32nd Annual Department of Defense Cost Analysis Symposium

Approximating the Probability Distribution of Total System Cost

Paul R. Garvey
Chief Scientist
The Economic and Decision Analysis Center
The MITRE Corporation

From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.

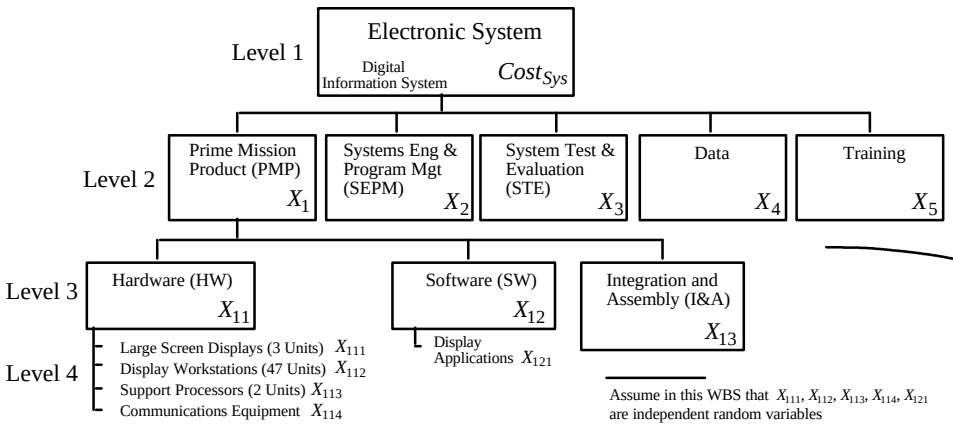
MITRE

Main Points

- The probability distribution of a system's total cost can *often* be approximated analytically by a "simple formula"
- Analytical approximations provide insights into cost risk drivers and other influences not readily seen in empirical distributions (e.g., Monte Carlo simulations) of a system's total cost
- Five circumstances are presented for identifying when the probability distribution of a system's total cost can be approximated analytically
- The normal or the lognormal distribution will almost always provide a good analytical approximation to the underlying distribution function of a system's total cost

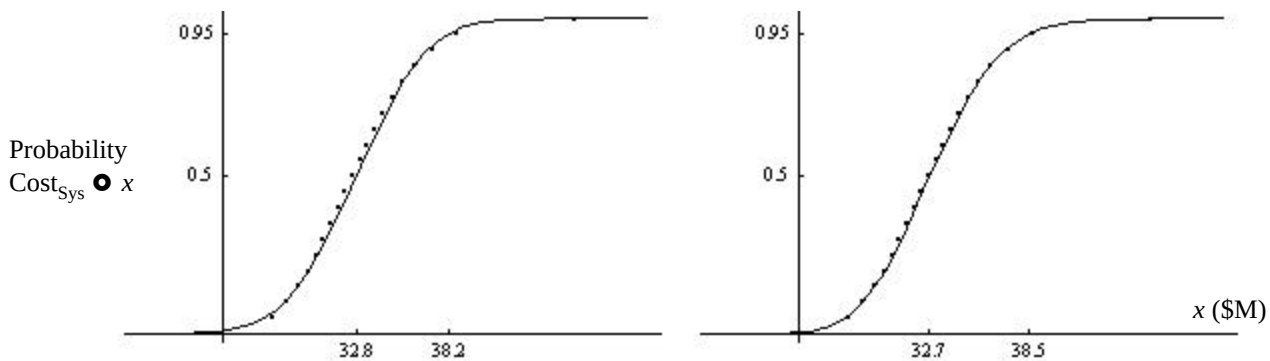
From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.

A Work Breakdown Structure Perspective



$$Cost_{Sys} = Cost_{PMP} + \sum_{i=2}^{n} X_i \quad (6-88)$$

where $Cost_{PMP} = X_1$



Solid Line- Assumed Normal
Dots- Monte Carlo Simulation

Solid Line- Assumed LogNormal
Dots- Monte Carlo Simulation

From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.

MITRE

Circumstance A

Circumstance A

If (in equation 6-88) the distribution function of $Cost_{PMP}$ *is normal and* $X_2, X_3, X_4, \ldots, X_n$ *are linear functions of* $Cost_{PMP}$, *such as* $X_i = a_i\, Cost_{PMP}$ *where* $a_i \geq 0$ $(i = 2, \ldots, n)$, *then the distribution function of* $Cost_{Sys}$ *is normal with mean*

$$(1 + a_2 + a_3 + \ldots + a_n) E\left(Cost_{PMP}\right)$$

and variance

$$(1 + a_2 + a_3 + \ldots + a_n)^2 Var\left(Cost_{PMP}\right)$$

Circumstance A is a direct consequence of the following:

Proposition 6-1 If X is a normal random variable and $Y = a\,X$, where a is a constant, then the distribution function for Y is normal with mean $a\,E(X)$ and variance $a^2 Var(X)$.

From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.

MITRE

Circumstance B

Circumstance B

If (in equation 6-88) $Cost_{PMP}$ *and* $X_2, X_3, X_4, \ldots, X_n$ *are independent random variables and each are normally distributed, then the distribution function of* $Cost_{Sys}$ *is normal with*

$$mean \quad E(Cost_{PMP}) + \sum_{i=2}^{n} E(X_i) \text{ and variance } Var(Cost_{PMP}) + \sum_{i=2}^{n} Var(X_i)$$

Circumstance B is a direct consequence of the following:

Proposition 6-2 If $X_1, X_2, X_3, \ldots, X_k$ are independent *normally* distributed random variables and $Y = X_1 + X_2 + X_3 + \ldots + X_k$, then, regardless of the size of *k*, the distribution function of *Y* is normal with mean $\sum_{i=1}^{k} E(X_i)$ and variance $\sum_{i=1}^{k} Var(X_i)$.

From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.

MITRE

Circumstance C

Circumstance C

Suppose (in equation 6-88) $Cost_{PMP}$ *is normal and* $Cost_{PMP}, X_2, X_3, X_4, \ldots, X_n$ *are independent random variables. Furthermore, suppose* $X_2, X_3, X_4, \ldots, X_n$ *are not necessarily each normally distributed. If the number of cost element costs in the sequence* $X_2, X_3, X_4, \ldots, X_n$ *is sufficiently large with no* X_i $(i = 2, \ldots, n)$ *dominating in standard deviation, then the distribution function of* $Cost_{Sys}$ *is approximately normal with*

$$mean \quad E(Cost_{PMP}) + \sum_{i=2}^{n} E(X_i) \text{ and variance } Var(Cost_{PMP}) + \sum_{i=2}^{n} Var(X_i)$$

Circumstance C is a combination of the central limit theorem and proposition 6-2. To see this, recall from equation 6-88 $Cost_{Sys}$ can be written as

$$Cost_{Sys} = Cost_{PMP} + \sum_{i=2}^{n} X_i$$

If the distribution function for $Cost_{PMP}$ is normal and the distribution function of the sum $\sum_{i=2}^{n} X_i$ is approximately normal (by the central limit theorem), then $Cost_{Sys}$ is the sum of two normally distributed random variables. In circumstance C, $Cost_{PMP}$ and $\sum_{i=2}^{n} X_i$ are independent. Thus, from proposition 6-2, the distribution function of $Cost_{Sys}$ is normal.

From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.

Circumstance D

Circumstance D

Suppose (in equation 6-88) $Cost_{PMP}$ *is normal. Suppose the sequence* $X_2, X_3, X_4, \ldots, X_n$ *contains some cost element costs correlated to* $Cost_{PMP}$ *(with correlation coefficient* r_{Cost_{PMP}, X_i}*) and some that are uncorrelated to* $Cost_{PMP}$*. Furthermore, suppose those that are uncorrelated to* $Cost_{PMP}$ *are mutually independent random variables. If the number of* X_i*'s* $(i \geq 2)$ *uncorrelated to* $Cost_{PMP}$ *is sufficiently large, with none dominating in standard deviation, then the distribution function of* $Cost_{Sys}$ *is approximately normal with*

$$\text{mean } E(Cost_{PMP}) + \sum_{i=2}^{n} E(X_i)$$

$$\text{and variance } Var(Cost_{PMP}) + \sum_{i=2}^{n} Var(X_i) + 2\sum_{i=2}^{n} r_{Cost_{PMP}, X_i} s_{Cost_{PMP}} s_{X_i}$$

If the X_i*'s* $(i \geq 2)$ *uncorrelated to* $Cost_{PMP}$*, which are also mutually independent, are each normally distributed, then the distribution function of* $Cost_{Sys}$ *is approximately normal with mean and variance given above; here it is not required there be sufficiently many of these* X_i*'s.*

From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.

MITRE

Circumstance E

Circumstance E

Suppose (in equation 6-88) $Cost_{PMP}, X_2, X_3, X_4, \ldots, X_n$ *are independent random variables. Furthermore, suppose* $Cost_{PMP}, X_2, X_3, X_4, \ldots, X_n$ *are* <u>*not*</u> *necessarily each normally distributed. If the number of cost element costs in the sequence* $Cost_{PMP}, X_2, X_3, X_4, \ldots, X_n$ *is sufficiently large with none dominating in standard deviation, then (by the central limit theorem) the distribution function of* $Cost_{Sys}$ *is approximately normal with*

$$\textit{mean } E(Cost_{PMP}) + \sum_{i=2}^{n} E(X_i) \textit{ and variance } Var(Cost_{PMP}) + \sum_{i=2}^{n} Var(X_i)$$

In all but the last circumstance discussed above, the distribution function for $Cost_{PMP}$ was given to be normal. This is common in electronic systems. The normality of $Cost_{PMP}$ is primarily driven by the central limit theorem, where $Cost_{PMP}$ typically reflects the sum of many *independent* hardware and software costs.

From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.

MITRE

In Summary

- Cost analysts should understand the mathematical relationships they define in a system's WBS to see whether analytical approximations to the probability distribution function of $Cost_{Sys}$ can be argued
- A class of circumstances exist that indicate when a *normal* distribution approximates the probability distribution function of $Cost_{Sys}$; recognizing when these circumstance apply, can save time and resources needed for complex Monte Carlo simulations
- It is typical to have a mixture of independent and correlated cost element costs in a system's WBS; howver, because of the central limit theorem, the greater the number of independent cost element costs the more it is that the distribution function of $Cost_{Sys}$ is approximately normal

From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.

In Summary

- The central limit theorem is very powerful; it does not take many independent cost element costs for the distribution function of $Cost_{Sys}$ to move towards normality - even when correlation is present between some cost element costs
- Such a move is evidenced when
  - A sufficient number of independent cost element costs are summed
  - When no cost element's cost distribution has a much larger standard deviation than the standard deviations of the other cost element cost distributions
- When conditions in the WBS result in $Cost_{Sys}$ being positively skewed, then the *lognormal* often approximates the probability distribution function of $Cost_{Sys}$

From Garvey, P. R., *"Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective"*, published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 1999.