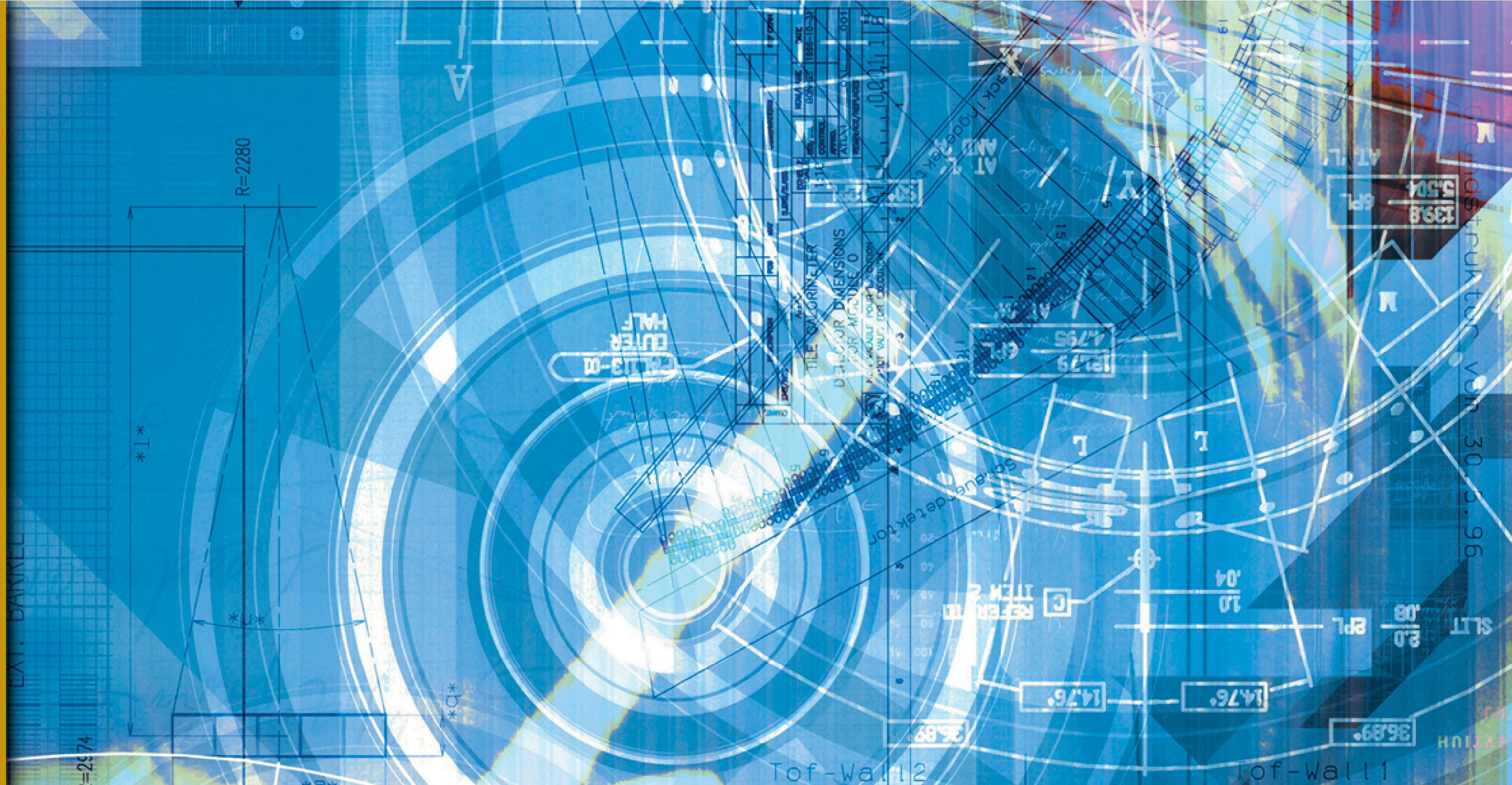


Using Method of Moments in Schedule Risk Analysis

Presented at 2012 SCEA/ISPA Joint Annual Conference and Training Workshop, June 26-29, 2012

presented by:

Raymond P. Covert rcovert@mcri.com

CRITICAL THINKING. SOLUTIONS DELIVERED.

Outline

- Introduction
- Method of Moments Approach to SRA
- Example Schedule
- Probabilistic Task Duration
- Correlating Probabilistic Task Durations
- Criticality Index
- Summary
- References
- Acronyms and Symbols

Introduction (Slide 1 of 2)

- Program schedule is a critical tool of program management
- Program schedules based on discrete estimates of time lack necessary information to provide 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) are defined at the task level in an effort to quantify the uncertainty of each task element

Introduction (Slide 2 of 2)

- Schedule network structure provides 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
- There are major drawbacks:
  - Correlated random variables representing probabilistic schedule durations can lead to unwieldy correlation matrices
  - The amount of time required to perform a meaningful SRA can be very time consuming
- This presentation will offer an alternative statistical approach that will eliminate some of these drawbacks

Outline

- Introduction
- Method of Moments Approach to SRA
- Example Schedule
- Probabilistic Task Duration
- Correlating Probabilistic Task Durations
- Criticality Index
- Summary
- References
- Acronyms and Symbols

CRITICAL THINKING. SOLUTIONS DELIVERED.

Analytic Approach to SRA

- Another method is an analytic approach using the **method of moments (MOM)**, 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
- MOM 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

Presented at the 2012 SCEA/ISPA Joint Annual Conference and Training Workshop - www.iceaaonline.com

Shortcomings of MOM and Solution

- Shortcomings of MOM as used in the Program Evaluation and Review Technique (PERT) when applied to SRA and observed by Hulett [Ref. 1] include problems with…
  - calculating the PDFs where tasks merge (i.e., merge bias)
  - calculating a criticality index
- Journal literature from the mathematical and electrical engineering community (specifically the IEEE transactions on VLSI) [Ref. 2] demonstrates how these problems can be overcome, and provides formulae for computing the distributions formed by merging parallel schedule tasks
- The Solution: My presentation today recommends a sound methodology that can overcome these problems and can calculate the schedule PDF using MOM

Reduced Number of Correlation Coefficients Required

- 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 correlation matrices
  - A typical statistical simulation requires the user to define PDFs for each task with random duration as well as a correlation coefficient, p, between each pair of random durations (i.e., detail tasks)
  - If the number of tasks is N, The number of correlations required will be N(N-1)/2
  - So, a schedule network with 100 elements will require the analyst to define (100*99)/2 = 9900/2 = 4950 correlations
  - 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

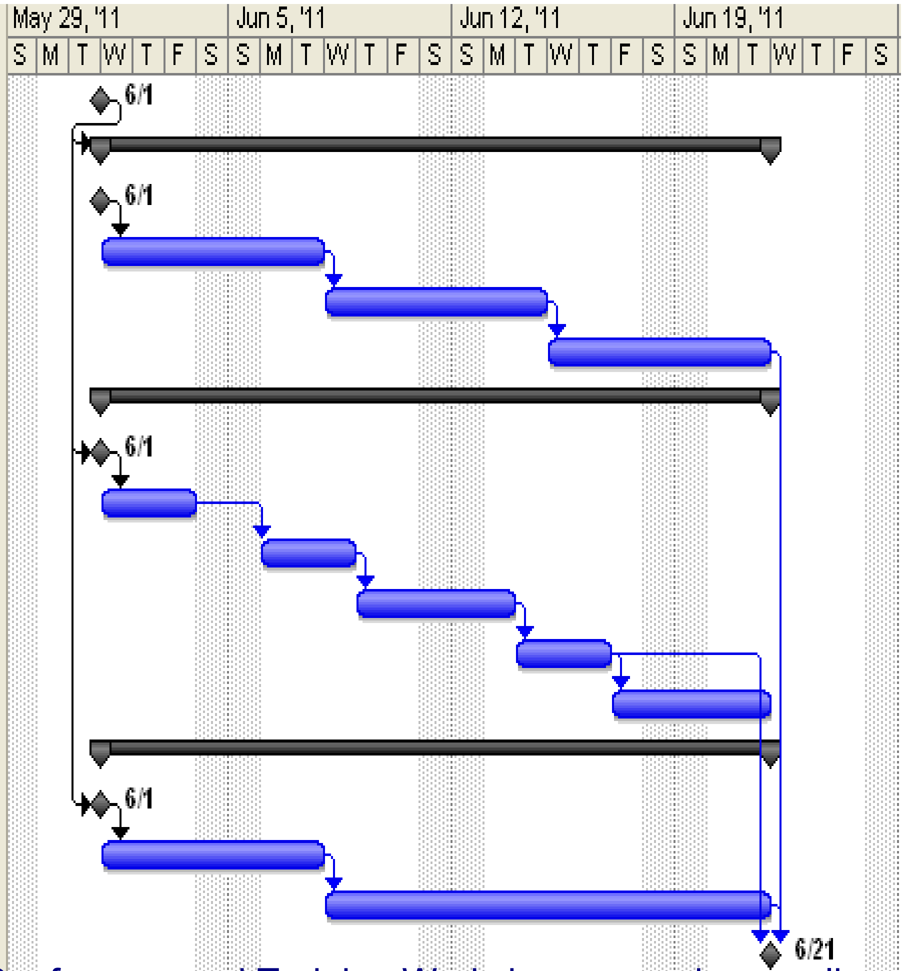
Outline

- Introduction
- Method of Moments Approach to SRA
- Example Schedule
- Probabilistic Task Duration
- Correlating Probabilistic Task Durations
- Criticality Index
- Summary
- References
- Acronyms and Symbols

Example Schedule

- The example schedule below will be used to demonstrate how to apply MOM to SRA

		Task Name	Duration	Start	Finish	Free Slack	Predi
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



Outline

- Introduction
- Method of Moments Approach to SRA
- Example Schedule
- Probabilistic Task Duration
- Correlating Probabilistic Task Durations
- Criticality Index
- Summary
- References
- Acronyms and Symbols

MOM Technique Applied to SRA

- MOM is a technique by which the moments of a distribution are calculated from the moments of its constituents
- It has been successfully applied in the cost community to analytically calculate the mean and variance of the sums of probabilistic costs of work breakdown structure (WBS) elements modeled as RVs
- In the schedule case, this is equivalent to calculating the moments of summary tasks from the probabilistic schedule durations of serial tasks

Mean and Variance of Individual Tasks

- If durations of individual tasks are modeled as random variables (RVs) using triangular distributions, the task's mean and variance can be calculated using simple formulae
- 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
  - The mean, μ, of a triangular distribution *T(L,M,H)* is

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

  - and its variance, σ^2, is

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

Mean and Variance of Summary Tasks

- The mean and variance of summary tasks can be calculated using

$\mu_{Total} = \sum_{i=1}^{N} \boldsymbol{\mu_i}$, where:

μ_i= the mean of task *i*

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

$\boldsymbol{\sigma^2_{Total}} = \sum_{k=1}^{N} \boldsymbol{\sigma^2_k} + \mathbf{2} \sum_{k=2}^{N} \sum_{j=1}^{k-1} \boldsymbol{\rho_{jk}\sigma_j\sigma_k}$, where:

σ^2_{Total} = the variance of the statistical sum of *N* serial 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

$\sum_{k=1}^{N} \sigma^2_k$ = the sum of the variances of the *N* serial task schedule

Probabilistic Task Durations

- If we assume:
  - The triangular distributions *T(L,M,H)* of subtasks shown below
  - Inter-element correlation between all subtasks $\rho = 0.2$
  - Tasks A, B and C start on the same date
- We can calculate the moments of the durations of the three parallel summary tasks A, B and C

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

μ and σ of Summary Task Durations using MOM (slide 1 of 2)

- Mean (μ) and sigma (σ) (i.e., the square root of the variance) of the summary tasks derived using the equations on slides 13-14

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

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

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

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

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

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

μ and σ of Summary Task Durations using MOM (slide 2 of 2)

- The results of the calculations

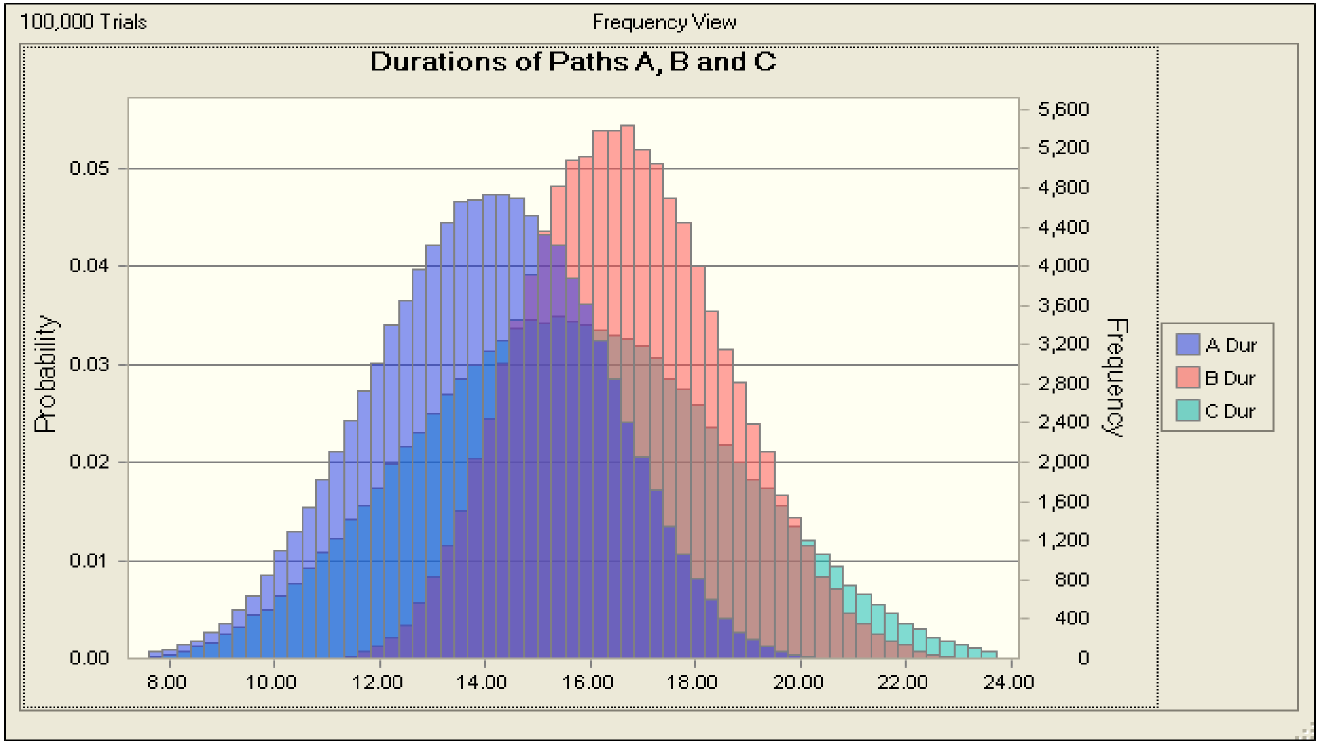
Summary Task	μ, days	σ, days
A	$\mu_A = 14.00$	$\sigma_A = 2.11$
B	$\mu_B = 16.67$	$\sigma_B = 1.87$
C	$\mu_C = 15.67$	$\sigma_C = 2.87$

- A sample 100,000 trial Monte Carlo simulation matches these results as shown below

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	$\mu_C = 15.67$	$\sigma_C = 2.88$

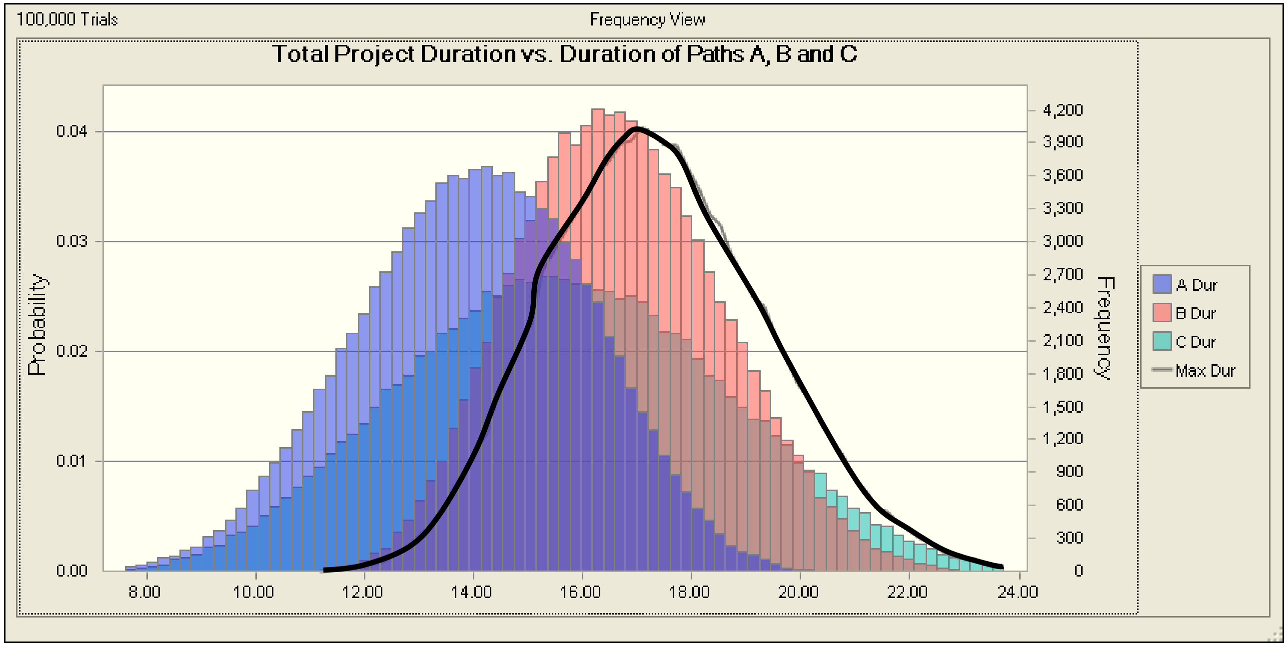
Shapes of Summary Task PDFs

- The shapes of the three summary task distributions produced from the Monte Carlo simulation appear to be Gaussian
  - Since the three summary tasks start on the same day, this chart represents the distribution of durations from start to finish



Total Duration and Finish Date

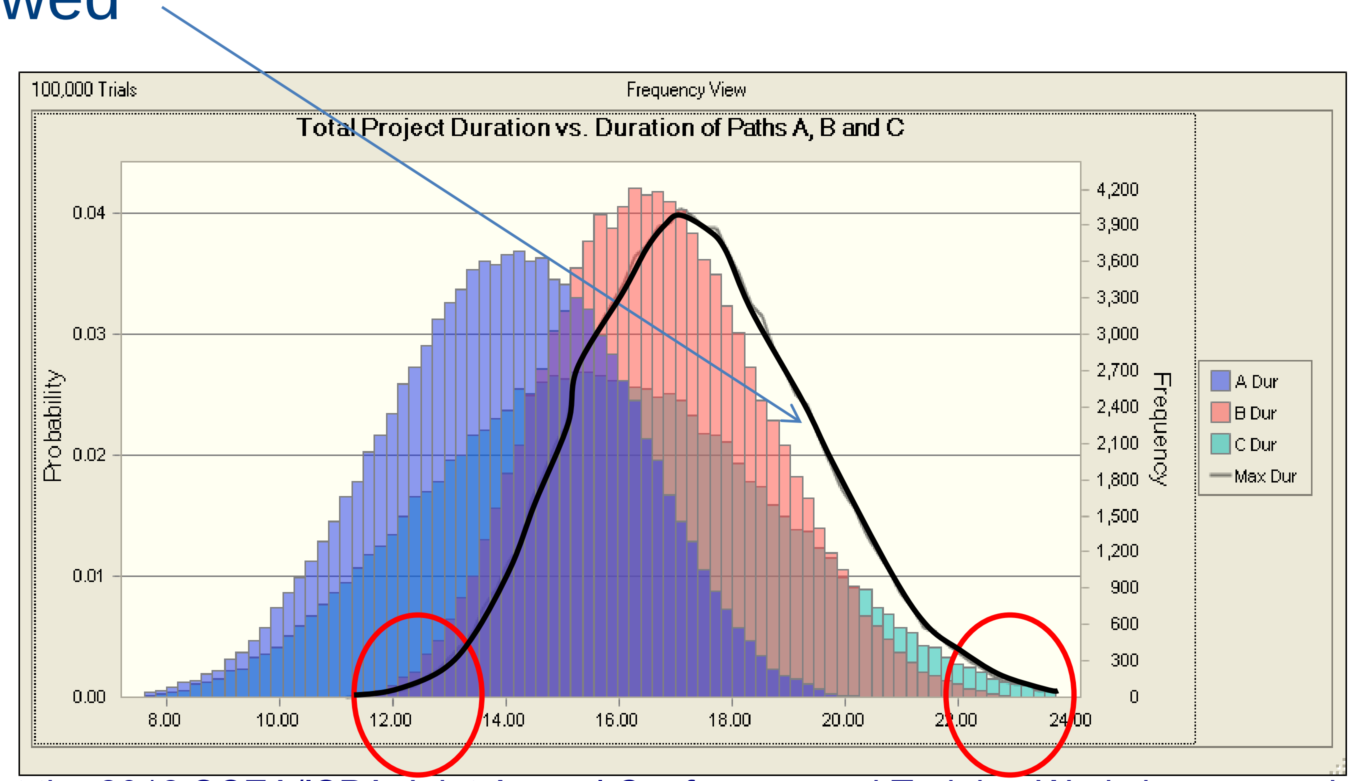
- The total duration of a summary task (or of the project) is defined by the probabilistic start and end dates of the project
  - If, as in this example, the start date is a discrete date, then the total duration is defined by the duration between the discrete start date and the probabilistic finish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



  - The probabilistic total duration in our example is shown as a black line

Probabilistic Total Duration

- 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



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))
- This is an important consideration because it allows us to deal with the problem of finding the **moments of the maximum of distributions in pairs**
- Nadarajah [Ref 2] has provided a method of calculating the first two moments of the max and min of two correlated Gaussian distributions

Critical Thinking. Solutions Delivered.

PDF of Max of Two Gaussian Distributions

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

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

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

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

- With the first two moments

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

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

where

$$\theta = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad \text{and} \quad \sigma^2 = E[(X-\mu)^2] = E[X^2] - \mu^2$$

Outline

- Introduction
- Method of Moments Approach to SRA
- Example Schedule
- Probabilistic Task Duration
- Correlating Probabilistic Task Durations
- Criticality Index
- Summary
- References
- Acronyms and Symbols

Correlations between Summary Tasks (Slide 1 of 3)

- The correlation, r, between any two summary tasks is not the same as the correlation between any two lowest level tasks, ρ
  - The inter-summary-task correlations and also the summary-task sigmas depend on ρ
  - In fact the pairwise summary task correlation, r, can be quite different
  - This is useful when we know the correlation between schedule summary task elements but not the correlation between individual tasks from different summary levels to each other but do have an idea of correlation between higher-level summary level task elements [Ref. 3]
- By solving the partitioned matrix problem, we can relate r_{AB} to the ρ for pairwise subtasks in A and B

$$r_{AB} = \frac{\rho \begin{bmatrix} \sigma_{A1}\sigma_{B1} + \sigma_{A1}\sigma_{B2} + \sigma_{A1}\sigma_{B3} \\ +\sigma_{A1}\sigma_{B4} + \sigma_{A1}\sigma_{B5} + \sigma_{A2}\sigma_{B1} + \sigma_{A2}\sigma_{B2} \\ +\sigma_{A2}\sigma_{B3} + \sigma_{A2}\sigma_{B4} + \sigma_{A2}\sigma_{B5} + \sigma_{A3}\sigma_{B1} \\ +\sigma_{A3}\sigma_{B2} + \sigma_{A3}\sigma_{B3} + \sigma_{A3}\sigma_{B4} + \sigma_{A3}\sigma_{B5} \end{bmatrix}}{\sigma_A \sigma_B} = \frac{\rho \sum \sigma_{Ai} \sum \sigma_{Bj}}{\sigma_A \sigma_B}$$

or,

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

Correlations between Summary Tasks (Slide 3 of 3)

- Using the values from slide 16, we can calculate r_{AB}, which is

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

- When we compare this calculation with the correlation coefficient calculated from the 100,000 trial Monte Carlo simulation shown below, we have excellent agreement

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 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 below

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

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

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

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

  - The correlation between the maximum of two (or more) summary tasks and another summary task is more difficult to determine, however it should be* approximately equal to the mean of the

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

* In this case

Correlation of Maximums of Summary Tasks

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

- 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 below
  - The approximation works well when the number of tasks in each summary level are approximately the same

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		

Note: If the PDF of max(A,B) ~ A, then use $r_{max(A,B),C} = r_{AC}$, likewise if the PDF of max(A,B) ~ B, , then use $r_{max(A,B),C} = r_{BC}$.

Outline

- Introduction
- Method of Moments Approach to SRA
- Example Schedule
- Probabilistic Task Duration
- Correlating Probabilistic Task Durations
- Criticality Index
- Summary
- References
- Acronyms and Symbols

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:

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

$$CI_B = P(B > max(A,C))$$

$$CI_C = P(C > max(B,C))$$

Finding the Criticality Index

- 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
  - Where the mean is $\mu_{X-A} = \mu_X - \mu_A$
  - and sigma is $\sigma_{X-A} = \sqrt{\sigma_X^2 + \sigma_A^2 - 2r_{XA}\sigma_X\sigma_A}$
- 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

Comparing Calculated CIs

- These results are compared to those calculated form the 100,000 trial Monte Carlo simulation
  - The results agree rather well

Case	CI, %		Difference	
	MoM	Monte Carlo	MoM – MC	(MoM-MC)/MC, %
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		

Presented at the 2012 SCEA/ISPA Joint Annual Conference and Training Workshop - www.iceaaonline.com

Outline

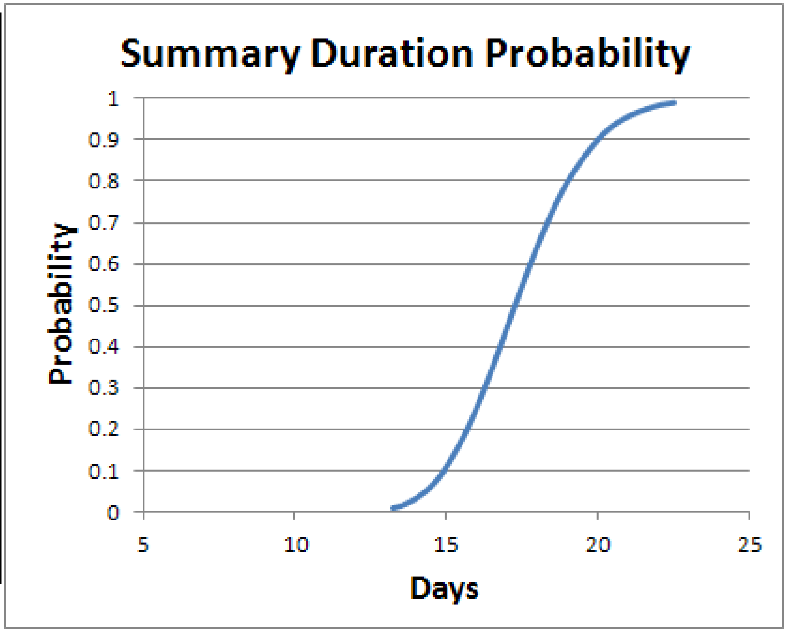
- Introduction
- Method of Moments Approach to SRA
- Example Schedule
- Probabilistic Task Duration
- Correlating Probabilistic Task Durations
- Criticality Index
- Summary
- References
- Acronyms and Symbols

Summary (Slide 1 of 2)

- MoM-based SRA inputs and outputs for sample schedule

INPUTS | OUTPUTS

Task	Subtask	L	M	H	μ	σ	CI	ρ
		DURATION, Days						
A		6	-	21	14.000	2.106	0.06	
	A1	2	5	7	4.667	1.027	0.06	0.2
	A2	2	5	7	4.667	1.027	0.06	
	A3	2	5	7	4.667	1.027	0.06	
B		10	-	25	16.667	1.871	0.60	
	B1	2	3	5	3.333	0.624	0.60	0.2
	B2	2	3	5	3.333	0.624	0.60	
	B3	2	3	5	3.333	0.624	0.60	
	B4	2	3	5	3.333	0.624	0.60	
	B5	2	3	5	3.333	0.624	0.60	
C		7	-	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	17.378	1.990	1.00	



MCR
CRITICAL THINKING.
SOLUTIONS DELIVERED.

Summary (Slide 2 of 2)

- This paper demonstrates a method of analytically determining the schedule PDF using MOM that solves two problems with MOM used in PERT:
  - Calculating the PDFs and their statistics where tasks merge (i.e., merge bias), and
  - Calculating a criticality index
- Using MOM in SRAs provides <u>instantaneous solutions</u> with <u>minimal definition</u> of task distribution parameters and correlation matrices

Outline

- Introduction
- Method of Moments Approach to SRA
- Example Schedule
- Probabilistic Task Duration
- Correlating Probabilistic Task Durations
- Criticality Index
- Summary
- References
- Acronyms and Symbols

References

1. Hulett, D., *Practical Schedule Risk Analysis*, Gower, Burlington, VT, 2009.
2. 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.
3. Covert, R., “Correlations in Cost Analysis”, 2006 Annual SCEA Conference, Tysons Corner, VA, June 13-16, 2006.

Outline

- Introduction
- Method of Moments Approach to SRA
- Example Schedule
- Probabilistic Task Duration
- Correlating Probabilistic Task Durations
- Criticality Index
- Summary
- References
- Acronyms and Symbols

Acronyms and Symbols

$\phi(\cdot)$	Probability density function
$\Phi(\cdot)$	Cumulative distribution function
$\max(\cdot)$	Maximum 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
PERT	Program Evaluation and Review Technique
RV	Random Variable
SRA	Schedule Risk Analysis
VLSI	Very Large Scale Integration