Using Method of Moments in Schedule Risk Analysis

Raymond P. Covert

MCR, LLC

rcovert@mcri.com

BACKGROUND

A program schedule is a critical tool of program management. Program schedules based on discrete estimates of time lack the necessary information to provide a robust determination of schedule uncertainty and therefore the risk that the proposed schedule will be completed on time. To determine the risk that a proposed, discrete schedule will meet a particular schedule and to find the probable critical paths (i.e., the criticality index), a probabilistic schedule risk analysis (SRA) is performed.

SRA is a process by which probability density functions (PDFs), usually triangular, are defined at the task level in an effort to quantify the uncertainty of each task element. The network structure provides the relationships of successor and predecessor tasks that define the mathematical problem to be solved. Typically, a statistical sampling technique (e.g., Monte Carlo sampling) is used to find a final, probabilistic project end date and criticality index (CI).

THE PROBLEM

The primary method of performing SRA is through a statistical simulation such as a Monte Carlo simulation, however there are major drawbacks: Correlated random variables representing probabilistic schedule durations can lead to unwieldy correlation matrices, and the amount of time required to perform a meaningful SRA can be very time consuming. Another method is an analytic approach using the method of moments, whereby the moments (e.g., the mean and variance) of the input distributions defined at the task level are used to determine the moments of the probability distribution representing the task finish date. This method has been and continues to be used in the cost risk analysis community to nearly instantaneously and precisely determine the primary moments of WBS summations. Shortcomings of MOM when applied to SRA include problems with calculating the PDFs where tasks merge (i.e., merge bias) and calculating a CI (Hulett – 2011). However Hulett fails to recognize and capitalize on the capabilities of MOM beyond statistical summation.

DIFFERENCES BETWEEN STATISTICAL SIMULATION AND MOM

There are fundamental differences in the way statistical simulation and MOM operate. Both techniques rely on the definition of a problem (i.e. a schedule network), specification of input probability distributions (the "inputs") and definition of correlation coefficients between the inputs. Beyond that point the two methods are very different.

A statistical simulation requires a large amount of trials to accurately estimate a series of resulting probability density functions. During each trial a statistical simulation samples all of the correlated random variables, calculates a set of outputs of interest ("forecasts"), and saves them. After many trials (e.g., 10,000 trials), the simulation uses the saved forecasts to provides histograms and a set of statistics for each forecast for which we are interested.As the size of the schedule network grows, and consequently the number of probabilistic inputs, the input and computing time needed to perform a SRA using statistical simulation grows exponentially.

MOM approaches the SRA problem analytically, rather than by brute force, by using series of statistical equations to calculate the exact moments (i.e., mean, variance, skewness, kurtosis) of forecasts. In the SRA problem, this requires calculation of probabilities of the duration of serial and parallel tasks. The probabilistic duration of tasks with serial relationships rely on statistical summation much akin to the summation of costs in a WBS (Young - 1992). The duration of a set of merged parallel tasks require knowledge of the maximum ("max") and sometimes the minimum ("min") of the probability distributions in question.Since MOM relies on statistical equations, results can be obtained nearly instantaneously.

Economy of Correlation Assumptions

A typical statistical simulation requires the user to define PDFs for each task with random duration as well as a correlation coefficient, ρ , between each pair of random durations (i.e., the detail tasks) to avoid unreasonably small forecast variance due to omission of correlation (Book - 1999). Given all tasks should have some non-zero correlation, and if the number of tasks is N, the number of correlations required will be N(N-1)/2. So, a schedule network with 100 detailed task elements will require the analyst to define (100*99)/2 = 9900/2 = 4950 correlations.

MOM uses statistical equations and not a statistical simulation to calculate moments, so calculating the sum (or max) of multiple summary tasks only requires using the calculated moments and pairwise correlation coefficients of the summaries rather than all of the correlation coefficients between each detailed task. If all tasks in a schedule network were arranged serially and without summary tasks, neither which are practical assumptions, there would be no difference in the number of pairwise correlations required for either a statistical simulation or MOM approach. However, creation of summary tasks and merged parallel paths require fewer definitions of correlation coefficients when using MOM. This is because MOM only requires definition of correlation coefficients between detail tasks and summary tasks with predecessor and successor relationships since it is not a simulation. This feature of MOM results in a reduction of the number of correlation coefficients required to effectively perform a SRA.

For example, if there are M = 20 summary tasks with L= 5 detail tasks each, then MOM only requires M+M*(L*L-1)/2 = 20+20*(5*4/2) = 220 correlations. MOM requires fewer definitions of correlation than does a statistical simulation.

Accurate Moments and PDF Shape

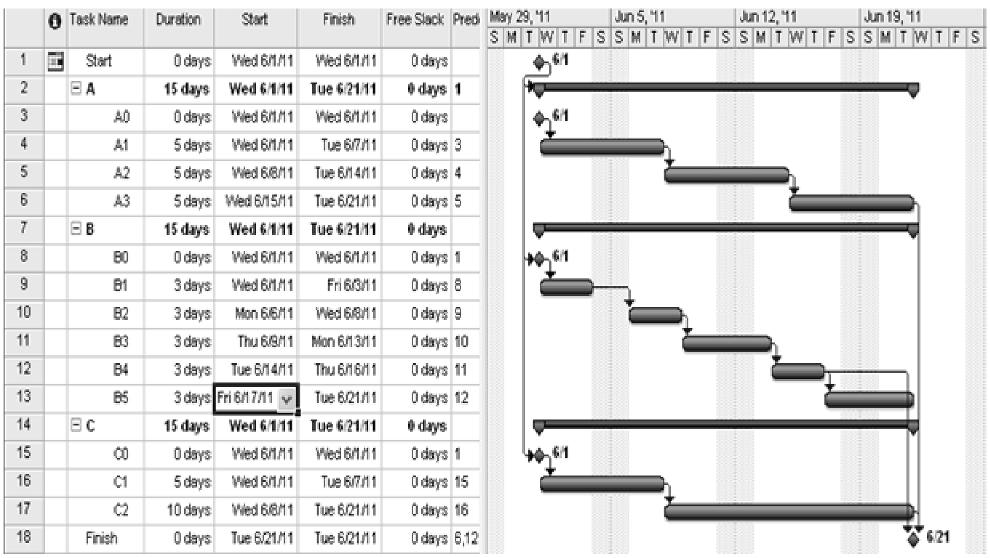
While MOM can provide the exact moments of a resulting forecast distribution using formulae, it is difficult to determine the exact shape of a forecast distribution. Since we have little *a priori* knowledge of the exact shape of the forecast, estimates of the percentiles (or other quantiles) of the forecast distribution will be an inherent source of inaccuracy. While MOM may provide exact moments, a set of guided assumptions must be made to choose an appropriate forecast shape to allow the estimation of quantiles.

On the other hand, the best a statistical simulation can provide is an estimate of moments and shape, and relies on an accurate technique of sampling random variables (Garvey - 1999, Iman& Conover- 1982, Lurie & Goldberg - 1998). In certain situations (i.e., multiple probabilistic branching, large strings of serial tasks) small errors in sampling correlated (even when ρ =0) random variables can produce significant errors (Book - 1999).

EXAMPLE SCHEDULE

The example schedule in Figure 1 is used to demonstrate the how MOM techniques can be used in a SRA to calculate total duration of the project and the CI of the three parallel summary tasks A, B and C.[1]

Figure 1 Example Schedule



	Task Name	Duration	Start	Finish	Free Slack	Pred
1	Start	0 days	Wed 6/1/11	Wed 6/1/11	0 days	
2	**A**	**15 days**	**Wed 6/1/11**	**Tue 6/21/11**	**0 days**	**1**
3	A0	0 days	Wed 6/1/11	Wed 6/1/11	0 days	
4	A1	5 days	Wed 6/1/11	Tue 6/7/11	0 days	3
5	A2	5 days	Wed 6/8/11	Tue 6/14/11	0 days	4
6	A3	5 days	Wed 6/15/11	Tue 6/21/11	0 days	5
7	**B**	**15 days**	**Wed 6/1/11**	**Tue 6/21/11**	**0 days**	
8	B0	0 days	Wed 6/1/11	Wed 6/1/11	0 days	1
9	B1	3 days	Wed 6/1/11	Fri 6/3/11	0 days	8
10	B2	3 days	Mon 6/6/11	Wed 6/8/11	0 days	9
11	B3	3 days	Thu 6/9/11	Mon 6/13/11	0 days	10
12	B4	3 days	Tue 6/14/11	Thu 6/16/11	0 days	11
13	B5	3 days	Fri 6/17/11	Tue 6/21/11	0 days	12
14	**C**	**15 days**	**Wed 6/1/11**	**Tue 6/21/11**	**0 days**	
15	C0	0 days	Wed 6/1/11	Wed 6/1/11	0 days	1
16	C1	5 days	Wed 6/1/11	Tue 6/7/11	0 days	15
17	C2	10 days	Wed 6/8/11	Tue 6/21/11	0 days	16
18	Finish	0 days	Tue 6/21/11	Tue 6/21/11	0 days	6,12

For each summary task, a different number of serially-linked detailed tasks are defined with a total deterministic duration of fifteen days. Also, the first serial task in each summary task starts on the same day, so the discrete end date for all three parallel tasks is on the same day.

Detail and Summary Task Duration Statistics

If durations of individual tasks are modeled as random variables (RVs) using triangular distributions[2], the task's mean (μ) and variance (σ^2) can be calculated using simple formulae.

[1] Microsoft Project was used to create the graphic of the schedule, however the MOM calculations and statistical were performed in Microsoft Excel. This avoids the shortcoming of many schedule risk simulators that report duration statistics in work days rather than calendar days.

Triangular distributions are uniquely defined by three parameters: *L*, the lowest possible value, *M*, the most likely value or mode; and *H*, the highest possible value. Themean of a triangular distribution *T(L,M,H)* is

$$\mu[T(L,M,H)] = \frac{(L+M+H)}{3}, \quad (1)$$

andthe varianceis

$$\sigma^2[T(L,M,H)] = \frac{(L^2+M^2+H^2-LM-LH-MH)}{18}. \quad (2)$$

The mean ($\mu_1 Total$) and variance (σ^2_{Total}) of summary tasks can be calculated using equations 3 and 4.

$$\mu_{Total} = \sum_{i=1}^{N} \mu_i, \text{ where:} \quad (3)$$

μ_i = the mean of task i

N = the number of serial task schedule elements being statistically summed

and

$$\sigma^2_{Total} = \sum_{k=1}^{N} \sigma_k^2 + 2\sum_{k=2}^{N}\sum_{j=1}^{k-1} \rho_{jk}\sigma_j\sigma_k, \text{ where:} \quad (4)$$

σ^2_{Total} = the variance of the statistical sum of N serial task schedule elements

ρ_{jk} = the correlation between durations of tasks j and k

σ_j, σ_k = the standard deviations of durations of tasks j and k

$\sum_{k=1}^{N} \sigma_k^2$ = the sum of the variances of the N serial task schedule elements

[2] Triangular distributions were assumed in this case, however the formulae to calculate the mean and variance of uniform, beta, normal and lognormal distributions can be easily substituted.

$2\sum_{k=2}^{N}\sum_{j=1}^{k-1}\rho_{jk}\sigma_j\sigma_k$ = the sum of the pairwise covariances of the N task schedule durations

The parameters for the triangular distributions *T(L,M,H)* of detailed tasks (A1, A2, …C2) of the example schedule networkis shown in Table 1. The inter-element correlation between all subtasks is defined to be ρ = 0.2. Assuming the first detailed tasks A1, B1 and C1 start on the same date,the moments of the durations of the three parallel summary tasks A, B and C can be easily calculated.

Table 1 Probabilistic Task Durations

Task	Subtask	L	M	H	μ, days	σ, days
A	A1-A3	2	5	7	4.667	1.027
B	B1-B5	2	3	5	3.333	0.624
C	C1	2	5	10	5.667	1.650
C	C2	5	10	15	10.000	2.041

The calculations of the mean (*m*) and sigma (*s*) (i.e., the square root of the variance) of the three summary tasks derived using equations 3 and 4.

$$\mu_A = \sum_{i=1}^{3}\mu_{Ai} = 3(4.667) = 14.00$$

$$\mu_B = \sum_{i=1}^{5}\mu_{Bi} = 5(3.333) = 16.67$$

$$\mu_C = \sum_{i=1}^{2}\mu_{Ci} = 5.667 + 10.000 = 15.67$$

$$\sigma_A = \sqrt{\sum(\sigma_{Ai})^2 + 2\rho\sum\sigma_{Ai}\sigma_{Aj}} = \sqrt{3(1.027)^2 + 2(0.2)(3)(1.027)^2} = 2.11$$

$$\sigma_B = \sqrt{\sum(\sigma_{Bi})^2 + 2\rho\sum\sigma_{Bi}\sigma_{Bj}} = \sqrt{5(0.624)^2 + 2(0.2)(10)(0.624)^2} = 1.87$$

$$\sigma_C = \sqrt{\sum(\sigma_{Ci})^2 + 2\rho\sum\sigma_{Ci}\sigma_{Aj}} = \sqrt{(1.650)^2 + (2.041)^2 + 2(0.2)(1.650)(2.041)} = 2.87$$

The results of the calculations are summarized in Table 2.

Table 2 Summary Task Duration Statistics using MOM

Summary Task	μ, days	σ, days
A	$\mu_A = 14.00$	$\sigma_A = 2.11$
B	$\mu_B = 16.67$	$\sigma_B = 1.87$
C	$\mu_C = 15.67$	$\sigma_C = 2.87$

A sample 100,000 trial statistical simulation using Latin hypercube sampling was run to re-create these results. The results, shown in Table 3, are a close match with the MOM-calculated Table 2. At this point, MOM calculates exact results, so any differences in the values in Tables 2 and 3 are due to statistical simulation error. The calculated Pearson correlation coefficients between the 100,000 samples of detailed tasks ranged from 0.1947 to 0.2064. While this may seem like a small amount, it contributes to the error of the statistical simulation.

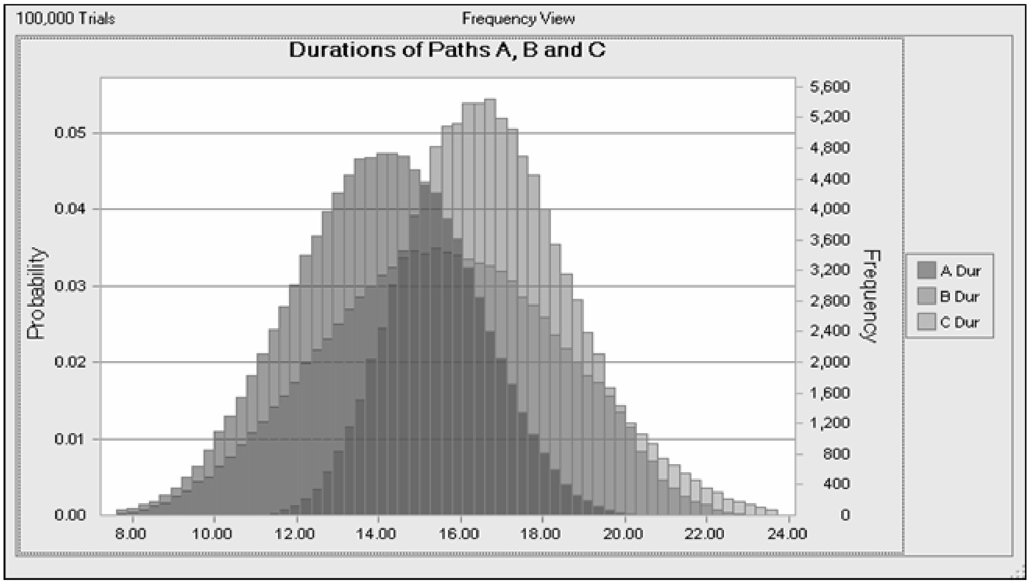
Table 3 Summary Task Duration Statistics using Statistical Simulation

Summary Task	μ, days	σ, days
A	$\mu_A = 14.00$	$\sigma_A = 2.11$
B	$\mu_B = 16.67$	$\sigma_B = 1.87$
C	$\mu_C = 15.67$	$\sigma_C = 2.88$

Shape of Summary Task PDF

The shapes of the three summary task distributions produced from the Monte Carlo simulation (Figure 2) appear to be Gaussian. Since the three summary tasks start on the same day, the durations in Figure 2 represent the distribution (in work days) of durations from start to finish.

Figure 2 Shapes of Summary Task PDFs

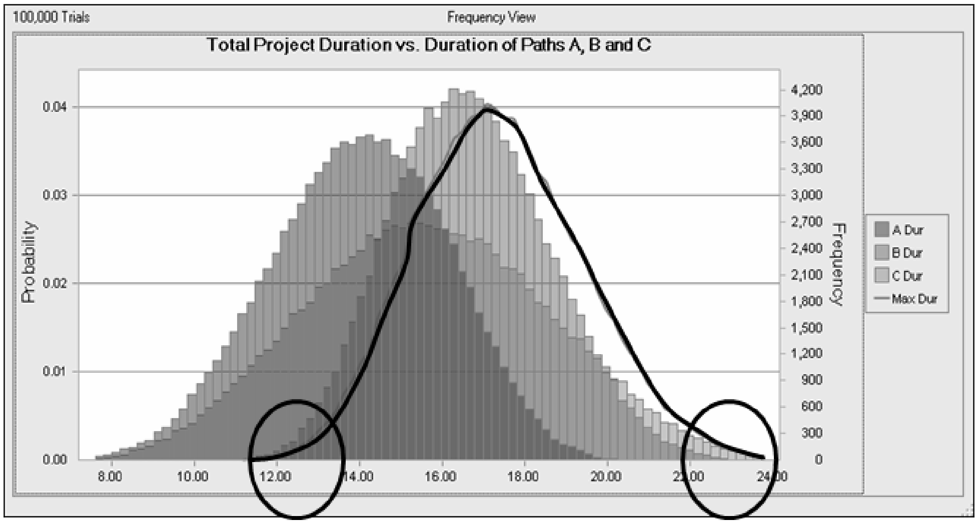


The total duration of a summary task (or of the project) is defined by the probabilistic start and end dates of the project. If, as in this example, the start date is a discrete date, then the total duration is defined by the duration between the discrete start date and the probabilistic finish date.

Shape of Project Duration PDF

The probabilistic total duration of the project in our example is shown in Figure 3 as a black line, and its characteristics are defined by the maximum durations of the three paths, A, B, and C. Note the bounds of this PDF are defined by the greatest duration of paths A, B and C and that the distribution is skewed.

Figure 3Probabilistic Total Duration



Probabilistic Finish Date

The finish date of the example schedule is defined by the latest end date of these three tasks, which is defined by max(A,B,C). This is equivalent to max(max(A,B),C) and max(A,max(B,C), which is an important consideration because it allows us to deal with the problem of finding the moments of the maximum of distributions in pairs. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*(Nadarajah and Kotz-2008) provides a method of calculating the first two moments of the max and min of two correlated Gaussian distributions.[3]

The PDF of $X=\max(X_1,X_2)$ is $f(x) = f_1(-x) + f_2(-x)$, where (5)

$$f_1(x) = \frac{1}{\sigma_1}\varphi\left(\frac{x+\mu_1}{\sigma_1}\right)\Phi\left(\frac{r(x+\mu_1)}{\sigma_1\sqrt{1-r^2}} - \frac{x+\mu_2}{\sigma_2\sqrt{1-r^2}}\right)$$

$$f_2(x) = \frac{1}{\sigma_2}\varphi\left(\frac{x+\mu_2}{\sigma_2}\right)\Phi\left(\frac{r(x+\mu_2)}{\sigma_2\sqrt{1-r^2}} - \frac{x+\mu_1}{\sigma_1\sqrt{1-r^2}}\right)$$

Where $\varphi\ (\cdot)$ and $\Phi\ (\cdot)$ are the PDF and the cumulative distribution function (CDF) of the standard normal distribution, respectively.

The first two moments are

[3] The integrated circuit industry has a deep interest in scheduling methods and routines which stems from the need to calculate signal transit and arrival times at nodes in integrated circuit paths.

$$E[X] = \mu_1 \Phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) + \mu_2 \Phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) + \theta\varphi\left(\frac{\mu_1 - \mu_2}{\theta}\right) \qquad (6,7)$$

$$E[X^2] = (\sigma_1^2 + \mu_1^2)\Phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) + (\sigma_2^2 + \mu_2^2)\Phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) + (\mu_1 + \mu_2)\theta\varphi\left(\frac{\mu_1 - \mu_2}{\theta}\right)$$

where

r = correlation between tasks 1 and 2,

$\theta = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2}$, and (8)

$\sigma^2 = E[(X - \mu)^2] = E[X^2] - \mu^2$ (9)

Defining Correlation Coefficients between Summary Tasks

The correlation, r , between any two summary tasks is not the same as the correlation between any two detailed tasks,ρ . The inter-summary-task correlations and also the summary-task sigmas depend on(, but r and (can be quite different.This fact is useful when we (1) know the correlation between schedule summary task elements but not the correlation between individual tasks from different summary levels or (2) know thecorrelation between individual tasks from different summary each other but do know the correlation between higher-level summary level task elements (Covert – 2006). In either case, we can make an educated guess of the value of r or we can calculate its exact value.

By solving the partitioned matrix problem, we can relate r_{AB} to the r for pairwise subtasks in A and B

$$\rho \sum \square \Sigma$$

or,

$$r_{AB} = \frac{\rho \, \Sigma\sigma_{Ai} \, \Sigma\sigma_{Bj}}{\sqrt{\Sigma(\sigma_{Ai})^2 + 2\rho \, \Sigma\sigma_{Ai}\sigma_{Aj}} \, \sqrt{\Sigma(\sigma_{Bi})^2 + 2\rho \, \Sigma\sigma_{Bi}\sigma_{Bj}}} \qquad (10)$$

Using the values from Tables 1 and 2, we can calculate r_{AB}, (using equation 10) which is

$$r_{AB} = 0.2 * \frac{3.082 * 3.118}{2.106 * 1.871} = 0.2 * \frac{9.610}{3.939} = 0.488 .$$

Comparing this calculation with the correlation coefficient calculated from the 100,000 trial Monte Carlo simulation shown in Table 4, we have excellent agreement. Again, any error lies in the statistical sampling process since the calculation of r_{AB} is exact.

Table 4 r_{AB} Calculated from Statistical Simulation

Trial values	A Dur	B Dur		
1	15.93	14.89	r_{AB} =	0.490664
2	12.10	15.57		
100000	14.09	17.17		

Forming the Correlation Matrix

Applying the same method we used to find r_{AB}, we can find r_{AC} and r_{BC} to form the correlation matrix of summary tasks shown in Table 5.

Table 5Calculated Correlation Matrix for Summary Tasks

r	A	B	C
A	1.000	0.488	0.377
B	0.488	1.000	0.429
C	0.377	0.429	1.000

The calculations used to create Table 5 are:

$$r_{AB} = 0.2 * \frac{3.082 * 3.118}{2.106 * 1.871} = 0.488$$

$$r_{AC} = 0.2 * \frac{3.082 * 3.691}{2.106 * 2.870} = 0.377$$

$$r_{BC} = 0.2 * \frac{3.118 * 3.691}{1.871 * 2.870} = 0.429$$

The formulae to compute the exact correlation between the maximum of two (or more) summary tasks and another summary task are not known, so its value must be determined through heuristic assumption. In the example problem, all probabilistic summary path durations are approximately equal, so no single summary task distribution is equal to the max of all distributions - in fact they all share in the definition of the shape and statistics of the total project duration. In this case, the correlation between the maximum of two (or more) summary tasks and another summary task should be approximately equal to the mean of the off-diagonal correlation coefficients shown above, which is

$$\tilde{r} = \frac{r_{AB} + r_{AC} + r_{BC}}{3} = \frac{0.488 + 0.377 + 0.429}{3} = 0.431 .$$

The approximation is very close to the result obtained by direct calculation of the correlation coefficient of $r_{MAX(A,B),C}$ obtained from the 100,000 trial Monte Carlo simulation, shown in Table 6.

Table 6Calculation of $r_{MAX(A,B),C}$ from Statistical Simulation

Trial values	A-B	C Dur		
1	15.93	13.49	$r_{MAX(A,B),C}$ =	0.439152
2	15.57	19.05		
100000	17.17	15.64		

As stated earlier, the approximation works well when the number of detailed tasks in each summary level is approximately the same, and when each summary task has roughly the same probabilistic duration. When the maximum of two PDFs is dominated by one summary task (i.e., PDF of max(A,B) is approximately the PDF of A), its PDF is roughly the same as that summary task. In that case, we can substitute the PDF of the max with the PDF of the dominant summary task and use the correlations calculated in Table 5. The rule applied to this example problem is shown below.

If the PDF of max(A,B) ~ A, then $r_{MAX(A,B),C} \sim r_{A,C}$,

If the PDF of max(A,B) ~ B, , then $r_{MAX(A,B),C} \sim r_{B,C}$, and so on.

CRITICALITY INDEX

The criticality index (CI) is the probability a particular task's path will be on the critical path, or the probability one path will have a longer duration than the others. In our example, there are three potential critical paths, each with its own CI as defined as:

$$CI_A = P(A > max(B,C)) \qquad (11)$$

$$CI_B = P(B > max(A,C)) \tag{12}$$

$$CI_C = P(C > max(B,C)) \tag{13}$$

Using the notation for the maximum of distributions to be X, then P(A>X) is the same as P(X<A) and therefore P(X-A<0). Since the probability of a difference of numbers is an integral function from -∞ to +∞, then the probability the difference between these functions is less than zero is the integral function from -∞ to 0. The mean of this distribution is

$$\mu_{X-A} = \mu_X - \mu_A, \tag{14}$$

and sigma is

$$\mu_{X-A} = \sqrt{\sigma_X^2 + \sigma_A^2 - 2r_{XA}\sigma_X\sigma_A} \tag{15}$$

We can use formulas to calculate the max(PDF) with the calculated values for means, standard deviations and correlation coefficients derived earlier to compute the CIs for the summary tasks A, B and C. These results are compared to those calculated form the 100,000 trial statistical simulation. Again,the results agree rather well. The percentage difference in the CI calculated form MOM and statistical simulation for summary task A appear to be exaggerated due to the small value of CI.

Table 7 Comparing CIs from MOM and Statistical Simulation

Case	CI, %		Difference	
	MOM	Sim	MOM – Sim	(MOM-Sim)/Sim, %
A max	6.1	5.3	0.8	13%
B max	59.7	60.3	-0.6	-1%
C max	34.2	34.4	-0.2	-1%
Total	100.0	100.0		

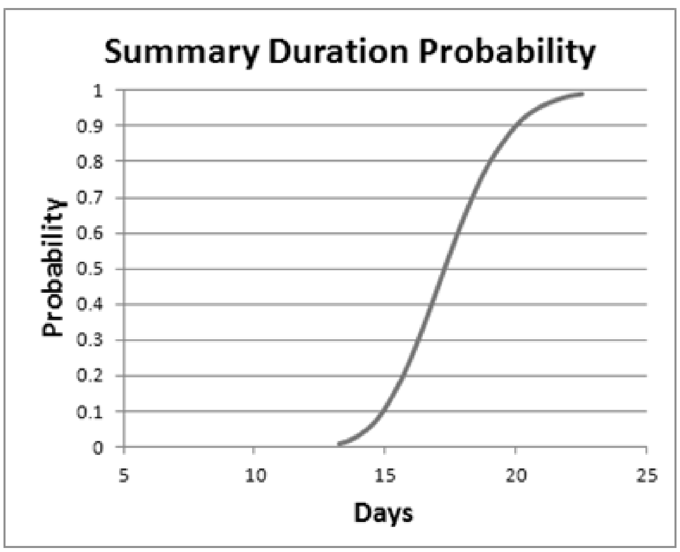
INPUTS AND OUTPUTS

The inputs and outputs of the MOM-based SRA are shown in Table 8 and Figure 4. Very few inputs are needed to perform the SRA, and it provides instantaneous and accurate results. MOM can be easily extended beyond this simple example, and will be more practical to use for SRAs with large schedule problems.

Table 8 MOM-based SRA Inputs and Outputs for Sample Schedule

INPUTS	OUTPUTS							
		DURATION, Days						
Task	Subtask	L	M	H	μ	σ	CI	ρ
A		6	15	21	**14.000**	**2.106**	**0.07**	0.2
	A1	2	5	7	4.667	1.027	0.07	
	A2	2	5	7	4.667	1.027	0.07	
	A3	2	5	7	4.667	1.027	0.07	
B		10	15	25	**16.667**	**1.871**	**0.59**	0.2
	B1	2	3	5	3.333	0.624	0.59	
	B2	2	3	5	3.333	0.624	0.59	
	B3	2	3	5	3.333	0.624	0.59	
	B4	2	3	5	3.333	0.624	0.59	
	B5	2	3	5	3.333	0.624	0.59	
C		7	15	25	**15.667**	**2.870**	**0.34**	0.2
	C1	2	5	10	5.667	1.650	0.34	
	C2	5	10	15	10.000	2.041	0.34	
SUMMARY		6		25	17.379	1.990	1.00	
Calendar Days		9		34	24			

Figure 4Plot of Probabilistic Duration of Project Using MOM



SUMMARY

This paper demonstrates a method of analytically determining the schedule PDF using MOM that solves two problems with MOM used in a SRA: calculating the PDFs and their statistics where tasks merge (i.e., merge bias); and calculating a criticality index. Journal literature from the mathematical and electrical engineering community (specifically the IEEE transactions on VLSI) demonstrates how these problems can be overcome, and provides formulae for computing the distributions formed by merging parallel schedule tasks.

Additional assumptions pertaining to the correlation of merged parallel paths allow the use of MOM in SRAs that provide instantaneous solutions with minimal definition of task distribution parameters and correlation matrices.

The author believes the method proposed in this paper can improve SRA by reducing the number of required inputs to perform a SRA and will also reduce the time required to find results. This should provide schedule analysts with a tool that can more practically perform SRA on schedule networks with a large number of elements.

ACRONYMS AND SYMBOLS

$\phi(\cdot)$	Probability density function
$\Phi(\cdot)$	Cumulative distribution function
$\max(\cdot)$	Maximum function
$\min(\cdot)$	Minimum function
θ	Standard deviation of difference of two random variables
r_{ij}	Correlation coefficient between summary schedule elements i and j
ρ_{ij}	Correlation coefficient between individual tasks i and j
CDF	Cumulative Distribution Function
IEEE	Institute for Electrical and Electronic Engineering
MOM	Method of Moments
PDF	Probability Density Function

RV	Random Variable
SRA	Schedule Risk Analysis
VLSI	Very Large Scale Integration

REFERENCES

Hulett, D., *Practical Schedule Risk Analysis*, Gower, Burlington, VT, 2011.p.15-48

Young, P. H. , "FRISK- Formal Risk Assessment of System Cost", AIAA 92-1054, Presented at 1992 Aerospace Design Conference, Irvine, CA, 3-6 February1992.

Book, S., "Why Correlation Matters in Cost Estimating", 32nd Annual DoD Cost Analysis Symposium, Williamsburg, VA, Feb. 1999. 58 charts.

Garvey, P., "Why Cost Analysts should use Pearson Correlation, rather than Rank Correlation", 32nd Annual DoD Cost Analysis Symposium, Williamsburg, VA, Feb. 1999.

Iman, R.L., and Conover, W.J., "A Distribution-free Approach to Inducing Rank Correlation among Input Variables," Communications in Statistics - Simulation, Computation, Vol. 11, No. 3(1982), pages 311-334.

Lurie, P.M. and Goldberg, M.S., "A Method for Simulating Correlated Random Variables from Partially Specified Distributions," Management Science, Vol. 44, No. 2, February 1998, pages 203-218.

Nadarajah, S. and Kotz, S., "Exact Distribution of the Max/Min of Two Gaussian Random Variables", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, (16:2), Feb. 2008, p.210-212.

Covert, R., “Correlations in Cost Analysis”, 2006 Annual SCEA Conference, TysonsCorner, VA, June 13-16, 2006.69 charts.