



Interactive Probabilistic Risk Analysis for Construction Engineering and Management

Dr. Jing Du, University of Texas at San Antonio

Aug 2013 - Present Assistant Professor/Department of Construction Science/UTSA

Jun 2011 - Aug 2013 Sr. Cost Analyst/Zachry Holdings, Inc.

Aug 2008 - May 2012 PhD in Construction Management/Michigan State University

Dr. Yilmaz Hatipkarasulu, University of Texas at San Antonio

Dr. Rui Liu, The University of Texas at San Antonio

Interactive Probabilistic Risk Analysis for Construction Engineering and Management

Jing Du, Ph.D.¹, Yilmaz Hatipkarasulu, Ph.D.² and Rui Liu, Ph.D.³

¹ Department of Construction Science, The University of Texas at San Antonio, San Antonio, TX; PH (210) 458-3053; email: jing.du@utsa.edu

² Department of Construction Science, The University of Texas at San Antonio, San Antonio, TX; PH (210) 458-3099; email: yilmaz.karasulu@utsa.edu

³ Department of Construction Science, The University of Texas at San Antonio, San Antonio, TX; PH (210) 458-3054; email: rui.liu@utsa.edu

Abstract

Risk analysis is critical to the success of construction projects. Traditionally, probabilistic risk analysis is based on the Monte Carlo simulation. Because of the difficulty of teaching and implementing the Monte Carlo simulation, probabilistic risk analysis has not been widely used in the industry, although it is more desirable. This paper developed an interactive probabilistic risk analysis tool called RISK (Real-time Interactive Simulation Kit) that makes the Monte Carlo simulation of project risks unprecedentedly easy. RISK was tested in a capstone class at the University of Texas at San Antonio. Result showed that RISK is an effective tool in teaching risk analysis for construction engineering and management. It is also expected to improve the utilization of the probabilistic risk analysis in the industry by providing graduates with relevant skills.

Introduction

Risk analysis is an important concept in construction engineering and management (CEM), as for most construction projects many of the controlling parameters, processes and activities are often stochastic, uncertain and poorly understood (Jing Du & Bormann, 2012; J. Du & El-Gafy, 2011, 2012). The main objective of risk analysis is to identify and quantify the risks associated with a particular project, and to predict the calculated impacts on schedule and cost. The Construction Industry Institute Research Project 280 recommends a three-level risk management framework (Construction Industry Institute, 2012), which includes:

- Risk identification: to identify a list of risk items from a risk register;
- Deterministic risk analysis: to develop a single-point estimate of the risks' impacts and to calculate the expected value.
- Probabilistic risk analysis: to analyze risk through probabilistic distribution estimates of potential impacts.

Realizing the level-3 risk analysis requires Monte Carlo simulation that is often supported by third-party software such like @RISK or Oracle Crystal Ball. The set-up of these software packages is rather cumbersome, requiring expertise in statistics as well as simulation. As a result, although the level-3 risk analysis is considered more desirable, it is rarely being used in real-world projects (Construction Industry Institute, 2012).

This paper introduces a Cloud based on-line risk analysis tool called RISK (Real-time Interactive Simulation Kit) that can be used to teach construction students how to perform the level-3 risk analysis, without any third-party software. To use this tool, a user defines risk items (i.e., names, likelihoods, impacts as three-point estimates or triangular distributions) via a web based risk register. For each defined risk item a random number tuple will be generated immediately according to the triangular distribution to represent possible impacts of the risk item. If two risk items are correlated, methods for preserving the correlation will be used such as Cholsky decomposition. Then an additive operation will be performed on the Cloud based computing server between the random number tuples of two risk items. The probability density function (PDF) curve of the aggregated random number tuples will be displayed instantaneously to represent cumulative impacts of two risk items. This process will be repeated for all risk items and outcomes will be recalculated every time when any part of the risk register is updated. Once the user finishes the parameterization of the last risk item, the impact of all risk items will be displayed instantaneously.

This tool was tested in a Capstone class as a component of the risk management practice. Students were asked to develop an on-line risk register for their capstone project and to calculate the probability of meeting the budget and schedule. With only the basic statistical knowledge, students were able to perform the level-3 risk analysis and to interpret the results independently. It was found the proposed tool to be an effective educational approach in teaching risk analysis for construction engineering and management. It is also expected to improve the utilization of the level-3 risk analysis in the industry by providing graduates with relevant skills. The reminder of this paper introduces RISK and the implementation example.

RISK: Real-time Interactive Simulation Kit

The foundation of RISK is an interactive Monte Carlo simulation paradigm called Simulate As You Operate (SAYO).

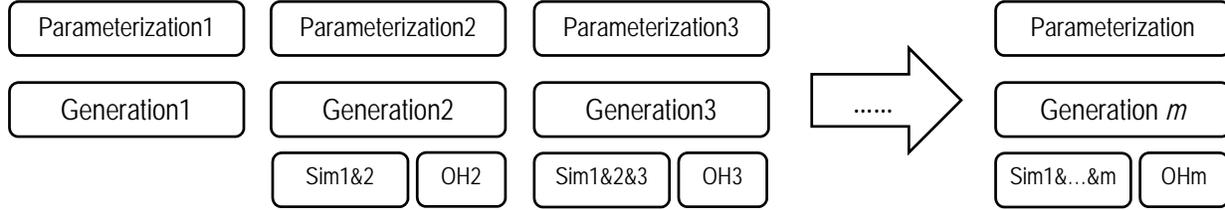
An Interactive Monte Carlo Simulation Paradigm - SAYO

The total probabilistic risk simulation time can be divided into four components:

- **Parameterization time (PT or pt_i):** The time spent by the user to parameterize the model. For example, the user defines the probability density functions (PDFs) of the model inputs;
- **Random number generation time (GT or gt_i):** The time spent by the system to generate or retrieve random numbers for the simulation according to the arbitrary distributions defined by the user;
- **Simulation time (ST or st_i):** The pure time spent by the system to perform the actual simulation tasks; and
- **Overhead (OH or ot_i):** Simulation involves lots of data fetching and processing operations and transferring. The data needs to be read and saved in the computer memory hierarchy frequently. Typically, a lot of time is required to transfer data between central processing unit (CPU) and the main memory, between CPU and secondary storage (hard disk), off-line storage and tertiary storage (e.g. tape drives), and among different hierarchical levels of the memory system. From the database operating standpoint, time is also required to perform database operations such as database initialization, read/write, insert, update,

delete, merge and indexing etc. The time consumed in such operations does not directly contribute to the probabilistic risk simulation, and thus can be called overhead.

The four components of a probabilistic risk simulation can be executed in parallel to improve the efficiency. The strategy of parallel execution is called “*Simulate As You Operate*” or SAYO. The extent of SAYO, i.e., the level to which the four components can be concurrently executed varies for different types of problem, as of Fig 1 (A).



(A) Proposed interactive probabilistic risk simulation process



(B) Traditional Monte Carlo simulation process

Fig. 1 Interactive Monte Carlo simulation versus traditional Monte Carlo simulation

For each parameterized variable, n random numbers (n =number of simulation trials) are generated together at one time, which becomes a random number tuple (RNT) for sub-simulations (while traditional method generates only one random number at one time). The time needed for each sub-simulation depends on the maximum of the above three. Assuming the parameterization takes the longest time then total time required for the interactive probabilistic risk simulation is:

$$TT_{cd} = \sum_{i=1}^m \max\{pt_i, gt_i, (st_{i+1} + ot_{i+1})\} = \sum_{i=1}^m pt_i = PT \quad (1)$$

where, m equals to the number of model variables and PT denotes the total time for parameterization. Using the traditional probabilistic risk simulation method, such as Monte Carlo simulation, the random number generation process and simulation won't be started until all the model variables have been parameterized (Fig 1 B). Simulation is then divided into n trials, wherein one random number is generated for each model variable in each simulation trial. The generated random numbers of m variables will then be used to perform one simulation and yield one result of the model. This process will be repeated for n times and statistical inferences may be made upon n simulation results. In this process, each trial requires one instance of overhead time and thus the total overhead time needed for traditional method is $(n \times OH)$, where OH is the overhead time for one simulation trial. Therefore, the total time required for traditional method can be defined as:

$$TT_{tra} = PT + GT + ST + n \times OH \quad (2)$$

As shown in Equation 2, SAYO approach saves $GT + ST + n \times OH$ in simulation time over traditional method.

Based on SAYO, RISK was developed. RISK includes a *DigitSource* which contains true random numbers generated by physical processes such as Quantum process; a *DigitBank* that stores user's history of parameterized models, model inputs and model simulation outputs; a *Model Evaluation* module that assigns the modeling, parameterizing and updating tasks to the other modules, and divides an entire problem to a set of sub-problems for enabling parallel and instant computation using grid computing; a distribution filter that can converted uniformly distributed random numbers to random numbers that follow arbitrary distributions; a *Temporary Storage* or cache that stores the sub-models and corresponding variables; a *Cloud based Grid Computing* facility that finish the computing tasks assigned; a *Synthesizing Module* that synthesizes the simulation results of sub-problems; and finally a web based user interface which is either in tabular or click-and-point format.

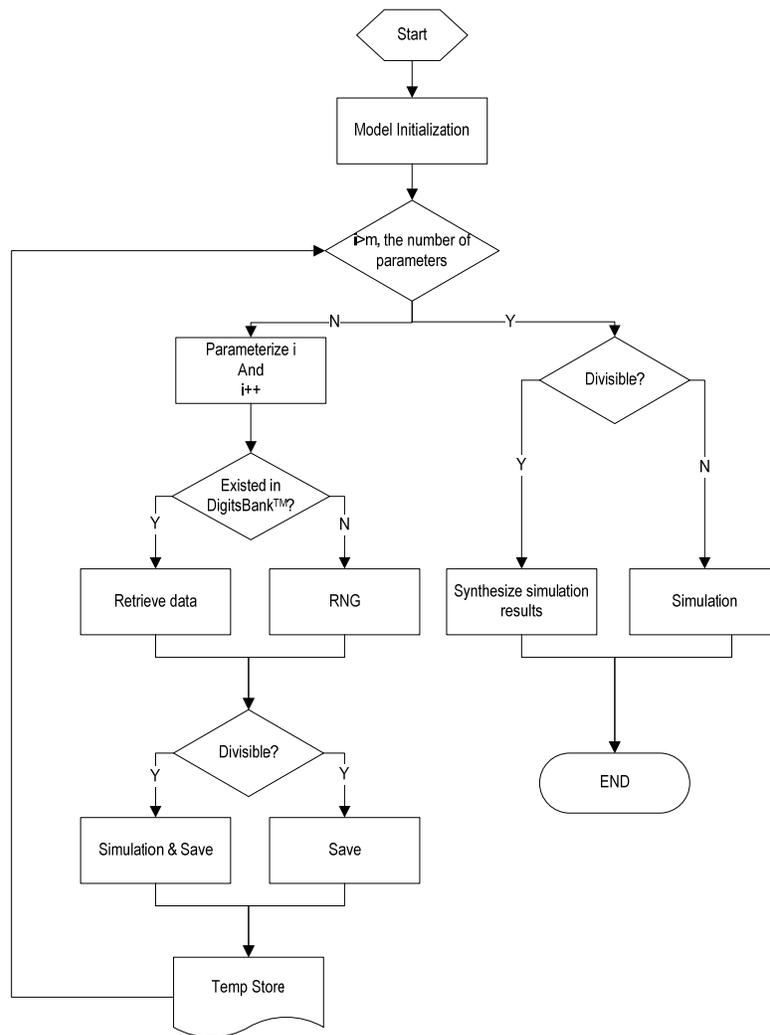


Fig 2. RISK simulation process

Fig 2 illustrates the operational process of RISK. The user first defines the risk model for a given problem through the web based UI. Then the user parameterizes at least one model variable (a random variable), which will be sent to the Model Evaluation module instantaneously. After receiving the distribution parameters (parameterization), the Model Evaluation module first

checks if there are any random number tuples (RNTs) in the DigitBank from past modeling and simulation, and follows the defined distribution. If any tuples are found, Model Evaluation module moves the RNTs to the temporary storage for future computation. If there are no existing RNTs that follow the defined distribution, Model Evaluation module utilizes the source random numbers, DigitSource, to generate random numbers following the defined distribution and saves them as RNTs in the temporary storage for future computation. The uniformly distributed true random numbers will be converted into random numbers that follow arbitrary distributions through a Random Number Filter, based on existing random number generation methods such as inverse F method, Acceptance-Rejection method, or Markov Chain Monte Carlo method. After the user defined and parameterized the first two variables, a simulation starts immediately in the cloud based grid computing facility while the user is parameterizing the third variable, and result will be saved as a random number tuple for variable 1 and 2, and RNT1&2. Then the random tuple of the third parameterized variable, or RNT3, will be aggregated with RNT1&2, which gives RNT1&2&3, while the user is still parameterizing the fourth variable. This process will be repeated until RNT1&2&3&...&M is obtained, when the user is very much likely just done with the parameterization of the last variable m . It is important to note that all the interim simulation and updating processes are executed concurrently in parallel with the user parameterization process. If all the model variables have been simulated, the temporary simulation outputs will be synthesized to the final simulation result and shown in the web based UI. The users may select to modify only part of the risk model, such as the cases in a what-if scenario analysis. In this case, RISK will hold the model information such as random number tuples used in the simulation and the simulation results of unaffected part of the model fixed, and only repeat the process described above on the affected part of the model. The final results will be synthesized to reflect the change in state of the model.

Fig 3 demonstrates the user interface of RISK. The major components include the user input area (Area A) and the result demonstration area (Area B). In addition, the heat map of risk analysis can be shown in Area C, which locates risk items in different sections according to the likelihood and impacts. Risk items can also be ranked in RISK based on the overall impacts on the project (Area D).

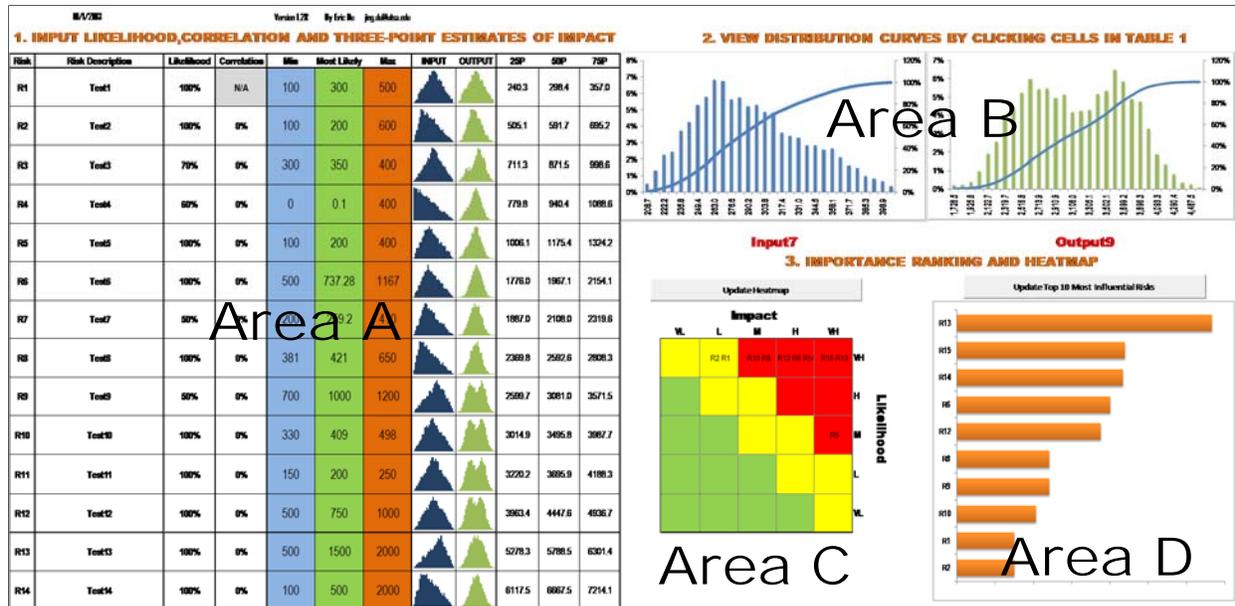


Fig. 3 User interface of RISK

Case Study

RISK was tested in an undergraduate capstone class at the University of Texas at San Antonio. The project was a parking facility with estimated budget of \$4.5 million dollars. The students were asked to develop a risk register and perform probabilistic risk analysis using RISK on this project. The following steps were followed by the students:

- (1) Identify the list of risk items
- (2) Determine the likelihood of risk items
- (3) Estimate the impacts of risk items
- (4) Performance interactive Monte Carlo simulation using RISK
- (5) Interpret the simulation results

The remainder of this paper summarizes the probabilistic risk analysis process enabled by RISK.

Identify the list of risk items

Risk analysis begins with identifying a list of relevant risk items. The Construction Industry Institute has identified 107 potential risks throughout the entire project lifecycle (Construction Industry Institute, 2007). Fig 4 illustrates the categorization of these risk items. This list was shown to the class, followed by a classroom discussion was initiated to walk through significant risk items. For example, one common risk of local projects is different site conditions. If site conditions are materially different from those indicated in the contract or those can be reasonably expected, dramatic cost increase may happen, caused by severe delays and mitigation costs. Another potential risk is the omissions and errors in the project documents. They will lead to increased RFI's, change orders, conflicts, disputes and ultimate affect project duration and cost. In the similar manner, causes and consequences of main risk items have been discussed in class. Due to the scope

of the capstone project, the discussion was focused on the risks relevant to project execution (Fig 4). The class was finally required to develop up to 50 risk items.

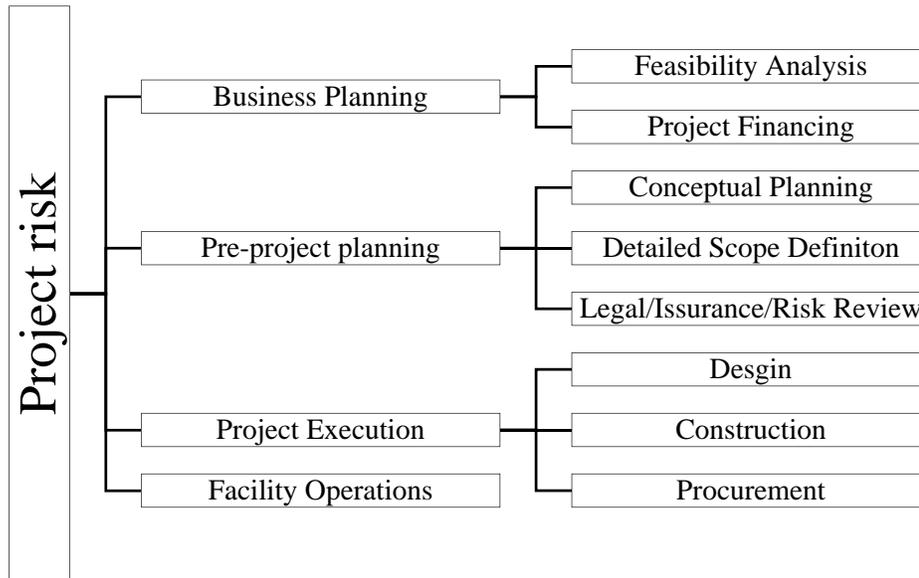


Fig. 4 Project risk categorization; adapted from (CII 2007)

Determine the likelihood of risk items

Then the class was asked to determine the likelihood of each of the identified risk items. Likelihood is shown as a percentage from 0% to 100%, with 100% to represent definitive events and 0% to represent impossible events. Reaching a consensus is critical for the risk analysis process, but the authors found that students had very different perceptions on the likelihood of the same risk items. Therefore, the Delphi method was introduced to the class to help with the class discussions (Linstone & Turoff, 1975). Nonetheless, the students were still given the right to choose their own ways to determine the likelihood. Some students actually proposed creative approaches to decide the likelihood. For example, in order to quantify the likelihood of “severe weather conditions”, some students proposed the use of historical weather data, such like the one shown in Fig 5. Severe weather conditions include high temperature, raining days, storms and other unexpected conditions. The dates with defined severe weather conditions were marked based on public accessible weather records. Then the frequency of severe weather days in given months was converted into the likelihood of “severe weather conditions”. For example, if 5 days in August are highlighted to have severe weather conditions based on history, then the likelihood for severe weather conditions in August will be $5/31=16\%$.

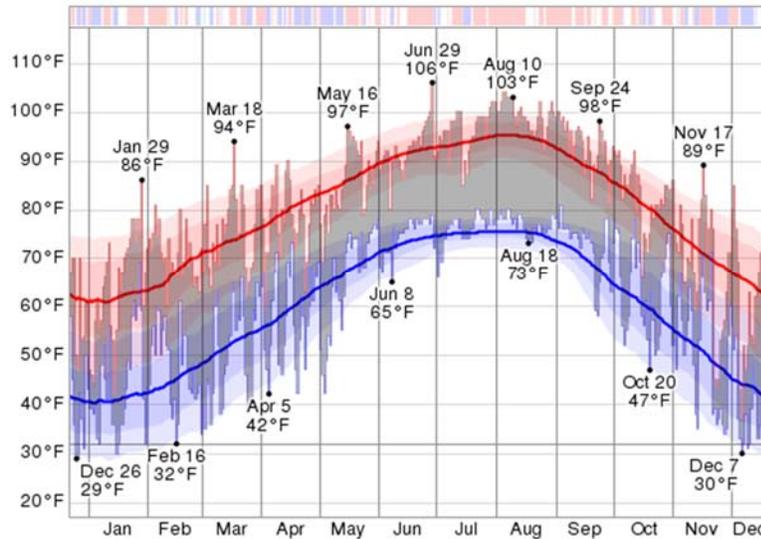


Fig. 5 Historical temperature record; (<http://weatherspark.com/history/31554/2013/San-Antonio-Texas-United-States>)

Estimate the impacts of risk items

Upon the occurrence of a risk item, it has negative impacts on cost or schedule (Construction Industry Institute, 2012). The class was asked to use the three-point estimate technique to quantify the impacts of identified risks on both schedule and cost, including the minimum, the most likely and the maximum impacts. Three-point estimate is used to rebuild triangular distribution (Creese, 2013), and is popular in CEM areas. For example, if severe weather happens, at minimum the project will be delayed by 1 day, at maximum 5 days, and most likely, 2 days. Then the three-point estimate of severe weather is (1, 2, 5) for the impact on project schedule. Once again, we found that students had very divergent perceptions. As a result, they were encouraged to use Delphi method or simply, voting. Other advanced statistical method to combine opinions have also been introduced in class, such as the Axiomatic Approach proposed by Clemen and Winkler (1999).

Perform interactive Monte Carlo simulation

After determining the list of risk items, and corresponding likelihoods and three-point estimates of impacts, a risk register is developed. A risk register is a document detailing all identified risks, including description, cause, likelihood of occurrence, impacts on objectives, owners and current status (Construction Industry Institute, 2012). The students were instructed to use RISK to performance interactive Monte Carlo simulation on the developed risk register. Because of the SAYO approach, the Monte Carlo simulation was performed in the “What You See Is What You Get” (WYSIWYG) manner. Fig 6 shows an example of applying RISK. Starting with an empty RISK table (step 0), the students can input risk description, likelihood, correlation (between two risk items), and the three-point estimate of risk impacts (Min, Most Likely and Max) sequentially. Once the necessary information has been typed in for the first risk item, RISK will automatically compute the distribution function and show the results simultaneously as distribution curves and critical percentile values (step 1). Then the student may repeat input for the second risk item. Upon finish, RISK will calculate the aggregated result simultaneously, and show the results in real time

(step 2). Similarly, right after the student typed in the information of the third risk item, aggregated result of three risk items is shown simultaneously (step3). As shown in the “output” column of step 3, the overall impact of three example risk items on the project schedule is a dual-peak distribution. We found that such an interactive simulation process, compared to the traditional Monte Carlo simulation, has significantly encouraged communications between students, and active in class discussions. Students can in fact “try” different scenarios and visualize the overall impacts on the project in real time. Without any significant difficulty, students were able to handle this tool.

Step 0:

Risk	Risk Description	Likelihood	Correlation	Min	Most Likely	Max	INPUT	OUTPUT	25P	50P	75P
R1	Example 1		N/A						#N/A	#N/A	#N/A
R2	Example 2								0.0	0.0	0.0
R3	Example 3								0.0	0.0	0.0

Step 1:

Risk	Risk Description	Likelihood	Correlation	Min	Most Likely	Max	INPUT	OUTPUT	25P	50P	75P
R1	Example 1	80%	N/A	3	5	10			4.0	5.3	6.7
R2	Example 2								4.0	5.3	6.7
R3	Example 3								4.0	5.3	6.7

Step 2:

Risk	Risk Description	Likelihood	Correlation	Min	Most Likely	Max	INPUT	OUTPUT	25P	50P	75P
R1	Example 1	80%	N/A	3	5	10			4.0	5.3	6.7
R2	Example 2	60%	0%	4	5	8			5.3	8.5	11.5
R3	Example 3								5.3	8.5	11.5

Step 3:

Risk	Risk Description	Likelihood	Correlation	Min	Most Likely	Max	INPUT	OUTPUT	25P	50P	75P
R1	Example 1	80%	N/A	3	5	10			4.0	5.3	6.7
R2	Example 2	60%	0%	4	5	8			5.3	8.5	11.5
R3	Example 3	30%	0%	10	15	30			6.2	11.1	20.5

Fig. 6 Interactive Monte Carlo simulation using RISK

Interpret simulation results

Finally, students performed RISK based Monte Carlo simulation on their developed risk registers on the team basis. And the results were shared in class to initiate discussions. Before discussions, important statistical concepts have been introduced to the students including PDFs (Probability Density Functions) and CDFs (Cumulative Density Functions). More importantly, the statistical meanings of these concepts have been linked directly to the problem of risk analysis. For example, students were told that CDFs can be used to quantify the probability of success.

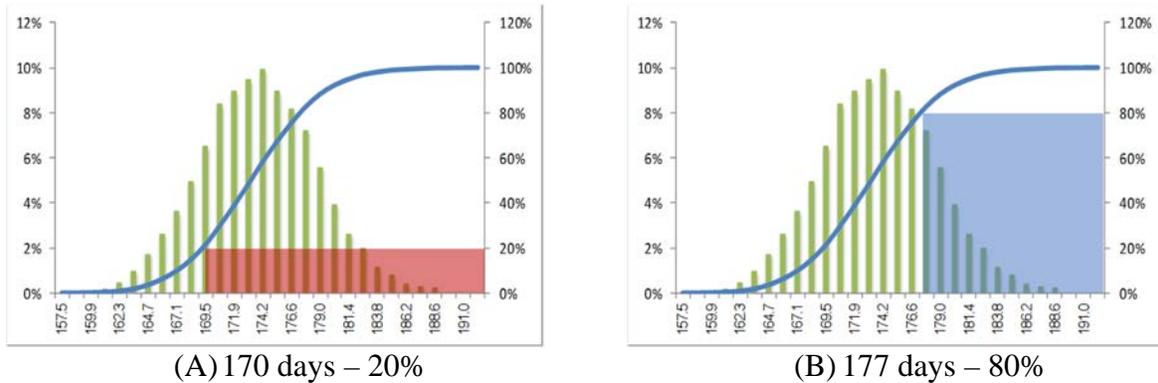


Fig.7 Discussion of the risk analysis result

Fig 7 shows a real case in the discussion. A team performed a RISK based simulation on their risk registrar to determine the probability of success of their estimated project duration. Simulation showed that there was only a 20% chance to meet their estimated duration (170 days). Instead of presenting the duration estimation directly to the owner, they were asked to extend their duration by 7 days (177 days), which increases the probability of success to 80%. Although this resulted in a longer project duration, bigger success chance can give more confidence to the owner. A discussion was followed about the psychology in bidding.

Conclusion

Risk analysis is critical to the CEM problems. Traditionally, Monte Carlo simulation is used to perform probabilistic risk simulation. It is often supported by third-party software such like @RISK or Oracle Crystal Ball. The set-up of these software packages is cumbersome, requiring expertise in statistics and simulation. We developed an interactive risk simulation tool called RISK to simplify probabilistic risk simulation. It was tested in a capstone class at the University of Texas at San Antonio. Result showed that RISK is an effective educational tool in teaching risk analysis for construction engineering and management. It is also expected to improve the utilization of the level-3 risk analysis in the industry by providing graduates with relevant skills.

References

1. Clemen, R., & Winkler, R. (1999). Combining Probability Distributions From Experts in Risk Analysis. *Risk Analysis*, 19(2), 187-203.
2. Construction Industry Institute, C. (2007). Contracting to Appropriately Allocate Risk: Construction Industry Institute.
3. Construction Industry Institute, C. (2012). Applying Probabilistic Risk Management in Design and Construction Projects: Construction Industry Institute.
4. Creese, R. C. (2013). *Converting Point Estimates for Cost-Risk Analysis*. Paper presented at the 120th ASEE Annual Conference and Exposition, Atlanta, GA.
5. Du, J., & Bormann, J. (2012). Improved Similarity Measure in Case Based Reasoning with Global Sensitivity Analysis: An Example of Construction Quantity Estimating. *Journal of Computing in Civil Engineering*.
6. Du, J., & El-Gafy, M. (2011). Feasibility Analytical Mapping (FAM) for the Bidding Decision: A Graphic Bidding Decision Making Model Based on Multidimensional Scaling and Discriminant Analysis. *International Journal of Construction Education and Research*, 7(3), 198-209.
7. Du, J., & El-Gafy, M. (2012). Virtual Organizational Imitation for Construction Enterprises: Agent-Based Simulation Framework for Exploring Human and Organizational Implications in Construction Management. *Journal of Computing in Civil Engineering*, 26(3), 282-297.
8. Linstone, H. A., & Turoff, M. (1975). *The delphi method*: Addison-Wesley Reading, MA.