AIAA-92-1054

FRISK– FORMAL RISK ASSESSMENT OF SYSTEM COST ESTIMATES

Philip H. Young

Resource Analysis Department
The Aerospace Corporation
P.O. Box 92957
Los Angeles, CA 90009-2957

Abstract

Cost elements in a Work-Breakdown Structure (WBS) of a system design are defined and analyzed as to low, best-estimate, and high cost. A triangular cost-distribution is postulated for each cost element, and the resulting mean and variance are derived. Dependencies among cost elements are specified by a correlation matrix. The means, variances, and correlations of the individual WBS-element costs are combined to yield the mean and variance of total cost. The total-cost distribution is then fit to a lognormal distribution with total-cost mean and variance.

I. Introduction

Risk assessment is an important aspect of understanding the cost of a proposed system design. Typically, low, best-estimate, and high costs for each of the several cost elements in a Work-Breakdown Structure (WBS) are estimated on the basis of a technical-risk assessment, after which a statistical distribution, such as a triangular, is postulated for each element cost. Means, variances, and typical percentiles can be derived from the statistical distribution. Dependencies among cost elements can be quantified in terms of a correlation matrix. Then the distribution of the sum of the element costs is determined, typically by a Monte Carlo sampling technique.

The Formal Risk (FRISK) method is an analytical, rather than a Monte Carlo based, cost-risk model. FRISK evaluates the total cost-distribution of a system design, given its WBS. Thus FRISK provides a simple (entirely analytic) tool to provide estimates of, for example, the 50th, 70th, and 90th percentiles of system cost of a proposed system design, in addition to the so-called "best-estimate" cost.

The mathematical structure of FRISK will be described in the following section, together with a complete list of inputs necessary for applying the method.

II. Ground Rules

Assumptions underlying FRISK's structure are as follows:

1. Cost variation of each element of a WBS is expressed in terms of the triangular distribution.

2. Total cost of the system design can be approximated by the lognormal distribution. (Historical results, one of which is presented here, indicate that this assumption is reasonable.) Mean and variance of the approximating lognormal distribution are calculated by a closed-form expression, as in Garvey[4].

FRISK requires the following input data:

1. System Input

   N = Number of cost elements in the WBS of the system design

2. Cost Element Inputs

$C_i^{(MIN)}$ = Minimum cost for WBS element i

$C_i^{(MAX)}$ = Maximum cost for WBS element i

m_i = Best-estimate (mode) cost for WBS element i

$\rho_{i_1 i_2}$ = Total (as opposed to "partial") correlation matrix elements for pairs of WBS elements, $i_1 = 1, ..., N$; $i_2 = 1, ..., N$,

where

$\rho_{ii} = 1$ (diagonal elements are all 1), $i=1, ..., N$;

$-1 \leq \rho_{i_1 i_2} \leq 1$ (all correlations lie between -1 and 1), $i_1 \neq i_2$;

$\rho_{i_1 i_2} = \rho_{i_2 i_1}$ (matrix is symmetric),

$i_1 \neq i_2$;

Finally, the total correlation matrix must be positive-definite.

III. Total Cost of System Design

To determine the distribution of total system cost, the mean and variance of each WBS cost element must be calculated. The expression for the mean is:

$$\mu_i = \left\{ C_i^{(MIN)} + m_i + C_i^{(MAX)} \right\} /3 \quad (1)$$

and, for the variance,

$$\sigma_i^2 = \left\{ C_i^{(MIN)^2} + m_i^2 + C_i^{(MAX)^2} - m_i\, C_i^{(MIN)} - m_i\, C_i^{(MAX)} - C_i^{(MIN)}\, C_i^{(MAX)} \right\} /18 \quad (2)$$

An illustrative triangular density function is shown in Figure 1. The exact analytic form is given in Abramson and Young[1].

If $\tilde{C}_i$ represents, for each i, the triangularly distributed random variables whose mean and variance are defined above, the total cost for the system design can be written as:

$$\tilde{C}_T = \sum_{i=1}^{N} \tilde{C}_i , \quad (3)$$

representing the sum of the costs of the individual WBS elements.

Using (1) and (2), the mean and variance of $\tilde{C}_T$ defined in (3) are, respectively,

$$\mu_T = E\,(\tilde{C}_T) = \sum_{i=1}^{N} \mu_i \quad (4)$$

and

$$\sigma_T^2 = V\,(\tilde{C}_T) = \sum_{i=1}^{N} \sigma_i^2 + 2 \sum_{i_1 > i_2} \sigma_{i_1} \sigma_{i_2} \rho_{i_1 i_2} \quad (5)$$

Note that if $\rho_{i_1 i_2} = 0$ for $i_1 \neq i_2$,

then (5) becomes

$$\sigma_{TI}^2 = \sum_{i=1}^{N} \sigma_i^2 \quad \text{(the "independent" case)}$$

Also, if $\rho_{i_1 i_2} = 1$ for all i_1 and i_2, (5) becomes

$$\sigma^2_{TD} = \left(\sum_{i=1}^{N} \sigma_i \right)^2$$

(the "totally dependent" case)

In this case DET $(\rho_{i_1 i_2}) = 0$.

If $\rho_{i_1 i_2} \geq 0$ for all i_1 and i_2, it follows that

$$\sigma^2_{TI} \leq \sigma^2_T \leq \sigma^2_{TD}$$

Thus, σ^2_{TI} and σ^2_{TD} are the minimum and maximum values of σ^2_T, respectively, meaning that $\tilde{C}_T$ has minimum dispersion when $\sigma^2_T = \sigma^2_{TI}$ and maximum dispersion when $\sigma^2_T = \sigma^2_{TD}$.

In all cases, the mean (μ_T) is the same as in (4), as it is independent of $\rho_{i_1 i_2}$.

The next step in the development of the analytic procedure for evaluating total system cost is to model the exact (albeit unknown) distribution of $\tilde{C}_T$ by a lognormal distribution, using μ_T and σ^2_T from equations (4) and (5). The lognormal distribution (see Cramér[3]) has two parameters (P_T and Q_T). From Cramér[3], it can be shown that:

$$P_T = .5 \text{ LOG} \left\{ \mu^4_T \,/\, \left(\mu^2_T + \sigma^2_T \right) \right\} \qquad (6)$$

$$Q_T = \left\{ \text{LOG} \left(1 + \sigma^2_T \,/\, \mu^2_T \right) \right\}^{1/2} \qquad (7)$$

If a lognormal distribution, on the other hand, has parameters as defined in (6) and (7), its mean and variance agree with the mean and variance of the exact distribution of $\tilde{C}_T$. For an illustration of the lognormal distribution, see Figure 2.

The density function of the total-system-cost lognormal distribution is:

$$f(x) = \text{EXP} \left\{ - \frac{\left(\text{LOG}(x) - P_T \right)^2}{2Q^2_T} \right\} \Big/ \left\{ Q_T \, x \, \sqrt{2\pi} \right\}, \qquad (8)$$

where $0 < x < \infty$

In the special case that a lognormal distribution has parameter values $P_T = 0$ and arbitrary Q_T, it follows from (6) and (7) that:

$$\mu_T = \text{EXP} \left(Q^2_T / 2 \right)$$

$$\sigma^2_T = \text{EXP}\,(Q^2_T) \left\{ \text{EXP}\,(Q^2_T) - 1 \right\}$$

$$\text{MED}\,(\tilde{C}_T) = 1$$

(i.e, the median total system cost is one)

which are well known relationships (see, for example, Book and Young[2]).

Finally, the cumulative distribution of $\tilde{C}_T$ is, from (8)

$$\Pr(\tilde{C}_T \leq c) = \int_0^c f(x)\, dx \qquad (9)$$

$\Pr(\bar{C}_T \leq c)$, from (9), can be used to evaluate any percentile cost desired, as illustrated in the following section.

IV. Illustrative Example

To illustrate the comparison between the lognormal distribution fit of total cost to the "exact" distribution (as derived by Monte Carlo sampling), consider the following simple case.

The input parameters are: $N = 4$; $C_i^{(MIN)} = (i - 1)$, $C_i^{(MAX)} = i$, $m_i = i - .5$, for $i = 1, \ldots, 4$.

The total correlation matrix has the form:

$$\left(\rho_{i_1 i_2}\right) = \begin{bmatrix} 1 & .9 & .8 & .7 \\ .9 & 1 & .6 & .5 \\ .8 & .6 & 1 & .4 \\ .7 & .5 & .4 & 1 \end{bmatrix}$$

In this example, the total cost-distribution derived by Monte Carlo sampling is shown in Figure 3. The corresponding lognormal distribution evaluated by equation (9) is shown in Figure 4. Note the fairly close correspondence between the two distributions, as indicated by their 50th, 70th, and 90th percentiles.

V. Conclusions

FRISK is an easy-to-use and robust technique for evaluating system design cost-risk, asking the user to characterize the system design in a simple and straightforward manner. It also can be used to do cost-risk comparisons of multiple system designs, assuming that the total cost distribution for each design can be adequately characterized by the lognormal distribution (as in Abramson and Young[1]).

Since the mathematical structure of FRISK is so simple (only a few, relatively straightforward equations are evaluated), it is amenable for programming on a personal computer.

If the triangular distribution assumption should be judged inadequate to represent cost-element variation for certain problems, it is a fairly easy matter to substitute other distributions that may be more appropriate. Finally, distributions other than the lognormal can be used to model the total cost distribution. Specifically, it has been found by the author that the beta distribution (see Cramér[3]) also matches simulated results very well.

VI. References

1. Abramson, Robert L. and Philip H. Young, "FRISKEM - Formal Risk Evaluation Methodology", ORSA/TIMS National Conference, Philadelphia, PA, 29-31 October 1990.

2. Book, S. A. and P. H. Young, "Optimality Considerations Related to the USCM-6 'Ping Factor' ", ICA/NES National Conference, Los Angeles, CA, 20-22 June 1990.

3. Cramér, H., "Mathematical Methods of Statistics", Princeton, 1946.

4. Garvey, P. R., "A General Analytic Approach to System Cost Uncertainty", <u>Cost Estimating and Analysis - Tools and Techniques</u>, Springer-Verlag, June 1990.