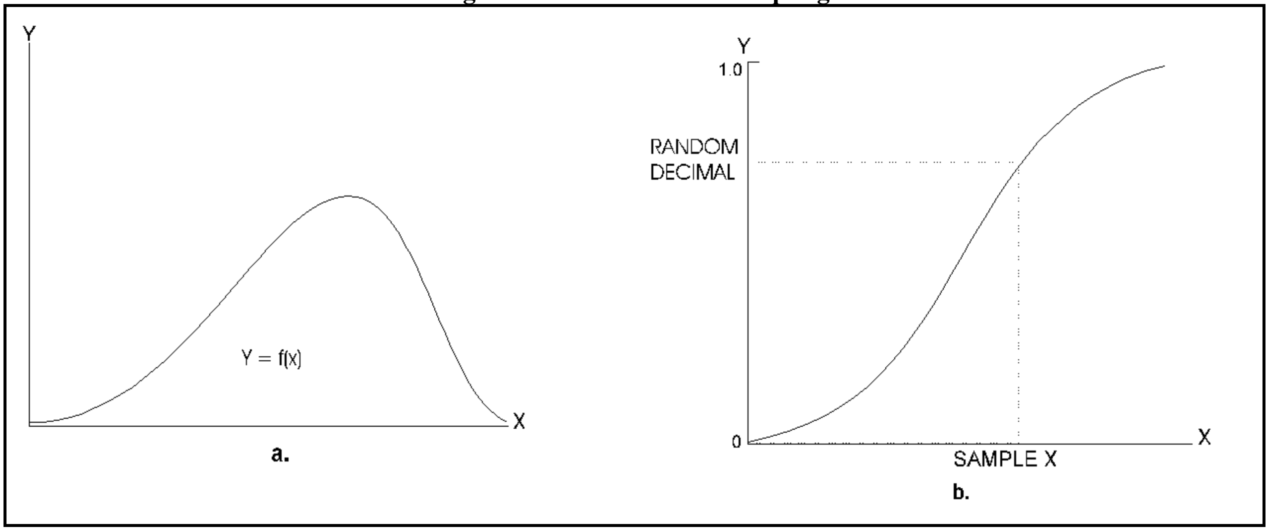
Table 8.4 Characteristics of Known Variables

Distribution	Skewness	Kurtosis
Uniform	0	1.8
Triangular	-.565 to +.565	2.4
Beta	Any Value	1.8
Normal	0	3

Monte Carlo Simulation

An alternative to the summation approach is to use the Monte Carlo Simulation Technique. With this approach, the distribution defined for each cost element (using beta, triangular, or an empirical distribution) is treated as a population from which several random samples are drawn. For example, a single cost element has been estimated and its uncertainty described as shown in A of Figure 8.8. From the probability density function, Y=f(X), a cumulative distribution is plotted, as shown in B of Figure 8.8. Next, a random decimal between zero and one is selected and located along the Y axis. By projecting horizontally from this random decimal location to the cumulative curve, the corresponding value of X can be determined. This value is considered as one sample of X for this specific cost element. A different random decimal is chosen for the next cost element and repeated until all cost elements have been sampled once. The sample values are summed to a total cost, and then the entire process is repeated again. This procedure is repeated several times (100-1000). The result is a normal distribution of random total costs that can be described by its mean and standard deviation and portrayed in the same manner as Figure 8.3 (Paul F. Dieneman, *Estimating Cost Uncertainty Using Monte Carlo Techniques*.)

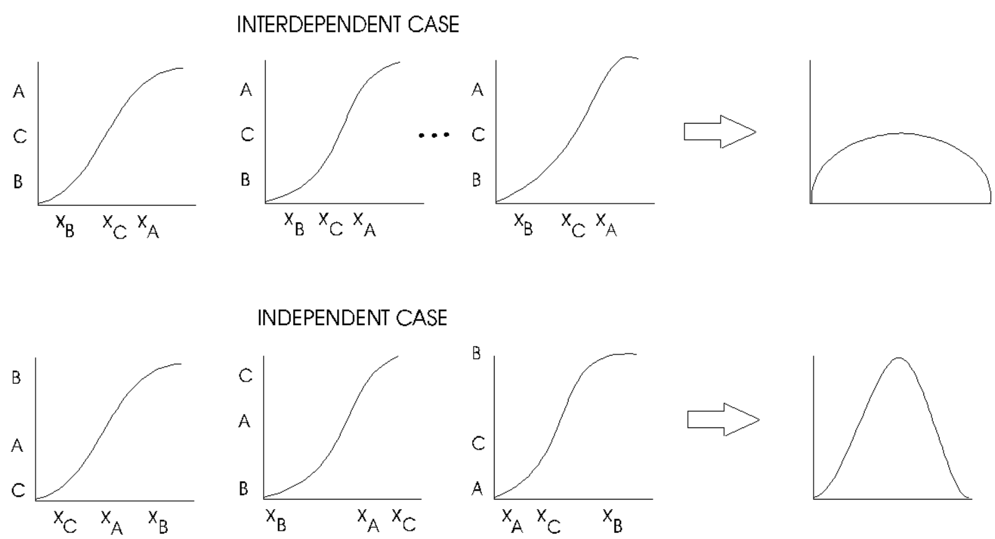
Figure 8.2 Monte Carlo Sampling



Again, the question of independence versus interdependence arises. The previous discussion of Monte Carlo assumed total independence. The opposite extreme is to assume total interdependence. A solution for this is to use the same random decimal for one pass through each of the cost elements. The sum of these observations is uniformly additive and results in a flatter, more rectangular distribution than in the independent case. The process for sampling in both the independent and interdependent cases, along with the resulting total cost distribution shapes, is illustrated in Figure 8.9.

Realistically, it is quite unlikely that a total system cost consists either of completely interdependent or independent cost elements. Nor does there appear to be a consensus on which assumption to make. One position holds that the only estimating errors meeting the criteria of randomness are cost uncertainties; and therefore, the assumption of independence is reasonable for cost uncertainty only. Interdependence appears to be more of a concern when cost and requirements uncertainties are considered jointly, or when requirements uncertainty is considered alone. That is, requirements variations tend to be viewed more like bias errors than the noise normally associated with randomness. (E. H. Yates et. al., *A Method for Deriving Confidence Estimates in Cost Analysis*)

Figure 8.3 Independent and Interdependent Assumptions for Monte Carlo Simulation



This concludes the discussion of the methods for dealing with uncertainty. The discussion was not intended to be exhaustive, but rather to provide an insight into the how and why of selected methods in prominent use. Section 8.4 discusses two commercially available software packages that can be employed for cost risk simulation.

8.4 Cost Uncertainty Models

@RISK Simulation Software

@RISK is an analysis and simulation direct add-in to standard spreadsheet programs (Microsoft EXCEL or Lotus 1-2-3) allowing the analysis of business and technical situations impacted by risk. The user replaces uncertain values in the spreadsheet with @RISK functions that represent a range of possible values, such as total profits or outputs. @RISK recalculates the spreadsheet hundreds or thousands of times, each time selecting random numbers from the @RISK functions entered. The result is a distribution of possible outcomes and the probabilities of getting those results.

@RISK employs both the Monte Carlo and Latin Hypercube simulation techniques to combine all the uncertainties identified in a system model. Risk analysis in @RISK is a quantitative method that seeks to determine the outcomes of a decision as a probability distribution. Thirty-seven different types of distributions are supported including: tbeta, binomial, chi square, Pareto, triangular, and Weibull. Up to 32,000 user-defined iterations per simulation are possible.

High-resolution graphics are used to present the output distributions from the @RISK simulations. Histograms, cumulative curves, summary graphs for cell ranges, zooming, and graphic overlays are all supported.

@RISK is available through Palisade Corporation of Newfield, New York, (607) 564-9993 or www.palisade.com. A demonstration and tutorial is available for download.

Crystal Ball Simulation Software

Crystal Ball is a fully integrated add-in program for Lotus 1-2-3 for Windows, Microsoft EXCEL for Windows or Microsoft EXCEL for Macintosh. Crystal Ball works with information the user provides about the uncertain inputs to the spreadsheet model. These assumptions are the cells that would be modified in a manual "what-if" analysis. For each assumption, a range of possible values (or a probability curve) is defined that reflect what is known about that value. There are 16 pre-defined curves, as well as a custom distribution capability that allows the user to assign probability distribution functions (PDFs) to cells in the spreadsheet. With a Graphical User Interface, Crystal Ball gives users the capability to perform risk analysis based on Monte Carlo simulations or Latin Hypercube sampling. As such, Crystal Ball has many similarities to the @RISK product discussed above. The Crystal Ball analyses are summarized in a graph showing the probability for each result. The capability to produce customizable charts, trend charts, overlay charts, and sensitivity charts is also provided to aid further analysis of the data.

Crystal Ball is available through Decisioneering, Inc., of Denver, Colorado, (800) 289-2550 or www.decisioneering.com. A free evaluation copy is available for download.

This concludes the discussion of uncertainty models. Again, the use of any model requires a clear definition of what uncertainty is to be treated and how the specific model satisfies the requirement.

8.5 Qualitative Indices of Uncertainty

Up to this point, the methods of treating uncertainty have all resulted in a quantitative adjustment or refinement of the point estimate. However, the use of qualitative indices has been proposed as a method of communicating a cost estimate's goodness, accuracy, or believability. Most of the indices are based upon the quality of the data and the quality of the estimating methodology.

For instance, John D. Hwang proposed the rating scheme using a two-digit code with ratings of 1 to 5 for data and for methodology, with a 1 representing highest quality and a 5 representing lowest quality. Thus, a rating of 1,1 would reflect that the estimate was the result of the highest quality level for both. The complete scoring system is shown in Table 8.5.

Table 8.5 Two Digit Confidence Index

Rating	Methods	Rating	Data
1	The basic method used to perform this analysis is exceptionally well documented and time tested; one or more other techniques have been used to verify the estimate provided.	1	Very complete, well-authenticated, highly relevant data, such as recent contractor actual costs, official catalog prices, etc. have been used.
2	The basic method used to perform this analysis is well documented, but no double-check or authentication has been possible.	2	The data used generally are relevant and from a reputable source; however, they are incomplete, preliminary, or not completely current.
3	The basic method used to perform this analysis has been documented, but has not been widely used or approved.	3	The data used have been obtained from official or standard sources; however, notable inconsistencies, lack of currency, or gaps in data reduce the confidence in the estimate.
4	A highly arbitrary method of analysis has been used.	4	The data used to make the estimate are highly suspect, of doubtful relevance, very sparse in quantity, and characterized by major inconsistencies.
5	The analysis is almost pure guesswork, and little or no confidence can be placed in it.	5	An almost total lack of current, reliable, relevant data makes the cost estimate completely uncertain.

The value of such qualitative indices appears to be their use as a sort of broad gauge for the manager to use in understanding the makeup of uncertainty. That is, such a qualitative index could be used to get a feel for what portion of the uncertainty is related to cost and what portion to data.

8.6 Summary

This chapter examined the terms, concepts, and approaches involved in analyzing cost risk and uncertainty. It should be clear from the chapter discussion that a consideration of risk and uncertainty is an integral part of the estimating process. For a more detailed discussion of cost and uncertainty analysis, read *Improving Cost Risk Analyses* by Fred Biery, David Hudak, and Shishu Gupta. This article can be found in the Spring 1994 edition of Society of Cost Estimating & Analysis' *Journal of Cost Analysis*.

8A. GAMMA FUNCTION TABLE

The values of the Gamma function used in the Beta distribution can be found using the following table. Note that (n+1) is equal to n (n), which allows the determination of gamma values greater than those contained in the table. For integer values, the gamma value can be found in (n)=(n-1)!

and for half integers by (m)=(m-1)! (π) -- $\frac{5}{2} \bullet \frac{3}{2} \bullet \frac{1}{2} \bullet \pi$.

Values of Γ (n) = $\int_0^\infty$ e-zxn-1dx; Γ (n+1) = n Γ (n)

n	**Γ (n)**	**n**	**Γ (n)**		**Γ (n)**	**n**	**Γ (n)**
1.00	1.00000	1.25	.90640	1.50	.88623	1.75	.91906
1.01	.99433	1.26	.90440	1.51	.88659	1.76	.92137
1.02	.98884	1.27	.90250	1.52	.88704	1.77	.92376
1.03	.98355	1.28	.90072	1.53	.88757	1.78	.92623
1.04	.97844	1.29	.89904	1.54	.88818	1.79	.92877
1.05	.97350	1.30	.89747	1.55	.88887	1.80	.93138
1.06	.96874	1.31	.89600	1.56	.88964	1.81	.93408
1.07	.96415	1.32	.89464	1.57	.89049	1.82	.93685
1.08	.95973	1.33	.89338	1.58	.89142	1.83	.93969
1.09	.95546	1.34	.89222	1.59	.89243	1.84	.94261
1.10	.95135	1.35	.89115	1.60	.89352	1.85	.94561
1.11	.94740	1.36	.89018	1.61	.89468	1.86	.94869
1.12	.94359	1.37	.88931	1.62	.89592	1.87	.95184
1.13	.93993	1.38	.88854	1.63	.89724	1.88	.95507
1.14	.93642	1.39	.88785	1.64	.89864	1.89	.95838
1.15	.93304	1.40	.88726	1.65	.90012	1.90	.96177
1.16	.92980	1.41	.88676	1.66	.90167	1.91	.96523
1.17	.92670	1.42	.88636	1.67	.90330	1.92	.96877
1.18	.92373	1.43	.88604	1.68	.90500	1.93	.97240
1.19	.92089	1.44	.88581	1.69	.90678	1.94	.97610
1.20	.91817	1.45	.88566	1.70	.90864	1.95	.97988
1.21	.91558	1.46	.88560	1.71	.91057	1.96	.98374
1.22	.91311	1.47	.88563	1.72	.91258	1.97	.98768
1.23	.91075	1.48	.88575	1.73	.91466	1.98	.99171
1.24	.90852	1.49	.88595	1.74	.91683	1.99	.99581
						2.00	1.00000

NOTE:
For large positive values of x, Γ (x) approximates Stirling's asymptotic series

$$\text{xze-z}\sqrt{\frac{2\pi}{x}}\left[1+\frac{1}{12x}+1/288x^2-\frac{139}{51{,}840}\text{x3}-\frac{571}{2{,}488{,}320}\text{x4}+\ldots\right]$$