Using Bayesian Belief Networks with Monte Carlo Simulation Modeling

Dr. Marina Dombrovskaya
Table Of Contents

- Introduction: Monte Carlo Simulation and Bayesian Belief Networks
- Bayesian Belief Networks and Cost Estimating Modeling
- Bayes’ Theorem
- Bayesian Belief Networks within a Monte Carlo Simulation Model
- Summary
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 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 represented by a table (matrix) with $2^N$ entries.
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

- 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?
Table Of Contents

- Introduction: Monte Carlo Simulation and Bayesian Belief Networks
- Bayesian Belief Networks and Cost Estimating Modeling
- Bayes’ Theorem
- Bayesian Belief Networks within a Monte Carlo Simulation Model
- Summary
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 interdependencies 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

- Introduction: Monte Carlo Simulation and Bayesian Belief Networks
- Bayesian Belief Networks and Cost Estimating Modeling
- Bayes’ Theorem
- Bayesian Belief Networks within a Monte Carlo Simulation Model
- Summary
Conditional Probability

- **Conditional probability** is the probability of an event given that another event already occurred.

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} \]

- \[ P(A|B) + P(\bar{A}|B) = 1 \]

- In general, \( P(A|B) \neq P(B|A) \)

- **Chain Rule:** \( P(E_1E_2\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|P_2,P_1)\cdots P(E_n|E_1,\cdots E_{n-1}) \)

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

\[ P(A = \{HH\}|B = \{HH, HT\}) = \frac{1/4}{2/4} = \frac{1}{2} \]
Bayes’ Theorem

Given events $A$ and $B$ such that $P(B) \neq 0$

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

- $P(A)$ 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*

The *posterior* can also be expressed as:

$$\text{posterior} = \frac{\text{prior} \cdot \text{likelihood}}{\text{evidence}}$$
Bayesian Belief Network and Conditional Probabilities

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

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

\[ P(N_1, N_2, N_3, N_4) = P(N_1)P(N_2|N_1)P(N_3|N_2)P(N_4|N_2,N_3) \]

Since Bayes’ Theorem is in some sense reversible, BBNs provide a way to reverse the logic without changing the network.
Bayesian Belief Network: Example

- **Inclement Weather**
  - T: 30%
  - F: 70%

- **Unfavorable Road Conditions**
  - IW: T
    - T: 90%
    - F: 10%
  - UR: F
    - T: 40%
    - F: 60%

- **Vehicle Replacement**
  - UR: T
    - T: 10%
    - F: 90%
  - VR: F
    - T: 1%
    - F: 99%

- **Tire Replacement**

- **Timely Maintenance**
  - T: 80%
  - F: 20%

- **Transition Probabilities**

- **Table:**
  - | UR | TM | T | F |
  - | T  | T  | 60%| 40%|
  - | F  | T  | 5% | 95%|
  - | T  | F  | 90%| 10%|
  - | F  | F  | 20%| 80%|

**ICEAA**

Booz | Allen | Hamilton
Bayesian Belief Network: Example

- Calculate probability of Tire Replacement (TR) given Inclement Weather conditions (IW) and absence of Timely Maintenance (TM).

\[ P(\overline{TR}|IW, \overline{TM}) = \frac{P(IW, TM|TR)P(TR)}{P(IW, TM)} = \frac{P(IW|TR)P(TM|TR)P(TR)}{P(IW)P(TM)} \]

\[ P(IW|TR) = \frac{P(TR|IW)P(IW)}{P(TR)} \]

\[ P(TM|TR) = \frac{P(TR|TM)P(TM)}{P(TR)} = \frac{P(TR|TM, UR)P(UR)+P(TR|TM, UR)P(UR)}{P(TR)} P(TM) \]
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

\[
P(TR|IW, \overline{TM}) = \frac{P(IW|TR)P(\overline{TM}|TR)P(TR)}{P(IW)P(TM)} = \frac{36.5\%*20.8\%*56.1\%}{30\%*20\%}
\]

- \( (TR|IW, \overline{TM}) = 70.9\% \)
Table Of Contents

- Introduction: Monte Carlo Simulation and Bayesian Belief Networks
- Bayesian Belief Networks and Cost Estimating Modeling
- Bayes’ Theorem
- Bayesian Belief Networks within a Monte Carlo Simulation Model
- Summary
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

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

### Risks

<table>
<thead>
<tr>
<th>Risks</th>
<th>Name</th>
<th>Description</th>
<th>Distribution Parameters</th>
<th>Prob of Occur</th>
<th>Distribution Type</th>
<th>Param 1</th>
<th>Param 2</th>
<th>Param 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TR</td>
<td>Tire Replacement</td>
<td>Triangular</td>
<td>0.2</td>
<td></td>
<td>100</td>
<td>500</td>
<td>700</td>
</tr>
<tr>
<td>2</td>
<td>VR</td>
<td>Vehicle Replacement</td>
<td>Triangular</td>
<td>0.01</td>
<td></td>
<td>400</td>
<td>1000</td>
<td>1500</td>
</tr>
</tbody>
</table>

Total estimated cost of vehicle maintenance: $2,674.84
Bayesian Belief Network for Risk Factors of Military Vehicle Maintenance Model

- **Inclement Weather**
  - T: 30%
  - F: 70%

- **Unfavorable Road Conditions**
  - IW: T: 90%, F: 10%
  - UR: T: 10%, F: 99%
  - VR: T: 10%, F: 90%

- **Timely Maintenance**
  - TM: T: 80%, F: 20%

- **Vehicle Replacement**
  - T: 90%, F: 10%

- **Tire Replacement**
  - T: 90%, F: 10%

- **Risk Factors Table**
<table>
<thead>
<tr>
<th>UR</th>
<th>TM</th>
<th>T</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>5%</td>
<td>95%</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>90%</td>
<td>10%</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>20%</td>
<td>80%</td>
</tr>
</tbody>
</table>
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
### Monte Carlo Simulation with risk factors modeled as independent events

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$2,567.55</td>
</tr>
<tr>
<td>Median</td>
<td>$2,536.74</td>
</tr>
<tr>
<td>Variance</td>
<td>$142,931.81</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>$378.06</td>
</tr>
<tr>
<td>Coefficient of Variation</td>
<td>14.72%</td>
</tr>
<tr>
<td>Min</td>
<td>$1,526.66</td>
</tr>
<tr>
<td>Max</td>
<td>$4,110.91</td>
</tr>
<tr>
<td>Range</td>
<td>$2,584.25</td>
</tr>
<tr>
<td>Standard Error</td>
<td>$11.96</td>
</tr>
</tbody>
</table>

### Monte Carlo Simulation with BBN modeling risk factors

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$2,824.26</td>
</tr>
<tr>
<td>Median</td>
<td>$2,751.78</td>
</tr>
<tr>
<td>Variance</td>
<td>$253,897.46</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>$503.88</td>
</tr>
<tr>
<td>Coefficient of Variation</td>
<td>17.84%</td>
</tr>
<tr>
<td>Min</td>
<td>$1,755.47</td>
</tr>
<tr>
<td>Max</td>
<td>$4,588.57</td>
</tr>
<tr>
<td>Range</td>
<td>$2,833.10</td>
</tr>
<tr>
<td>Standard Error</td>
<td>$15.93</td>
</tr>
</tbody>
</table>
Table Of Contents

› Introduction: Monte Carlo Simulation and Bayesian Belief Networks
› Bayesian Belief Networks and Cost Estimating Modeling
› Bayes’ Theorem
› Bayesian Belief Networks within a Monte Carlo Simulation Model

› Summary
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


For further information . . .

Marina Dombrovskaya
Senior Consultant

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