SCHEDULE RISK

by

M.R. Flint [1]

Abstract

To correctly account for risk within a work schedule can be a time consuming and difficult task to do in a valid and defensible manner, given the complex, networked nature of most schedules. To provide also the functionality to cater for interdependence of component times poses a quantum leap in that difficulty. Yet, as this paper contends, interdependence of component times exists in most schedules and should not be ignored. The consequences could prove to be expensive.

This paper addresses the subject of 'schedule risk' and how it may be determined in a valid and defensible manner. ***Key Words****: risk, risk analysis, schedule risk, Total Schedule Time, interdependence, correlation.*

INTRODUCTION

Purpose and Scope

1. This paper addresses the subject of 'schedule risk' and how it may be determined in a valid and defensible manner. In doing so, it addresses an underlying management objective to have the best estimate of the statistical distribution that best represents the behaviour of the Total Schedule Time, as a stochastic variable. Having this information, allows management to properly assess the risk associated with the schedule and the impact thereof on other objectives, particularly project cost.

2. Schedule risk is represented by a probability distribution about its Mode, for the Total Schedule Time. Likewise, the risk associated with each activity comprising the schedule should be represented in the same way, given that the Total Schedule Time is a function of its component activities.

3. As an organised program of activities, a schedule may be a simple series of activities, a set of parallel activities or a complex combination of both, ie a network of activities. Because the latter is almost always the case in practice, risk analysis needs to follow an orderly process, one step at a time, from start to finish through the schedule.

4. Given details of a network, the process will be to determine (by estimation or assumption), in sequence:

- a probability distribution of time for each activity comprising the network;
- the extent of correlation between activities;
- the critical path of the network;
- the most likely value for the Total Schedule Time (Mode of the critical path);
- the probability distribution for the total time of the critical path; and
- the Mean of the distribution and its Range for the specified Confidence Limits.

Background

5. See Annex A for relevant background to this paper.

Terminology

6. See Glossary at Annex B.

---

[1] *Mr Flint was a private consultant with many years experience as a maintenance engineer and logistician in the RAAF, some seven years as a director in Defence concerned with policy for Defence-related industry and six years as a director in Defence of capital equipment acquisitions. From 1996, he worked with major suppliers in preparation of responses to Australian Department of Defence Requests for Tender; and with project offices within the Australian DOD, often in the discipline of Life Cycle Costing and risk management.*

7. Abbreviations and other terms used herein are listed below:

TT	Total Schedule Time (of the critical path).
T	Time of an individual activity in the schedule network
TTMode	the Mode of a TT distribution
TTMean	the Mean of a TT distribution (triangular distribution assumed) = (TTLow + TTMode + TTHigh) / 3
TTMax	the highest value in the TTRange.
TTMin	the lowest value in the TTRange.
TTRange	TTMax – TTMin
Confidence Level (CL)	the specified percentage of all values in a TT distribution or, conversely, the probability that the TT is within the TTRange.
Distribution	the assumed statistical distribution for a variable.
Mode	the most frequently occurring value in a single node distribution.
Sigma	Standard Deviation (SD) of a statistical distribution.
SigmaLow	Standard Deviation (SD) of left-hand side of triangular distribution
SigmaHigh	Standard Deviation (SD) of right-hand side of triangular distribution
Variance	a standard measure of spread of a statistical distribution; equal to the square of the Standard Deviation.
VarianceLow	Variance of left-hand side of triangular distribution
VarianceHigh	Variance of right-hand side of triangular distribution
RangeLow	TTMode -TTMin, in absolute values (assumed = 3* SigmaLow)
RangeHigh	TTMax - TTMode, in absolute values (assumed = 3* SigmaHigh)
%RangeLow	(TTMode -TTMin)/TTMode*100
%RangeHigh	(TTMax - TTMode)/TTMode*100
Co-variance	The relative variance between two distributions (i,j) being summed; represented by the formula: $\Sigma\Sigma$[Sigma (i) * Sigma (j) * Rho (i,j)], i<j
Rho(i,j)	Coefficient of Correlation between a pair of distributions (i, j).
Distribution Pair	Any of the nC2 pairs of distributions (i,j).
nC2	The number of combinations of n entities taken two at a time. nC2 = FACT(n)/(FACT(n-2)*FACT(2)).

DISCUSSION

Details of the Network

8. Details of the network will comprise the normal information required by a project management tool such as Microsoft Project, being at least:

- the start date for the schedule;
- the title of each activity;
- the most likely value for duration of each activity (plus the unit of time);
- estimate of minimum duration of each activity;
- estimate of maximum duration of each activity; and
- precedents for start and finish of each activity

9. Earliest start dates and latest finish dates are optional.

10. This information should be presented in a table in or importable as a table to a spreadsheet tool.

Probability Distribution of Activity Times

11. In most instances, analysts will assume that the probability distributions for activity times are either Triangular or Beta, but other shapes may be used if data is available to support the choice. Once selected, the same form should be used for all activities. Both distributions use the same Minimum,

Mode and Maximum estimates (three-point estimate); but formulae for calculations of Means and Standard Deviations will be different and resultant values of Means and Standard Deviations may be different, depending on the skewness of the distribution.

12. This paper assumes Triangular distributions.

Correlation Between Activities

13. A principal contention of this paper is that risk analysis, be it of schedules (or costs), cannot validly ignore the phenomenon of correlation[2] between components of schedule (or cost). Unfortunately, too many textbooks and risk modelling tools tend to ignore correlation by assuming total independence of all components.[3]

14. Correlation between a pair of variables (activity time in this case) may be assumed to have the value:

- zero (activity times are assumed independent);
- +1 or -1 (activities are assumed totally interdependent, positively or negatively); or
- a value between –1 and +1 (some degree of interdependence).

15. One should not assume away the existence of correlation. Would a reader seriously suggest that scheduled activities in a network are totally independent? While some pairs may indeed be independent, it is contended here that most pairs will be interdependent (correlated to some extent). For example, if there is a risk of a building boom, resources will become scarcer, which would affect many related activities in the same way.

16. Most scheduling tools and even simulators tend to ignore the existence of correlation, ie assume total independence of variables. Under this assumption, a series of activity distributions would be summed as follows:

- TTMode (z) = Σ Mode (i), i = 1,..n activities;
- TTVariance (z) = Σ Variance (i), i = 1,..n activities; and
- TTRange (z) = $\sqrt{}\ \Sigma$ Range (i), i = 1,..n activities.
- TTMean (z) = = (TTMin (z) + TTMode (z) + TTMax (z)) / 3 (triangular distribution only).

17. Certain textbooks[4] that discuss scheduling, particularly PERT, use the foregoing formula for variance. This is a simplifying assumption that gives an invalid result, simply because the assumption of total independence is invalid. The same invalid assumption is found in texts discussing cost risk analysis. This assumption can seriously underestimate the variance in the Total Schedule Time. As an example, for 'n' identical activities in series (identical in Mode and Variance), it can be shown that the ratio of the respective total variances (and ranges) for the two assumptions (total independence to total interdependence) can be as great as of $1/\sqrt{n}$. For a series of 20 activities on the critical path, the common assumption of total independence would produce a variance and range of some 4.47 ($\sqrt{20}$) smaller than the opposite assumption and a Standard Deviations (SD) about 2.1 times smaller. Note that these values are in absolute terms. A better indication of risk is to express a range as a percentage of the Mode (%RangeLow and %RangeHigh).

18. In practice, neither of the extreme assumptions is ever correct; the answer is somewhere in between the two. However, providing for correlation between activities in a statistically valid manner poses considerable difficulties. Assessment and application of correlations is highly subjective and time

---

[2] Throughout this paper, correlation is taken to mean a degree of interdependence between components of schedule or cost and vice-versa. However, it is acknowledged that correlation measures only an 'apparent' relationship between two variables and does not claim to establish a cause and effect relationship as such.

[3] Some risk modelling tools make some provision for correlation but the validity of algorithms used is difficult to determine because of proprietary intellectual property rights..

[4] For example, Project management, a Systems Approach to Planning, Scheduling and Controlling, H. Kerzner, John Wiley & Sons, 1998, Page 662.

consuming, even by experienced analysts. In addition, the mathematics (statistical applications), while well founded theoretically and not difficult in themselves, require computer algorithms and power to do the number-crunching involved.

19. However, because of the mathematical relationship that can be shown to exist between the outer and inner bounds of the triangular distribution (total correlation and zero correlation respectively), reasonably valid approximations may be developed to estimate the range of the total activity time (Minimum to Maximum), for rough estimates of intermediate values of overall levels of correlation. For example, if the analyst believes, a priori, that there is a medium degree of overall correlation between activities, the estimated bounds (range) may be taken as the mid-points between the outer and inner bounds (maximum and minimum ranges respectively). This method of approximation would be adequate for most applications.[5]

Critical Path

20. One may safely assume that the Total Schedule Time is the summation of the activities on the critical path of a network. However, the valid determination of the distribution for Total Schedule Time is not a trivial matter.

21. To identify the critical path of a complex schedule network and its distribution, one needs to:

- first identify all of the nodes in the network, including the start and finish nodes;
- identify the branches in parallel between each pair of nodes;
- identify and quantify (three-point estimates) the activities in series in each branch;[6]
- determine the activity time distribution for each branch, between each pair of nodes;
- determine the activity time distribution between each pair of nodes, by selecting the largest of the parallel distributions; and
- except for the first, repeat the foregoing steps through the network from one node to the next, from start to end node.

22. At this point, one would have established the critical path and its distribution.

23. In the foregoing Dot Point 6, what is meant by 'the largest of the parallel distributions'? Given that all distributions are triangular and assuming total independence, the largest of 'i' parallel distributions defines the resultant distribution of the elapsed time at Node(j). The triangular distribution is determined by:

- a Minimum = Max [Minimum (i)];
- a Mode = Max [Mode (i)]; and
- a Maximum = Max [Maximums (i)]

[5] In a similar way, several levels, say low, medium, high, may be defined for correlation and respective spreads about the Mode determined.

[6] Risk may be estimated as the Maximum and Minimum values about the Mode but this can be very subjective. An alternative approach, which is used in cost risk analysis, is to define categories of schedule risk as plus and minus percentages of the Mode. A minimum of three and a maximum of five categories are recommended, say Very Low, Low, Medium, High, Very High, each with its associated plus and minus per cent value.

24. If the parallel activities are assumed to be totally interdependent, the distribution of the largest may be approximated by:

- a Minimum = Max [Minimum (i)];
- a Mode = Average [Mode (i)]; and
- a Maximum = Max [Maximums (i)].

25. If the parallel activities are assumed to be partially interdependent (some degree of correlation), the distribution of the largest may be approximated by:

- a Minimum = Max [Minimum (i)];
- a Mode = Average {Average [Mode (i)], Max [Mode (i)]}; and
- a Maximum = Max [Maximums (i)].

26. In the cases of interdependence (some correlation), it should be noted that the resultant distribution of parallel activities entering Node(j) may not in fact be triangular but may take any unimodal form between triangular, through beta to uniform (rare), depending on the assumed distributions of parallel activities and extent of correlation between pairs of activity paths.[7]

27. Once all parallel paths between nodes have been expressed as single times and distributions, the critical path may be either completed (if end node receives parallel paths) or be a series of internodal times. If the latter, the series has to be summed to a total time and distribution.

28. Determination of the probability distribution for the total time of the critical path can be cumbersome. In the general case, the following formulae need to be used in the summation of a series of 'i' activity distributions:

- Mode (z) = Σ Mode (i), i = 1,...n
- Rho (i,j) = {-1 to 1}, for all pairs [i,j]
- Variance (z) = Σ Variance (i) + 2 * ΣΣ [Sigma (i) * Sigma (j) * Rho (i,j)], i<j
- Sigma (z) = √ Variance (z)
- Low (z) = Mode (z) - x * SigmaLow (z)
- High (z) = Mode (z) + x * SigmaHigh (z)
- Range (z) = [Mode (z) - x * SigmaLow (z)] to [Mode (z) + x * SigmaHigh (z)]
- [For x = 3, Confidence Level = 99.6 per cent]
- Mean (z) = (Low (z) + Mode (z) + High (z)) / 3
- where:
  - z = the summed distribution.
  - Mode (z) = TTMode
  - i = the first activity of a pair, in each of the (i, j) pairs addressed for correlation.
  - j = the second activity of a pair, in each of the (i, j) pairs addressed for correlation.
  - n = number of activity distributions to be summed.
  - Rho (i,j) = Coefficient of Correlation between a pair of distributions (i, j).
  - ΣΣ = Sum of [Sum of terms over i = 1 to (j-1)] over j = 2 to n.

29. Once the three-point distribution for the sum of the series is established, the mean and percentile values may then be calculated.

[7] At this point of investigation, the author has not determined, through a literature search or otherwise, the theoretical distribution of parallel activities that involve correlation. Further work is required on this aspect.

Mean and Range - Critical Path

30. The range of a triangular distribution is assumed to embrace six Standard Deviations and 99.6 per cent of values.

31. However, where the triangular distribution is skewed significantly to the right, which is normally the case in practice for both total costs and Total Schedule Time, a very useful approximation[8] is to consider the distribution as two separate triangular distributions, one either side of the Mode. Each of the ranges (negative and positive) may be considered to cover three Standard Deviations. Percentile values can be determined accordingly.

32. The Mean of the distribution may also be calculated using the standard formula for a triangular distribution.

Conclusions

33. To correctly account for risk within a work schedule can be a time consuming and difficult task to do in a valid and defensible manner, given the complex, networked nature of most schedules. To provide also the functionality to cater for interdependence of component times poses a quantum leap in that difficulty. Yet, as this paper contends, interdependence of component times exists in most schedules and should not be ignored. The consequences could prove to be expensive.

34. Notwithstanding the mathematical difficulties and the subjective, albeit skilled nature of schedule risk analysis, useful approximations can be formulated to take the pain out of the process.

M.R. Flint
Director
Acquisition and Logistics Study Centre
Canberra

29Dec21

[8] *Virtually all textbooks on statistics give formulae for the Variance, Standard Deviation and Skewness, among other measures, for a range of standard distributions. However, the author is yet to find one that explains how these parameters should be interpreted for skewed distributions. One can easily see how Variance and Standard Deviation can have practical application for symmetrical distributions but, for asymmetrical distributions, their interpretation is not quite so intuitive. Just what does the Standard Deviation (and Variance) mean for a skewed distribution, other than a relative measure of dispersion? For a symmetrical distribution, one can measure cumulative probability as a function of Standard Deviations from the mean. But that does not work for skewed distributions, except in special cases where an assumed distribution can be transformed to a symmetrical distribution, eg the Log-Normal. In the meantime, the risk associated with an activity time or cost has to be expressed in some way. Until a substantiated answer is found to the question, this author has adopted the approximation of considering the triangular distribution as two separate triangular distributions, one either side of the Mode, with each of the ranges (negative and positive) being considered to cover three Standard Deviations.*

ANNEX A TO
alsc_FA_Pay Scheduling_2.2.doc

BACKGROUND TO SCHEULE RISK

1. Risk is ever present in all human endeavour, an important one of which is the acquisition and ownership of capital equipment systems. Management objectives for this endeavour and the attendant risks thereto may be grouped into five essential categories:

   a. Technical Capability (system physical, functional and performance objectives);

   b. Through Life Support (system support physical, functional and performance objectives);

   c. Commercial (contractual, financial and industry support objectives);

   d. Schedule (production and delivery objectives); and

   e. Cost (capital and life cycle cost objectives).

2. These categories are recommended and often used as the primary sub-divisions of a Request for Tender (RFT), the structure of the Statement of Work (SOW), the Work Breakdown Structure (WBS) and as the primary criteria for evaluation of tenders. It is imminently logical to use the WBS as the basis of risk analysis and risk management tools and databases normally cater for this.

3. Hence these categories are also the principal risk categories

4. Specialist personnel will normally undertake risk analysis respect of each category and there may be some attempt made to cross-relate these risk evaluations and to aggregate them into an overall risk evaluation. Considerable work remains to be done in this area, however.

5. Unfortunately, too many textbooks and risk modelling tools assume, invalidly, that all components of a schedule (a complex network of activities) are totally independent of each other, ie they assume zero correlation between any pair of activities. This simplifying assumption cannot be sustained and could have serious consequences by way of inaccurate estimates of schedule and cost risk for important projects.

6. Although the alternative assumption of some extent of interdependence (correlation) is more time consuming and difficult mathematically to do in its fullest version, and is based on sound and accepted statistical theory, useful approximations that would be adequate for most applications can be developed to cater for correlation.

ANNEX B TO
alsc_FA_Pay Scheduling_2.2.doc

GLOSSARY

DEFINITIONS

Term	Definition
Level of Risk	the Expected Loss, due to a Risk Event; the expected detrimental impact on an Entity, due to a Risk Event; a measure of the Likelihood of a Risk Event (Threat) * Consequences.
Risk	the chance of something happening that will have an impact upon objectives. (AS/NZS 4360:2004) the possibility of a future event, should it occur will have an effect on project objectives including cost, schedule and technical. (Risk Management Lexicon Version 1.0, Project Management Institute).
Risk Analysis	the evaluation of a situation, environment or set of conditions to determine the technical, financial or business risks inherent in the venture or mission.
Schedule	a program of networked activities; an organisation of activities and their associated estimated execution times.
Schedule Risk	the probable range of values that the Total Schedule Time may take, as a stochastic variable, given the program of individual activities comprising the schedule and the risk associated with each activity.
Total Schedule Time	the summation of durations of activities on the critical path of a schedule.