Using Bayesian Belief Networks with Monte Carlo Simulation Modeling

Dr. Marina Dombrovskaya

ICEAA Conference
Denver, CO
June, 2014

ICEAA

International Cost Estimating
and Analysis Association

Booz | Allen | Hamilton

Table Of Contents

Booz | Allen | Hamilton

International Cost Estimating
and Analysis Association

Introduction: Monte Carlo Simulation Modeling

▸ **Monte Carlo simulation** is a probabilistic method of modeling complex systems with many interrelated uncertain variables.

▸ It is a widely accepted technique in Cost Estimating for modeling cost uncertainty and performing risk analysis.

▸ MC simulation is based on repeated random sampling of probability distributions assigned to uncertain variables. After random sampling is performed, numeric results are combined according to assigned relationships, such as CERs.

▸ Modern Monte Carlo simulation tools have become very powerful and fast: used Booz Allen's *Argo* tool for Excel for Monte Carlo simulation model in this presentation
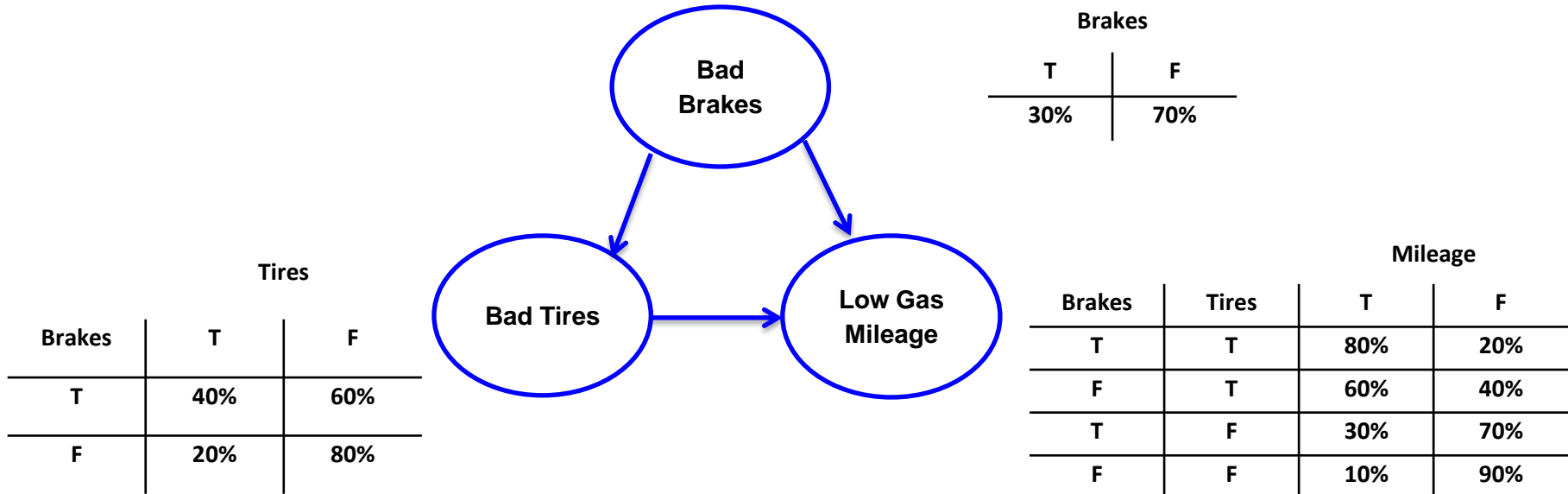
Introduction: Bayesian Belief Networks

▸ A **Bayesian Belief Network** (BBN) is a probabilistic model that represents random variables and dependencies among them with assigned Bayesian probabilities in a form of a directed acyclic graph.

▸ BBNs provide a visual representation of inter-dependencies among random variables and estimate probabilities of events that lack direct data.

▸ *Nodes* of the graph are random variables. *Directed edges* represent conditional dependencies between random variables with causal relationship in the direction of the edge.

▸ Each node has a *probability function* associated with it that takes in the values of the node's parent nodes and outputs conditional probability of the variable represented by the node.

Introduction: Bayesian Belief Networks

▸ A node that has no parents is an *independent* event.

▸ If two nodes are not connected by an edge, then the these nodes are *conditionally independent* of each other.

▸ If a node has **N** binary parents, then the probability function associated with this node can be representated by a table (matrix) with 2^N entries.

Booz | Allen | Hamilton

Bayesian Belief Network: Example

▸ Model of a relationship between risk of abnormal wear and tear on brakes and tires and low gas mileage in a car

Bad Brakes

Brakes	
T	F
30%	70%

Bad Tires

Low Gas Mileage

Tires

Brakes	T	F
T	40%	60%
F	20%	80%

Mileage

Brakes	Tires	T	F
T	T	80%	20%
F	T	60%	40%
T	F	30%	70%
F	F	10%	90%

▸ What is the probability of low gas mileage given that a car's brakes and tires are bad?

▸ What is the likelihood of having bad breaks if a car has low gas mileage?

Booz | Allen | Hamilton

Table Of Contents

International Cost Estimating
and Analysis Association

Booz | Allen | Hamilton

Monte Carlo Simulations and Cost Estimating Modeling

▸ There is inherent uncertainty in cost estimating models: uncertainty about point estimate cost and schedule estimates, probability of risk occurrence, uncertainty about risk impact.

▸ Monte Carlo simulation modeling is a highly effective method for modeling uncertainty and performing risk analysis within a cost estimating model.

▸ One of the main aspects of creating a rigorous Monte Carlo simulation cost estimate is the accuracy in defining uncertainty and risk parameters associated with the cost components of the model.

▸ It is equally important to assess and accurately represent inter-dependencies between uncertain variables and risks, which are measured via correlation.

Bayesian Belief Networks and Cost Estimating Modeling

‣ Since oftentimes historical data is insufficient for a rigorous statistical analysis, both probability distribution and correlation are commonly estimated via a subject matter opinion.

‣ However, inherent complexity of variable inter-dependencies is often overlooked during such estimates which could significantly affect results of Monte Carlo simulation model.

‣ Bayesian Belief Networks naturally model complex relationships among cost components and risks.

‣ For cost estimating models nodes of a BBN could be cost components or risks associated with cost components. Edges show causal relationships among risks and/or cost components.

# Combining BBNs with Monte Carlo Simulation Cost Estimating

▶ Since BBNs contain conditional probability information, it is natural to model posterior probabilities of random variables with a Monte Carlo simulation.

▶ In a Monte Carlo simulation we randomly sample independent random variable in a BBN, then follow the network direction to simulate conditional probabilities and impacts of dependent random variables.

▶ Easy to conduct what-if risk analysis: can compute conditional probabilities assuming certain risks are turned on or off.

Table Of Contents

Conditional Probability

▸ ***Conditional probability*** is the probability of an event given that another event already occurred.
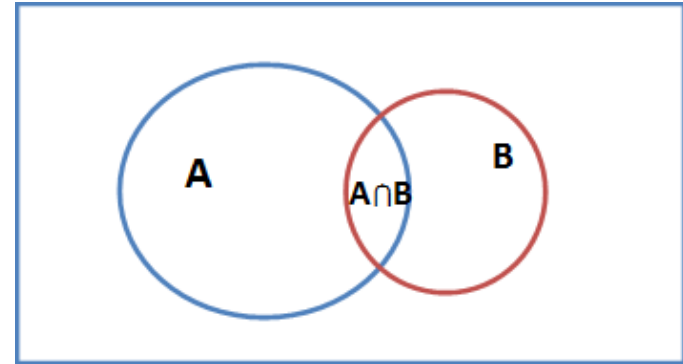
▸ $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

▸ $P(A|B) + P(\bar{A}|B) = 1$

▸ In general, $P(A|B) \neq P(B|A)$

▸ **Chain Rule**: $P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|P_2, P_1) \cdots P(E_n|E_1, \cdots E_{n-1})$

▸ Example: Toss a fair coin twice. What is the probability of getting two heads, given that the first toss results in heads?

$$P(A = \{HH\} | B = \{HH, HT\}) = \frac{1/4}{2/4} = \frac{1}{2}$$

Booz | Allen | Hamilton

Bayes' Theorem

▸ Given events **A** and **B** such that **P(B) ≠ 0**

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

▸ **P(A)** is the probability of event **A**, the **prior**

▸ **P(B)** is the probability of event **B**, the **evidence**

▸ **P(A|B)** is the conditional probability of event **A** given event **B**, the **posterior**

▸ **P(B|A)** is the conditional probability of event **B** given event **A**, the **likelihood**

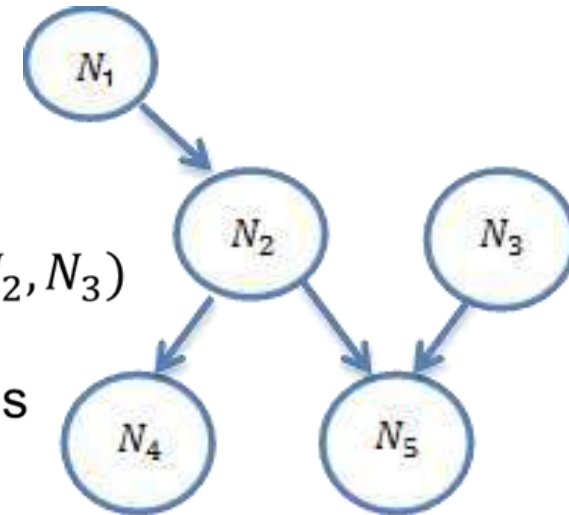$$posterior = \frac{prior \cdot likelihood}{evidence}$$

Bayesian Belief Network and Conditional Probabilities

▸ Given a Bayesian Belief Network, probability of network is given by the following equation

▸ $P(N_1, N_2, N_3, N_4) = \prod_{i=1}^{4} P(N_i | Parents(N_i))$

▸ $P(N_1, N_2, N_3, N_4) = P(N_1)P(N_2|N_1)P(N_3)P(N_4|N_2)P(N_5|N_2, N_3)$

▸ Since Bayes' Theorem is in some sense revirsible, BBNs provide a way to reverse the logic without changing the network.
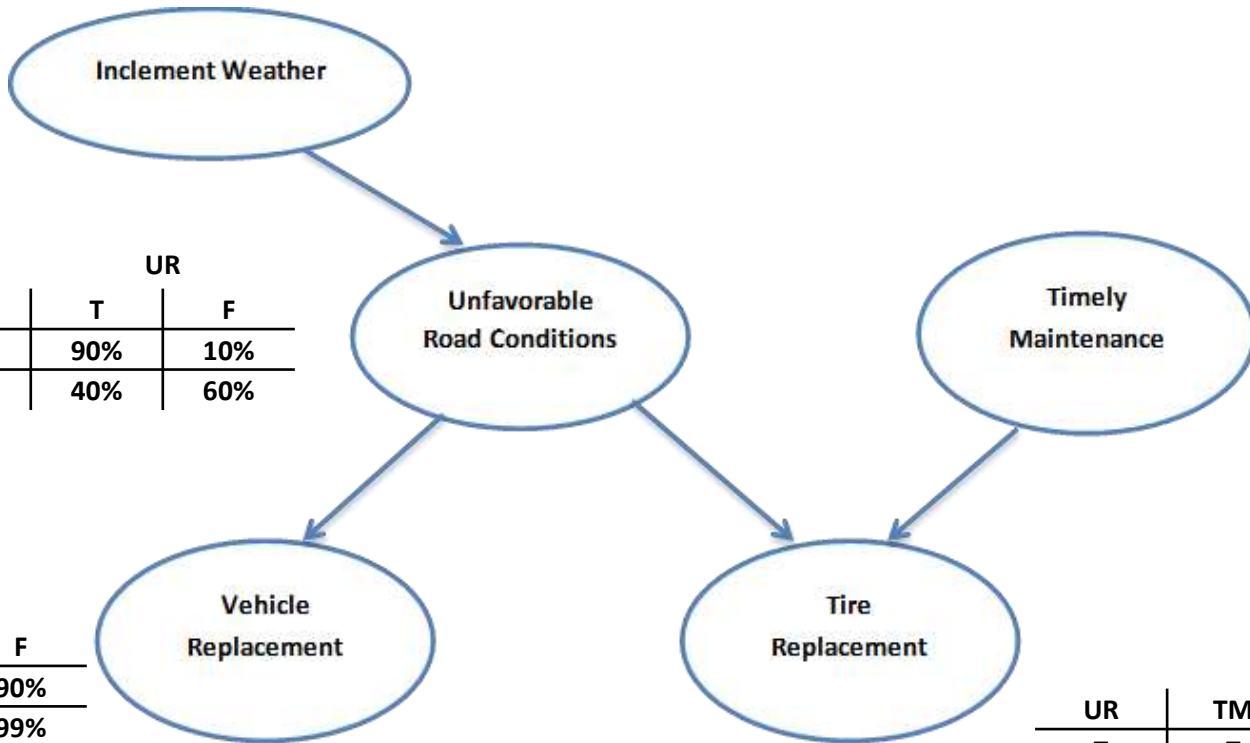
Booz | Allen | Hamilton

Bayesian Belief Network: Example

IW

T	F
30%	70%

UR

IW	T	F
T	90%	10%
F	40%	60%

TM

T	F
80%	20%

VR

UR	T	F
T	10%	90%
F	1%	99%

TR

UR	TM	T	F
T	T	60%	40%
F	T	5%	95%
T	F	90%	10%
F	F	20%	80%

Booz | Allen | Hamilton

Bayesian Belief Network: Example

▸ Calculate probability of Tire Replacement (TR) given Inclement Weather conditions (IW) and absence of Timely Maintenance (TM).
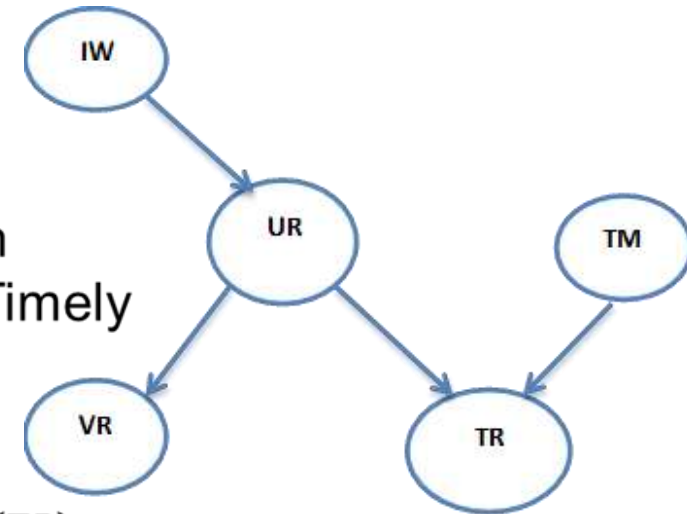
▸ $P(TR|IW,\overline{TM}) = \dfrac{P(IW,\overline{TM}|TR)P(TR)}{P(IW,\overline{TM})} = \dfrac{P(IW|TR)P(\overline{TM}|TR)P(TR)}{P(IW)P(\overline{TM})}$

▸ $P(IW|TR) = \dfrac{P(TR|IW)P(IW)}{P(TR)}$

▸ $P(\overline{TM}|TR) = \dfrac{P(TR|\overline{TM})P(\overline{TM})}{P(TR)} = \dfrac{P(TR|\overline{TM},UR)P(UR)+P(TR|\overline{TM},\overline{UR})P(\overline{UR})}{P(TR)}P(\overline{TM})$

Bayesian Belief Network: Example

▸ $P(TR|IW) = P(TR|IW, UR)P(UR|IW) + P(TR|IW, \overline{UR})P(\overline{UR}|IW)$

▸ $P(TR) = \sum P(TR|UR, TM)P(UR)P(TM)$

▸ $P(UR) = P(UR|IW)P(IW) + P(UR|\overline{IW})P(\overline{IW})$

▸ Now we can plug in values from probability tables

▸ $P(TR|IW, \overline{TM}) = \dfrac{P(IW|TR)P(\overline{TM}|TR)P(TR)}{P(IW)P(\overline{TM})} = \dfrac{36.5\% * 20.8\% * 56.1\%}{30\% * 20\%}$
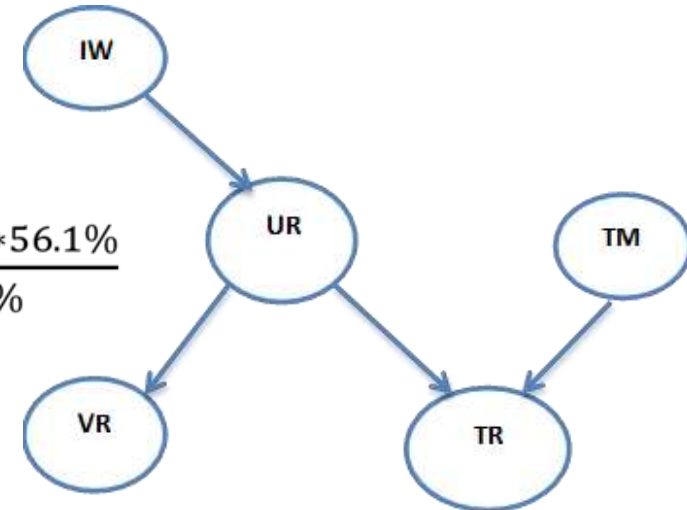
▸ $(TR|IW, \overline{TM}) = 70.9\%$

Table Of Contents

Booz | Allen | Hamilton

Incorporating BBNs into Monte Carlo Simulation Cost Estimate

▸ **Toy Cost Estimate Problem**: Estimate cost of yearly maintenance of a military vehicle given maintenance cost components, such as technical maintenance cost, personnel cost and storage cost, and risk factors, such as tire replacement and vehicle replacement.

▸ Create Monte Carlo simulation model in MS Excel using *Argo* - Monte Carlo simulation Excel tool.

▸ First, model risk factors of tire and vehicle replacement independently. For each risk factor probability of occurrence is modeled via a Bernoulli distribution and cost impact is modeled via a Triangular distribution.

▸ Second, model risk factors via BBN that we presented in a previous example. The only risk factors with impact were tire and vehicle replacement which were modeled via the same Triangular distributions as in independent case.

Monte Carlo simulation Cost Estimate for Military Vehicle Maintenance with Risk Analysis in Excel using Argo

Costs	Nme	Description	Column1	Distribution Type	Param 1	Param 2	Param 3	Cost Impact
1	Vehicle Maintenance	Total Cost per Year		Rollup				$358.92
1.1	Oil	Oil Price		Normal	3.8	0.5		$5.14
1.2	Changes	Oil Changes per year		Triangular	0	9	15	$7.02
1.3	Fill_ups	Number of fill ups per year		Triangular	20	35	52	$28.40
1.4	Fuel	Cost of fuel per fill up		Normal	10	2		$8.49
1.5	Brakes	Brake maintenance		Triangular	11	15	25	$16.34
1.6	Tire	Tire maintenance		Triangular	11	20	25	$15.16
1.7	Engine	Engine maintenance		Triangular	0	100	120	$50.17
2	Personnel	Total cost of maintenance personnel		Rollup				$1,381.02
2.1	Salary	Salary per FTE		Normal	100	12		$72.64
2.2	FTEs	Number of FTEs		Triangular	15	20	24	$19.01
3	Storage	Storage cost		Normal	300	45		$264.23

Risks	Name	Description	Cost Risk Distribution Parameters					
			Prob of Occur	Distribution Type	Param 1	Param 2	Param 3	
1	TR	Tire Replacement	0.2	Triangular	100	500	700	670.6640014
2	VR	Vehicle Replacement	0.01	Triangular	400	1000	1500	0

Total estimated cost of vehicle maintanance	$ 2,674.84

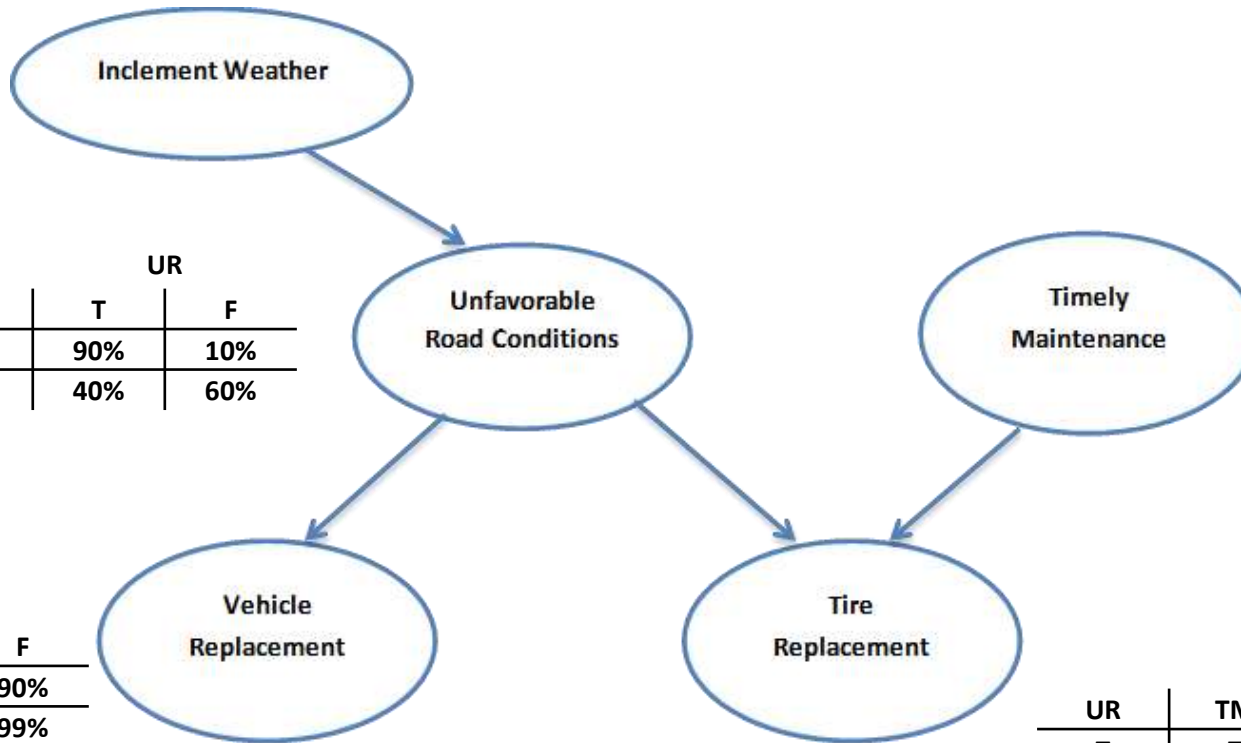# Bayesian Belief Network for Risk Factors of Military Vehicle Maintenance Model

IW

T	F
30%	70%

UR

IW	T	F
T	90%	10%
F	40%	60%

TM

T	F
80%	20%

VR

UR	T	F
T	10%	90%
F	1%	99%

TR

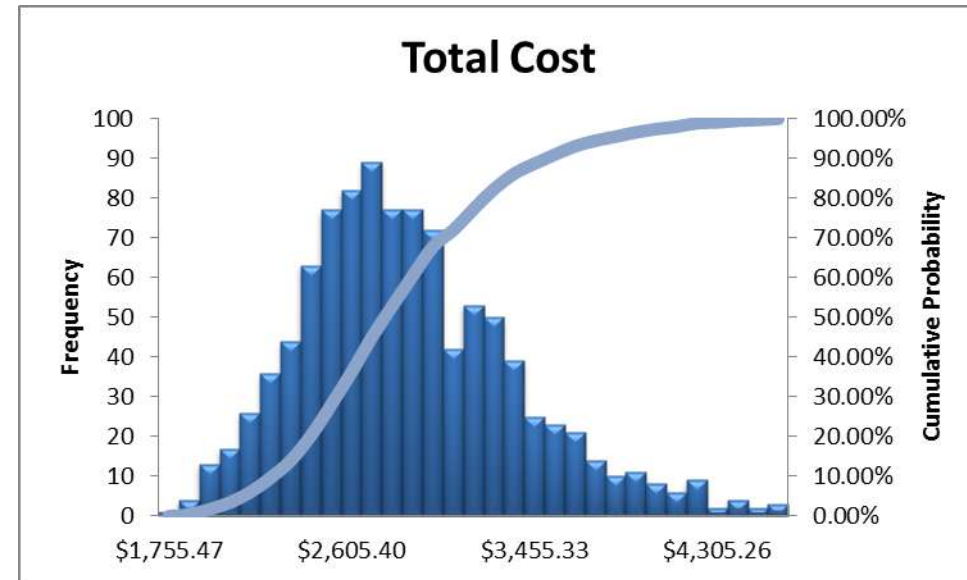| UR	TM	T	F
T	T	60%	40%
F	T	5%	95%
T	F	90%	10%
F	F	20%	80%

Results of Argo Simulation: Independent Risk Factors vs BBN

Monte Carlo Simulation with risk factors modeled as independent events

Monte Carlo Simulation with BBN modeling risk factors

Results of Argo Simulation: Statistics for Total Cost

Monte Carlo Simulation with risk factors modeled as independent events

Total Cost

Statistics	Values
Mean	$ 2,567.55
Median	$ 2,536.74
Variance	$ 142,931.81
Standard Deviation	$ 378.06
Coefficient of Variation	14.72%
Min	$ 1,526.66
Max	$ 4,110.91
Range	$ 2,584.25
Standard Error	$ 11.96

Monte Carlo Simulation with BBN modeling risk factors

Total Cost

Statistics	Values
Mean	$ 2,824.26
Median	$ 2,751.78
Variance	$ 253,897.46
Standard Deviation	$ 503.88
Coefficient of Variation	17.84%
Min	$ 1,755.47
Max	$ 4,588.57
Range	$ 2,833.10
Standard Error	$ 15.93

Table Of Contents

International Cost Estimating
and Analysis Association

Booz | Allen | Hamilton

Summary

▶ Bayesian Belief Networks offer methodology for modeling inter relational complexity within a cost estimating model providing both qualitative and quantitative approaches to the problem

▶ BBNs can account for more risk factors and inter-relationships among them

▶ BBNs with Monte Carlo simulation modeling provide flexibility for cost estimating and risk analysis

▶ BBNS are applicable for integrated cost, schedule and risk analysis

References

1. Arízaga, Javier F. O., "A methodology for project risk analysis using Bayesian belief networks within a Monte Carlo simulation environment", Digital Repository at the University of Maryland, 2007
http://drum.lib.umd.edu/bitstream/1903/6871/1/umi-umd-4362.pdf

2. Korb, Kevin B., Nicholson, Ann E., "Bayesian Artificial Intelligence", CRC press, 2010 http://www.csse.monash.edu.au/bai/book/BAI_Chapter2.pdf

3. Wikipedia, http://en.wikipedia.org/wiki/Bayesian_network

International Cost Estimating and Analysis Association

Booz | Allen | Hamilton

For further information . . .

Marina Dombrovskaya
Senior Consultant

Booz | Allen | Hamilton

Booz Allen Hamilton Inc.
8283 Greensboro Drive
McLean, VA 22102
(301)838 - 3698
Dombrovskaya_Marina@bah.com

Booz | Allen | Hamilton

Back-up Slides