2017 ICEAA Professional Development & Training Workshop
June 6 - 9, 2017 • Portland, OR

A "Common Risk Factor" Method to Estimate Correlations between Distributions

Presented by:

Marc Greenberg
Strategic Investment Division (SID)
National Aeronautics and Space Administration

Outline

- Correlation Overview
- Why Propose Another Correlation Method?
- Underlying Basis for "Common Risk Factor" Method
  - Concept of Mutual Information
  - Using the Unit Square to Estimate Mutual Information
- "Common Risk Factor" Method (for pair of activities)
  - Apply 7 Steps to Estimate Correlation between 2 Distributions
- Examples
  - Correlation of Durations for Two Morning Commutes
  - Correlation of Costs for Two WBS Elements of a Spacecraft
- Conclusion, Other Potential Applications & Future Work

Correlation Overview (1 of 3) [a]

- A statistical measure of association between two variables.
- It measures how strongly the variables are related, or change, with each other.
  - If two variables tend to move up or down together, they are said to be positively correlated.
  - If they tend to move in opposite directions, they are said to be negatively correlated.
- The most common statistic for measuring association is the Pearson (linear) correlation coefficient, ρ_P
- Another is the Spearman (rank) correlation coefficient, ρ_S
  - Used in Crystal Ball and @Risk

(a) Source: *Correlations in Cost Risk Analysis*, Ray Covert, MCR LLC, 2006 Annual SCEA Conference, June 2006

Correlation Overview (2 of 3) [a]

– Captured through mathematical relationships w/in cost model

– Specified by the analyst and implemented w/in cost model

– Correlations (or dependencies) between the uncertainties of WBS CERs are generally determined **subjectively**

- However, as we collect more data, more and more of these correlations are determined using historical data

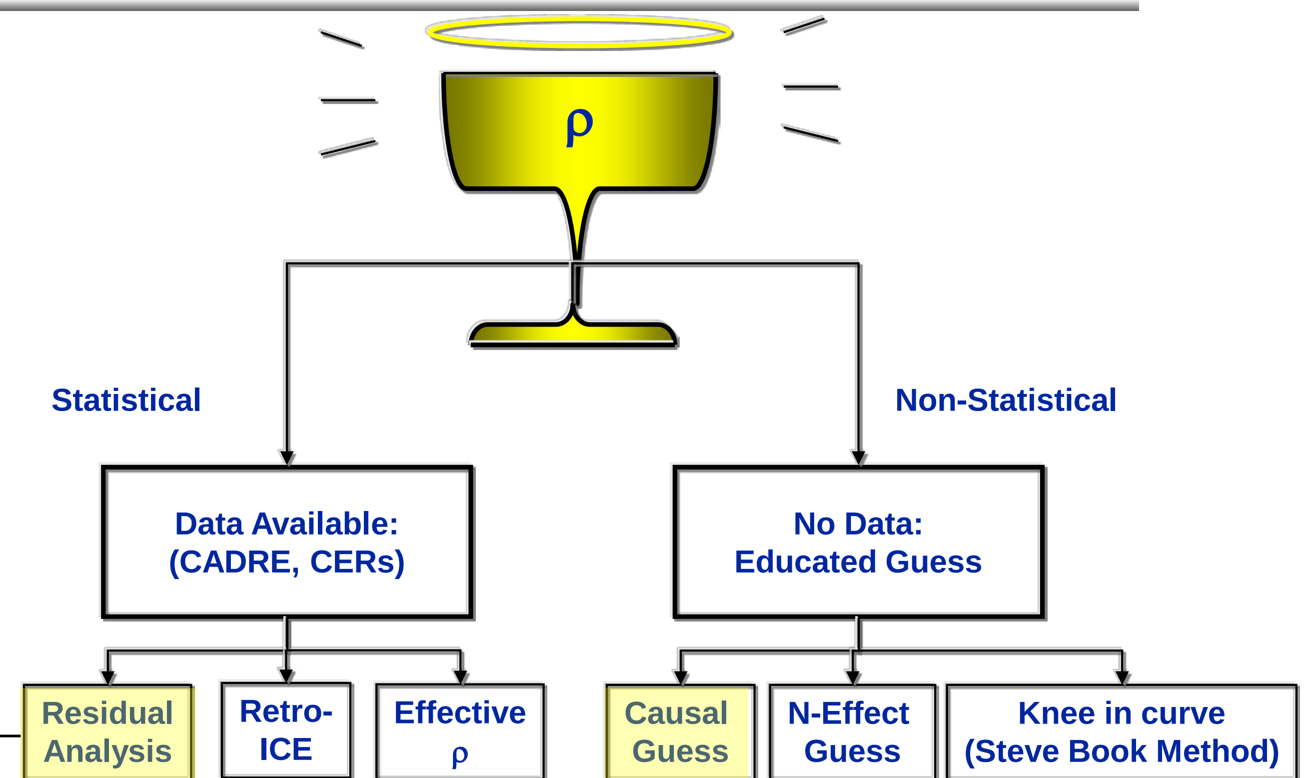
$$\sigma_{Total}^2 = \sum_{k=1}^{n} \sigma_k^2 + 2\sum_{k=2}^{n} \sum_{j=1}^{k-1} \rho_{jk}\sigma_j\sigma_k$$

… where ρ_{jk} is the correlation between uncertainties of WBS CERs j and k (notated at σ_j and σ_k, respectively)

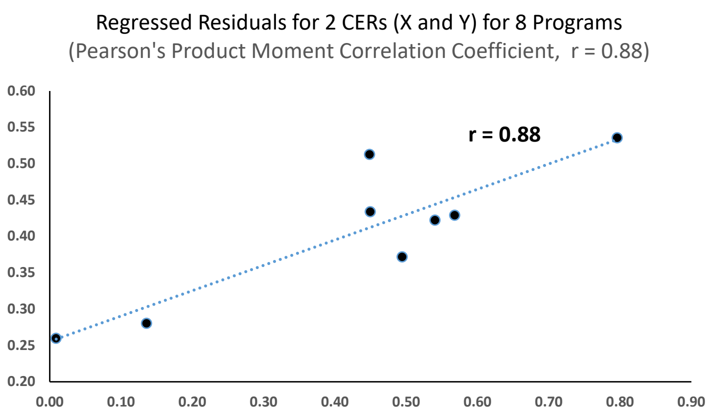
The remainder of this presentation will focus on how to calculate ρ_{jk}

(a) Source: *Joint Agency Cost Schedule Risk and Uncertainty Handbook (Sec. 3.2 & Appendix A)*, 12 March 2014

Correlation Overview (3 of 3) [a]



Example:



Example:

Strength	Positive	Negative
None	0	0
Weak	0.3	-0.3
Medium	0.5	-0.5
Strong	0.9	-0.9
Perfect	1	-1

(a) Schematic from *Correlations in Cost Risk Analysis*, Ray Covert, MCR LLC, 2006 Annual SCEA Conference, June 2006

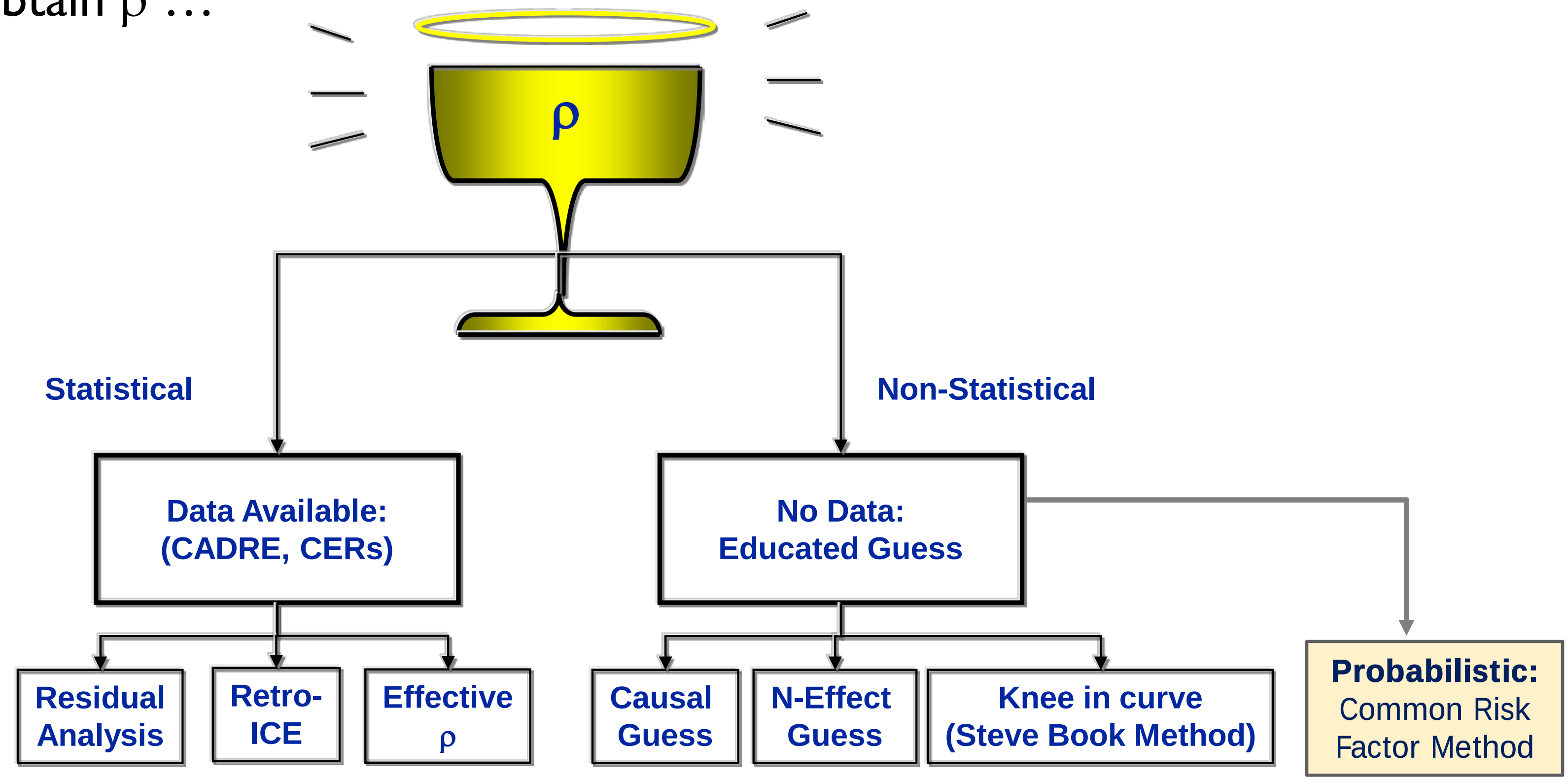
Why Propose *Another* Correlation Method?

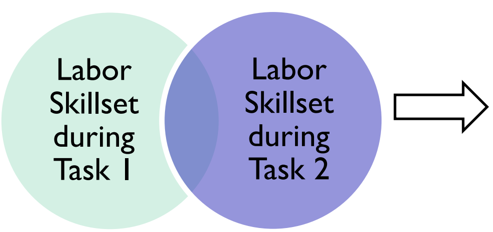
1. For statistical methods, lack of data makes it difficult to calculate robust Pearson's R or Spearman's Rho
   - Example: Residuals from previous slide produces Rho = 0.88. However, the residuals exhibit an "influential observation."
2. For non-statistical methods, there can be many issues:
   - "N-Effect" and "Knee-in-the-Curve" methods are not inherently intuitive to the non-practitioner.
   - Although "Causal Guess" method is simple and intuitive, the analyst and/or subject matter expert are still guessing.
   - Whenever parameters of 2 uncertainty distributions lack basis, the correlation between them is difficult to justify.

Unlike these other methods, the Common Risk Factor Method provides correlation between 2 uncertainties based upon common root-causes. Applying this method may lessen the degree of subjectivity in the estimate.

Correlation Overview (Revisited) [a]

2 paths to obtain ρ …





<u>"Common Risk Factor Method" Notional Example (Output Only)</u>:

Given Tasks 1 & 2 each have an apprentice welder, we expect added uncertainty in the duration of Tasks 1 & 2 due to the lack of skills for each "untested" welder.

Task 1: Max Duration will go up by 5 days due to adding P/T welder to team

Task 2: Max Duration will go up by 10 days due to adding F/T welder to team

Tasks 1 and 2 Correlation = 0.40, partly driven by common skillset in each task.

(a) Schematic from *Correlations in Cost Risk Analysis*, Ray Covert, MCR LLC, 2006 Annual SCEA Conference, June 2006

Outline

- Correlation Overview
- Why Propose Another Correlation Method?
- Underlying Basis for "Common Risk Factor" Method
  - Concept of Mutual Information
  - Using the Unit Square to Estimate Mutual Information
- "Common Risk Factor" Method (for pair of activities)
  - Apply 7 Steps to Estimate Correlation between 2 Distributions
- Examples
  - Correlation of Durations for Two Morning Commutes
  - Correlation of Costs for Two WBS Elements of a Spacecraft
- Conclusion, Other Potential Applications & Future Work

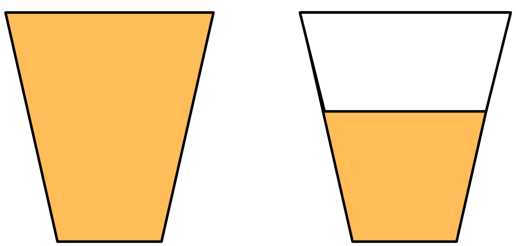
Concept of Mutual Information

- Whenever two objects share common features, these features can be perceived as "mutual information"

Binary string x: 0 0 0 1 0 1 1 1
Binary string y: 1 0 1 1 1 0 0 0

⇨ **Mutual information: = 2 / 8 or 0.25 or 25%**

16 oz. of OJ   8 oz. of OJ



The "least common denominator" is 8 oz. of OJ

⇨ **Mutual information: = 8 / 16 or 0.50 or 50%**

Mutual information can also be applied to risk factors that are common among a pair of uncertainty distributions.

Mutual Information between 2 groupings

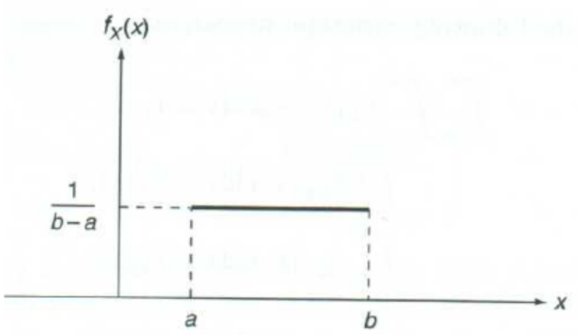
Weighted Ave: Mutual Information = Σ Weight * (Minimum (X,Y) / Maximum (X,Y))

Group X	Group Y	Minimum (X, Y)	Maximum (X, Y)		Weight	Wtd Mutual Information
16 oz.	8 oz.	8	16	8 / 16 = 0.50	16 / 32 = 0.50	0.50 x 0.50 = 0.25
4 oz.	12 oz.	4	12	4 / 12 = 0.333	12 / 32 = 0.375	0.333 x 0.375 = 0.125
4 oz.	4 oz.	4	4	4 / 4 = 1.00	4 / 32 = 0.125	1.00 x 0.125 = 0.125
		Sum:	32			0.50

Mutual Information between Group X and Y

The Unit Square: Meeting Times Example [a]

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{, for } a \le x \le b \\ 0 & \text{, otherwise} \end{cases}$$



Similarly, the girl's actions can be depicted as a single continuous random variable *Y* that takes all values over an interval a to b with equal likelihood.

In this example, the interval is from 0.0 to 1.0 hour. Therefore a = 0.0 and b = 1.0.

Notation for this uniform distribution is **U [0, 1]**

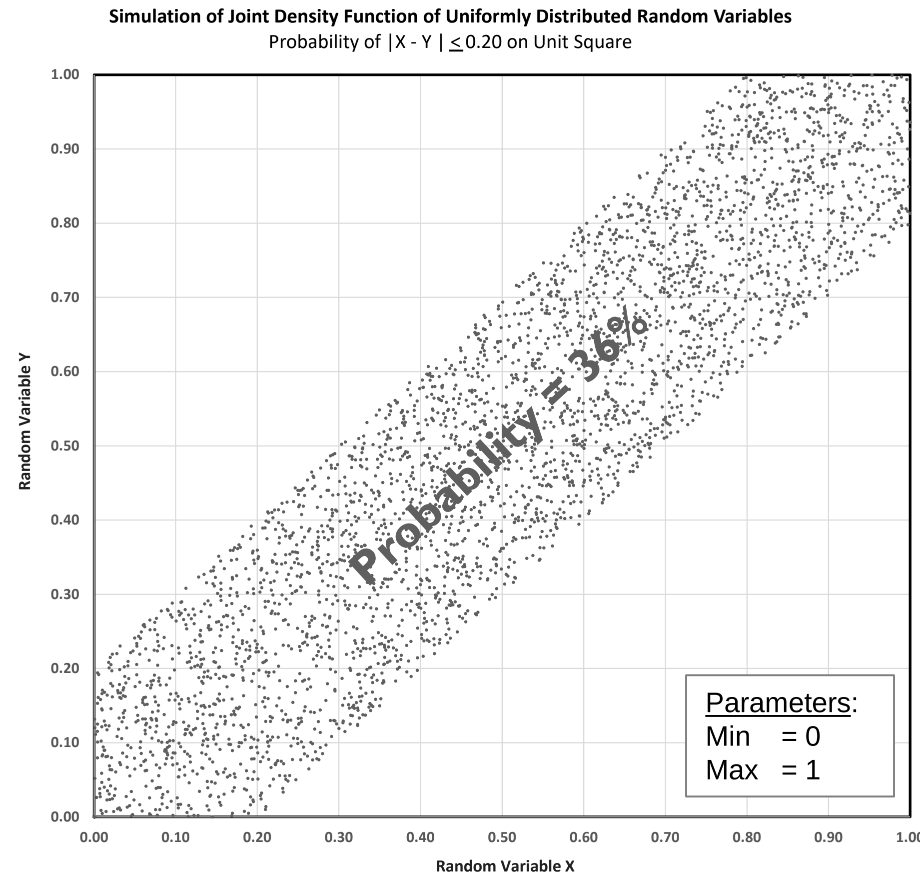
(a) K. Van Steen, PhD, *Probability and Statistics*, Chapter 2: Random Variables and Associated Functions

The Unit Square: Meeting Times Example

Iteration	rv (X)	rv (Y)	|X - Y|	|X - Y| ≤0.2?
1	0.142	0.318	0.176	1
2	0.368	0.733	0.365	0
3	0.786	0.647	0.138	1
4	0.375	0.902	0.528	0
5	0.549	0.935	0.386	0
6	0.336	0.775	0.439	0
7	0.613	0.726	0.113	1
:	:	:	:	:
:	:	:	:	:
9998	0.157	0.186	0.029	1
9999	0.384	0.991	0.607	0
10000	0.045	0.399	0.354	0
			Total =	3630

This simulation indicates that out of 10,000 trials, the boy and girl meet 3,630 times.

Probability they will meet = 0.363 or 36%



The Unit Square … Why do we Care?

Iteration	rv (X)	rv (Y)	|X - Y|	|X - Y| ≤ 0.2?
1	0.142	0.318	0.176	1
2	0.368	0.733	0.365	0
3	0.786	0.647	0.138	1
4	0.375	0.902	0.528	0
5	0.549	0.935	0.386	0
6	0.336	0.775	0.439	0
7	0.613	0.726	0.113	1
:	:	:	:	:
:	:	:	:	:
9998	0.157	0.186	0.029	1
9999	0.384	0.991	0.607	0
10000	0.045	0.399	0.354	0
			Total =	3630

X: 0.786 * 60 = 47 minutes. Arrives at 9:47am.
Y: 0.647 * 60 = 39 minutes. Arrives at 9:39am.

The girl arrives at 9:39am. The boy arrives at 9:47am. He arrived w/in the 12 minute (0.2 hr) time window. So they **do** meet.

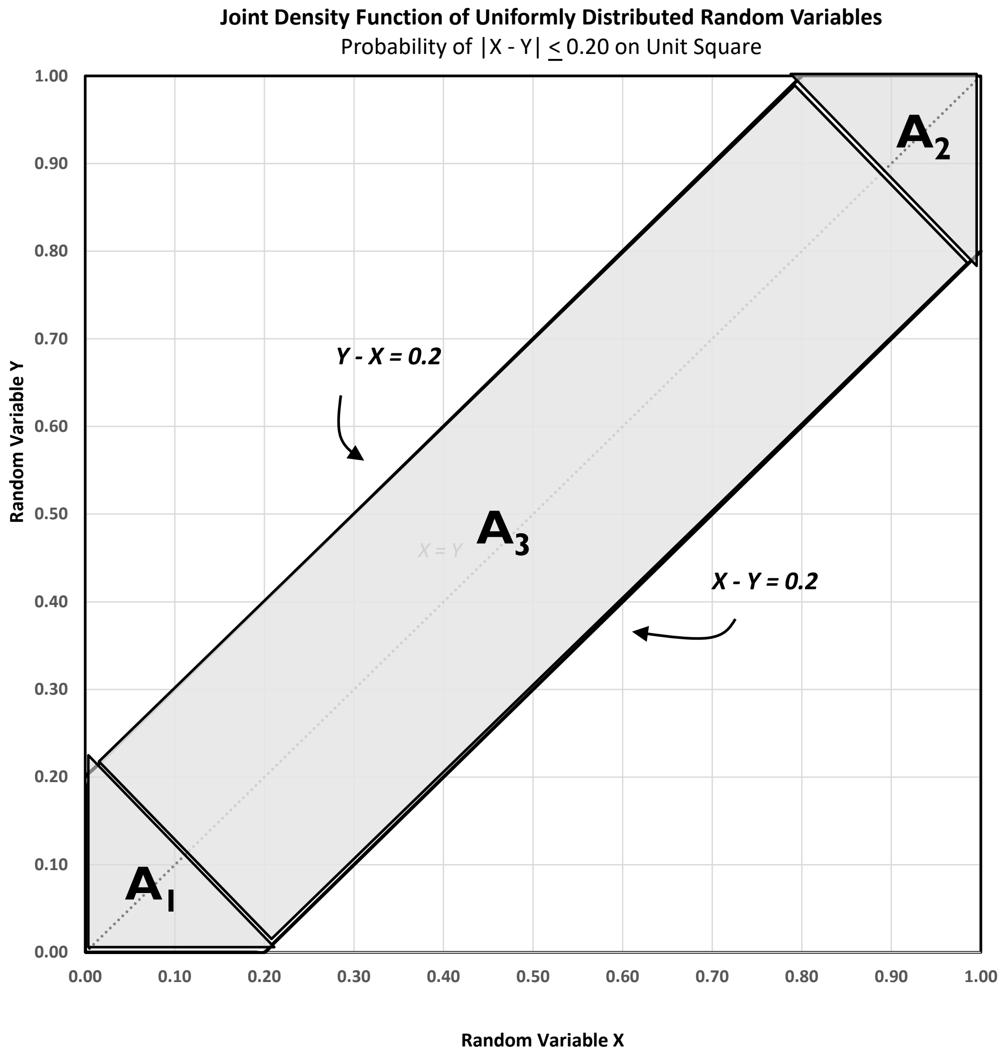
X: 0.375 * 60 = 22 minutes. Arrives at 9:22am.
Y: 0.902 * 60 = 54 minutes. Arrives at 9:54am.

The boy arrives at 9:22am. The girl arrives at 9:54am. She arrived after the 12 minute (0.2 hr) time window. So they **do not** meet.

Using 10,000 trials, the boy & girl meet 3,630 times.
Probability they will meet = **0.363**

Given that each person will "use up" 20% of their respective 1.0 hour time interval, we demonstrate the frequency (out of 10,000 trials) that the boy and girl are in "similar states" = Mutual Information

Unit Square: Geometric Estimate of Prob.



The Probability is Determined by Calculating the Area of the Shaded Region:

$A_1 = A_2 = 0.5\ (L) * (L) = 0.5\ L^2$

$A_3 = \text{sqrt}\ (2)\ (L) * \text{sqrt}\ (2)\ (1 - L)$

$= 2\ L\ (1 - L)$

$\text{Area} = A_1 + A_2 + A_3$

$\text{Area} = 0.5\ L^2 + 0.5\ L^2 + 2\ L\ (1 - L)$

$\text{Area} = L^2 + 2\ L\ (1 - L)$

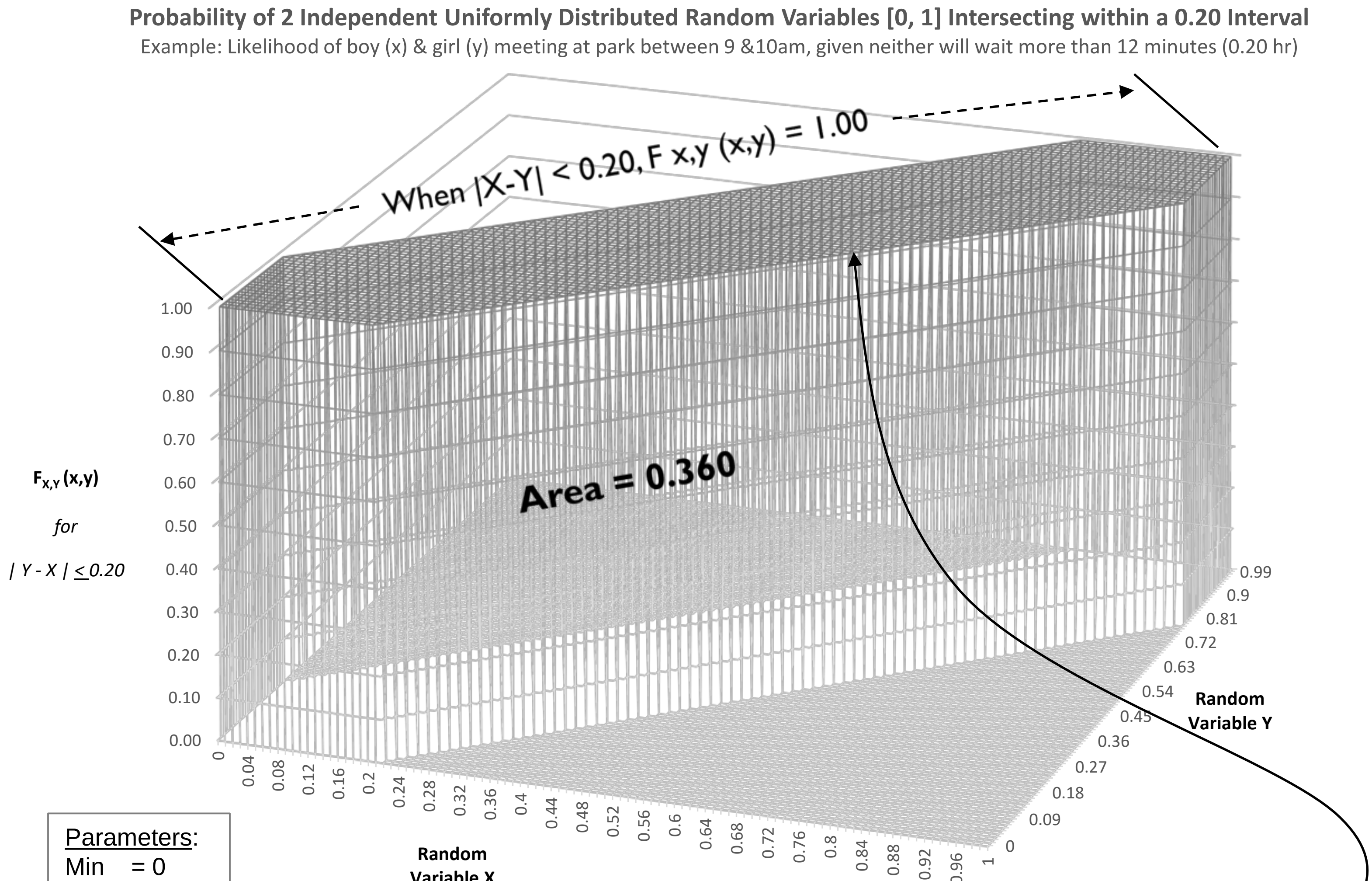
$\text{Area} = 0.20^2 + 2\ (0.20)\ (1 – 0.20)$

$\text{Area} = 0.360$

Note: This Probability is actually a Volume, not an Area …

Joint Distribution of Uniformly Distributed Random Variables X and Y

Under the condition to include only Fx,y(x,y) values in cases where |X – Y| ≤ 0.20



Parameters:
Min = 0
Max = 1

Given uniformly distributed random values, X and Y, for |X–Y| ≤ 0.20, the chance of a "meetup" remains constant along U[0,1]. Otherwise the joint probability = 0%

Outline

- Correlation Overview
- Why Propose Another Correlation Method?
- Underlying Basis for "Common Risk Factor" Method
  - Concept of Mutual Information
  - Using the Unit Square to Estimate Mutual Information
- "Common Risk Factor" Method (for pair of activities)
  - Apply 7 Steps to Estimate Correlation between 2 Distributions
- Examples
  - Correlation of Durations for Two Morning Commutes
  - Correlation of Costs for Two WBS Elements of a Spacecraft
- Conclusion, Other Potential Applications & Future Work

Common Risk Factor Method *(for 2 activities)*

(a) For methods on developing uncertainty distributions using risk factors, refer to “Expert Elicitation of a Maximum Duration using Risk Scenarios,” 2014 NASA Cost Symposium presentation, M. Greenberg

Ground Rules and Assumptions (1 of 2)

- If it is available, then this method can be used as a cross-check

- Future work will include efforts on negative correlation

- With > 5 common risk factors, SME has difficulty "separating" salient risk factors from all possible risk factors.
  - Risk factor pairs tend to become alike, producing correlations > 0.30
- As a general rule, risk factors contributing < 5% to overall uncertainty should be added into "Undefined" category

Ground Rules and Assumptions (2 of 2)

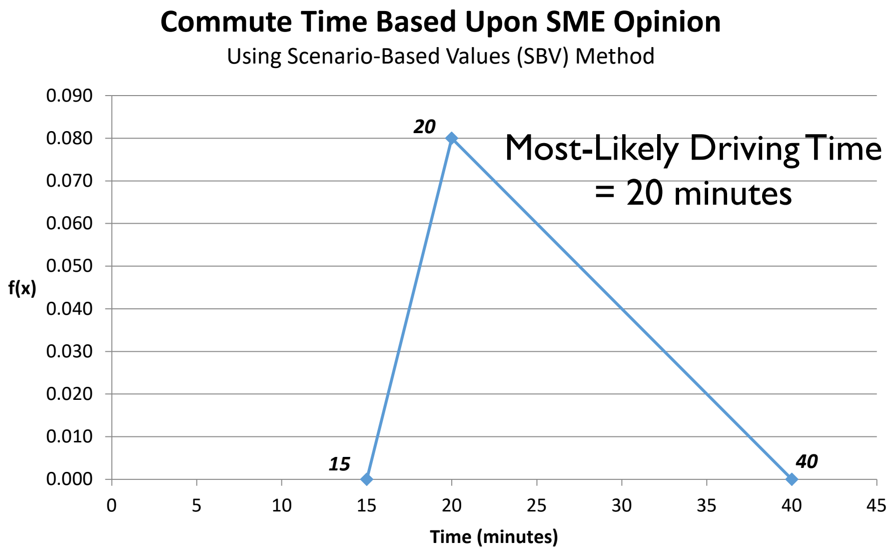
- Common risk factors are assumed to be correlated whenever the common risk factors are in a similar state. This occurs when each common risk factor has "overlapping" rv's along U[0,1]
  - Trial 98, Weather is *moderate* for both rv's X &Y => X & Y are Correlated
  - Trial 99, Weather is *moderate* for rv X, *severe* for rv Y = X & Y are not Correlated
- The least common denominator (LCD) of relative contributions of each common risk pair models each common risk factor as continuous rv's from 0 → min value, not anywhere along U[0,1]
  - Result is that LCD technique will produce lower correlation values

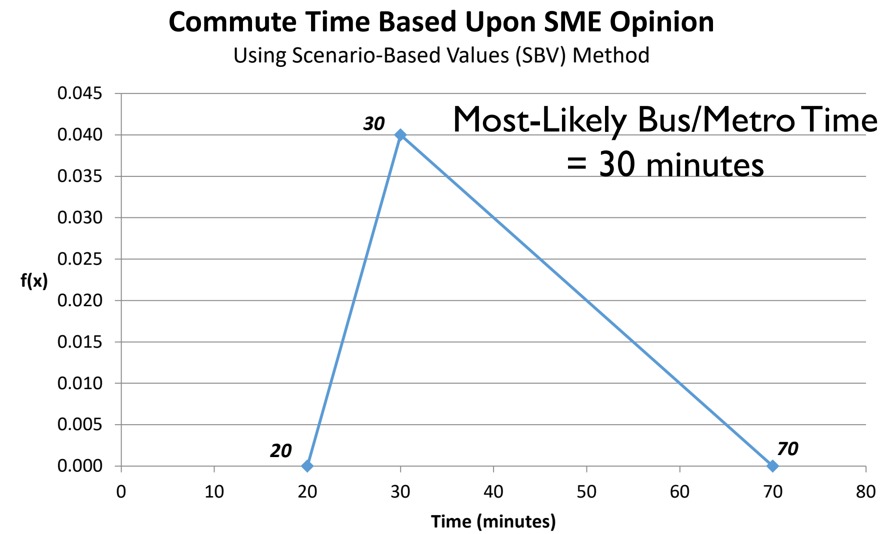
Example 1: Correlation of 2 Commute Durations

- Commute is from Commuter's Residence to anywhere in Washington, DC
- Maximum Commuting Distance for Phase A of the Study = 8 miles
- A person (X) commuting to work in DC from inside the beltway has a most-likely commute time of 20 minutes by car
- A person (Y) commuting into DC from inside the beltway has a most-likely commute time of 40 minutes by bus & metro
- To run the simulation for estimating total commute time, assume persons X and Y commutes have a medium correlation = 0.50.

Examples and Cases that Follow are Notional.
They are Provided to Demonstrate the Methodology.

Example 1: Commute Times





Bus/Metro: Potential 40 minute impact versus Most-Likely Bus/Metro Time

So what is the correlation between these two uncertainty distributions?

If we know the relative contributions of underlying risk factors for each distribution, we can calculate the correlation between these two distributions

Create Risk Reference Table (Step 1)

Objective	Means These are Primary Factors that can impact Objective
Maximize Average Speed from Residence to Workplace	Route Conditions
	# of Vehicles on Roads
	Mandatory Stops
	Efficiency
	Undefined

The utility of this Objective Hierarchy is to aid the Expert in:

(a) Establishing a Framework from which to elicit most risk factors,

(b) Describing the relative importance of each risk factor with respect to means & objective, and

(c) Creating specific risk scenarios

Create Risk Reference Table (Step 1 cont'd)

Objective	Means These are Primary Factors that can impact Objective
Maximize Average Speed from Residence to Workplace	Route Conditions
	# of Vehicles on Roads
	Mandatory Stops
	Efficiency
	Undefined

Q: What are some factors that could degrade route conditions?

A: *Weather, Road Construction, and Accidents*

Q: What influences the # of vehicles on the road in any given morning?

A: *Departure time, Day of the Work Week, and Time of Season (incl. Holiday Season)*

Q: What is meant by Mandatory Stops?

A: *By law, need to stop for Red Lights, Emergency Vehicles and School Bus Signals*

Q: What can reduce Efficiency?

A: *Picking the Bus or Metro Arriving Late, Bus Stopping at Most Stops, and Moving Below Optimal Speed (e.g. driving below speed limit).*

Create Risk Reference Table (Step 1 cont'd)

Objective	Means These are Primary Factors that can impact Objective	Risk Factors These are Causal Factors that can impact Means	Description (can include examples) Subject Matter Expert's (SME's) top-level description of each Barrier / Risk
Maximize Average Speed from Residence to Workplace	Route Conditions	Weather	Rain, snow or icy conditions. Drive into direct sun.
		Accidents	Vehicle accidents on either side of highway.
		Road Construction	Lane closures, bridge work, etc.
	# of Vehicles on Roads	Departure Time	SME departure time varies from 6:00AM to 9:00AM
		Day of Work Week	Driving densities seem to vary with day of week
		Season & Holidays	Summer vs. Fall, Holiday weekends
	Mandatory Stops	Red Lights	Approx 8 traffic intersections; some with long lights
		Emergency Vehicles	Incl. police, firetrucks, ambulances & secret service
		School Bus Signals	School buses stopping to pick up / drop off
	Efficiency	Bus/Metro Arriving Late	Bus arriving late. Metro arriving late.
		Bus Stopping at Most Stops	On rare occasion, will call someone during commute
		Moving below Optimal Speed	Bus or Car Driver going well below speed limit
	Undefined	Undefined	It's possible for SME to exclude some risk factors

This is the most time-intensive part of SME interview & serves as reference for the interview method being used.

Step 2. Estimate Risk Factor % Contributions

For each type of commute, respective SMEs ascribe the following "max" time impacts to 4 risk factors:

- Weather, Road Construction, Bus/Metro Arriving Late and Departure Time

Risk Factor	Max Impact vs Most Likely		
	Car	Bus/Metro	Total
Weather	4.0	2.0	6.0
Road Construction	10.0	8.0	18.0
Bus/Metro Arriving Late	0.0	26.0	26.0
Departure Time	6.0	4.0	10.0
Total Delay (minutes):	**20**	**40**	**60**

→

Contribution of Total	
Car	Bus/Metro
0.20	0.05
0.50	0.20
0.00	0.65
0.30	0.10
1.00	1.00

% Impact due to Realization of Given Risk

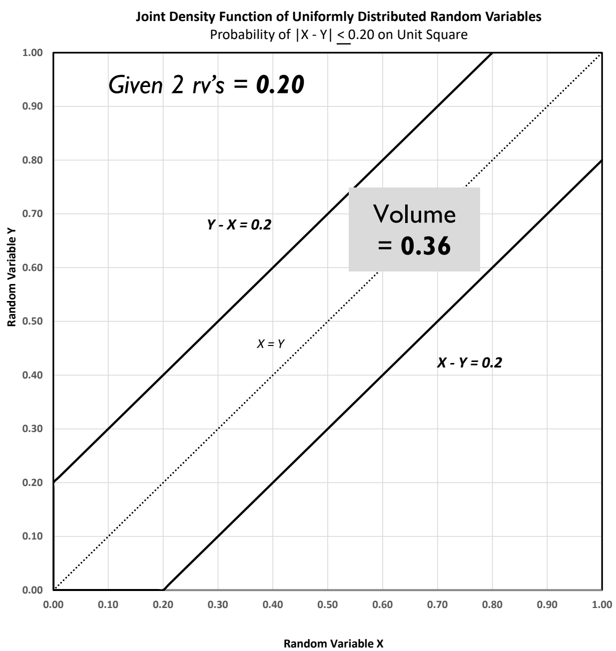
Note: These impacts can be elicited "ad-hoc" from the SME. Nevertheless, it is recommended to apply more structured methods during the SME interview for long-duration activities or ones with higher criticality indices. [a]

(a) For methods on developing uncertainty distributions using risk factors, refer to "Expert Elicitation of a Maximum Duration using Risk Scenarios," 2014 NASA Cost Symposium presentation, M. Greenberg

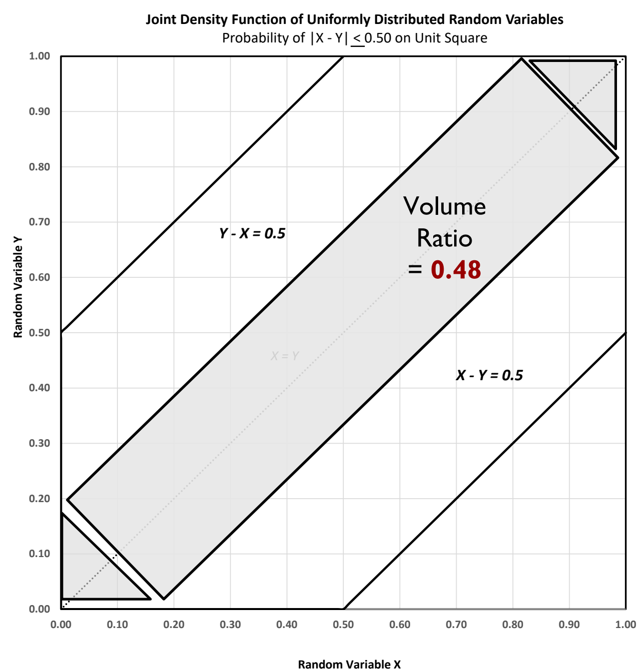
Correlation of a Risk Pair (Road Construction)

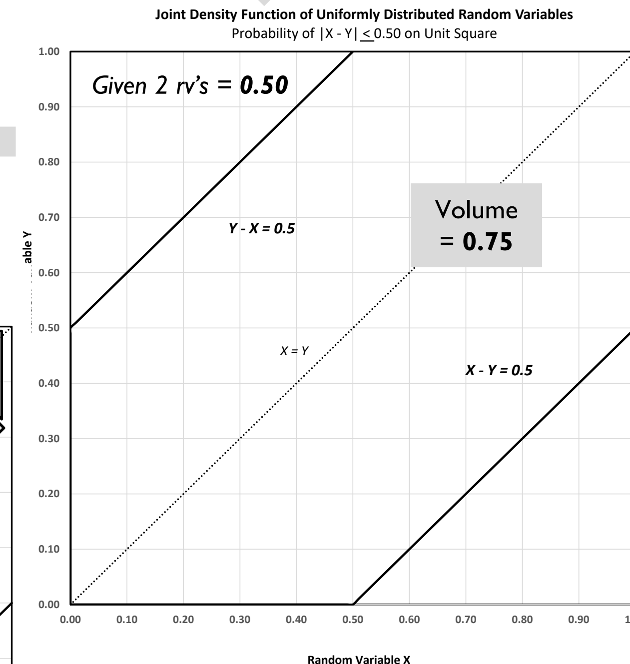
	Car	Bus/Metro
	0.20	0.05
Road Const	0.50	0.20
	0.00	0.65
	0.30	0.10

The "maximum possible" value of 0.50 is used to calculate a probability of 0.75 that rv's X and Y are in a similar "state."



Obtain mutual information by calculating volume ratio.





Correlation of this Risk Pair Indicates a "Relative" Volume = 0.36 / 0.75 = **0.48**

Common Risk Factor Method: Steps 3 – 7

Step 3. Min & Max Volumes Associated with Common Risk Factors

Step 4. Correlation (per risk factor pair) = Min Volume / Max Volume

Step 5. Weighting Factor for Each Min/Max = Max Volume divided by Sum of Max Volumes

Risk Factor	Contribution of Total		Calculated Volumes wrt		Min Volume	Max Volume	Min/Max Volume	Weighting Factor	Weighted Min/Max
	Car	Bus/Metro	Car	Bus/Metro					
Weather	0.20	0.05	0.360	0.098	0.098	0.360	0.271	***0.14***	0.039
Road Construction	0.50	0.20	0.750	0.360	0.360	0.750	0.480	***0.30***	0.144
Bus/Metro Arriving Late	0.00	0.65	0.000	0.878	0.000	0.878	0.000	***0.35***	0.000
Departure Time	0.30	0.10	0.510	0.190	0.190	0.510	0.373	***0.20***	0.076
Totals:	1.00	1.00	1.620	1.525	**0.648**	**2.498**	*0.281*	1.000	**0.259**

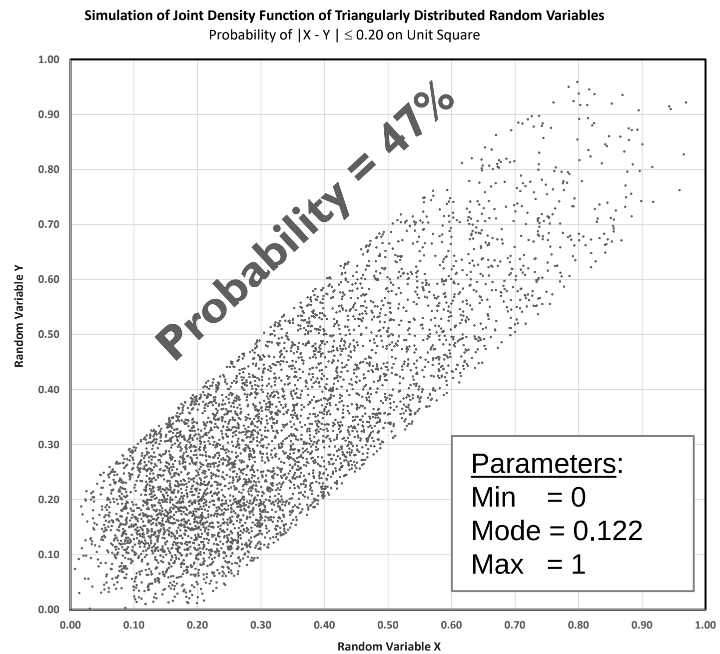
Step 6. Weight Correlation of Each Pair of Common Risk Factors

Step 7. Sum up Weighted Correlations to get total Correlation

The 0.26 correlation value reflects the mutual information (of common risks) between these 2 activities. The analyst's "Causal Guess" of 0.50 was not a reasonable estimate of correlation.

Non-Uniformly Distributed Risk Factors

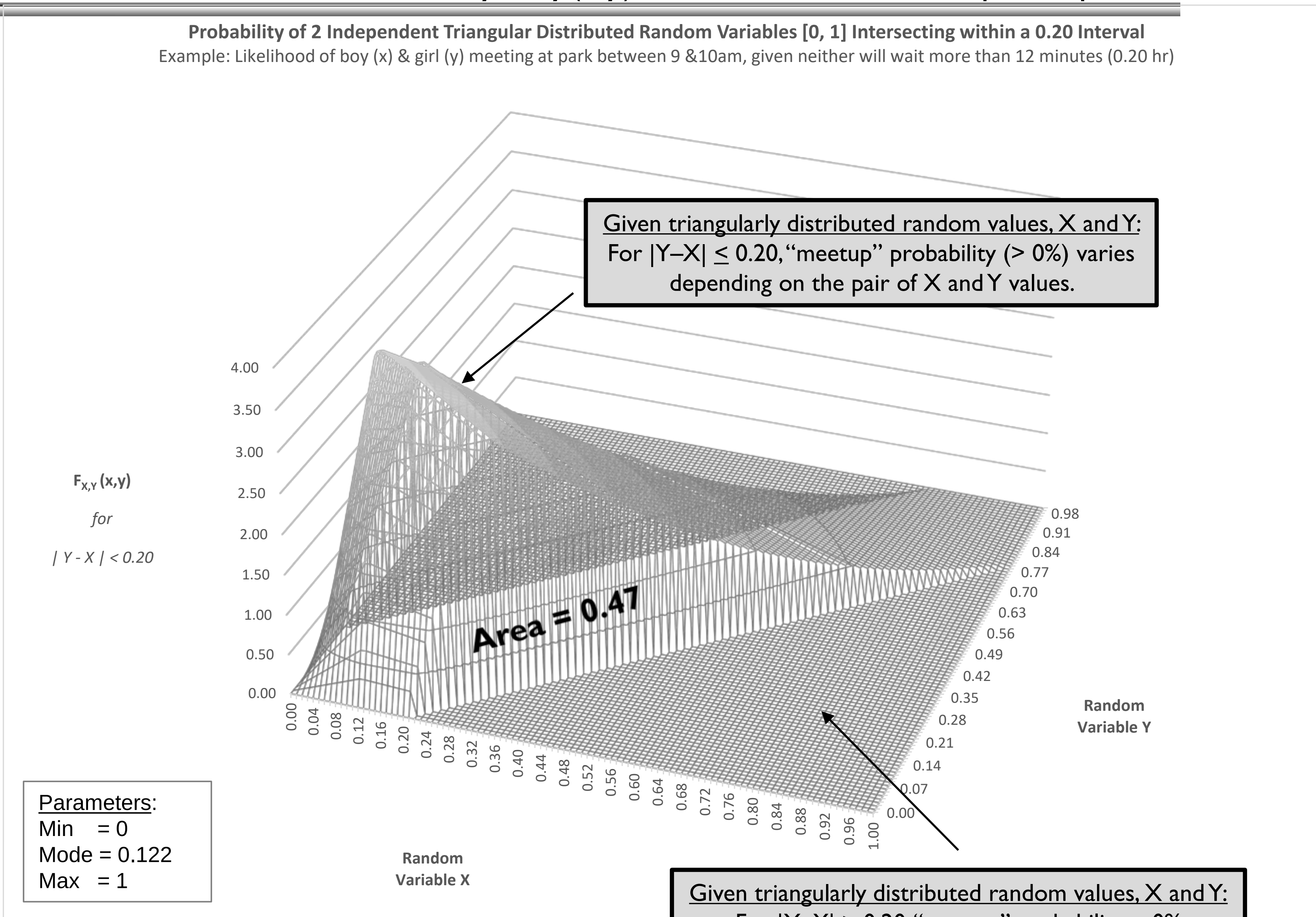
Iteration	rv (X)	rv (Y)	|X - Y|	|X - Y| < 0.2?
1	0.285	0.193	0.092	1
2	0.381	0.195	0.186	1
3	0.262	0.170	0.093	1
4	0.374	0.137	0.237	0
5	0.061	0.161	0.100	1
6	0.496	0.911	0.415	0
7	0.529	0.286	0.243	0
:	:	:	:	:
:	:	:	:	:
9998	0.144	0.677	0.532	0
9999	0.169	0.110	0.058	1
10000	0.654	0.513	0.141	1
			Total =	4730



When each risk factor is modeled as a **triangular distribution**, the simulation indicates the boy and girl meet 4,730 times out of 10,000 trials = 47% probability of meeting.
This probability increased due to added central tendency from each risk factor.

Joint Distribution of Triangularly Distributed Random Variables X and Y

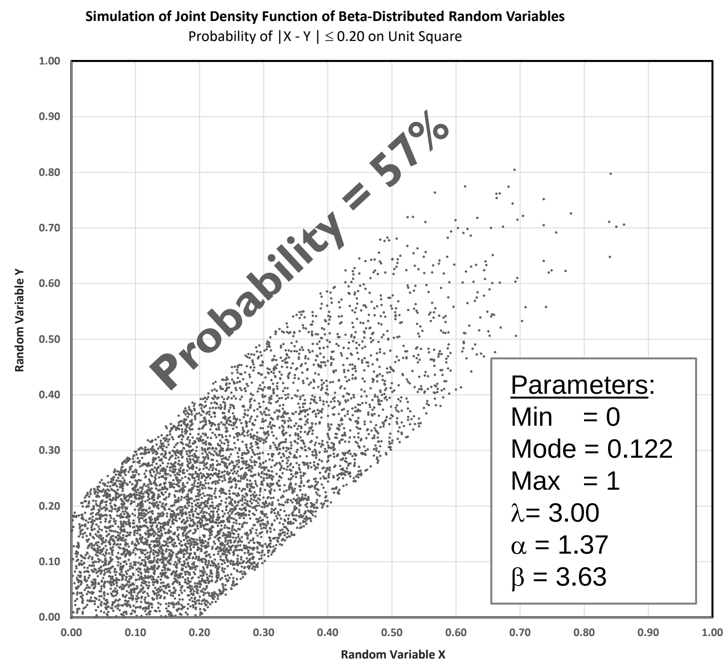
Under the condition to include only Fx,y(x,y) values in cases where |X – Y| ≤ 0.20



Given triangularly distributed random values, X and Y:
For |X–Y| > 0.20, "meetup" probability = 0%

Non-Uniformly Distributed Risk Factors

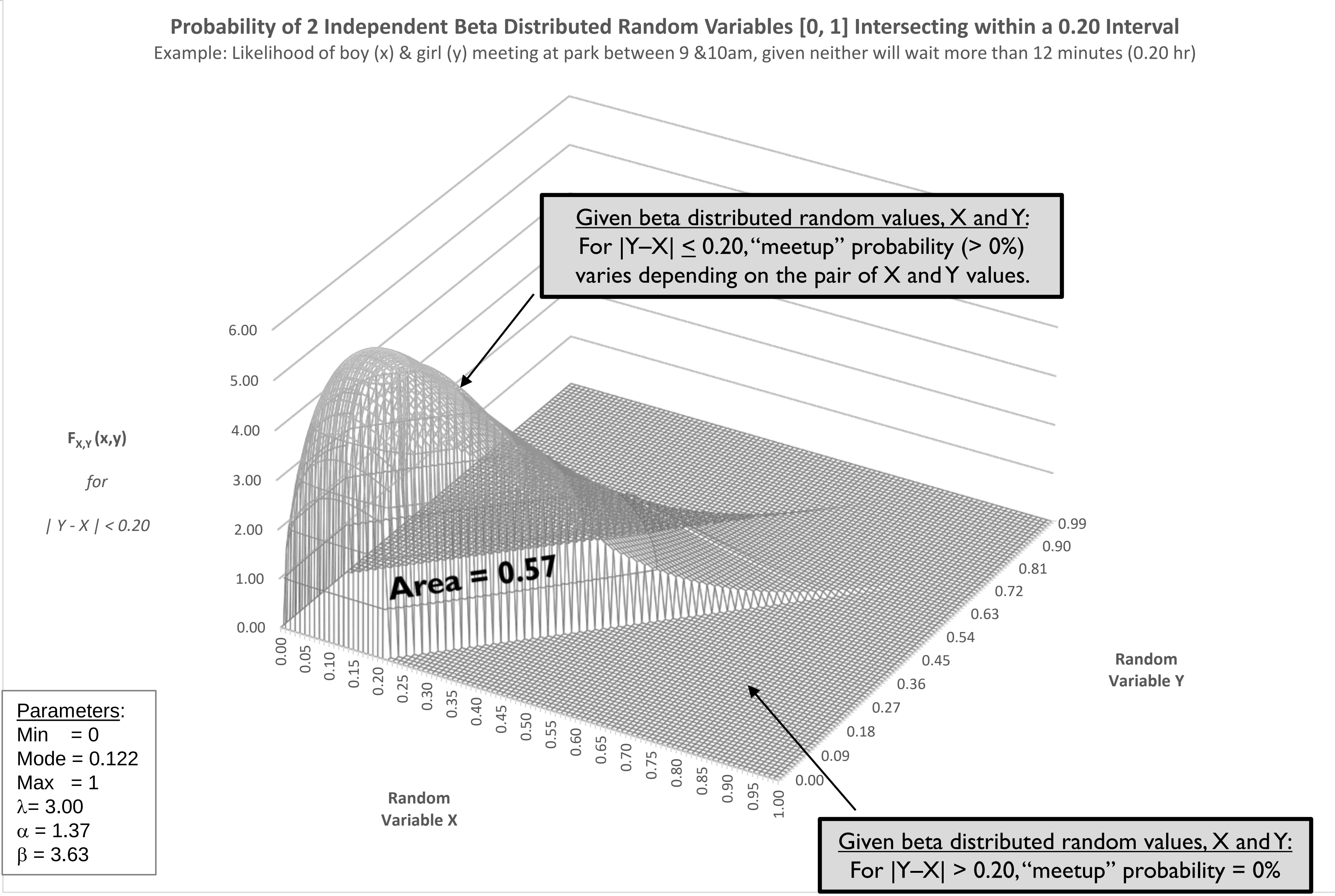
Iteration	rv (X)	rv (Y)	|X - Y|	|X - Y| < 0.2?
1	0.071	0.072	0.001	1
2	0.035	0.622	0.587	0
3	0.416	0.728	0.312	0
4	0.001	0.370	0.369	0
5	0.291	0.343	0.052	1
6	0.213	0.275	0.061	1
7	0.145	0.098	0.047	1
:	:	:	:	:
:	:	:	:	:
9998	0.292	0.019	0.273	0
9999	0.219	0.277	0.058	1
10000	0.251	0.154	0.096	1
			Total =	5725



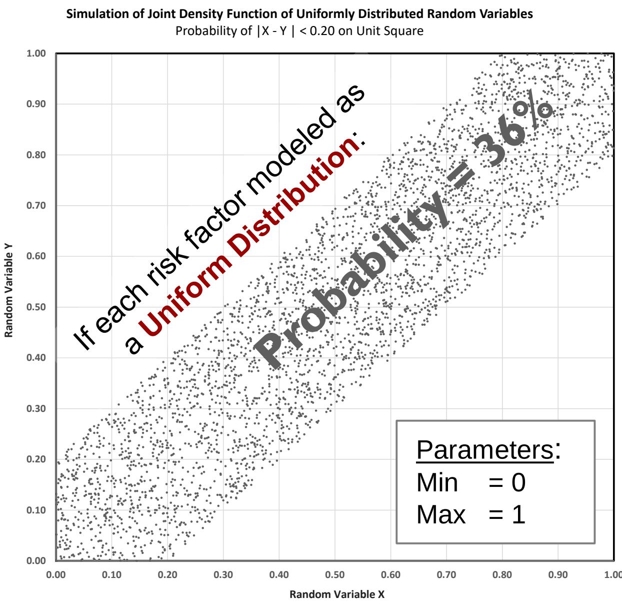
When each risk factor is modeled as a **beta distribution**, the simulation indicates the boy and girl meet 5,725 times out of 10,000 trials = 57% probability of meeting.
This probability increased due to added central tendency from each risk factor.

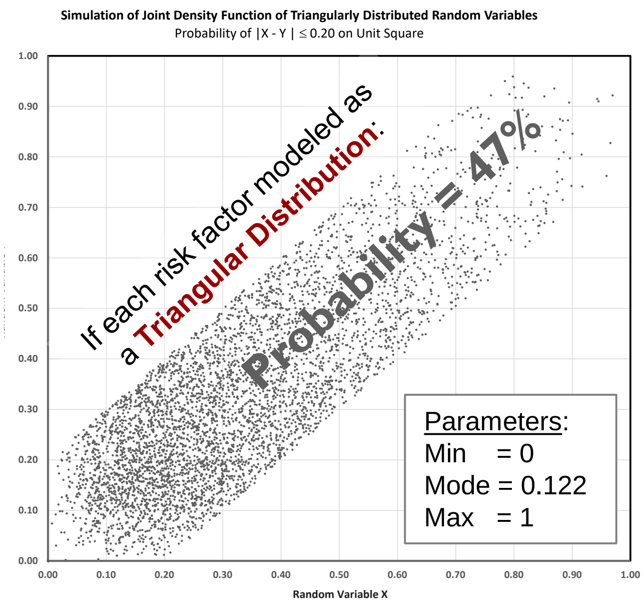
Joint Distribution of Beta Distributed Random Variables X and Y

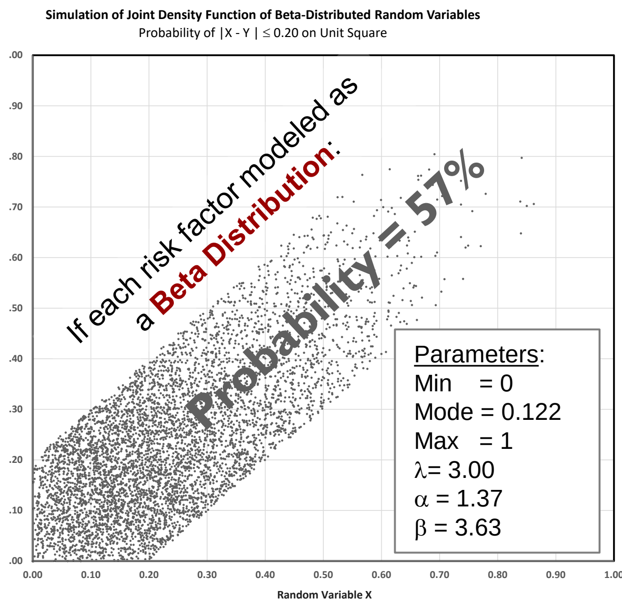
Under the condition to include only Fx,y(x,y) values in cases where |X – Y| < 0.20



Non-Uniformly Distributed Risk Factors







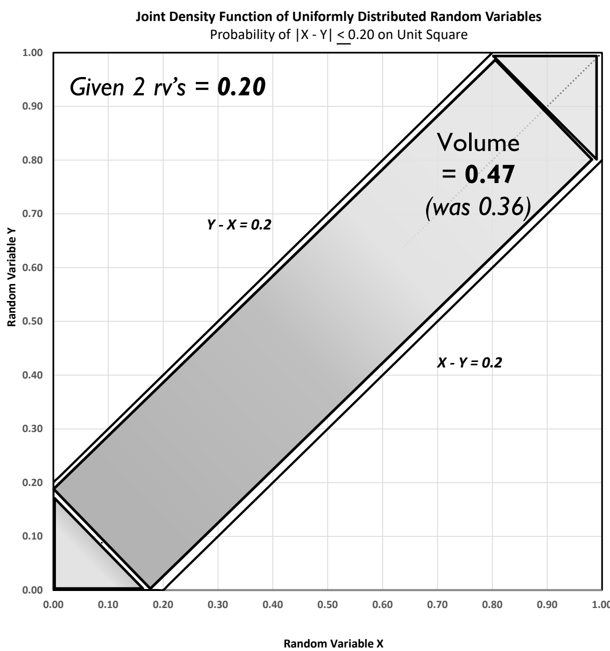
As the **central tendency** for a risk factor increases, so will the probability of meeting, resulting in a higher probability that rand. var. X and rand. var. Y are in a similar state.

As previously shown in Slide 26, this is just part of the correlation calculation. This "volume" must be compared to the "possible" volume of the risk pair …

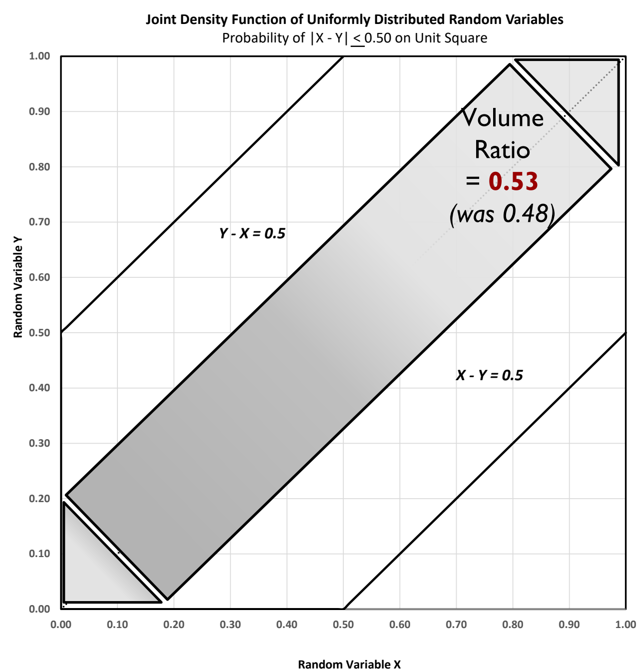
Correlation of a Risk Pair (Road Const'n): Triangular Risk Factor *

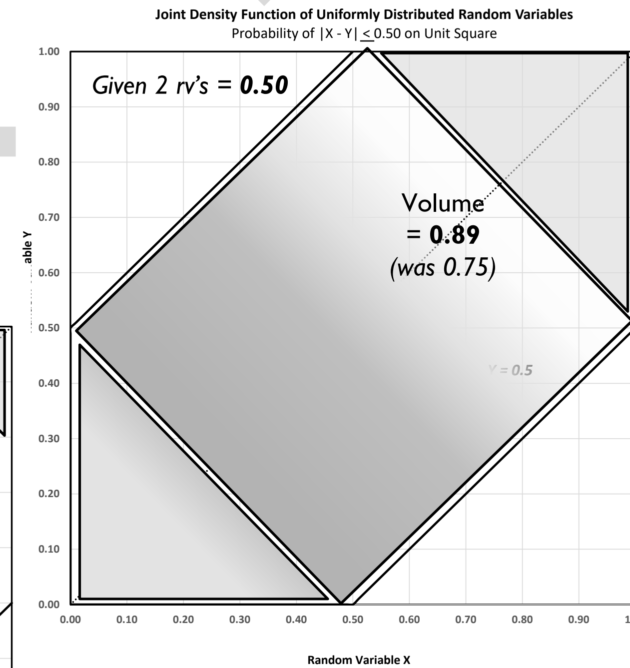
	Car	Bus/Metro
	0.20	0.05
Road Const	0.50	0.20
	0.00	0.65
	0.30	0.10

The "maximum possible" value of 0.50 is used to calculate a probability of 0.89 that rv's X and Y are in a similar "state."



Obtain mutual information by calculating volume ratio.





Correlation of this Risk Pair Indicates a "Relative" Volume = 0.47 / 0.89 = **0.53** *(was 0.48 for U[0,1])*

* *If this exercise were done for random variables that were beta distributed, the correlation of this risk pair would be = 0.57 / 0.93 =* **0.61**

Recommended Applications

Best for looking at Correlations for Distributions where Risk Impacts are of Most Concern …

- **Cost and Schedule Estimating**
  - Estimates early-on in Acquisition Life Cycle
    - Pre-Phase A, pre-Milestone A, etc. where <5 "top-level" risks tend to dominate
  - Technology Cost Estimating (TRL < 6)
  - Cross-check on data-driven Correlations ("Statistical")
  - Support Independent Estimates (and/or Assessments)
- **Technical Design and/or Assessment**
  - Assess Early-stage Risks in System Design & Test
  - Assess threats / barriers to Systems' Safety
  - Standing Review Board (SRB) Evaluations

Recap / Conclusion

In summary, this presentation covered:

- Current challenges that estimators have in specifying defensible correlations between uncertainty distributions
- The concept of modeling correlation based upon mutual information
- How the unit square can be used to estimate correlation
  - Depicted as an "intersection" in the unit square of two random variables.
    - Representing risks assumed to follow (a) uniform (b) triangular or (c) beta distributions.
- A 7-step method on how to estimate correlation based upon knowledge of risk factors common among the pair of uncertainty distributions
- Examples on how to apply the 7-step method (see more in Backup!)

Unlike other methods, the Common Risk Factor Method provides correlation between 2 uncertainties based upon common root-causes. Applying this method may lessen the degree of subjectivity in the estimate.

Presenter: Marc Greenberg

Title: Operations Research Analyst

Organization: NASA Strategic Investment Division

Office Location: NASA HQ, Washington, DC

Email: marc.w.greenberg@nasa.gov

Phone: 202.358.1025

Backup Slide Set #1 of 3

Continuation of Example 1

Results: Correlation of Commute Time Uncertainties

- Part 1 – All risk factors contribute to > 98% of uncertainty
- Part 2 – Account for "Unexplained Uncertainty" for each Commuting Uncertainty Distributions (Car and Bus/Metro)

- Improve % on-time arrivals of busses and metro trains
- Improve arrival frequency of busses and metro trains during holidays

Case A: Correlation of Commute Time Uncertainties

Risk Factor	Contribution to Commute Time Uncertainty (Car)	Contribution to Commute Time Uncertainty (Bus/Metro)	Min Volume	Max Volume	Correlation due to Common Risk Factor	Weighting Factor	Weighted Correlation
Weather	0.25	0.20	0.360	0.438	0.823	0.184	0.152
Accidents	0.34	0.18	0.328	0.564	0.580	0.238	0.138
Road Construction	0.26	0.12	0.226	0.452	0.499	0.191	0.095
Departure Time	0.15	0.10	0.190	0.278	0.685	0.117	0.080
Bus/Metro Arriving Late	0.00	0.40	0.000	0.640	0.000	0.270	0.000
Total:	1.00	1.00	*1.103*	*2.372*		1.000	**0.465**

Adding content bumps up Correlation from 0.26 to 0.465.

SME provides content on "Undefined" (a catch-all for "Unexplained Variation"):

Risk Factor	Contribution to Commute Time Uncertainty (Car)	Contribution to Commute Time Uncertainty (Bus/Metro)	Min Volume	Max Volume	Correlation due to Common Risk Factor	Weighting Factor	Weighted Correlation
Weather	0.20	0.14	0.260	0.360	0.723	0.147	0.106
Accidents	0.28	0.13	0.243	0.482	0.505	0.197	0.099
Road Construction	0.22	0.08	0.154	0.392	0.392	0.160	0.063
Departure Time	0.12	0.07	0.135	0.226	0.599	0.092	0.055
Bus/Metro Arriving Late	0.00	0.28	0.000	0.482	0.000	0.197	0.000
Undefined	0.18	0.30	0.328	0.510	0.000	0.208	0.000
Total:	1.00	1.00	*1.120*	*2.450*		1.000	**0.323**

Having undefined risk factors reduces Correlation from 0.465 to 0.32.

Case B: Correlation of Commute Time Uncertainties

Risk Factor	Contribution to Commute Time Uncertainty (Car)	Contribution to Commute Time Uncertainty (Bus/Metro)	Min Volume	Max Volume	Correlation due to Common Risk Factor	Weighting Factor	Weighted Correlation
Weather	0.20	0.16	0.294	0.360	0.818	0.152	0.124
Accidents	0.28	0.15	0.278	0.482	0.576	0.203	0.117
Road Construction	0.22	0.09	0.172	0.392	0.439	0.165	0.073
Departure Time	0.12	0.07	0.135	0.226	0.599	0.095	0.057
Bus/Metro Arriving Late	0.00	0.20	0.000	0.360	0.000	0.152	0.000
Undefined	0.18	0.33	0.328	0.551	0.000	0.233	0.000
Total:	1.00	1.00	*1.207*	*2.370*		1.000	**0.371**

The Risk Mitigation effort would slightly increase Correlation from 0.32 to 0.37.

This increase in Correlation (versus Case A) is due to an increase in Mutual Information between the common Risk Pairs (where BOTH values > 0)

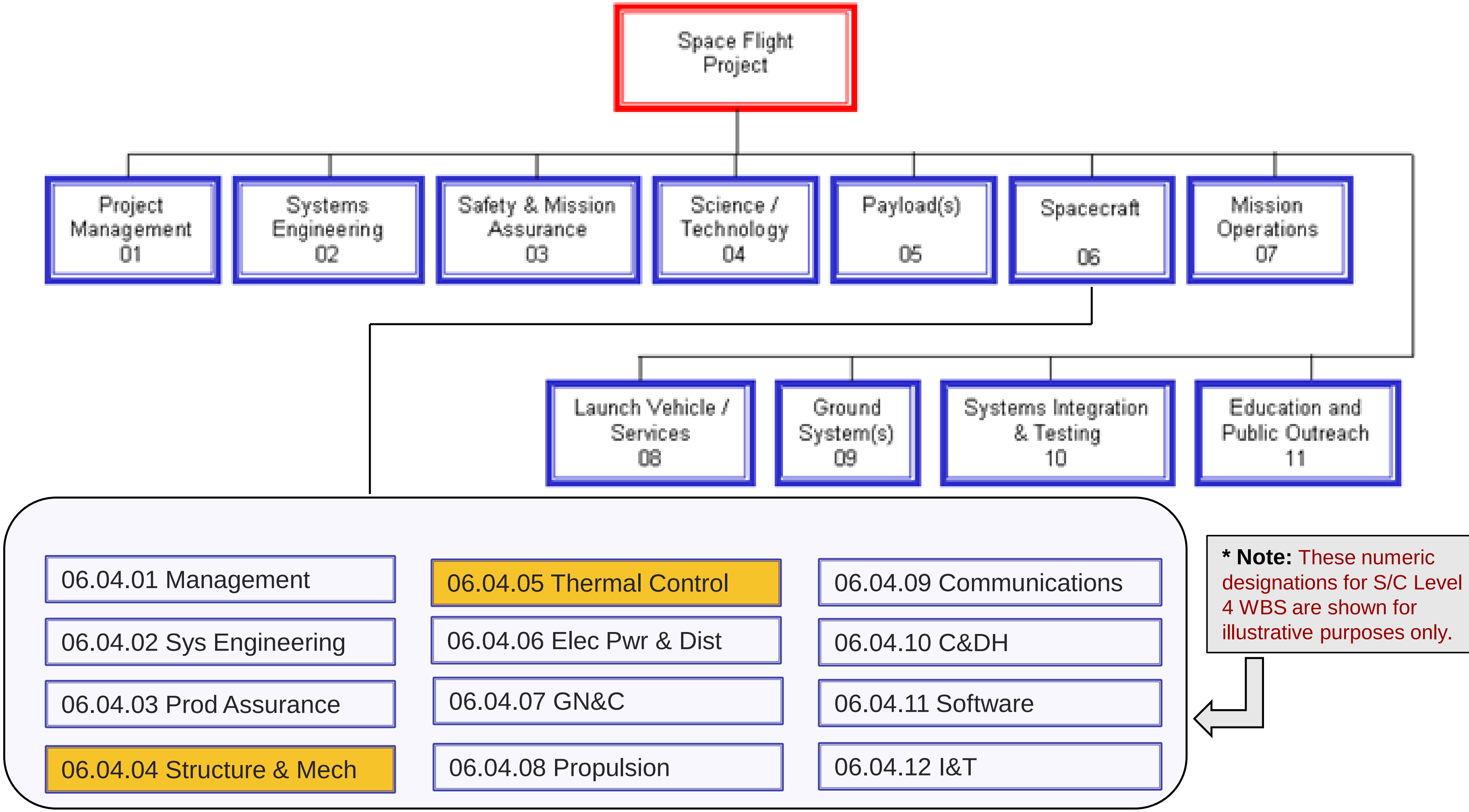
By reducing Bus/Metro's top "uncertainty driver," the dispersion for the Bus/Metro commute went down (not shown here).
At the same time, correlation between the distributions went up.

Backup Slide Set #2 of 3

Example 2

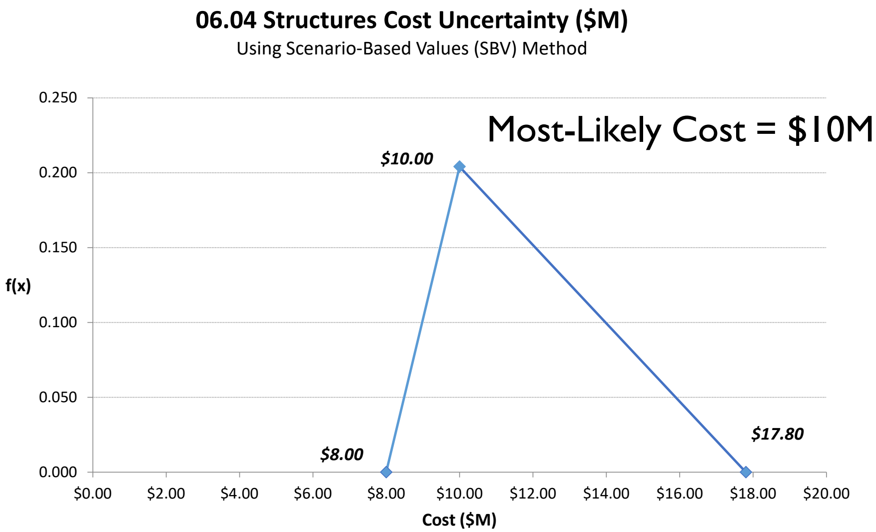
Space Flight Project WBS Standard Level 2 Elements

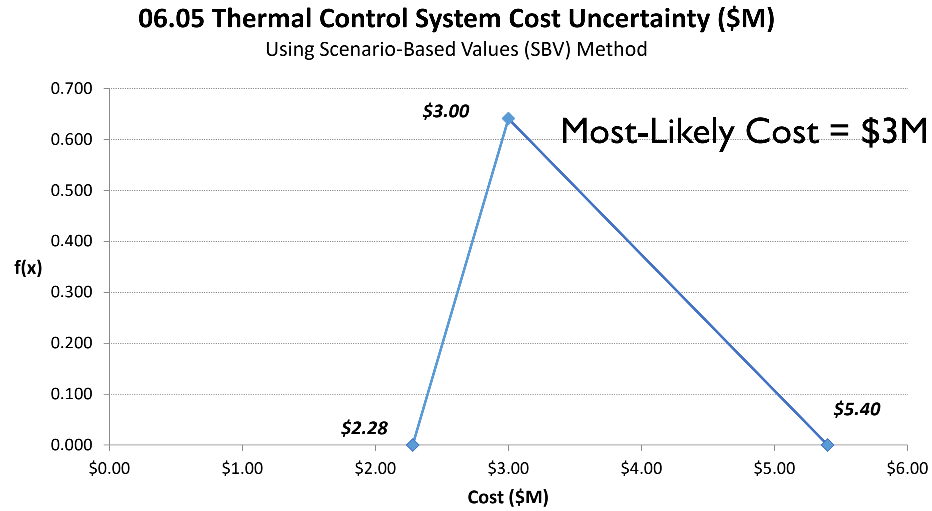
Ref: NPR 7120.5, Appendix G



The next notional example shows an estimate of correlation between pre-Phase A costs of S/C "Structure & Mech" and "Thermal Control"

Example 2: Spacecraft Cost Elements





Thermal Control Systems: Potential $2.4M impact versus Most-Likely Cost

So what is the correlation between these two uncertainty distributions?

If we know the relative contributions of underlying risk factors for each distribution, we can calculate the correlation between these two distributions

Create Risk Reference Table (Step 1)

Objective	Means These are Primary Factors that can impact Objective
Complete DDT&E for a Spacecraft that Meets Cost & Schedule Objectives	Complete Technical Design to Satisfy System (or Mission) Requirements
	Provide for Adequate Resources & Expertise for Program Execution
N/A	Undefined

The utility of this Objective Hierarchy is to aid the Expert in:

(a) Establishing a Framework from which to elicit most risk factors,

(b) Describing the relative importance of each risk factor with respect to means & objective, and

(c) Creating specific risk scenarios

Create Risk Reference Table (Step 1, cont'd)

Objective	Means These are Primary Factors that can impact Objective
Complete DDT&E for a Spacecraft that Meets Cost & Schedule Objectives	Complete Technical Design to Satisfy System (or Mission) Requirements
	Provide for Adequate Resources & Expertise for Program Execution
N/A	Undefined

Q: What could influence the successful completion of your Technical Design?

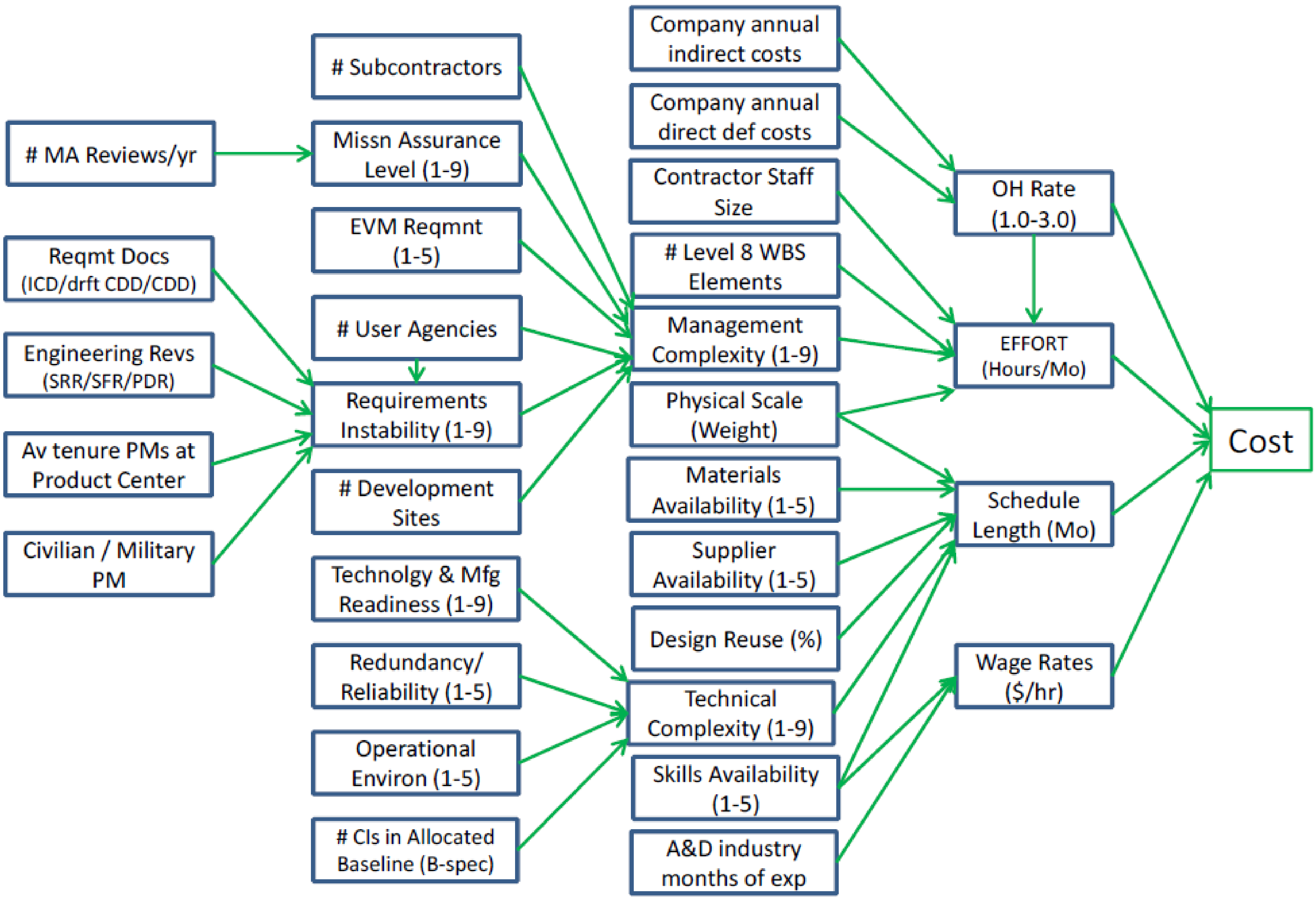
- *Design Complexity*
- *System Integration Complexity*
- *1 or more Immature Technologies*
- *Requirements Creep*
- *Skills Deficiency (Vendor)*

Q: What are threats and barriers for you getting adequate resources & expertise for Program Execution?

- *Lack of Programmatic Experience (NASA)*
- *Material Price Volatility*
- *Organizational Complexity*
- *Funding Instability*
- *Insufficient Reserves (Sched and/or Cost)*

(a) Next slide served as reference to assist in brainstorming process

Space Vehicle Development Cost "Causal Process"



Create Risk Reference Table (Step 1, cont'd)

Objective	Means These are Primary Factors that can impact Objective	Risk Factors (Primary) These are Causal Factors (aka "Threats" or "Barriers") that can impact Means	Description Subject Matter Expert's (SME's) top-level description of each Barrier / Risk
Complete DDT&E for a Spacecraft that Meets Cost & Schedule Objectives	Complete Technical Design to Satisfy System (or Mission) Requirements	Design Complexity	The complexity of designing certain aspects may be underestimated
		System Integration Complexity	We don't fully appreciate the challenges of system integration that will need to occur in 18 months
		1 or more Immature Technologies	There is a likelihood that we may need to incorporate certain components that are currently at TRL 6
		Requirements Creep	About 2/3 of these types of projects have experienced requirements creep in the past decade
		Skills Deficiency (Vendor)	The Vendor may lose some of it's "graybeards" over the next year, leaving a dearth in Technical Expertise
	Provide for Adequate Resources & Expertise for Program Execution	Lack of Programmatic Experience (NASA)	The Program Office staff has experienced a higher-than-usual turnover rate in the past year
		Material Price Volatility	The system includes exotic matls that, in the past, were subject to large price swings (largely due to low supply)
		Organizational Complexity	As of right now, there are 2 vendors, 4 sub-contractors, 3 NASA Centers and 1 university working on this project
		Funding Instability	Because this project is not an Agency priority, it is subject to funding cuts in any given year.
		Insufficient Reserves (Sched and/or Cost)	Because of the above risks, it's likely that project will not have sufficient schedule margin and/or cost reserves
N/A	Undefined	Undefined	In most cases, the SME will not be able to specify ALL risk factors that contribute to schedule / cost uncertainty

This is the most time-intensive part of SME interview & serves as reference for the interview method being used.

Step 2. Estimate Risk Factor % Contributions

For each cost, the SME ascribes the following "max" cost impacts to 5 risk factors:

- Systems Integration Complexity, Requirements Creep, Skills Deficiency (Vendor), Lack of Programmatic Experience (NASA) and Organizational Complexity

Risk Factor	Max Impact vs Most Likely shown by WBS in $M			Contribution of Total	
	06.04.04	06.04.05	Total ($M)	06.04.04	06.04.05
System Integration Complexity	$2.00	$0.45	$2.45	0.26	0.21
Requirements Creep	$1.50	$0.75	$2.25	0.19	0.36
Skills Deficiency (Vendor)	$0.80	$0.00	$0.80	0.10	0.00
Lack of Programmatic Experience (NASA)	$1.00	$0.30	$1.30	0.13	0.14
Organizational Complexity	$1.00	$0.00	$1.00	0.13	0.00
Undefined	$1.50	$0.60	$2.10	0.19	0.29
Total Cost Impact ($M):	**$7.80**	**$2.10**	**$9.90**	**1.00**	**1.00**

% Impact Due to Realization of Given Risk

Steps to Calculate Correlation Between These 2 Spacecraft WBS are the Same as Those Used for Example 1.

Common Risk Factor Method: Steps 3 - 7

Step 3. Min & Max Volumes Associated with Common Risk Factors

Step 4. Correlation (per risk factor pair) = Min Volume / Max Volume

Step 5. Weighting Factor for Each Min/Max = Max Volume divided by Sum of Max Volumes

Risk Factor	Contribution of Total		Calculated Volumes wrt		Min Volume	Max Volume	Min/Max Volume	Weighting Factor	Weighted Min/Max
	06.04.04	06.04.05	06.04.04	06.04.05					
System Integration Complexity	0.26	0.21	0.447	0.383	0.383	0.447	0.856	***0.20***	0.172
Requirements Creep	0.19	0.36	0.348	0.587	0.348	0.587	0.592	***0.26***	0.156
Skills Deficiency (Vendor)	0.10	0.00	0.195	0.000	0.000	0.195	0.000	***0.09***	0.000
Lack of Programmatic Experience (NASA)	0.13	0.14	0.240	0.265	0.240	0.265	0.905	***0.12***	0.108
Organizational Complexity	0.13	0.00	0.240	0.000	0.000	0.240	0.000	***0.11***	0.000
Undefined	0.19	0.29	0.348	0.490	0.348	0.490	**0.000**	***0.22***	0.000
Totals:	1.00	1.00	1.817	1.724	**1.318**	**2.223**	*0.392*	1.000	**0.436**

Step 6. Weight Correlation of Each Pair of Common Risk Factors

Step 7. Sum up Weighted Correlations to get total Correlation

The 0.44 correlation value reflects the mutual information (of common risks) **between Costs of WBS 06.04.04 and 06.04.05**

Case A: Correlation of Spacecraft Cost Uncertainties

Risk Factor	Contribution to WBS Cost Uncertainty (06.04.04)	Contribution to WBS Cost Uncertainty (06.04.05)	Min Volume	Max Volume	Correlation due to Common Risk Factor	Weighting Factor	Weighted Correlation
System Integration Complexity	0.26	0.15	0.278	0.447	0.621	0.191	0.119
Requirements Creep	0.19	0.42	0.348	0.664	0.524	0.284	0.149
Skills Deficiency (Vendor)	0.10	0.00	0.000	0.195	0.000	0.083	0.000
Lack of Programmatic Experience	0.13	0.10	0.190	0.240	0.792	0.103	0.081
Organizational Complexity	0.13	0.00	0.000	0.240	0.000	0.103	0.000
Undefined	0.19	0.33	0.348	0.551	0.000	0.236	0.000
Total:	1.00	1.00	*1.163*	*2.336*		1.000	**0.349**

The Risk Mitigation effort would decrease Correlation from 0.44 to 0.35.

This decrease in Correlation (versus Baseline) is due to an decrease in Mutual Information between the common Risk Pairs (where BOTH values > 0)

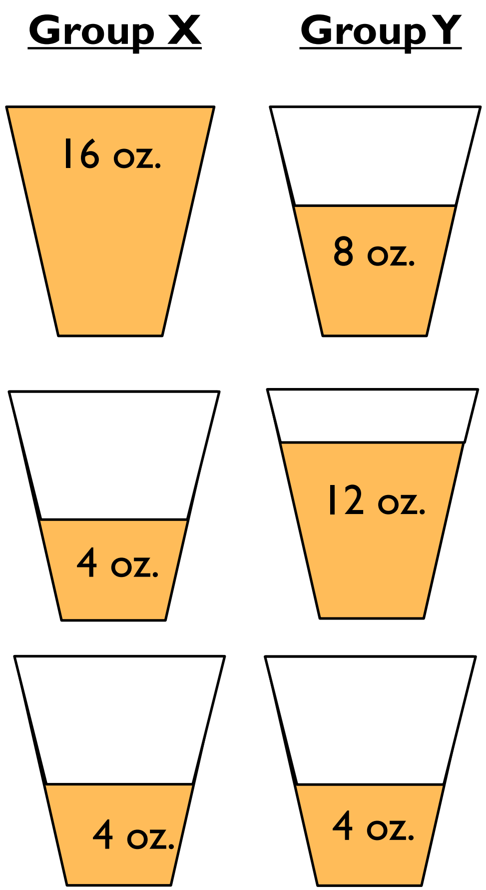
By reducing two "uncertainty drivers," the dispersion for the WBS 06.04.05 (Thermal Ctrl) went down (not shown here). **Also, correlation between the distributions went slightly down.**

Backup Slide Set #3 of 3

Depictions of Mutual Information

Mutual Information between 2 groupings

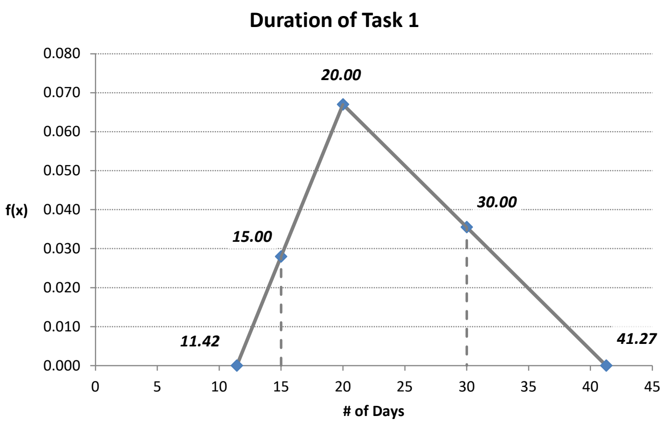
Method 1: Mutual Information = Σ Minimum (X,Y) / Σ Maximum (X,Y)



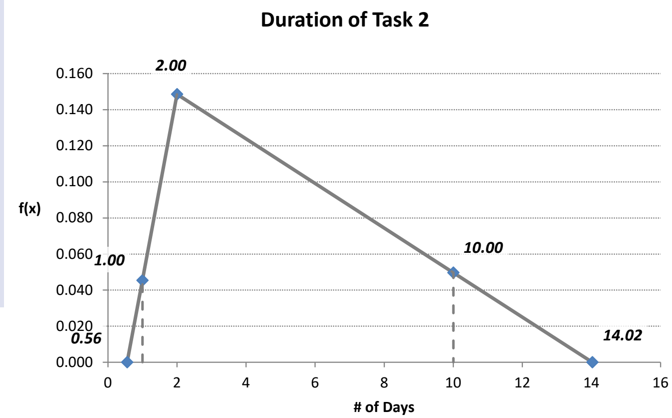
Minimum (X, Y)	Maximum (X, Y)	
8	**16**	**8 / 16 = 0.50**
4	**12**	**4 / 12 = 0.33**
4	**4**	**4 / 4 = 1.00**
Sum: 16	**32**	**16 / 32 = 0.50**

Mutual Information of Risk Factors

Mutual information can also be applied to risk factors that are common among a pair of uncertainty distributions.



The more "similar" the 2 weather contributions (to their respective task uncertainties), the higher the % of mutual information.



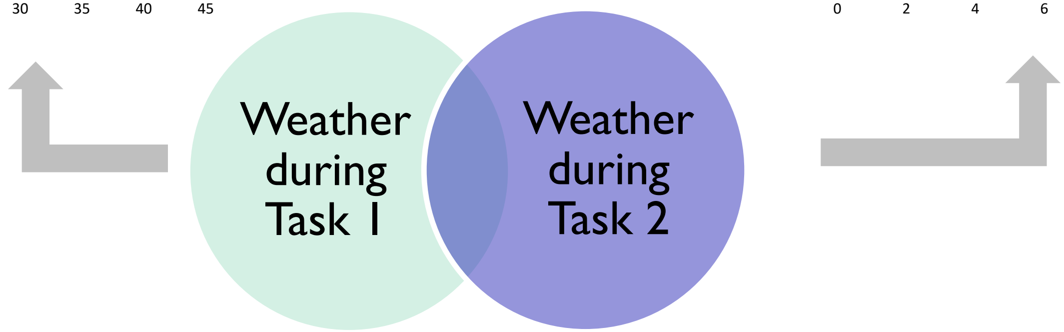


*Illustration showing **Weather** as a risk factor attributed to duration uncertainties for Tasks 1 and 2.*
(This common risk factor reflects mutual information between Tasks 1 & 2)

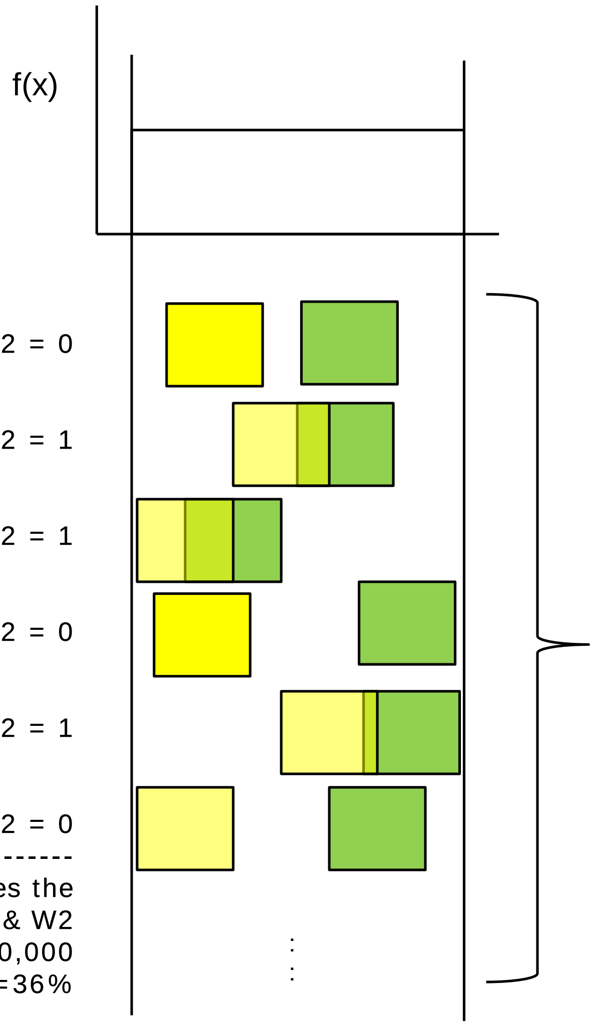
Depiction of 2 Uniformly Distributed RVs Intersecting …

Given "Weather" for Distributions 1 and 2

W1   W2

Weighting for each continuous random variable = 0.2

For the following interval U[0,1], how often would W1 and W2 be in a "similar" state?



Weather is not similar between W1 and W2 = 0

Weather is similar between W1 and W2 = 1

Weather is similar between W1 and W2 = 1

Weather is not similar between W1 and W2 = 0

Weather is similar between W1 and W2 = 1

Weather is not similar between W1 and W2 = 0

After 10,000 iterations, the total # of times the "state" of weather is similar between W1 & W2
= 3,600 / 10,000
=36%

Therefore:
After 10,000 iterations, W1 and W2 will overlap approximately 3,600 times. In other words, W1 and W2 are expected to be in similar states about 36% of the time.

Another way of describing this is that, when given a common pair of risk factors (each with equal "weighting" of 0.20), they have a 36% chance of being in a similar "condition" or "state."