Risk Modeling & Measurement in Construction

Developing a reasonably reliable way of determining best & worst case scenarios for construction project budgets & schedules is key to healthy & competitive project management.

ALI TOURAN

Keeping projects on time and within budget are two of the most important functions of construction project management. Estimates of project cost and duration are based on the knowledge of the estimators and schedulers, experience and data from similar projects completed previously, and a large number of assumptions made regarding productivity rates and material prices.

Almost every project component that consumes time and/or money is prone to some chance variation. Some items such as material prices, when a vendor has guaranteed prices, have a lower chance of variability. Other items, such as various labor productivity rates that can be sensitive to many factors (for example: weather, temperature, state of economy, unions involved and location), have a much higher chance of variation and can impact the project duration and cost. Risk measurement and analysis is one process of developing a logical vehicle for predicting the extent of these variations and possibly forecasting both worst and best case scenarios for the project budget and schedule.

Almost every party involved in the project needs to perform its own kind of risk analysis. The owner, public or private, needs to assess the amount of uncertainty in the project cost and schedule in order to make plans for seeking project funding. Multi-year megaprojects are very sensitive to variations in project durations. The cost of money needed to finance these projects becomes prohibitively high as the project duration increases. The contractor needs to assess the risk and uncertainty so that a reasonable contingency for the project can be included in the bid, especially in competitive lumpsum contracts.

Careful evaluation of this contingency is important. A low estimate of the required contingency may get the contractor the job, but costs may increase after the project starts as time and cost variations develop an unfavorable impact on the project. A high or conservative estimate of contingency will put the contractor at a disadvantage because the bid may not be competitive enough to get the job. So, depending on who is interested in risk analysis, the objective may be different but the general approach is the

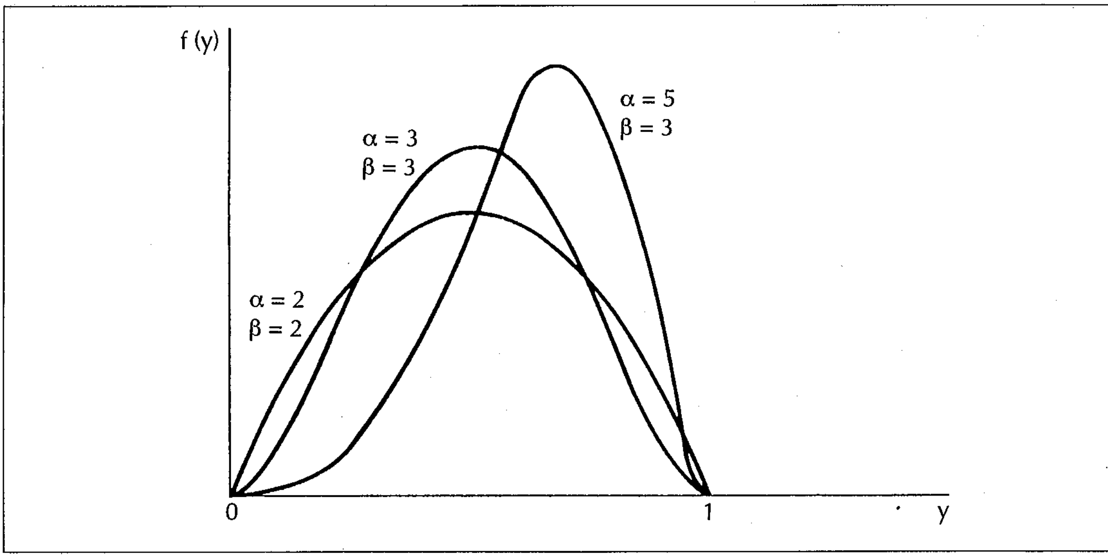

FIGURE 1. Graphs of beta distributions.

same — *i.e.,* to identify areas that are prone to uncertainty and to develop a model that can predict the combined effect of these areas on the project's budget and schedule.

Objectives

The more common methodologies that are available for modeling uncertainties in project cost and schedule are discussed here, and guidelines for their use are also provided. Various techniques that can be used for risk assessment are described and examples are provided without looking at strategies for identifying and minimizing risk. The Construction Industry Institute and Diekmann *et al.* provide more detail on the process of identifying sources of risk in a project.[1,2] Furthermore, only probabilistic approaches to evaluation of variations of project components are described here. Approaches that are based on specifying some deterministic safety margin for critical items based on the expertise of the seasoned personnel are not considered here, although these approaches, in many cases, tend to work better than the more sophisticated probabilistic analysis.

General Approach

The general approach in assessing risk (or the probability of cost or schedule overrun/underrun) is to treat various components of the project, especially those components that are expected to vary greatly, as random variables. The underlying assumptions in both probabilistic scheduling and estimating are very similar. In almost every case, a model is developed for predicting the project cost or schedule. Since this model is a function of several random variables (those components of cost or schedule that have a fair chance of variation and that are expected to contribute to the total project uncertainty), it is itself a random variable.

If the distribution of the random variable that is used to model total project cost or total project duration can be estimated, then the probabilities associated with various levels of confidence regarding meeting a specific deadline or a prescribed budget level can be computed. The problem is that, in many cases, it would be very difficult (if at all possible) to analytically find the distribution of the random variable representing the total project cost or schedule. That is why, in many cases, a simulation analysis is conducted to arrive at the cumulative distribution function (CDF) of the total cost or schedule.

The following factors may affect the analysis outcome:

- The choice of statistical distributions and parameters used to model individual project components
- The choice of the mathematical model for

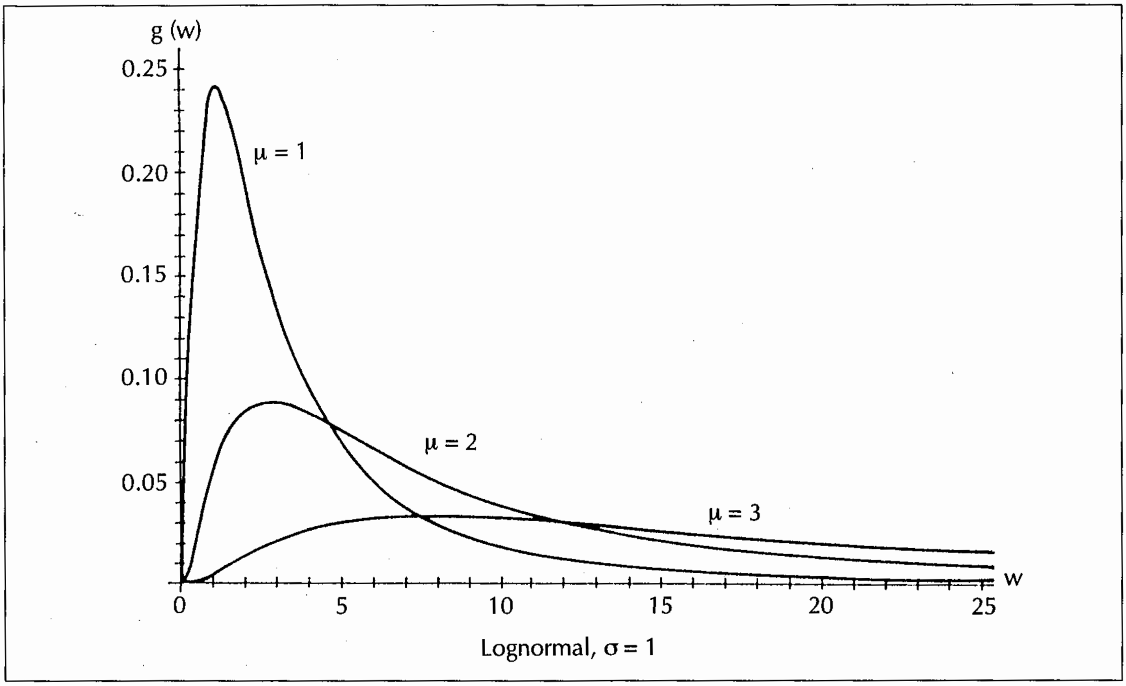

FIGURE 2. Graphs of lognormal distributions.

the total project cost or schedule

- The choice of analytical technique used to solve the predictive model

Underlying Statistical Distributions

As mentioned earlier, the general approach in assessing uncertainty in construction projects is to treat the project components with a high potential for variability as random variables. Therefore, an activity's duration that traditionally is estimated with a single number, or a unit cost item that the estimator usually estimates based on the information available deterministically, is modeled as a set of random variables with specified means and variances.

In most cases, the specification of a distribution type is also needed in order to be able to conduct a probabilistic analysis. Almost always, a well-known theoretical statistical distribution is used to model the item's variability. Statistical distributions are employed because they are well-known, usually fully documented and, therefore, easier to work with and to evaluate. Given the variety of statistical distributions available, there is adequate choice in selecting a reasonable distribution for modeling a certain parameter's variability.

In the past three decades, research has been conducted on the nature of construction cost and duration distributions. Several features of cost and duration distributions have been identified. For example, it is understood that the distribution should preferably have confined limits, should only take positive values in the ranges of interest, should be unimodal and may be skewed (unsymmetrical).[3] For example, the developers of the Program Evaluation and Review Technique (PERT), a probabilistic network-based scheduling technique, have suggested using a beta distribution to model activity duration times.[4] Beta is a unimodal distribution with confined lower and upper bounds and can take several shapes depending on the distribution's shape factors (see Figure 1). It provides a flexible means for modeling activity duration times. PERT has been in use since the late 1950s.

Teicholz found out that the cycle times of construction equipment (*e.g.*, scrapers) follow a lognormal distribution.[5] The observations of O'Shea *et al.* and Gaarslev later supported this conclusion.[6,7] Lognormal is a unimodal distribution that can take only positive values, and is skewed to the right (see Figure 2).

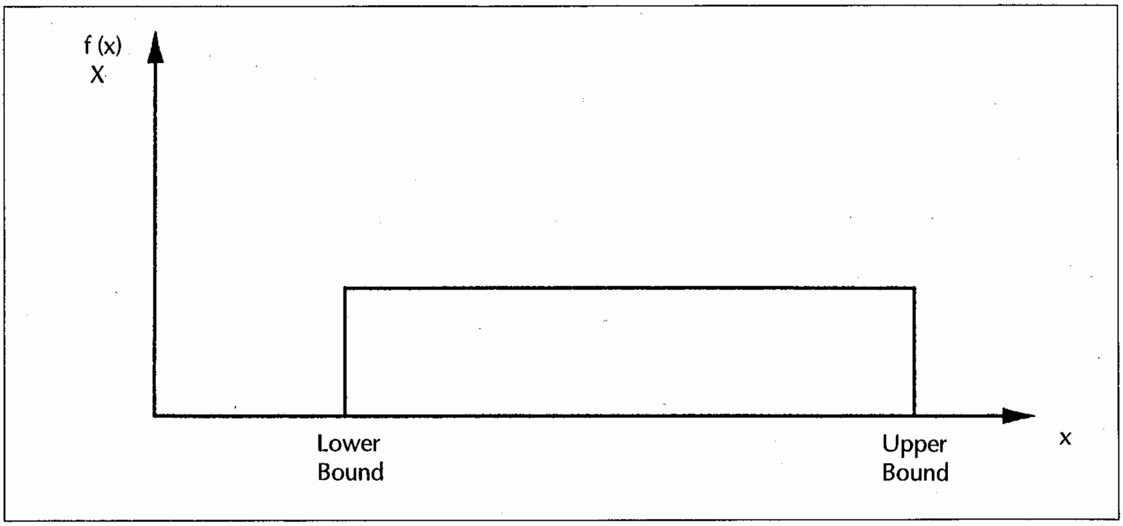

FIGURE 3. Uniform distribution density function.

More recent studies found that the cost items (such as overhead, concrete, electrical, mechanical, *etc.*) in low-rise office buildings (two to four stories) are lognormally distributed.[8,9] Other researchers have considered uniform (see Figure 3) and triangular (see Figure 4) distributions for modeling cost or duration.[10]

General Guidelines for the Selection of Distribution. A broad set of guidelines can be established for specifying distributions based on the characteristics of each distribution.

If the amount of data regarding a component is very limited, or if the component is expected to vary within a very narrow range, then a uniform distribution can be used since there is no preference regarding the most likely value of the distribution. An advantage of uniform distribution is its simplicity and its ease of visualization.

If the range is appreciable and some data are available regarding the most likely value of the distribution, then a triangular distribution may be advantageous. For example, if the estimator feels that the unit labor cost of concrete formwork is \$1.75/square foot (sf), but may vary between \$1.25/sf and \$2.05/sf, then a triangular distribution with a minimum value of 1.25, a maximum value of 2.05 and the most likely value of 1.75 may be a proper choice.

If the estimator thinks, on the other hand, that the same unit cost varies between \$1.5 and \$1.75, then a uniform distribution with a minimum value of 1.50 and a maximum value of 1.75 may be used. This approach would mean that it is equally likely that the unit cost of formwork takes any value between \$1.50 and \$1.75/sf.

Both beta and lognormal distributions resemble the triangular distribution in the sense that the data are grouped around a mode and the distribution is not necessarily symmetrical. In fact, in PERT scheduling, the scheduler defines a beta distribution for each activity duration by specifying a lower bound, an upper bound and a most likely value.

The use of an empirical distribution to model a random component is another approach that is sometimes taken. In this case, a histogram of data collected previously on the component is used to model the component's variation. The use of empirical distributions generally requires a computer simulation for arriving at the function that represents the total cost or schedule.

Project Schedule

PERT Approach. The most common approach in probabilistic scheduling is PERT, where every activity is modeled as a random variable distributed according to a beta distribution. The total project duration is computed along the network's critical path (the longest path) by adding the means of the activities on the critical path. According to the Central Limit Theorem

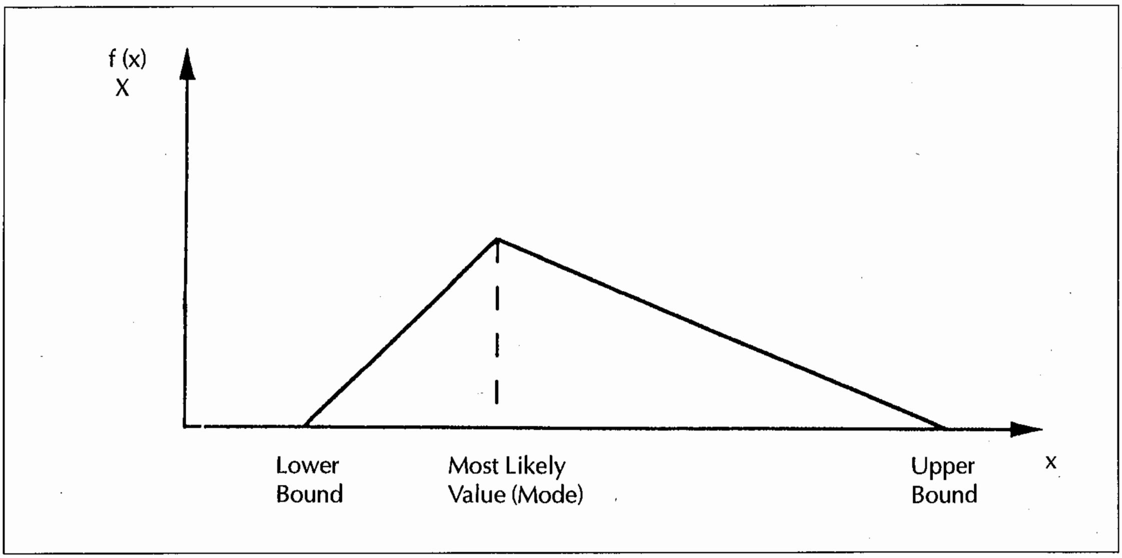

FIGURE 4. Triangular distribution density function.

(CLT), the sum of several independent and identical random variables is a random variable with an approximately normal distribution. The mean of this normal random variable is the sum of the means of the individual random variables and the variance of the total is the sum of the variances of the individual random variables. In this way, the total project duration is modeled as a normal distribution and its parameters can be conveniently estimated from the activity data. If activity durations are not independent, then the use of the CLT is not theoretically justified. Moder *et al.* provide a more detailed explanation of the application of PERT.[11]

The CLT can be used if the number of activities that contribute to the total project duration (*i.e.*, the activities on the critical path) is relatively "large." Although some statisticians have suggested that the number of random variables should be larger than 30 (for example, Devore[12]), experience indicates that with numbers larger than ten, reasonable approximations to normal distribution can be expected.[13]

The other concern in applying the CLT to PERT is that, in some cases, several paths in the project are almost as long as the critical path. In these cases, it is possible that the shorter paths that happen to have larger variances than the critical path will become critical. In such cases, the key questions become:

- To what path should the CLT be applied?
- Which path is actually the longest?

One suggested solution has been to use the Monte Carlo simulation in analyzing these cases. This issue has been discussed under "merge event bias problem" in various publications.[11]

Monte Carlo Simulation Technique. In the Monte Carlo simulation approach in scheduling, a random number is generated on a computer to generate a duration for each activity using its distribution. These numbers are used to schedule the network and then the total project duration is computed. In this process, the activities on the critical path (the sequence of activities with the longest total duration) are identified. This process of generating random numbers according to various activity distributions is repeated many times (from several hundred times to several thousand times) and every time the critical activities are identified. Then, a criticality index is computed for each activity that reflects the probability that a certain activity can become critical. This criticality index is simply the ratio of the number of times an activity was on the critical path to the total number of simulation runs. In this way, the activities with a high probability of becoming critical are identified. This method can help management allocate a proper level of attention to these components of the project.

Many factors affect the choice of methodology in risk analysis. Two examples are presented in order to illustrate some of these concerns.

Example 1. Figure 5 shows a network schedule proposed for a major tunneling project that was employing several tunnel boring machines (TBMs). As can be seen, the network is developed at a macro level, without getting into much detail.[14] The project was the proposed construction of an 83-kilometer (km) long elliptical ring for the main collider of the Colorado Superconducting Super Collider project. It was suggested that ten TBMs be used to bore ten 8.3-km tunnel segments, each of which having its own access shaft. It was also assumed that the shafts would be divided into two groups of five. Five shafts would be constructed and then the shaft sinking crew would move to a new location for sinking five other shafts. This scheme introduced five shaft location-to-location moves. Each tunnel boring activity could start after its shaft was constructed, and the TBM was delivered and assembled. The total project was divided into 55 activities.

A probabilistic analysis was desirable because it was important to evaluate the probabilities associated with the various durations for the project. The PERT approach could not be used because the longest path in the network consisted of only four activities. It was not sure that the CLT could be used with as few as four activities when the activity distributions are not necessarily normal. Therefore, the shape of the network can impact the method of analysis. Generally, short networks with many parallel paths are not suitable for PERT analysis because the assumption of normality for the total duration will be questionable and the critical path may change as various parallel paths may become critical.

A Monte Carlo simulation was conducted on the network depicted in Figure 5 by specifying statistical distributions for various activities in the network and by generating random numbers according to these distributions. The underlying distributions were specified by collecting data on similar projects and interviewing tunneling experts.[15-17] Based on these efforts, several types of distributions were assumed for the activities.

A simulation model was developed and the model was run for 400 iterations[11] using an available simulation software package.[18] At every iteration, the computer generated duration times for each of the 55 activities in the network, added the durations of the serial activities, compared the length of various paths and saved the duration of the longest path. In this way, for every iteration, a total project duration was computed. These iterations were then used to compute the CDF and probability density function (PDF) of the total project duration (see Figure 6).

The results show that if a combination of triangular and lognormal distributions are assumed for activity duration times (the third combination in Figure 6), then there is a 50 percent chance that the project will be completed in about 180 weeks. The project duration, if a confidence level of 80 percent is sought, can be estimated at 250 weeks.

Now that a model is developed for the total project schedule, it is possible to perform various kinds of sensitivity analyses and investigate the impact of individual activity durations on the total project schedule. The reason for using various statistical distributions for modeling activity durations was that, given the data available, it was difficult to prefer lognormal distribution to the normal distribution for tunnel boring and vice versa. By using a combination of reasonable distributions, a family of CDFs were generated that defined an envelope for the CDF of the total project duration. There is some inconvenience in using the model because there is not a direct mathematical solution. In addition, for any further analysis, the software and the model developed to study the impact of various changes on the project schedule would have to be examined.

This example was used only to illustrate the use of Monte Carlo technique in network-based scheduling. It is possible that other more effective construction plans could be conceived for the project described. Modelers simply built a simulation model based on a conceptual plan developed by a task force. A Monte Carlo model could then be developed to reflect the new construction plan and schedule.

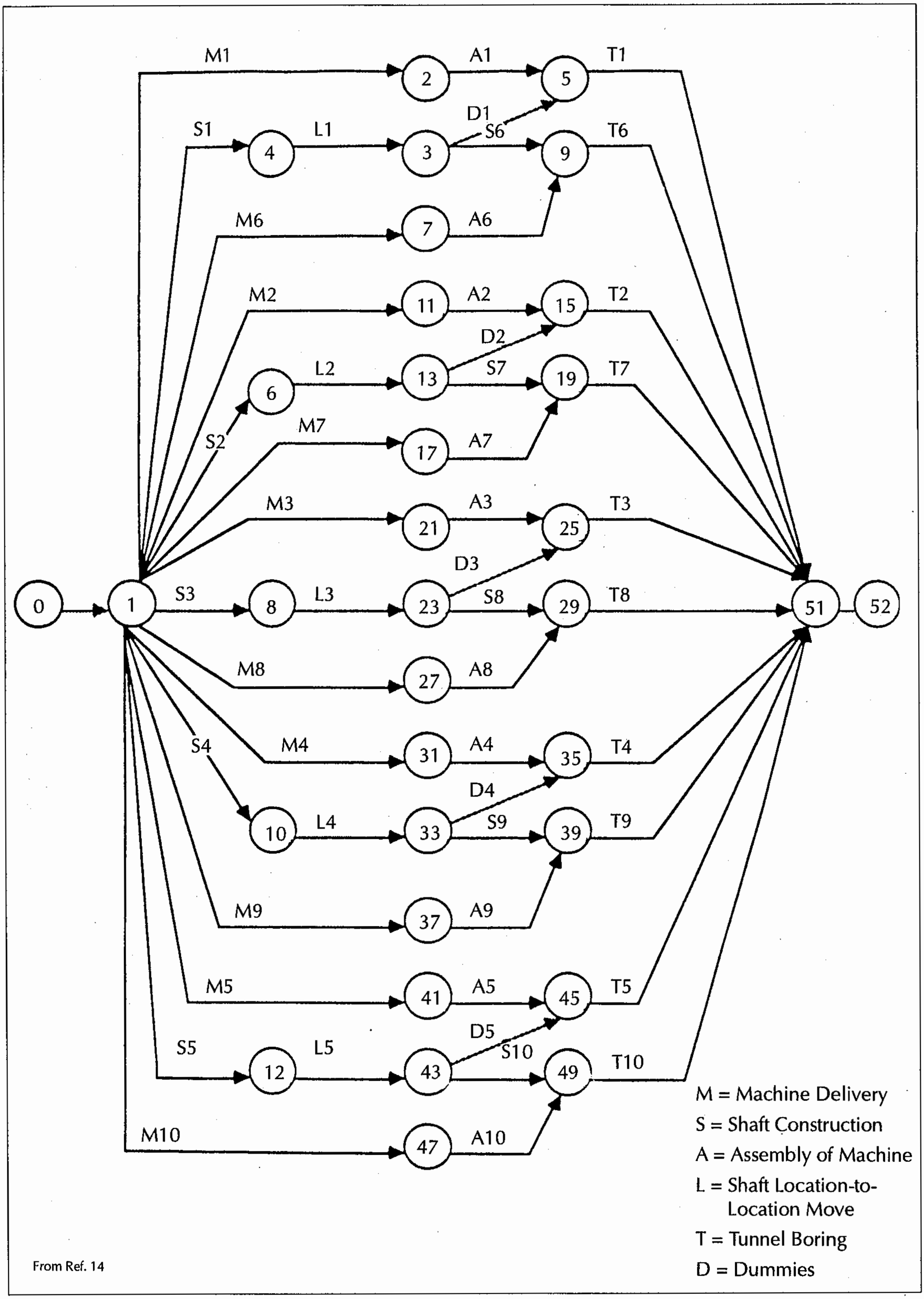

FIGURE 5. The PERT network for Example 1.

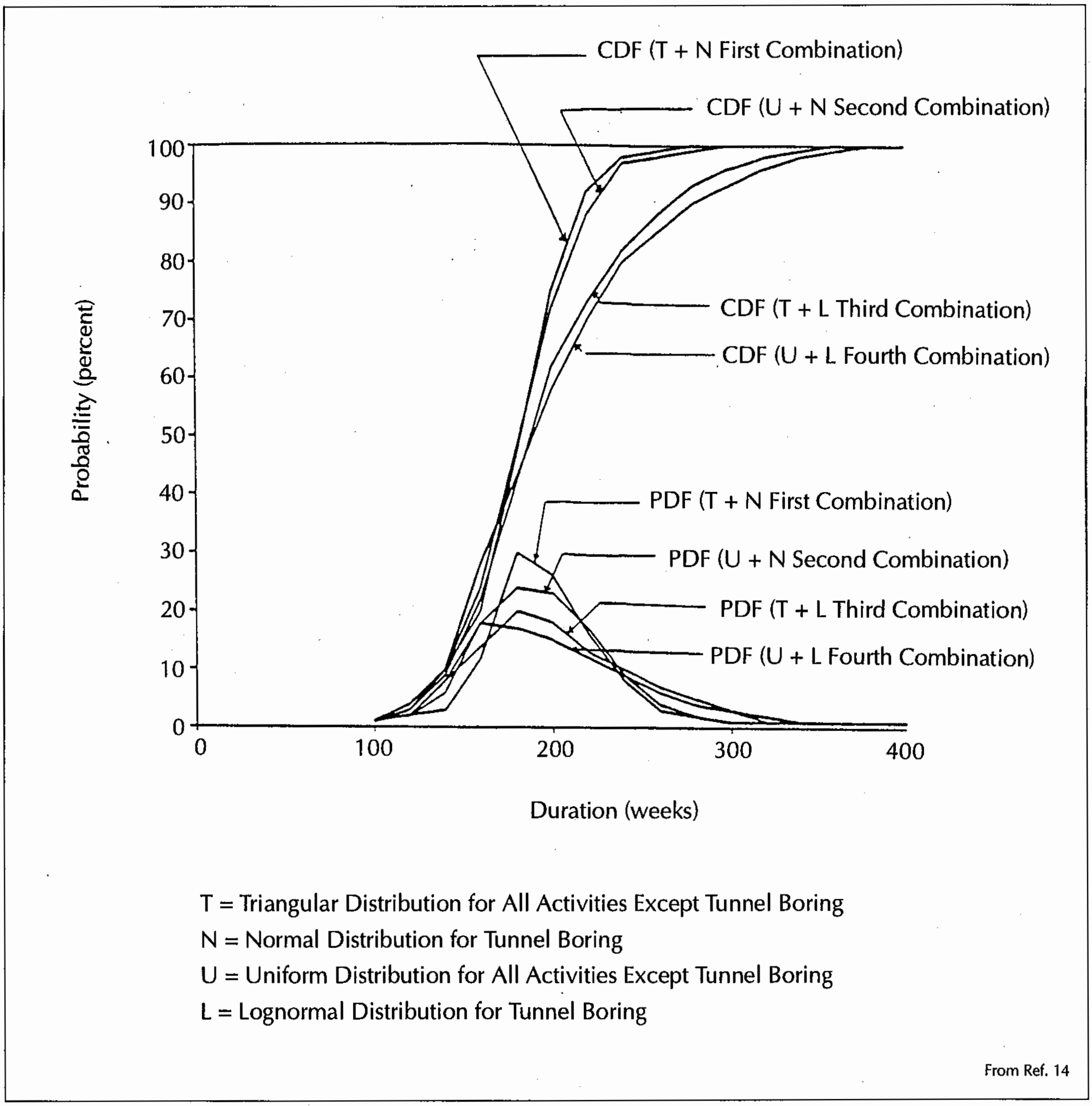

FIGURE 6. Summary of risk analysis results for Example 1.

Example 2. Risk assessment and analysis for project duration do not necessarily have to be tied to a scheduling network. Large portions of the schedule may not be of interest to top management or may not reveal a large potential for variability. In such cases, it will be wise to focus on specific areas where variations in duration can have a strong impact on the project. A risk model that was developed as part of the Concept Design Report for the Massachusetts Water Resources Authority (MWRA) Inter Island and Outfall tunnels is examined.[19] Again, it is emphasized that this model is used solely as an example and is not used as a defense or critique of the model.

One objective of the study was to develop a CDF for the total duration of tunnel boring for the Inter Island and Outfall tunnels. It was argued that within the Deer Island Treatment Plant and Facilities, the Outfall tunnel was on the critical path and, moreover, the activity with highest potential for variability was tunnel boring. Therefore, it was sensible to conduct a risk analysis on the tunnel boring operation. The tunnel duration consisted of several components all of which were computed according to the following procedure:

> The time to tunnel in a certain rock type, with a certain quality and with a certain water inflow, is equal to the length of the tunnel segment divided by the TBM achieved rate in the same type of rock with the same quality and water inflow. The TBM achieved rate is defined as the product of the TBM utilization rate (the time the machine is boring as a proportion of the total working hours) and the TBM penetration rate (instantaneous penetration rate) in the same type of rock with the same quality and water inflow.

The model has been simplified so that it can be discussed here in a reasonable space and, at the same time, preserve the essentials of the approach taken in the actual study.

The following criteria are considered:

$$T_{ijk} = L_{ijk}/(P_{ijk}\, U_{ijk}) \qquad 1$$

In Equation 1, T_{ijk} is the time required to tunnel segment denoted by ijk; i is the rock type; j is rock quality (excellent, good, fair, poor or altered — based on its Rock Quality Designation [RQD]); k is the water inflow rate (high, medium, low — based on permeability); L_{ijk} is the length of the tunnel in a certain rock, with a certain quality and water inflow rate; P_{ijk} and U_{ijk} are the TBM penetration rate and utilization rate at the given conditions, respectively.

$$L_{ijk} = L\, W_{ijk}\, R_i\, Q_{ij} \qquad 2$$

In Equation 2, L is the total tunnel length; R_i is the probability that rock type i is encountered; Q_{ij} is the probability that rock of quality j is encountered given that the rock type is i; and W_{ijk} is the probability of having the water inflow rate k, given that the rock type is i and the rock quality is j.

From Equation 2, it is clear that:

$$\sum_{k} W_{ijk} = 1$$

and also:

$$\sum_{i}\sum_{j}\sum_{k} L_{ijk} = L$$

In Equation 1, P_{ijk} and U_{ijk} are both random variables that provide ranges for the TBM utilization and penetration rates under assumed i, j and k conditions. Every random variable has to be identified with a distribution and the relevant parameters. In the actual study, two sets of computations were carried out. In one, uniform distributions were assumed for every random variable. In the second, triangular distributions were assumed for every random variable. For the uniform distributions, ranges of distributions were estimated based on the available information, experience and expert opinions. For the triangular distributions, the most likely value of every distribution was estimated in addition to the distribution range.

For example, the TBM penetration rate in Argillite, in excellent rock conditions (RQD > 96), was estimated to vary between 10.1 and 14.1 feet per hour (ft/hr). The most likely value for this rate was estimated as 12.1 ft/hr. Also, it was assumed that water inflow will only affect the utilization rate rather than the penetration rate. Therefore, the specified ranges for TBM penetration were assumed to be valid regardless of water inflow conditions. In this way, a triangular distribution or a uniform distribution was completely specified for the penetration rate in Argillite in excellent conditions. The same approach was used to estimate the ranges of distributions in order to model penetration rates with other qualities of Argillite or with other types of rock that were expected to be encountered in the tunneling operation. It is clear that a large number of random variables had to be specified in order to estimate the various times required to tunnel a segment.

For computing the CDF of the total tunnel duration, a Monte Carlo simulation approach was utilized. A computer program was developed that sampled various statistical distributions specified by the modelers in order to pick up values used in Equation 1. Every random variate specified was sampled once. Values of T_{ijk} were computed depending on the i, j and k that was sampled. The $T_{tot} = \sum T_{ijk}$ for each segment was computed to provide total number of hours required for tunnel boring. This process of sampling the distributions was repeated 100,000 times and every time a T_{tot} was computed. These T_{tot} values were used to con-

```
                    S U M M A R Y   R E P O R T

  SIMULATION PROJECT BSCE RISK                      BY TOURAN

  DATE  5/30/1991                                   RUN NUMBER10000 OF10000

  CURRENT TIME    .0000E+00
  STATISTICAL ARRAYS CLEARED AT TIME    .0000E+00

              **STATISTICS FOR VARIABLES BASED ON OBSERVATION**

                    MEAN     STANDARD  COEFF. OF  MINIMUM   MAXIMUM
                    VALUE    DEVIATION VARIATION   VALUE     VALUE

 DURATION          .205E+03  .338E+02  .165E+00  .139E+03  .309E+03
1                          **HISTOGRAM NUMBER 1**
                               DURATION

 OBS  RELA   UPPER
 FREQ FREQ  CELL LIM 0        20        40        60        80        100
                     +    +    +    +    +    +    +    +    +    +    +
   0  .000  .130E+03 +                                                 +
  88  .009  .145E+03 +                                                 +
 742  .074  .160E+03 +****                                             +
 ***  .119  .175E+03 +*******   C                                      +
 ***  .161  .190E+03 +*********        C                               +
 ***  .165  .205E+03 +*********                C                       +
 ***  .160  .220E+03 +*********                        C               +
 ***  .118  .235E+03 +*******                                C         +
 835  .083  .250E+03 +****                                        C    +
 539  .054  .265E+03 +***                                           C  +
 320  .032  .280E+03 +**                                              C+
 179  .018  .295E+03 +*                                                C
  57  .006  .310E+03 +                                                 C
   0  .000  .325E+03 +                                                 C
   0  .000  .340E+03 +                                                 C
   0  .000  .355E+03 +                                                 C
   0  .000    INF    +                                                 C
 ---                 +    +    +    +    +    +    +    +    +    +    +
                     0        20        40        60        80        100
```

FIGURE 7. The results of simulation analysis for Example 2.

struct a CDF for the total tunnel duration. Using this CDF, various confidence levels could be computed for the completion of the tunneling operation. It is apparent that any existing correlations among the model parameters in adjacent tunnel segments were neglected. Kim provides further discussion of tunnel risk analysis.[20] The issue of correlation among random variables and its impact on the analysis outcome is further discussed in the project cost section below.

In order to illustrate the process of risk assessment, a much simplified scenario of the above problem is presented and two methods of solution are presented: simulation approach and direct analytical approach.

A Monte Carlo simulation study is conducted on a simplified version of Example 2 and these computations are carried out with hypothetical data. It is assumed that estimating the duration time required for tunneling a 1,000-foot segment in a certain rock under specific conditions is of interest. The duration time can be modeled as:

$$T = L/(P\,U) \qquad 3$$

In Equation 3, L is equal to 1,000 feet; and P and U are random variables that portray variations in the expected TBM penetration and utilization rates, respectively. Furthermore, it is assumed that both P and U are independent and follow a uniform distribution. The bounds of the distributions may be estimated by doing a literature search, examining historical data or consulting experienced personnel. It is assumed that P may be any number between 8 ft/hr and 12 ft/hr, and U may be between 40 and 60 percent.

A simple Monte Carlo simulation model was developed using a simulation software development package.[18] The simulation was run for 10,000 times. At every run T was computed. A histogram of T and cumulative values of the T distribution were computed (see Figure 7). Using this distribution, various confidence levels can be investigated. According to analysis results, the average time to bore the tunnel was 205 hours, with a standard deviation of 33.8 hours. From the cumulative curve, it can be deduced that there is almost a 70 percent chance that the project can be completed within 220 hours. On the other hand, the probability of finishing the project in 175 hours is only about 20 percent.

In some cases, when the model is sufficiently simple, it may be advantageous to attempt to develop a CDF for the desired function (tunnel duration in this example) analytically. If U and P are assumed to be independent (the assumption that was made in the simulation approach as well), then it will be easy to compute the mean and variance of T:

$$T = L\,(1/P)(1/U) \qquad 4$$

T is a function of the two independent random variates U and P (see Equation 4). Values of the mean and variance of T can be computed by:

$$E[T] = \frac{L}{(b_P - a_P)\,(b_U - a_U)}\,Ln\left(\frac{b_P}{a_P}\right) \times Ln\left(\frac{b_U}{a_U}\right) \qquad 5$$

$$Var[T] = L^2\{\frac{1}{a_P b_P a_U b_U} - \frac{[Ln(b_P/a_P)]^2\,[Ln(b_U/a_U)]^2}{(b_P - a_P)^2\,(b_U - a_U)^2}\} \qquad 6$$

In Equations 5 and 6, b_P, a_P, b_U and a_U stand for the limits of uniform distributions for P and U (with b denoting the upper bounds and a denoting the lower bounds); $E[T]$ is the expected value or mean of duration; and $Var[T]$ is the variance of the duration time. (Please refer to the top box on page 46 for the derivation of Equations 5 and 6.) By substituting the values of L, a_P, b_P, a_U and b_U in Equations 5 and 6, $E[T]$ was computed to be 205 hours and $Var[T]$ as 1,172. Noting that the standard deviation is the square root of variance, the standard deviation of T is computed as 34.2 hours. Note that the results obtained from the analytical method are theoretically accurate. By comparing these results with simulation results, it can be seen that while simulation estimates the mean accurately, there is some error in estimating the standard deviation. This error could be reduced by running the simulation many more times or by checking the random number generating routines that the simulation software uses. However, from a practical point of view, the simulation results appear to be sufficiently accurate.

The CDF of T can be computed with a little more effort also. Equations 7 and 8 give the CDF of the tunnel duration for various values of t. (For the derivation of Equations 7 and 8, refer to the bottom box on page 46.) $F_T(t)$ is the probability of finishing the tunnel boring within t hours:

for $t < 1{,}000/(0.4*12) = 208.3$ *hours*

$$F_T(t) = 9 - \frac{1250}{t}Ln(12) - \frac{1250}{t} + \frac{1250}{t}Ln\left(\frac{1667}{t}\right) \qquad 7$$

for $t > 208.3$ *hours*

$$F_T(t) = -\frac{1250}{t}Ln\left(\frac{2500}{t}\right) + \frac{1250}{t}Ln(8) + \frac{1250}{t} - 3 \qquad 8$$

As an example of using these equations, $F(175)$, or the probability of finishing the tunneling within 175 hours, is computed as 0.208 using Equation 7. Or using Equation 8, $F(220) = 0.688$; *i.e.*, the probability of finishing the tunneling operation within 220 hours is 68.8 percent. As can be seen, these results are almost identical to the results obtained from simulation analysis. Figure 8 compares the CDFs for the total duration computed by simulation and analytical approaches.

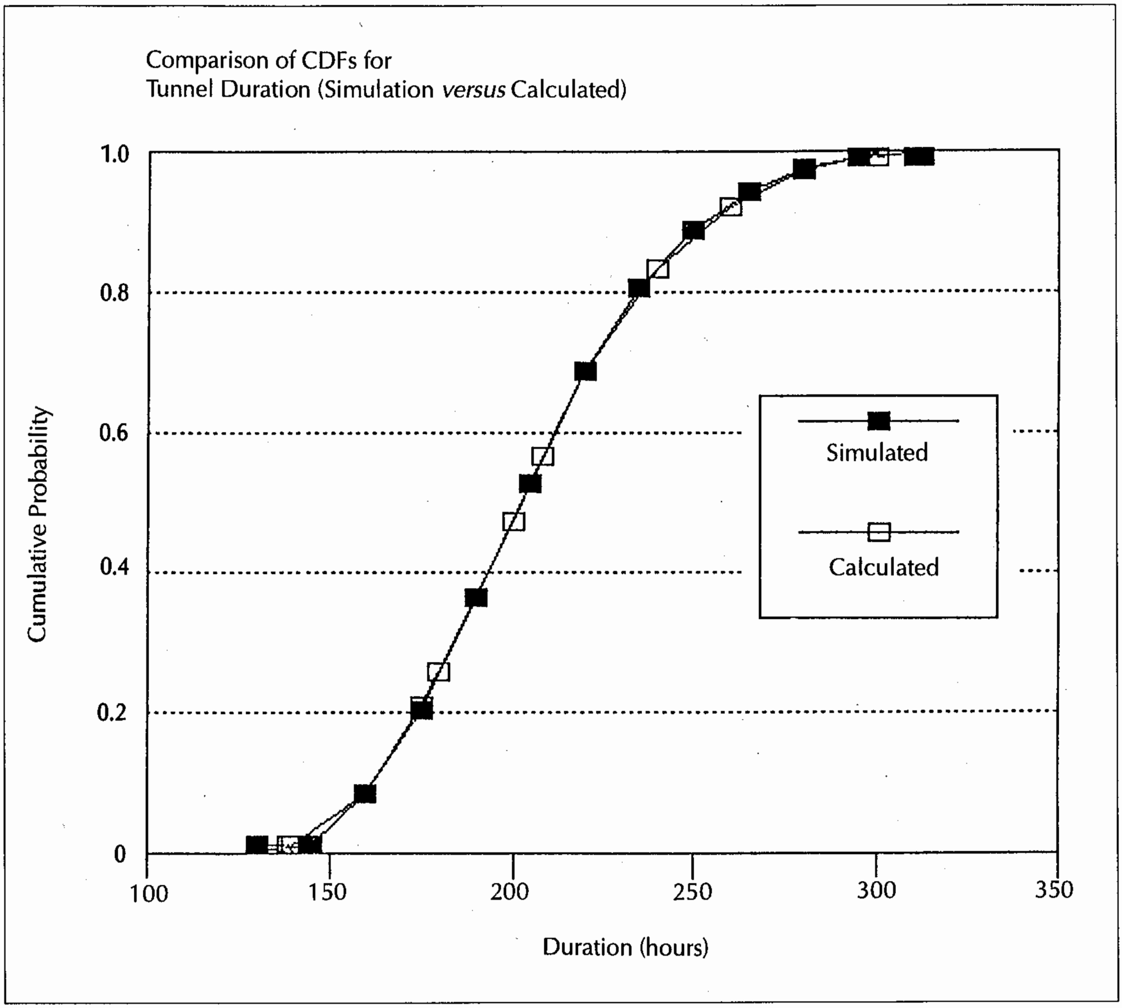

FIGURE 8. Comparison of the results of the analytical approach and the simulation approach.

Equation 4 gives the tunneling duration time under specific conditions. If the total tunnel length consists of several segments, and the duration of constructing each of those segments can be modeled according to Equation 4, then it can be assumed that the total duration, T_i, is approximately distributed according to a normal distribution (the Central Limit Theorem). Moreover, the mean of the total will be equal to the sum of the means of the individual total durations and the variance of the total will be equal to the sum of the variances of the individual total durations. It is interesting to note that in this case the CDF of the total (normal distribution) can be computed quickly and more conveniently than the case where there was only one tunnel segment and Equations 7 and 8 had to be derived.

It is generally worthwhile to try to find a direct solution before attempting a Monte Carlo approach. Equations 5 and 6 can serve as powerful means for sensitivity analysis and can easily allow the estimator to evaluate the impact of changing various parameters of the distributions on the total project mean and variance.

Examples discussed so far have illustrated typical applications of probabilistic analysis to duration estimation. It was observed that risk measurement can be done either by using traditional network-based schedules or by utilizing any appropriate relationship that realistically defines a duration or productivity rate. Also, in both cases, the use of a direct analytical approach or simulation modeling may be possible. All the examples cited above assume independence among variables. Analysis of correlated random variates is significantly more complicated than independent variates.

Project Cost

A common application of risk analysis in construction is to compute the CDF of the total project cost, which in turn can help the contractor specify the margins of safety needed in the levels of funding required. The CDF developed by the contractor can help the contractor arrive at a reasonable contingency sum and to allocate contingency to various project activities.[2,21,22] Again, the Monte Carlo simulation technique is commonly used in cost risk assessment.

The total project cost is modeled as a random variable that is the sum of several cost items, themselves being random numbers:

$$C_{tot} = \sum_{i=1}^{n} C_i \qquad 9$$

In Equation 9, C_{tot} is the total project cost; and C_i is a project cost component.

Obviously, if the cost variations of every small cost component that goes into a detailed estimate are considered, the approach would be impractical. Because of this, the individual cost components that are considered consist of the major items that generally appear on the estimate summary sheets and the recap sheets. Also, it is understood that most of the total cost variation is due to the variability of a limited number of components.[1,23]

Therefore, only those items with high potential for variation are considered as random variables and the rest of the items are assumed to be fixed. Curran defines a *critical variance* for the bottom line.[23] Any single component that has the potential of changing the project bottom line by more than this critical variance is considered a *critical component* and should be modeled as a random variable. Curran suggests the critical variance to be 0.5 percent of the project bottom line for conceptual estimates and 0.2 percent of the bottom line for detailed estimates. So, for example, in a $10,000,000 conceptual budget estimate, if any single component has the potential of changing the total cost by more than $50,000, then this component is considered critical. Furthermore, Curran suggests that in over 90 percent of the projects of all types, the number of critical items was fewer than 30.[23] Other cost items in the project, then, can be established as fixed values. C_{tot} in Equation 9 is then composed of a fixed and a random component. Since the various project cost components can have various distributions, accurate computation of C_{tot} involves the computation of a number of convolution integrals and becomes very lengthy.

Monte Carlo simulation can simplify the process if a computer and the appropriate software are available. It consists of generating random numbers according to C_i distributions, adding up these items, adding the fixed costs to these and, then, computing the total project cost. This procedure is repeated at least several hundred times, and every time a value for C_{tot} is computed. A histogram, and later a CDF, can be constructed with the values of C_{tot}. The CDF can then be used to estimate the probability of completing a project at, or below, a certain budget.

Problems with the Monte Carlo Approach. Although the Monte Carlo approach provides a straightforward means for probabilistic estimating, there are major limitations in its application. First, statistical distributions for the various cost components need to be established. Second, if the random numbers are not independent, their correlations should be fully documented for the correct implementation of the Monte Carlo technique.

One logical method for investigating the distribution type is to collect data from similar projects, assume a distribution and perform a proper test of goodness of fit in order to evaluate the hypothesis. In the absence of historical data, the same general guidelines regarding the choice of distribution described earlier can be used.

One of the more common sources of error in Monte Carlo simulation is that it is assumed that cost components are independent and changes in one cost component do not affect any other cost component. This assumption is clearly inappropriate in typical construction projects. However, it is assumed that if the correlation between variables is sufficiently small, the assumption of independence does not create large errors. Generally, disregarding the correlation between variables in a Monte Carlo simulation results in an underestimation of the total cost variance since the effect of covariances in computing the variance is neglected. It is important to emphasize that an accurate estimate of variance is crucial in probabilistic estimating because the estimator is interested in deviations from the mean estimate for computing the probability of cost overrun or underrun.

The accurate method of incorporating correlations is time-consuming and requires a great deal of data that is not always available. In some cases, depending on the underlying distributions, it is not possible to make an accurate analysis. One suggested method involves combining highly correlated cost items into a single cost item so that all the remaining cost items (some of which are a combination of several correlated cost items) can be considered to be independent variables.[24] For example, assume that a project cost consists of ten cost items (C_1 to C_{10}):[9]

$$C_{tot} = \sum_{n=1}^{10} C_i \qquad 10$$

Further, assume that there is reason to believe that items C_4, C_5 and C_6 are highly correlated and that C_9 is correlated with C_{10}. Define C' and C'' so that:

$$C' = C_4+C_5+C_6 \qquad 11$$

$$C'' = C_9+C_{10} \qquad 12$$

If the estimator can specify the underlying distributions and parameters of C' and C'', and if the rest of cost components can be assumed to be independent, then by rewriting Equation 10 as Equation 13, a Monte Carlo simulation can be conducted:

$$C_{tot} = C_1+C_2+C_3+C'+C_7+C_8+C'' \qquad 13$$

In Equation 13 all the items are assumed to be independent. Curran presents a hypothetical example to show the application of this method.[24] The problem is that, in many cases, it will be difficult and even unnatural to lump together various cost components and estimate their combined range, parameters and distribution.

For conducting an accurate analysis of the total cost variance, the joint density functions of the correlated cost components are needed. The PDF that the estimator or risk analyst specifies for a certain cost component is actually the marginal distribution of that cost component. In general, if different cost components are not independent, knowing the marginals of these random variables is not sufficient to obtain their joint density functions. Without the joint density function, the correlated random numbers cannot be generated for Monte Carlo simulation.

The case of multivariate normal distribution is an exception, however. If marginals of the multivariate normal distribution and the covariance matrix are available, then the joint density usually can be found and the analysis can be conducted. The cost components have to be normally distributed in this instance. Multivariate normal distribution can be transformed to multivariate lognormal.[25] Also, in special cases, approximations to analyze the correlated random variates can be employed.[9] However, the use of approximations comes with the cost of reduced accuracy. This level of detail in conducting risk analysis in construction, however, is almost never attempted in practice and the assumption of independence or the simpler method described above is all that is actually used.

Analytical Approach. As long as Equation 9 is used to model total project cost, the CLT may be an attractive alternative. The total cost will have an approximately normal distribution and the total mean and variance can be estimated from parameters of the various cost components. As discussed earlier, the user does not even need to specify the statistical distribution for the cost components. In most cases, the number of cost components considered for risk analysis is sufficiently large. Therefore, the assumption of normality in CLT is reasonable.

Comprehensive Cost Functions. Equation 9 is the simplest form of function that may be used for cost risk analysis. A more general model was suggested by Diekmann and is presented with slight modification in Equation 14:[26]

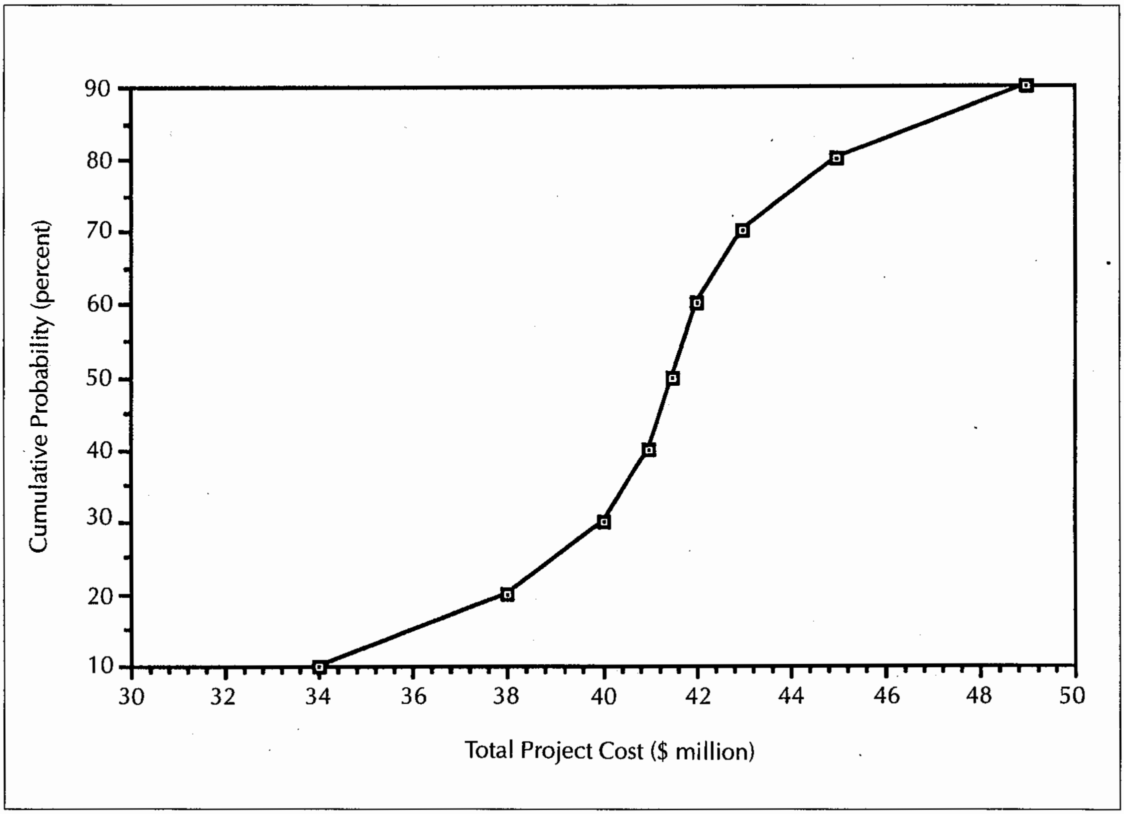

FIGURE 9. The CDF of the total project cost for Example 3.

$$C_{tot} = \sum_i [q_i (m_i + w_i l_i)] + \sum_j C_j \qquad 14$$

In Equation 14, the total cost is composed of i categories of work, and j indirect cost items; q_i is the work quantity in category i; m_i is the unit material cost of category i; l_i is the labor productivity rate (man-hours / q) for category i; w_i is the wage rate related to labor l_i; and C_j is the indirect cost item j.

Again, the Monte Carlo approach can be used to develop a CDF for C_{tot}. Any of the parameters described above may have variations that have to be considered in the analysis. An analytical solution may not always be convenient or even feasible, depending on the shape of the cost function. Computations become cumbersome, especially if reasonably complex and realistic distributions (such as lognormal or beta) are assumed for the parameters.

Most project cost functions can be modeled in a format similar to Equations 9 and 14. Several software packages are available that allow the user to conduct risk analysis on a personal computer (generally using a simulation approach). In using these packages, the user loses some flexibility in modeling, but the process becomes convenient and fast. Understanding underlying assumptions used in the development of these packages is extremely important in order to avoid errors in the interpretation of results.

Example 3. Assume that a project's budget (or target estimate) was approximately estimated at $42 million. Furthermore, assume that the project's critical components have been identified, their distributions and parameters specified and a Monte Carlo simulation was conducted using the general format of Equation 9. A cumulative distribution function for the project has been developed as presented in Figure 9. The computation of the CDF by Monte Carlo simulation technique is very similar to the method described in Example 2 and will not be repeated here. Figure 9 shows that there is a 60 percent chance of completing the project within

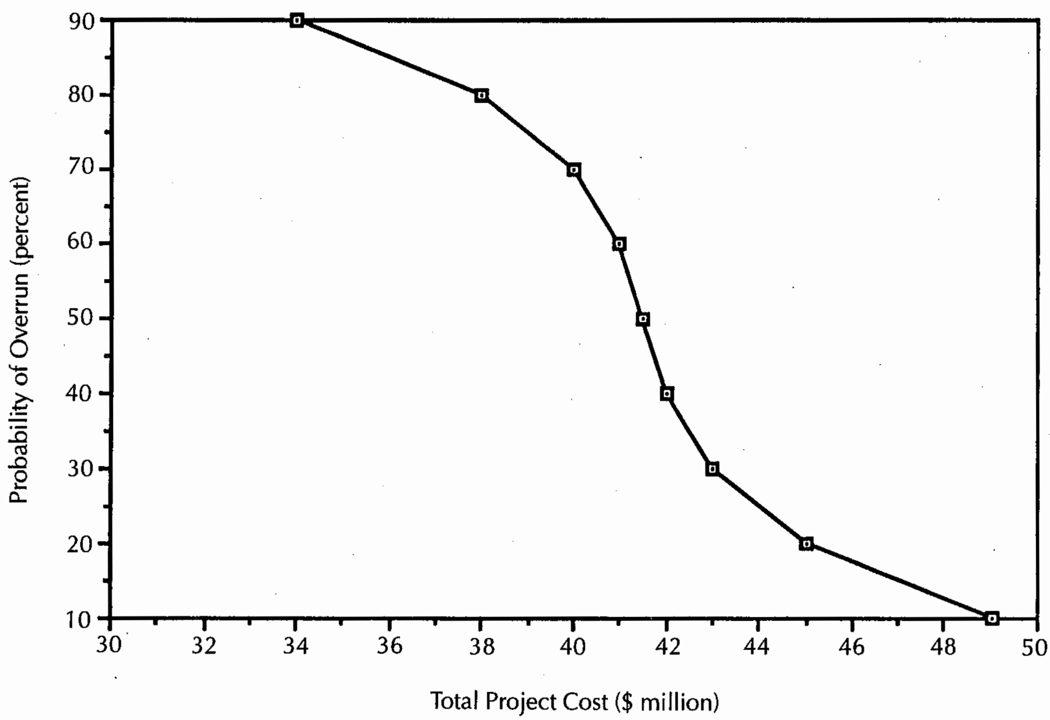

FIGURE 10. The overrun profile for the total project cost for Example 3.

the estimated or desired budget. If the owner is not comfortable with this confidence level and would prefer a confidence level of, say, 80 percent, then the required budget would be about $45 million. Some practitioners prefer to arrange the CDF of Figure 9 in a slightly different way and develop a so-called overrun profile for the project cost (see Figure 10).[1,23]

In Figure 10, the values of the y-axis are simply the complements of the values of y-axis of Figure 9. The same conclusions can be drawn from Figure 10. There is a 40 percent chance of budget overrun if the target estimate is $42 million, and there is a 20 percent chance of budget overrun if the target estimate is $45 million.

The same approach can be used by the contractor for arriving at a reasonable contingency sum for the project. The contractor can develop a CDF for project cost (excluding contingency or profit) and then choose a markup so that the probability of losing money on the project falls below a certain acceptable threshold.

Summary

The process of risk measurement and analysis in construction management, as related to project cost and schedule, has been described. The main emphasis was placed on the methodologies used to model and measure risk rather than the process of identification of risk prone components and areas. This latter issue, very important on its own right, deserves separate treatment.

The techniques used in risk measurement were divided into two major categories: the analytical and the simulation approaches. The methodologies discussed were further illustrated by using a few examples in order to provide a good understanding of the practice of risk modeling and measurement in construction.

ACKNOWLEDGMENTS — *The software used for the Monte Carlo simulations was SLAM-II. Figures 5 and 6 were taken from Reference 14. ASCE's cooperation in this regard is appreciated.*

ALI TOURAN *is an Assistant Professor of Civil Engineering at Northeastern University. He holds M.S. and Ph.D. degrees in Civil Engineering with an emphasis on construction management and engineering from Stanford University. His areas of expertise are construction cost and productivity analysis, simulation modeling, and heavy construction equipment and methods. He has performed research in concrete productivity, simulation and tunneling, material handling and bridge deck construction. He has lectured on construction process simulation, and data collection and analysis.*

REFERENCES

1. *Management of Project Risks and Uncertainties,* Construction Industry Institute, Publication No. 6–8, Austin, Texas, October, 1989.

2. Diekmann, J.E., Sewester, E.E., & Taher, K.H., *Risk Management in Capital Projects,* Construction Industry Institute Source Document 41, Austin, Texas, October, 1988.

3. Spooner, J.E., "Probabilistic Estimating," *Journal of Construction Eng. and Management,* ASCE, Vol. 100, No. CO1, March, 65–77, 1974.

4. *PERT Cost Systems Design,* DOD and NASA Guide, The Office of the Secretary of Defense and the National Aeronautics and Space Administration, U.S. Government Printing Office, Washington, D.C., June, 1962.

5. Teicholz, P., "A Simulation Approach to the Selection of Construction Equipment," *Technical Report No. 26,* Construction Institute, Stanford University, Stanford, Calif., 1963.

6. O'Shea, J.B., Slutkin, G.N., Shaffer, L.R., "An Application of the Theory of Queues to the Forecasting of Shovel-Truck Fleet Productions," *Construction Research Series,* No. 3, Dept. of Civil Eng., University of Illinois, Urbana, Ill., June, 1964.

7. Gaarslev, A., "Stochastic Models to Estimate the Production of Material Handling Systems in the Construction Industry," *Technical Report No. 111,* Construction Institute, Stanford University, Stanford, Calif., August, 1969.

8. Wiser, E.P., "Analysis of Cost Data in Commercial Construction Projects," report submitted to the Dept. of Civil Eng. of Northeastern University, in partial fulfillment of the requirements of the Degree of Master of Science, Boston, Mass., 1991.

9. Touran, A., & Wiser, E.P., "Monte Carlo Technique with Correlated Random Variables," *Journal of Construction Eng. & Mngmt.,* ASCE, Vol. 118, No. 2, June, 1992.

10. Mlakar, P.F., & Bryant, L.M., "Direct Range Cost Estimating," *Proceedings, AACE,* Boston, Mass., June, k.4.1–k.4.4, 1990.

11. Moder, J.J., Phillips, C.R., & Davis, E.W., *Project Management with CPM, PERT, and Precedence Diagramming,* 3rd ed., Van Nostrand Reinhold Co., New York, New York, 1983.

12. Devore, J.L., *Probability and Statistics for Engineering and the Sciences,* 3rd ed., Brooks/Cole Publishing Co., Pacific Grove, Calif., 1991.

13. Miller, R.W., *Schedule, Cost and Profit Control with PERT,* McGraw-Hill Inc., New York, New York, 1983.

14. Touran, A., & Banafa, A., "Probabilistic Scheduling in Tunneling," *Proceedings, 6th Conf. on Computing in Civil Eng.,* ASCE, Atlanta, Georgia, pp. 597–604, 1989.

15. Banafa, A.M., "SSC Tunnel Scheduling Under Uncertainty," report submitted to the Dept. of Civil Eng. of the Univ. of Colorado, in partial fulfillment of the requirements for the Degree of Master of Science, Boulder, Colorado, 1987.

16. Hunter, H.E., "Drilled Shaft Construction at Crown Point, New Mexico," *Proc., RETC,* Vol. 1, SME, Littleton, Colo., 544–565, 1979.

17. Wilson, J.W., "Shaft Sinking Technology and the Future Needs of the Mining Industry," *Proc. RETC,* SME, Littleton, Colo., 1976.

18. Pritsker, A.A.B., *Introduction to Simulation and SLAM-II,* 3rd ed., Systems Publishing Corp., West Lafayette, Ind., 1986.

19. "Tunnel Risk Assessment," Concept Design Report, Metcalf and Eddy, Boston, Mass., June, 1989.

20. Kim, Y.W., "Decision Support Framework in Adaptable Tunneling," dissertation submitted to the Dept. of Civil Eng. of Massachusetts Institute of Technology, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy, Boston, Mass., April, 1984.

21. Hackney, J.W., "Applied Contingency Analysis," *AACE Transactions,* B.2.1–B.2.4, 1985.

22. Jackson, R.L., Kinsey, J.W., Cotton, W.E., & Moore, S.K., "The Why and How of Contingency Management," *Transactions, AACE,* 9th Annual Meeting, Denver, Colo., N.0.1–N.0.5, 1985.

23. Curran, M.W., "Range Estimating," *Cost Estimating,* No. 3, Vol. 31, March, 18–26, 1989.

24. Curran, M.W., "Range Estimating Reduces Iatrogenic Risk," *Proceedings, American Assoc. of Cost Eng.*, Boston, Mass., June, k.3.1–k.3.3, 1990.

25. Johnson, M.E., & Ramberg, J.S., "Transformations of the Multivariate Normal Distribution with Applications to Simulation," *Los Alamos Sci. Lab. Tech. Report LA-UR-77-2595*, Los Alamos, N.M., 1978.

26. Diekmann, J.E., "Probabilistic Estimating: Mathematics and Applications," *Journal of Construction Eng. and Management*, ASCE, Vol. 109, No. 3, September, 297–308, 1983.

Derivation of Expected Value and Variance: Equations 5 & 6

Equation 5 is derived as follows:

Assuming independence between U and P:

$$E[T] = E\left[\frac{L}{PU}\right] = L.E\left[\frac{1}{P}\right].E\left[\frac{1}{U}\right] \qquad \textbf{5.1}$$

U and P are assumed to be uniformly distributed; the probability density function of U, $f_U(u)$ then becomes:

$$f_U(u) = \frac{1}{b_U - a_U} \qquad \textbf{5.2}$$

$$E\left[\frac{1}{U}\right] = \int_{a_U}^{b_U} \frac{1}{u} f_U(u)du = \int_{a_U}^{b_U} \frac{du}{(b_U - a_U)\, u} = \frac{1}{(b_U - a_U)} Ln(b_U/a_U)$$

Using a similar approach for computing $E[1/P]$, and substituting in Equation 5.1, Equation 5 is obtained as given.

Equation 6 is derived as follows:

$$Var[T] = E[T^2] - \{E[T]\}^2 \qquad \textbf{6.1}$$

$$Var[T] = Var[L/PU] = L^2\, Var[1/PU] = L^2.E[1/P^2U^2] - \{E[T]\}^2 \qquad \textbf{6.2}$$

Note that:

$$E[1/P^2U^2] = E[1/P^2].E[1/U^2]$$

Also:

$$E\left[\frac{1}{U^2}\right] = \int_{a_U}^{b_U} \frac{du}{(b_U - a_U)\, u^2} = \frac{1}{a_U \cdot b_U} \qquad \textbf{6.3}$$

$E[1/P^2]$ can be computed in a similar manner and by substituting these values and $E[T]$ in Equation 6.2, Equation 6 can be obtained.

Derivation of the CDF of Tunnel Duration: Equations 7 & 8

To derive Equation 7, note that $T = 1000/PU$ is a function of two independent uniform random variates, $U[0.4,0.6]$ and $P[8,12]$ with specified ranges. The limits of integration for computing the CDF of T changes when the value of T passes from 208.3 ($T = 1000/(0.4)(12)$).

For $T < 208.3$ *hours*:

$$F_T(t) = P[T < t] = P[(1000/PU) < t] = \frac{1}{(0.6 - 0.4)(12 - 8)} \int_{1000/(0.6t)}^{12} \left[\int_{1000/(Pt)}^{0.6} du\right] dp \qquad \textbf{7.1}$$

For $T > 208.3$ *hours:*

$$F_T(t) = P[T < t] = P[(1000/PU) < t] = 1 - \frac{1}{(0.6 - 0.4)(12 - 8)} \int_{8}^{1000/(0.4t)} \left[\int_{0.4}^{1000/(Pt)} du\right] dp \qquad \textbf{8.1}$$

Equation 7.1 leads to Equation 7 and Equation 8.1 leads to Equation 8.