DEPARTMENT OF NATIONAL DEFENCE

CANADA

OPERATIONAL RESEARCH AND ANALYSIS ESTABLISHMENT

DIRECTORATE OF MATHEMATICS AND STATISTICS

PROJECT REPORT NO. PR303

COST ESTIMATION USING THREE POINT ESTIMATES

by

E.J. Emond

OTTAWA, CANADA MAY 1985

Spare

DEPARTMENT OF NATIONAL DEFENCE

CANADA

OPERATIONAL RESEARCH AND ANALYSIS ESTABLISHMENT

DIRECTORATE OF MATHEMATICS AND STATISTICS

PROJECT REPORT NO. PR303

COST ESTIMATION USING THREE POINT ESTIMATES

by

E.J. Emond

OTTAWA, CANADA MAY 1985

DEPARTMENT OF NATIONAL DEFENCE

CANADA

OPERATIONAL RESEARCH AND ANALYSIS ESTABLISHMENT

DIRECTORATE OF MATHEMATICS AND STATISTICS

PROJECT REPORT NO. PR303

COST ESTIMATION USING THREE POINT ESTIMATES

by

E.J. Emond

Approved By ______________________
Director,
Mathematics and Statistics

OTTAWA, CANADA

May 1985

INTRODUCTION

1. This report describes a method of summing cost estimates to produce a probability distribution for the total cost. The method is well suited for practical use because the calculations are all done by computer and the analyst's input has been reduced to three easily understood estimates for each cost element. The probability distribution for the total cost which is the output of the method can be used to calculate the most likely total cost and the probable spread about this value. This information would be valuable to project management in comparing total costs of competing systems and in establishing contingency allowances.

2. An example from a real DND project is included. This example illustrates that the methods developed can be applied in practice and also shows the added information to be gained. An alternative method of allocating a contingency allowance is suggested. This method provides a rational link between estimated project cost and contingency allowance.

DESCRIPTION OF THE METHOD

3. The method is based on the assumption that the uncertainty in a cost element can be adequately described by a triangular distribution. As illustrated in Figure 1, this type of distribution is completely specified by three values: lowest cost, most likely cost, and highest cost.

4. The triangular distribution was chosen because it requires only three easily understood input values to specify, and yet is adaptable enough to accurately reflect the actual probability distribution of the final cost around the predicted most likely cost. While other distribution functions (such as the Beta distribution) also serve as good models of the uncertainty in the final cost, they are more difficult for a cost analyst to use because they require extra input values which are not easily understood or estimated.

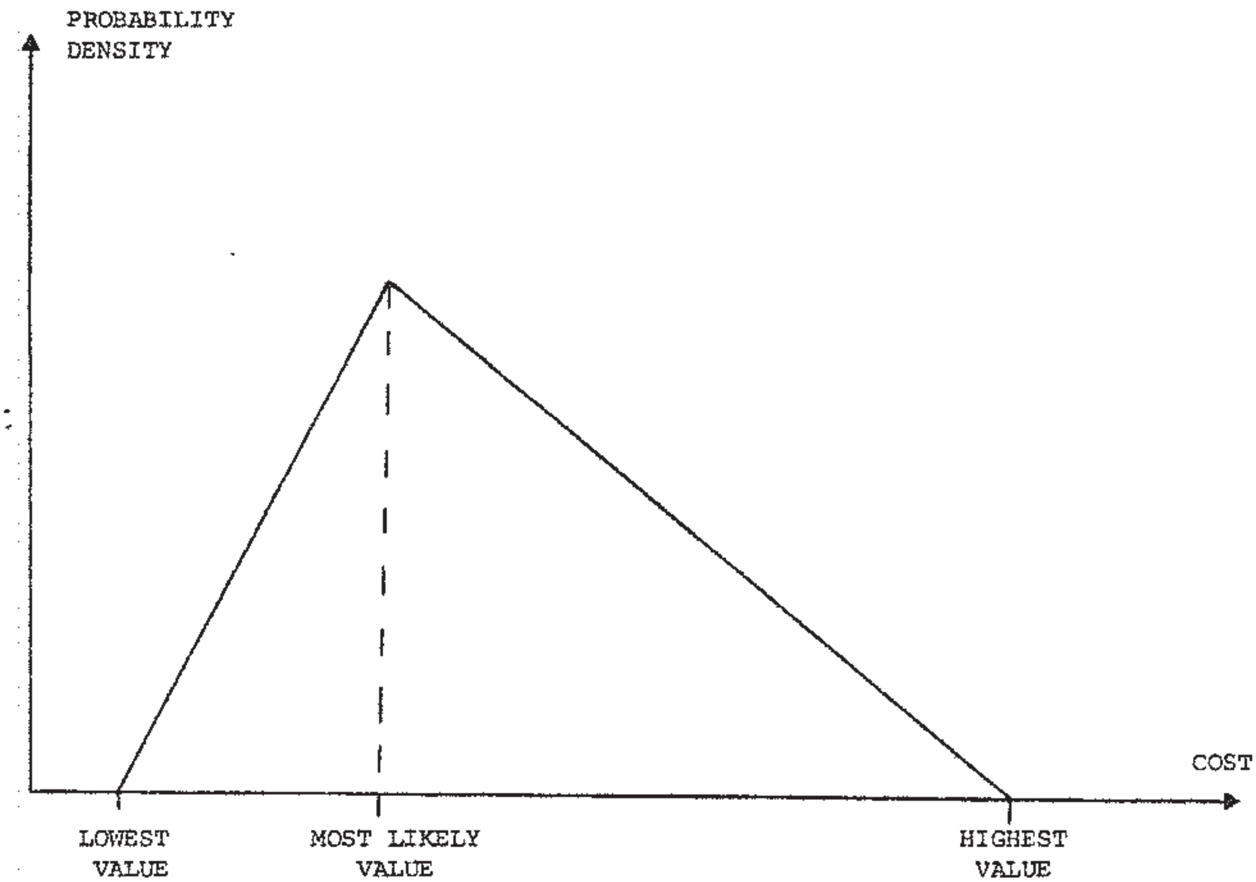


FIGURE 1

THE TRIANGULAR DISTRIBUTION

5. There are three other useful probability distribution functions which may be considered as degenerate cases of the three point triangular distribution. The first is the uniform or rectangular distribution in which there is no "most likely" cost so that each value between the lowest and highest is equally likely. The other two distributions are right-angled triangular distributions in which the most likely cost is equal to either the lowest or highest cost. These three distributions are shown in Figure 2. In the remainder of this paper, including examples, it is to be understood that these 3 distribution functions are included with the traditional triangular distribution whenever the term "triangular distribution" is used. Adoption of this convention greatly increases the utility and applicability of the method while causing no extra complexities.

COMBINING COST ESTIMATES

6. Having estimated the cost of two or more components of a system, it is almost always required to combine them to get a total cost. In order to do this, we first must assume that the individual cost components are independent. (This assumption will be discussed in detail later.) Given the most likely cost for each of two independent cost components, it is not true in general that the most likely total cost is the sum of the component most likely costs. This is a mathematical fact that is often ignored by or unknown to cost analysts. As an example, consider a system whose total cost is the sum of 3 individual components. The estimated lowest, most likely, and highest cost for each component is given in Table I. Independence is assumed.

TABLE I

EXAMPLE COMPONENT COSTS

	LOWEST VALUE	MOST LIKELY VALUE	HIGHEST VALUE
A.	$3.0 million	$ 7.0 million	$14.0 million
B.	$1.0 million	$ 1.5 million	$ 3.5 million
C.	$2.0 million	$ 3.0 million	$ 5.0 million
SUM	$6.0 million	$11.5 million	$22.5 million

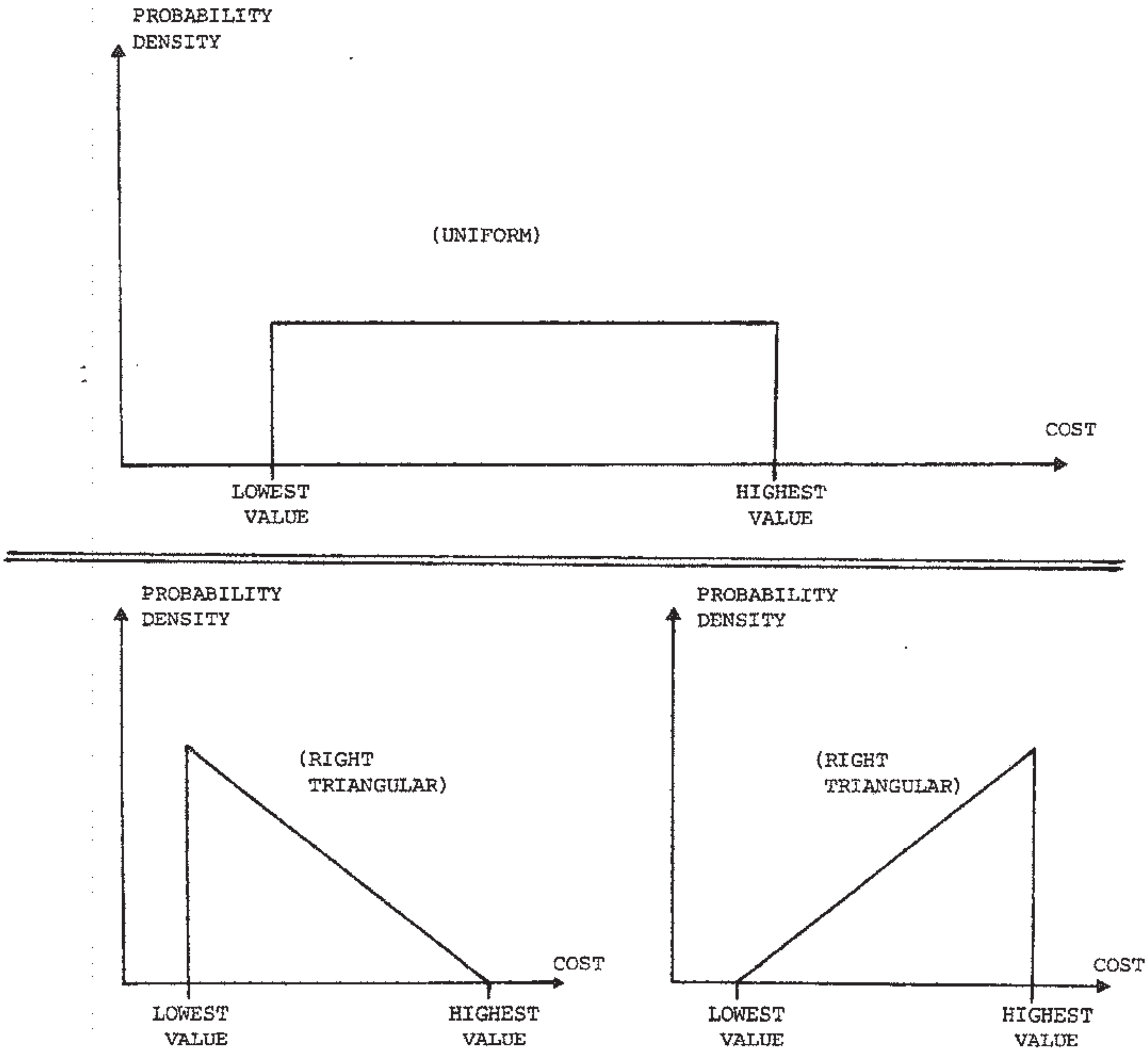


FIGURE 2

DEGENERATE TRIANGULAR DISTRIBUTIONS

7. The correct way to combine the three individual component costs is to sum the three triangular probability distributions to get the resultant probability distribution for the total cost. The most likely total cost as well as other pertinent statistics can then be read from the resultant distribution. It is not a simple matter to sum continuous probability distributions. The technique, called convolution, involves solving increasingly complicated integrals. (See Annex A.) Fortunately, it is fairly simple to evaluate these integrals numerically on a computer. For the example in Table I, the probability distribution for the total cost has been calculated using a computer to perform the numerical convolutions. Figure 3 shows the results.

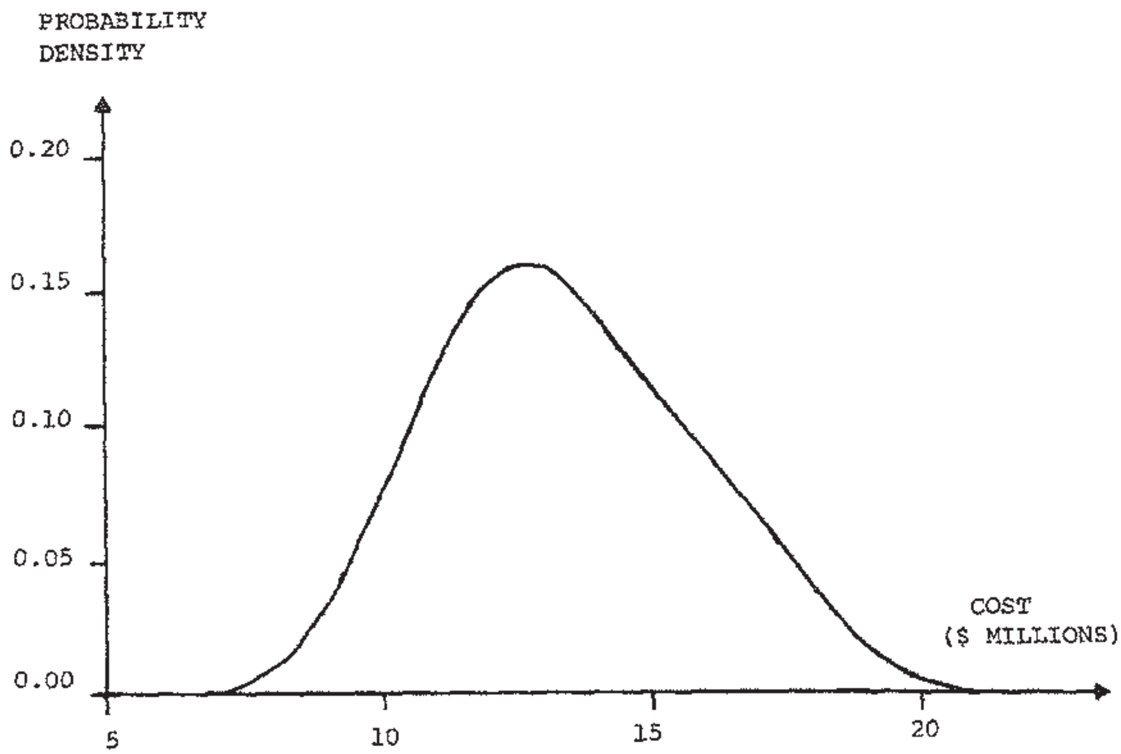
8. Note that the sum of the modes (most likely values) in Table I was $11.5 million but the mode in Figure 3 is $12.6 million. The sum of the three individual modes gives too low a value because the distributions are not symmetric; they are skewed to the right.

9. Having calculated the probability distribution for the total cost, we actually have much more information than simply the most likely total cost. For example, the expected or mean total cost is $13.3 million and the median or 50^{th} percentile cost is $13.1 million. Figure 4 shows the cumulative probability distribution curve for the total cost distribution of Figure 3. Using Figure 4, a confidence interval of any desired probability can be produced. For example, a 90% confidence interval for the total cost is $9.5 million to $17.5 million. Any desired percentile value can be obtained from Figure 4.

THE PROBLEM OF INDEPENDENCE

10. As stated earlier, individual cost elements must be independent in order to be combined using the method described earlier or any of the approximations in common use. Serious errors can result from combining cost elements which have been incorrectly assumed to be independent. Certain types of dependence can be dealt with.

.../6



Minimum value . . . . $ 6.0 million

Maximum value . . . . $22.5 million

Mode (most likely value) . . . $12.64 million

Median (50th percentile) . . . $13.14 million

Mean (average value) . . . $13.33 million

FIGURE 3

TOTAL COST DISTRIBUTION FOR THE EXAMPLE IN TABLE I

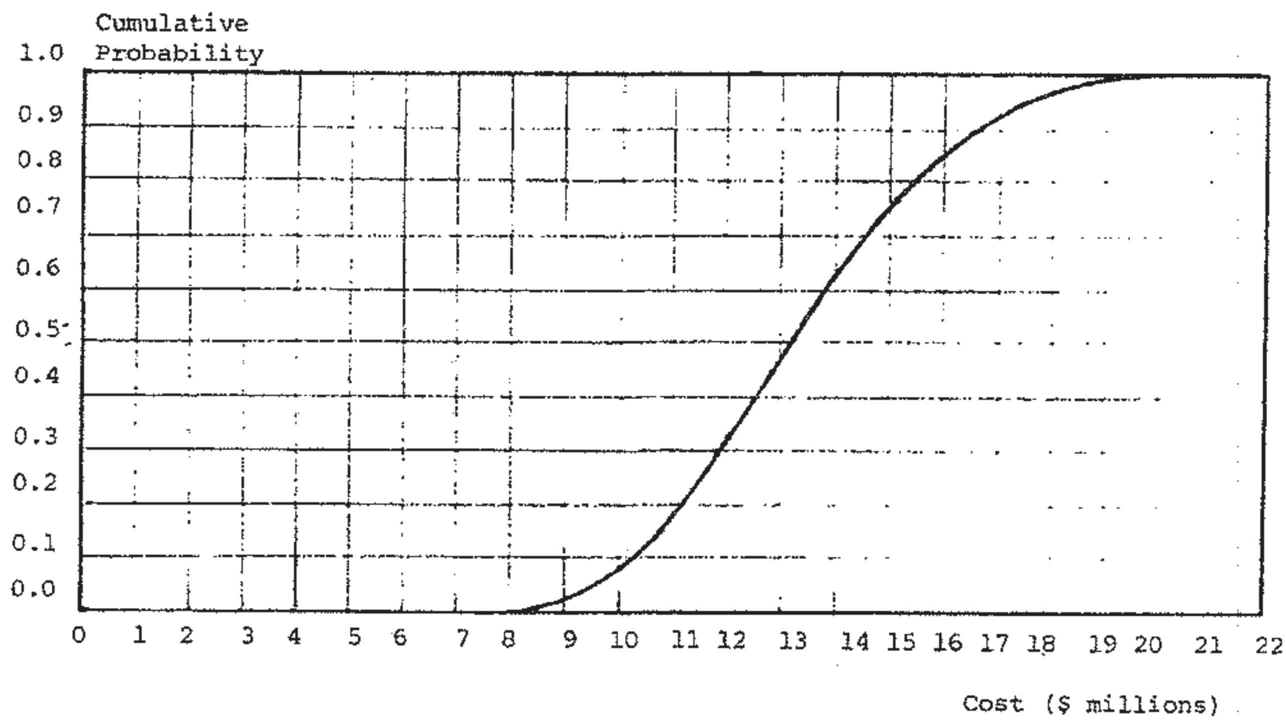


FIGURE 4

CUMULATIVE DISTRIBUTION FOR TOTAL COST OF TABLE I EXAMPLE

11. Given that two or more cost elements are not independent, it is often possible to handle this by redefinition into a single element which comprises the dependent ones. For example, in estimating the cost of a new computer, the cost of tape units may be related to the cost of disk units because spending more on either one implies spending less on the other. This could be handled by defining a cost element called "mass storage" which includes both tapes and disks.

12. Another type of dependence that can be dealt with is "perfect" or "near-perfect" dependence. In this type of dependence, two or more items have lowest, most likely, and highest cost distributions, but the nature of the items is such that the final or actual costs will be at nearly the same percentile points. That is, the final costs will be all low, medium, or high but not one at the lower end of its distribution and another at the medium or high end of its distribution. The solution in this case is to combine separately the cost distributions of the elements which have the perfect or near-perfect dependence into a single new element using a special technique. One such technique is to use a type of computer simulation in which a single randomly chosen value is used in each iteration to select values for the dependent cost elements. The resulting overall cost distribution for the perfectly dependent elements can then be treated as a single cost element. (It may be noted also that the analogous case of near-perfect negatively correlated dependence can be handled in a similar way.)

13. A common situation is one in which the final cost of one element is estimated as a percentage of another. If the percentage is exact, this situation is easily handled by adding the two elements together. For example if item B is estimated as 10% of item A, it is simply added into the cost of item A so that the lowest, most likely, and highest costs for item A are all increased by 10%. A more usual case is one in which the percentage is itself a random variable with lowest, most likely, and highest values. (Note that any of the degenerate triangular distributions are also allowed.) This case is handled by combining the two elements in a special way. It is necessary to assume that the "percentage" distribution

.../9

is independent of the cost distribution. If this is the case, it is fairly straightforward to combine the two elements using the same technique described earlier except that one is now multiplying random variables instead of adding. However, the same technique of convolution is still involved. As an example, consider the problem of combining the two elements given in Table II.

TABLE II

DEPENDENT COST ELEMENTS

	LOWEST VALUE	MOST LIKELY VALUE	HIGHEST VALUE
A.	\$ 3.0 million	\$ 7.0 million	\$14.0 million
B.	5% of A.	10% of A.	18% of A.

14. Figure 5 shows the original triangular distribution for item A and the resulting distribution when item B is combined. This resultant distribution is saved by the computer to be combined with any other cost elements in deriving the overall cost.

15. A final and more general type of dependence is one in which there is correlation between two or more cost elements. For example, suppose that the lowest, most likely, and highest costs have been estimated for items X and Y and that we wish to combine the two to get a total cost. Suppose also that the final cost of item Y is known to be linearly correlated with the final cost of item X. This means that we cannot use the usual convolution method to obtain the distribution of the sum of X and Y. A way to handle this case is presented below.

16. The simplest way to interpret the statement that random variable Y is correlated to random variable X is to assume that the expected value of Y given X is linear.

$$E(Y|X=x) = \alpha + \beta x$$

(α and β are constants.)

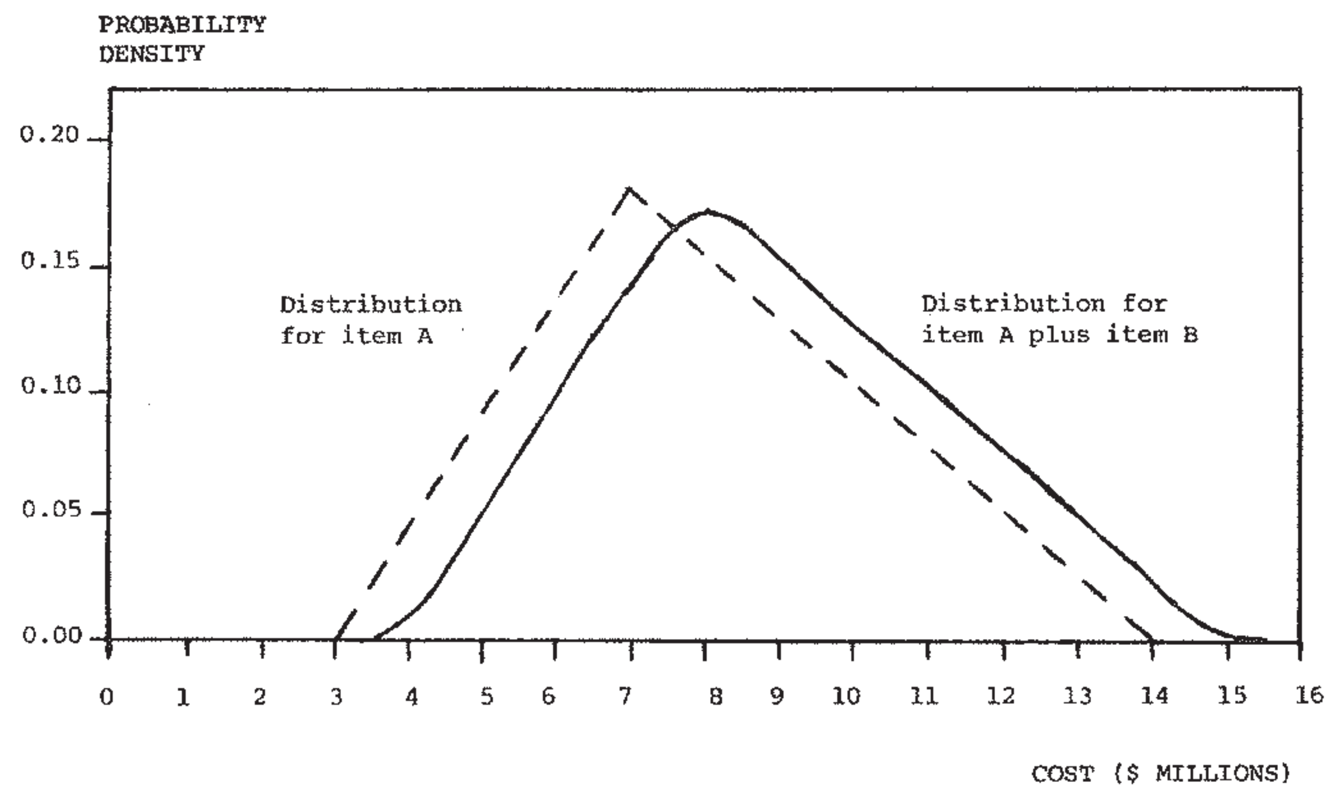


FIGURE 5

DISTRIBUTION FUNCTION FOR SUM OF DEPENDENT ELEMENTS OF TABLE II

17. The constants α and β may be written in terms of the moments of X and Y as follows.

$$E(Y|X=x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x) \qquad (1)$$

$\mu_x = E(X)$

$\sigma_x^2 = Var(X)$

$\mu_y = E(Y)$

$\sigma_y^2 = Var(Y)$

ρ = Correlation coefficient of Y and X. $(-1 \leq \rho \leq 1)$

18. If we further assume that the variance of Y given X is constant, we have the following expression for Var (Y|X).

$$Var(Y|X) = \sigma_y^2 (1-\rho^2) \qquad (2)$$

19. Equations (1) and (2) give the mean and variance of the distribution of Y given X. Using the estimated low, most likely, and high values for X and Y, we can calculate μ_x, σ_x, μ_y, and σ_y. It is necessary for the analyst to estimate the value of ρ, which is the correlation coefficient between X and Y. Note that if ρ is zero, X and Y are uncorrelated and E(Y|X) and Var(X|Y) are both unrelated to X. Similarly, if ρ is plus or minus one, the variance of Y given X is zero, and the value of Y is completely determined once X is known.

20. In order to estimate the distribution of the sum of X and Y, we must make an assumption about the distribution of Y given X. The most natural assumption is that this distribution is approximately Normal with mean and variance as given by equations (1) and (2). To derive the distribution of the sum of X and Y, we use a computer program in which for each iteration a random value for X is drawn from its triangular distribution. Given this value of X, a value of Y is chosen at random

.../12

from the distribution of Y given X which we have assumed is approximately Normal with mean and variance given by equations (1) and (2). The values of X and Y are summed and recorded for each iteration, thus building up the distribution of X + Y after many iterations.

21. As an example, consider the cost elements X and Y whose three point estimates are given in Table III.

TABLE III

CORRELATED COST ELEMENTS

	LOWEST COST	MOST LIKELY COST	HIGHEST COST
X	$ 1 million	$ 2 million	$ 4 million
Y	$ 1 million	$ 2 million	$ 3 million

22. Figures 6 and 7 show the cumulative distribution functions for various values of the correlation coefficient, ρ. In Figure 6, the three values of ρ used were: $\rho = 0.01$ (virtually independent), $\rho = 0.5$, and $\rho = 0.99$ (virtually perfect dependence). Note how the cumulative distribution curves flatten out as ρ increases from 0.01 to 0.99. This reflects the fact that positive correlation increases the variance of the sum of two random variables because the high (and low) values tend to occur together making very high (and very low) sums more probable. In Figure 7, the three values used for ρ were: $\rho = -0.01$ (virtually independent), $\rho = -0.50$, and $\rho = -0.99$ (virtually perfect dependence). In this case the cumulative distribution curves become steeper as ρ goes from -0.01 to -0.99, reflecting the fact that negative correlation decreases the variance of the sum of two random variables because a high value for either one tends to be coupled with a low value for the other, making very high and very low sums less probable.

A REAL EXAMPLE

23. To test the three point estimate method, it was decided to look for a suitable current or fairly recent capital expenditure project. The Sea King Helicopter replacement project proved to be a good choice.

.../13

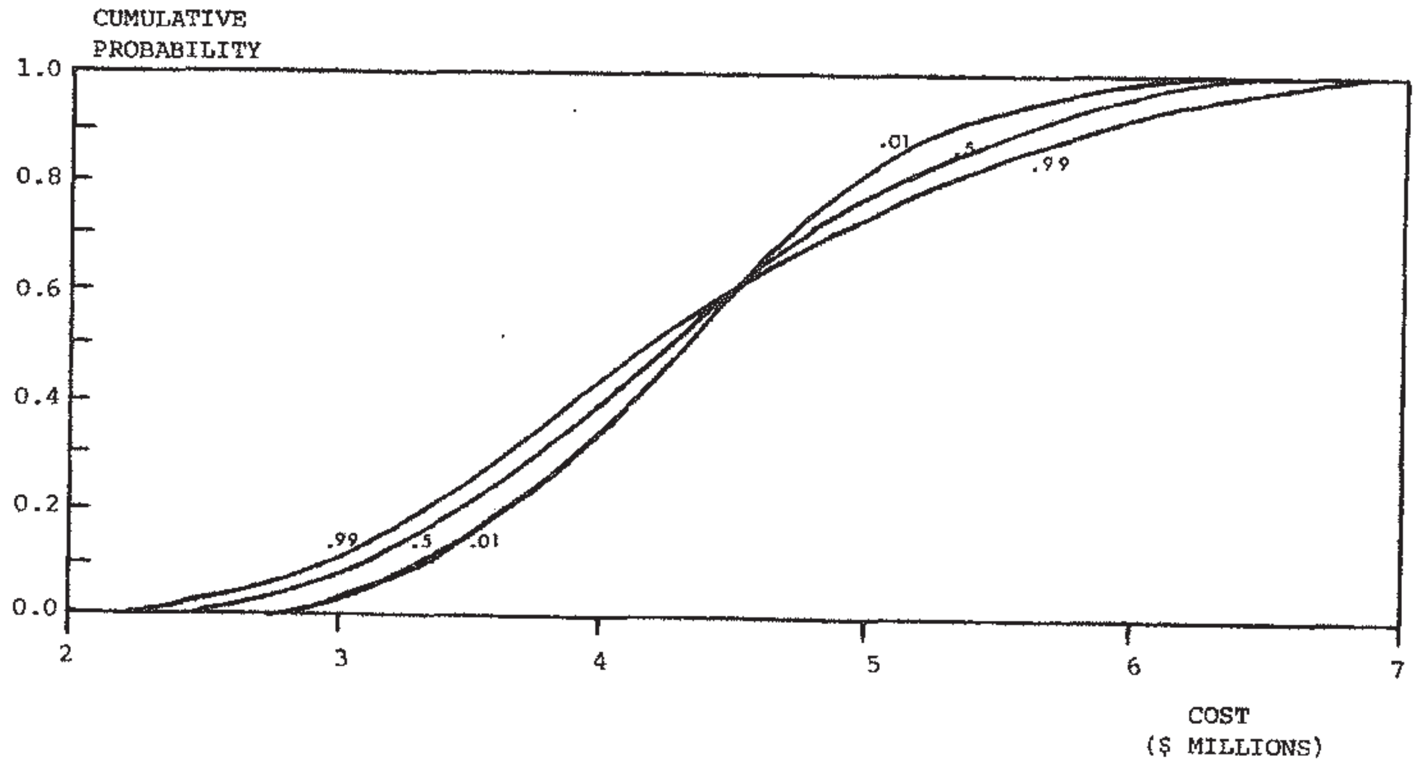


FIGURE 6

CUMULATIVE DISTRIBUTION CURVES FOR
CORRELATED COST ELEMENTS OF TABLE III

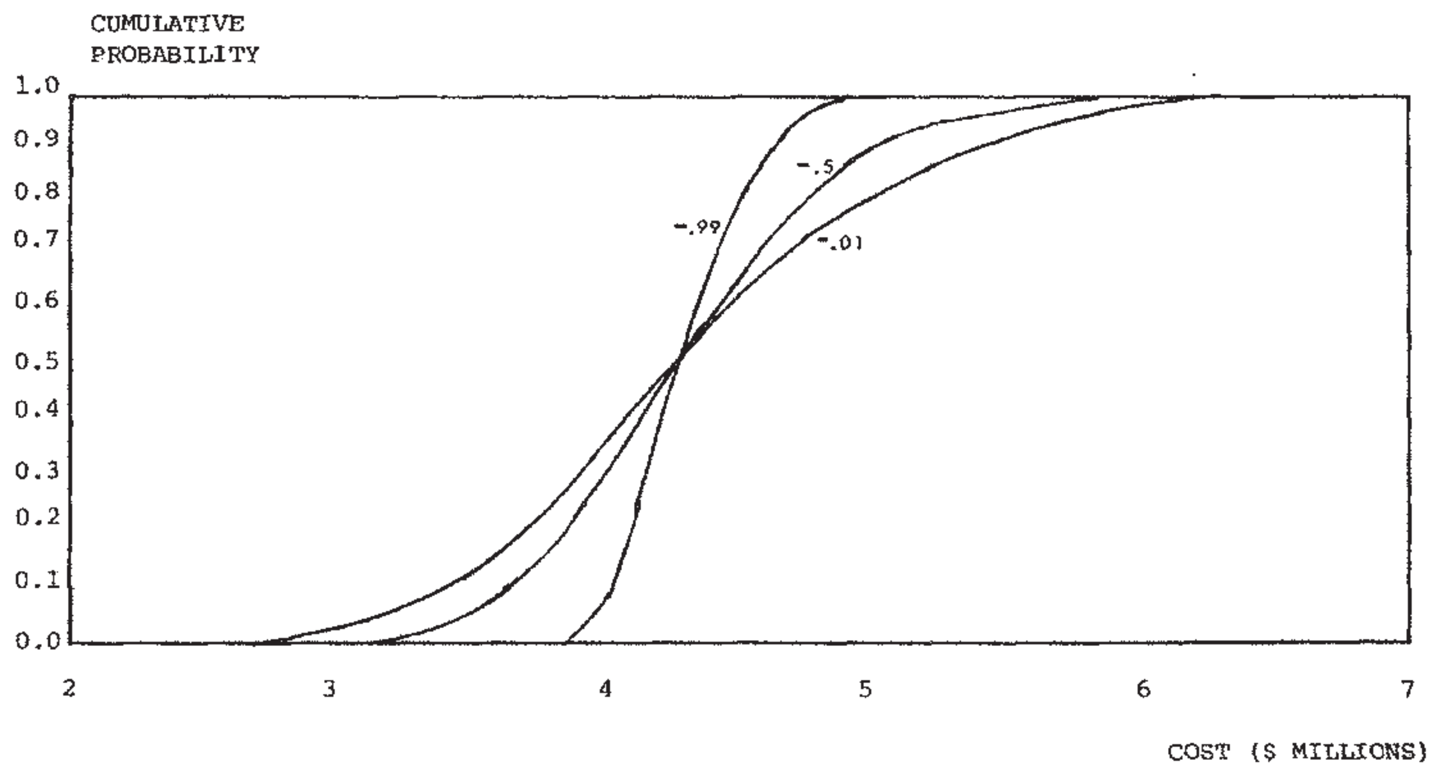


FIGURE 7

CUMULATIVE FREQUENCY CURVES FOR
CORRELATED COST ELEMENTS OF TABLE III

This project was at the stage in which capital cost estimates had been made for a set of well defined cost categories and for a variety of vehicle options. Also, the staff involved in making these estimates, DMAEM (6), were willing and able to provide the required information. The actual dollar values have been disguised to preserve data confidentiality.

24. The project was divided into nine cost categories, each of which contained from one to 15 cost elements. Overall, there were 33 cost elements which combined to produce the total cost.

25. For 25 of the 33 items low, most likely and high cost values were given. These were based on the analyst's judgement which in turn was based on research and expert knowledge. It was verified that the assumption of independence was reasonable in these 25 cases. This means that the eventual cost of any of the 25 elements was considered unlikely to be strongly influenced or correlated with any other. It should be noted that inflation and foreign currency exchange rates were not considered here.

26. Of the remaining 8 cost elements, 4 were given as a fixed percentage of the sum of various other items. These were handled as discussed earlier by combining them with the appropriate distributions. For example, consider an item which is estimated at a fixed rate of 10% of the "Prime System" cost. The "Prime System" cost distribution is calculated by combining its constituent cost items. The distribution is then adjusted by multiplying all the cost values by 1.10 to account for the related item.

27. The remaining 4 cost elements were also estimated as a percentage of items. However, in these cases, the percentage itself was estimated as a triangular distribution with low, most likely, and high values. As discussed earlier, a special technique was used to handle these cases.

28. Figure 8 shows the probability distribution for the total cost (of all 33 elements) for each of the two contenders. Note the increase in information provided by these distribution curves over a single value such as most likely cost. The difference in cost between the two contenders

.../16

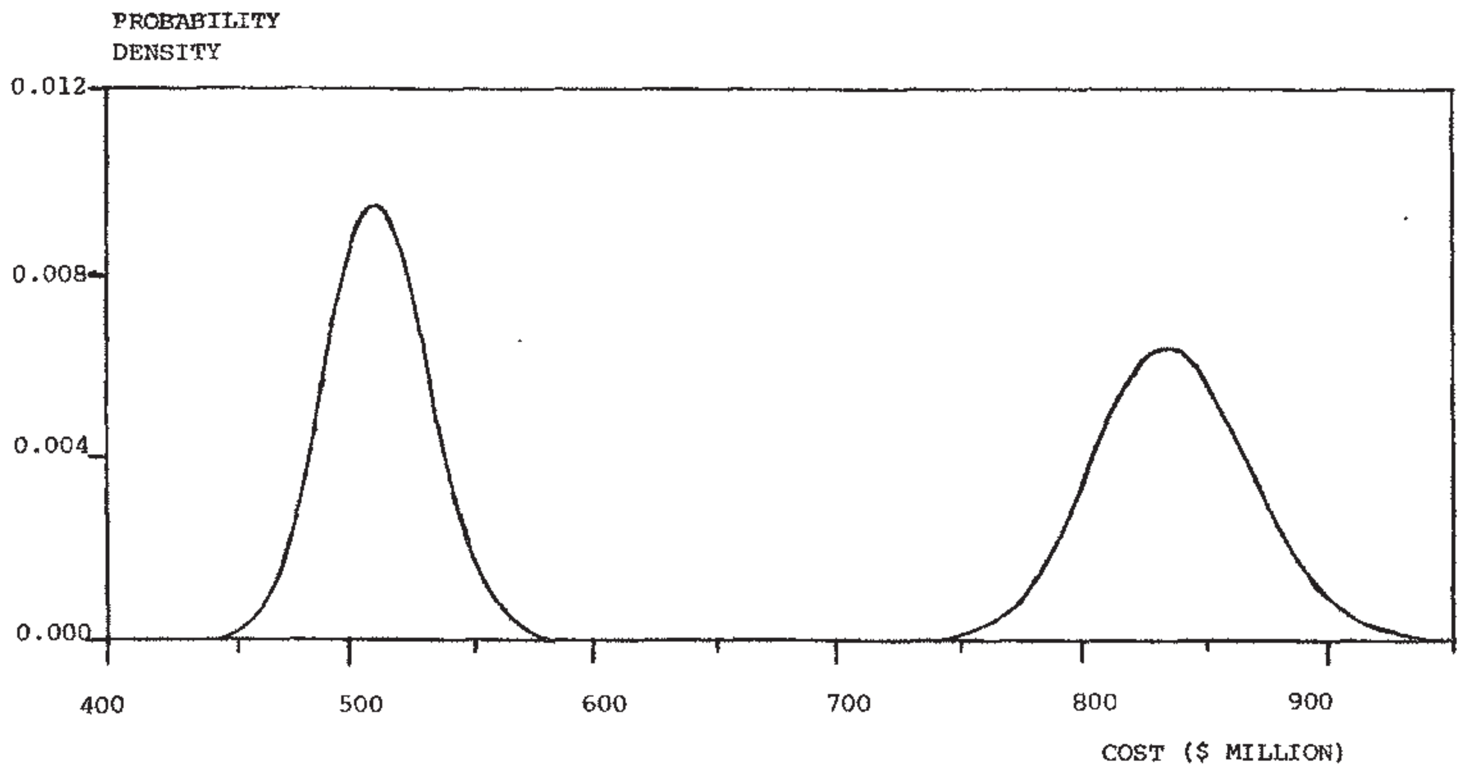


FIGURE 8

DISTRIBUTION FUNCTIONS FOR TOTAL COST OF TWO CONTENDERS can readily be seen, as can the monetary uncertainty or risk for each contender. The less expensive option has a most likely value of 510 million, and the more expensive option has a most likely value of 835 million, according to the methods of this paper. It is interesting to note that simply adding the most likely costs for each of the 33 cost elements gives values for the two options of 460 and 770 million. As mentioned earlier, the difference results from the fact that the sum of the modes (most likely values) of non-symmetrical distributions is in general not equal to the mode of the distribution of the sum.

29. The extra information on the likely spread of the final cost around the predicted most likely cost can be used as the basis for an alternative method of establishing a contingency allowance. Consider the less expensive contender in Figure 8. Given the most likely cost of 510 million, the conventional method is to add 10% for contingency, for a total project ceiling of 561 million. According to Figure 8, this value is at virtually the 100th percentile. In this case the 10% allowance practically guarantees that the project manager will have enough funds to provide for any upward random variation in project cost. An alternative method would be to set a project goal of 510 million, the most likely cost. There is approximately a 50 percent chance that the final cost will be less than this value. Also, according to Figure 8, there is approximately a 90 percent chance that the final cost will be less than 550 million. Thus a contingency fund of 40 million could be allocated with the expectation that this amount will not normally be required except in rare cases. Such a procedure would provide a rational link between project cost estimates and contingency allowances.

CONCLUSION

30. The three point estimate method (along with the techniques for handling dependencies) is a practical and easily applied procedure for estimating project cost totals. It provides more information than current methods, gives an explicit measure of the cost uncertainty, and provides a rational way of determining the amount of contingency funding.

ADDING RANDOM VARIABLES: CONVOLUTIONS

1. In order to explain the method of summing random variables, the discrete case will be discussed first because it is easiest. Let X and Y be discrete random variables. Thus X assumes the values $x_1, x_2, \ldots, x_m$ with probabilities $\alpha_1, \alpha_2, \ldots, \alpha_m$. Similarly Y assumes values $y_1, y_2, \ldots, y_n$ with probabilities $\beta_1, \beta_2, \ldots, \beta_n$. Suppose X and Y are independent and let Z = X + Y. Then we can calculate the probability distribution of Z as follows.

$$\text{Prob } (Z = z_0) = \sum_{i=1}^{m} \text{Prob}(X=x_i)\text{Prob}(Y = z_0 - x_i) \qquad (1)$$

The values which Z can assume are all possible sums $(x_i + y_j)$ where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. There are at most mn such values.

2. Another equivalent way to state equation 1 is given by equation 2.

$$\text{Prob}(Z = z_0) = \sum_{S} \alpha_i \beta_j \qquad (2)$$

S is the set of all pairs (i,j) where $x_i + y_j = z_0$. (Note that S may be empty.)

3. In the continuous case, the summation signs are replaced by integral signs and the probabilities are replaced by probability density functions. Let X be a continuous random variable with density function f(x), and let Y be a continuous random variable with density function g(y). Suppose again that X and Y are independent and let Z = X + Y. Analogous to equation 1, we can write equation 3 as follows.

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx \qquad (3)$$

4. In practice, it is rare that equation 3 can be solved explicitly. A common procedure is to approximate the continuous random variables by discrete ones and then use equation 2 to produce an approximation to $h(z)$. This procedure can be made as accurate as desired by using increasingly accurate discrete approximations to the continuous random variables.

BIBLIOGRAPHY

1. "Determining Cost Uncertainty in Bottoms-Up Estimating", R.L. Black and J.J. Wilder, Journal of Parametrics, June 1984.

2. "An Application of Subjective Probabilities to the Problem of Uncertainty in Cost Analysis", H.R. Jordan and M.R. Klein, U.S. Navy, November 1975.

3. "The Analysis, Management and Reporting of Risk, Maj B.F. Hough, ORAE Memorandum No. M118, April 1985.