A Systematic Approach for Empirical Sensitivity Analysis on Monte Carlo Cost Models

Matt Pitlyk
SCEA/ISPA Joint Annual Conference
June 26 – 29, 2012

Sensitivity Analysis

- A key component of cost risk analysis
- Performed on the input variables for Monte Carlo cost models
- Goal is to determine those variables that cause the most variation in the final distribution
- Identify the best candidates for risk mitigation plans

Traditional Sensitivity Analysis

- Ranks model inputs using their correlation coefficient with the overall cost
- Appropriate for linear models, but it is not sufficient to accurately identify the largest uncertainty drivers – can be fooled
- In the case of the non-linear models it is neither appropriate nor accurate. This is because a correlation coefficient measures the strength of the linear relationship between variables and so is not appropriate for models where the relationships between distributions are non-linear
  - Common non-linear models are learning curves and other models based on regression

Alternative Sensitivity Analysis

- An alternate method is the following systematic approach
  - Reduce the variation on a single input variable, rerun the simulation, and measures the effect on the variation of the final distribution
  - The variation of the input variable is then restored and the process is repeated on the next input variable
  - Once this process is complete for all inputs, the results are compared to determine those inputs that cause the greatest change in variation

- This method is appropriate for both linear and non-linear models because it does not rely on the linear correlation

- A great advantage of this approach is that it puts sensitivity analysis in terms of dollars which makes it easier to communicate

Reducing Variation

This presentation explores three ways to reduce variation on model inputs in order to measure their contribution to the overall variation

- Replace the input variable with a constant
  - Removes all variation associated with that variable
  - Choose the mean/median for symmetric distributions
  - Choice of a constant is harder for non-symmetric
  - If there is no correlation between the input and any other variables, use a constant

- Reduce the standard deviation of the input distribution to close to zero
  - This will preserve the correlation between variables during simulation and any secondary effects such as correlation by association*

* By "correlation by association" it is meant that two variables are correlated because they are both correlated to a third variable, in this case the variable whose variation is being reduced

Reducing Variation (continued)

- Simulate the distribution being measured to preserve correlation. Then, include a constant, rather than the distribution, when calculating the total cost
  - This method ensures the secondary effectives of inducing correlation are maintain, thus not altering the original model in that way, while also removing all the variation contributed directly by the variable being measured
  - An equivalent method would be to replace the distribution with a Bernoulli with probability of occurrence one and cost impact equal to the replacement constant

The experiments performed for this paper used this third option

Preserving correlation

Is it necessary to preserve correlation? The answer depends on what is truly trying to be investigated

- If the goal is to evaluate the results as if that input variable has no distribution, then the answer is that correlation should not be preserved
  - Ex. Risk mitigation, reducing scope of project
- If the goal is to determine how much variation is added to the original model, then correlation needs to be preserved
  - This is the goal of traditional sensitivity analysis
  - Preserving correlation keeps any secondary effects of correlation in the model

Experiment Ground Rules

In addition to the method discussed on the previous slide, the following ground rules were followed when performing the experiments for this presentation

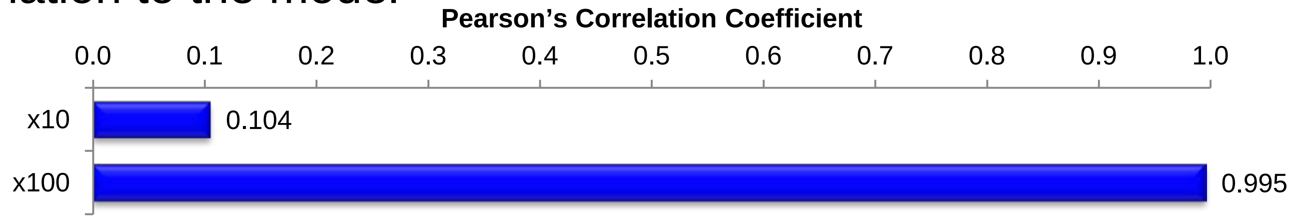
- All distributions used were symmetric in order to provide a clear choice for the replacement constant
- The 80^{th} percentile was used as the percentile of interest for the Alternative SA. As many government agencies are required to budget to a given percentile above the 50^{th}, it is useful to chose one to track throughout the presentation.
- Experiments were performed using Booz Allen Hamilton's Monte Carlo engine Eos. 15,000 iterations were run to minimize Monte Carlo sampling variation

Traditional SA vs Alternative SA – Linear: At least as good

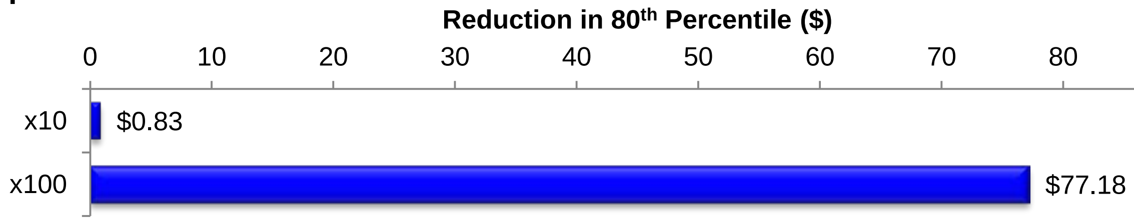
Consider the model:

Total Cost = N(10,1) * 10 + N(10,1) * 100

Traditional SA correctly identifies the second term as the one contributing the most variation to the model



Alternative SA also correctly identifies the second term as adding the most variation



Both correctly rank the inputs and indicate a large difference between the variation contributed by each input, but the alternative SA quantifies that difference in more tangible units: dollars

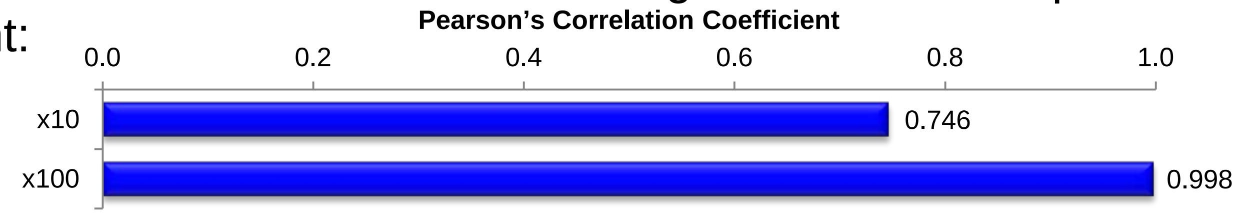
Booz | Allen | Hamilton

Traditional SA vs Alternative SA – Linear: Better

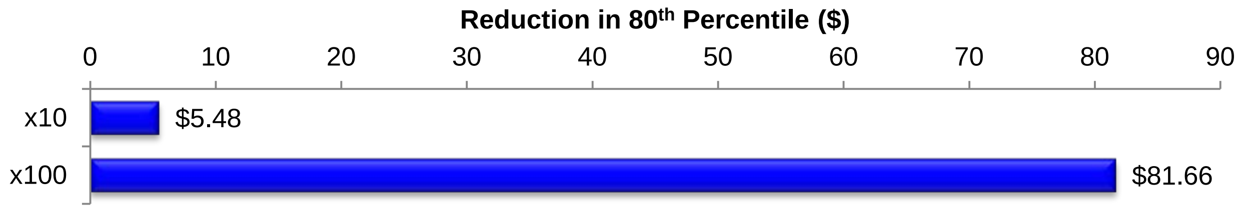
Now consider the same model but with a .7 correlation coefficient induced between the two distributions

Total Cost = N(10,1) * 10 + N(10,1) * 100, corr(X,Y) = .7

In this case Traditional SA ranks the distributions in the correct order but the correlation coefficients are much closer together than in the previous experiment:



Conversely, Alternative SA ranks them in the correct order but the *difference* between the measurements is insignificantly affected by the injected correlation (the difference is still about $75):



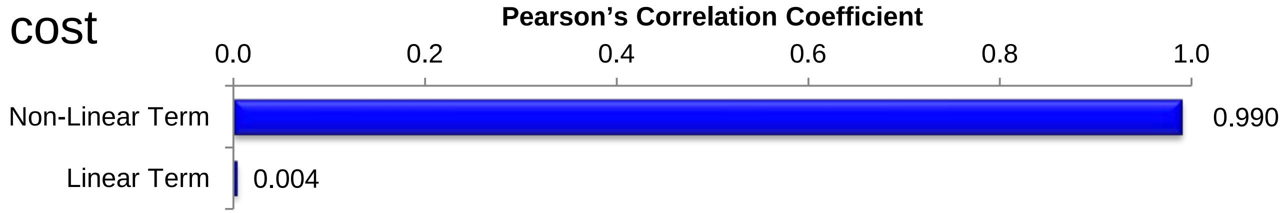
The x10 term is transitively correlated to the Total Cost via the x100 term. This leads to an inaccurate analysis in the traditional case.

Traditional SA vs Alternative SA – Non-Linear: At least as good

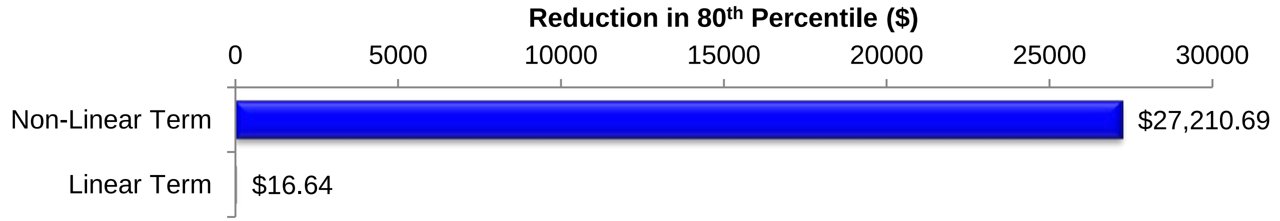
Consider the following non-linear model:

Total Cost = 100 * N(10,1)^3 + 100 * N(10,1)

Traditional SA identifies the second term as contributing almost all the variation to the total cost



Alternative SA also correctly identifies the second term as adding the most variation



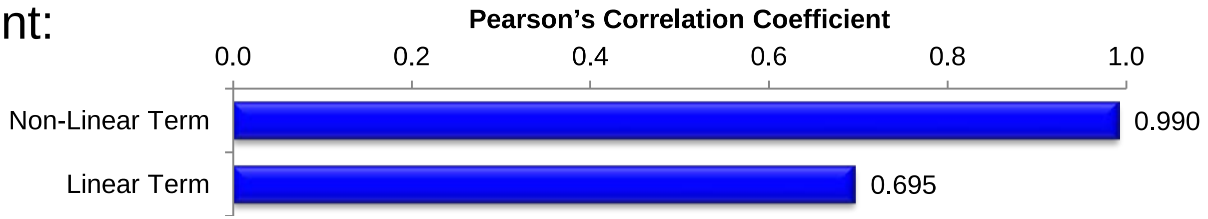
Both correctly rank the inputs and indicate a large difference between the variation contributed by each input, but the alternative SA quantifies that difference in more tangible units: dollars

Traditional SA vs Alternative SA – Non-Linear: Better

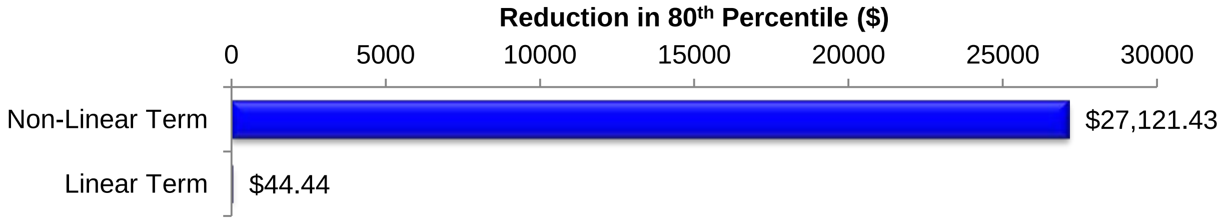
Now consider the same model but with a .7 correlation coefficient induced between the two distributions

Total Cost = 100 * N(10,1)^3 + 100 * N(10,1), corr(X,Y) = .7

In this case Traditional SA ranks the distributions in the correct order but the correlation coefficients are much closer together than in the previous experiment:



Conversely, Alternative SA ranks them in the correct order but the *difference* between the measurements is insignificantly affected by the injected correlation (the difference is still about $83):



Application to Discrete Risks

The final section of this presentation will describe an application of the Alternative SA method to discrete risks within a model in order to optimize a risk mitigation plan. There are two attributes of discrete risks that simplify the two issues raised in the first part of the presentation

- Discrete risks are rarely correlated
  - Because they are rarely correlated, there is no need to preserve correlation in the model and so replacing the distribution with a constant is valid
  - If correlation does happen to exist, it should not be preserved for risk mitigation investigations as discussed earlier

- Discrete risks are mitigated to zero impact
  - Because mitigated risks have their impacts reduced to zero, the choice of zero as the constant to replace the risk is clear

The Model

Regardless of how the main model (everything but discrete risks) is structured (linear, non-linear, sums, products, etc.), discrete risks are generally summed with the main model. For this reason the main model can be represented in a toy problem with single distribution and summed with the risks

- Most models with a sufficient number of summed distributions and non-perfect correlation will be approximately Normally distributed
  - This toy problem will use a Normal(10,1) to represent the main model

- The entire model will consist of the sum of the Normal(10,1) representing the main model and the five discrete risks listed on the next slide

The Model

Non-Risks	Distribution Type	Mean	SD
Main Model	Normal	10	1

Risks	Distribution Type	Prob of Occ	Min	Mode	Max
Risk 1	Triangular	10%	10	20	40
Risk 2	Triangular	30%	10	25	40
Risk 3	Triangular	40%	20	30	35
Risk 4	Triangular	70%	10	20	25
Risk 5	Triangular	30%	20	20	30

Total Model	
Total	Main Model + SUM(Risks 1-5)

Measuring Risk Impacts

The method for measuring the impact of discrete risks is the same as for uncertainty distributions

- Reduce variation on a risk by replacing it with a constant of zero
  - No need to preserve correlation as risks are rarely correlated
  - Zero is the correct constant since risks are mitigated to a zero impact
- Rerun the simulation
  - Depending on the program used, you can build your model to produce this result along side the total
- Measure the change in variation
  - Record the 80^{th} percentile (or percentile of interest)

Measuring Risk Impacts

This table shows the 80^{th} percentile that results from removing each risk by itself. The conditional formatting helps highlight the risks that reduce the 80^{th} percentile the most. The risks reducing the 80^{th} percentile the most are good candidates for a risk mitigation strategy.

Remove 1 Risk	
Risk 1	$ 66.60
Risk 2	$ 59.51
Risk 3	$ 55.85
Risk 4	$ 59.78
Risk 5	$ 60.77

This process can be extended to investigate the affect of removing multiple risks from the model in the search for a optimal mitigation strategy. Since risk impacts are added to the main model, the order of removing them is irrelevant.

Identifying Risks to Mitigate

These charts show the five ways to remove one risk (same as the previous slide but now colored in relation to all the combinations), the ten ways to remove two risks, and the ten ways to remove three risks, and the resulting 80th percentile after each combination is removed form the original model.

Original Model	
80th Percentile	$ 72.63

Remove 1 Risk	
Risk 1	$ 66.60
Risk 2	$ 59.51
Risk 3	$ 55.85
Risk 4	$ 59.78
Risk 5	$ 60.77

	Remove 2 Risks				
	Risk 1	Risk 2	Risk 3	Risk 4	Risk 5
Risk 1		$ 57.75	$ 53.85	$ 57.80	$ 59.00
Risk 2			$ 49.08	$ 41.89	$ 55.47
Risk 3				$ 38.85	$ 49.21
Risk 4					$ 43.16
Risk 5					

Greatest reduction in the 80th percentile

	Remove 3 Risks														
	Risk 1					Risk 2					Risk 3				
	Risk 1	Risk 2	Risk 3	Risk 4	Risk 5	Risk 1	Risk 2	Risk 3	Risk 4	Risk 5	Risk 1	Risk 2	Risk 3	Risk 4	Risk 5
Risk 1															
Risk 2			$ 45.11	$ 40.25	$ 53.95										
Risk 3				$ 36.81	$ 44.11				$ 32.89	$ 31.51					
Risk 4					$ 41.37					$ 39.27					$ 34.79
Risk 5															

Duplicate combinations are only shown once

Optimizing Risk Mitigation

This information (reduction in the 80th percentile) may not be enough to decide on an optimized risk mitigation strategy for a budget. It may be necessary to account for the cost to mitigate each risk. Therefore, the next step is to pair up each combination with the total cost to mitigate those risks. In most cases the total cost will be the sum of the costs to mitigate each risk in the combination.

Risk	Cost to Mitigate
Risk 1	$ 10.00
Risk 2	$ 12.00
Risk 3	$ 15.00
Risk 4	$ 7.00
Risk 5	$ 11.00

Risk Combinations

Risks Removed	Total Cost to Mitigate
1,2	$ 22.00
1,2,3	$ 37.00
1,2,5	$ 33.00
2,3,4	$ 34.00
3,4,5	$ 33.00

- Assigning costs to each risk produces a “total cost to mitigate” for each combination of risks removed from the model
- Conditional formatting can be use to “black out” risk mitigation combination that cost above a certain dollar amount, keeping the options within budget

Ranking Risk Mitigation Combinations

Once costs to mitigate have been calculated for each combination of risks, rankings can be created based on

- Lowest percentile, then lowest cost
- Lowest cost, then lowest percentile
- Lowest Cost Reduction/Cost to Mitigate Ratio*

Risks	80th Percentile	Cost to mitigate
2,3,5	$ 31.52	$ 38
2,3,4	$ 32.94	$ 34
3,4,5	$ 34.82	$ 33
3,4	$ 38.94	$ 22
2,4,5	$ 39.23	$ 30

Risks	80th Percentile	Cost to mitigate
4	$ 59.91	$ 7
1	$ 66.80	$ 10
5	$ 60.67	$ 11
2	$ 59.61	$ 12
3	$ 55.80	$ 15

Risks	Ratio
4	$ 1.80
4,5	$ 1.63
2,4	$ 1.61
3,4	$ 1.53
2,3,4	$ 1.16

* The Cost Reduction/Cost to Mitigate Ratio is the reduction in the 80^{th} percentile caused by the removal of the risks divided by the cost to remove those risks

Choosing the Optimal Risk Strategy

The rankings can then be used to choose the optimal risk strategy based on a budget limit or a requirement such as best value buying.

Risks	80th Percentile	Cost to mitigate
2,3,5	$ 31.52	$ 38
2,3,4	$ 32.94	$ 34
3,4,5	$ 34.82	$ 33
3,4	$ 38.94	$ 22
2,4,5	$ 39.23	$ 30

With a budget limit, find the first combination of risks in this ranking table that fits under the budget. Ex: If budget limit = 33.5, mitigate risks 3, 4, & 5.

If there is a best value requirement, find the first combination of risks on this table that is within your budget

Risks	Ratio	Cost to mitigate
4	$ 1.80	$ 7
4,5	$ 1.63	$ 18
2,4	$ 1.61	$ 19
3,4	$ 1.53	$ 22
2,3,4	$ 1.16	$ 34

Conclusion

The method examined in this presentation is an alternative to Traditional Sensitivity Analysis. Unlike the Traditional method which relies on the Pearson's (linear) correlation coefficient between an input distribution and total cost, the Alternative method removes the variation from the input and measure the effect on the percentile of interest in dollars. This method has two main advantages. The first is that it is appropriate and accurate for both linear and non-linear models. The second is that it quantifies the results in dollars which is extremely useful for the analyst.