Triangular Distributions and Correlations

The simple math behind triangular distributions and correlations in Monte Carlo simulations

Jennifer Lampe  Jeffrey Platten

June 9, 2015
San Diego, CA
<table>
<thead>
<tr>
<th>Agenda</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Types of distributions</td>
<td>3</td>
</tr>
<tr>
<td>Simple trigonometry</td>
<td>6</td>
</tr>
<tr>
<td>The risk adjusted mean</td>
<td>14</td>
</tr>
<tr>
<td>Using triangular distributions in Monte Carlo simulations</td>
<td>16</td>
</tr>
<tr>
<td>Correlation</td>
<td>21</td>
</tr>
</tbody>
</table>
Why use a triangular distribution?

- Triangular distributions are often used in estimating cost risks because the math is relatively simple and because it nearly approximates a lognormal distribution.

The difference in Monte Carlo output will be insignificant.
Some (mathematically) possible distributions
What is a triangular distribution?

- A triangular area which visually represents mathematically the likelihood of possible outcomes by defining
  - lower limit (a)
  - upper limit (b)
  - mode (c) or most likely
- Implies that the likelihood increases consistently (on a straight line) as the estimate approaches the mode from either side
- The area of the triangle is 1 (or 100%, if a, b and c are percents)
- The base of the triangle is b-a
  - So the height of the triangle is \( \frac{2}{b-a} \)

\[
\text{Height} = \frac{2}{b-a}
\]
Triangular distribution simple trigonometry

Height = \frac{2}{b-a}

Why is Height = 2/(b-a) ?
Because Area of \( \Delta = 1 \)
and Area of \( \Delta = \frac{1}{2} \) Base \cdot Height
= \frac{1}{2} (b-a) \cdot \frac{2}{(b-a)} = 1

Let \( r' \) be a random number between 0 and 1.
\( x' \) is the calculated estimate in which there is \( r' \) probability that the actual outcome will be less than \( x' \) and
(1-\( r' \)) probability that the actual outcome will be greater than \( x' \)

a = minimum expected value
b = maximum expected value
c = most likely value
Triangular distribution simple trigonometry

For small values of $r'$, which fall into the triangular distribution to the left of $c$:

- $r' \leq \frac{1}{2}(c-a)\frac{2}{(b-a)}$
- $r' \leq \frac{c-a}{(b-a)}$

the line defining the left edge of the triangle is all that’s needed to calculate $x'$

$\text{Height} = \frac{2}{(b-a)}$

$a = \text{minimum expected value}$
$c = \text{most likely value}$
$b = \text{maximum expected value}$
Calculate the formula for the line defining the left edge of the triangle. This is needed to calculate $x'$

$$y = \text{intercept} + \text{slope} \cdot x$$

We can get a formula for the slope of the line because we know $(a,0)$ and $(c, \frac{2}{b-a})$ are two points on the line.

$$\text{Slope} = \frac{\text{rise}}{\text{run}} = \frac{2}{(b-a) \cdot (c-a)}$$

and we can calculate the intercept by plugging in $(a,0)$

$$0 = \text{intercept} + 2/[(b-a)(c-a)] \cdot a$$

$$\text{Intercept} = -\frac{2a}{[(b-a)(c-a)]}$$

$$y = -\frac{2a}{[(b-a)(c-a)]} + \frac{2x}{[(b-a)(c-a)]}$$

$$y = 2(x-a)/[(b-a)(c-a)]$$

$a =$ minimum expected value

$c =$ most likely value

$b =$ maximum expected value

Height $= \frac{2}{(b-a)}$
Triangular distribution simple trigonometry

For a random $r'$, which is small [$\leq (c-a)/(b-a)$] solve for $x'$
$r'$ represents an area in the triangular distribution

$r' = \frac{1}{2} \text{Base} \cdot \text{Height}$
$r' = \frac{1}{2} (x'-a) \cdot y'$
$r' = \frac{1}{2} (x'-a) \cdot 2(x'-a)/[(b-a)(c-a)]$
$r' = (x'-a)^2/[(b-a)(c-a)]$
$(x'-a)^2 = r'(b-a)(c-a)$
$x'-a = \text{SQRT}[r'(b-a)(c-a)]$
$x' = \text{SQRT}[r'(b-a)(c-a)] + a

$a = \text{minimum expected value}$
$c = \text{most likely value}$
$b = \text{maximum expected value}$
Triangular distribution simple trigonometry

For large values of $r'$, which fall into the triangular distribution to the right of $c$:

$$r' > \frac{1}{2}(c-a)\frac{2}{(b-a)}$$
$$r' > \frac{c-a}{(b-a)}$$

the line defining the right edge of the triangle is needed to calculate $x'$, in addition to the left edge.

For ease of calculation, define: $s' = 1 - r'$

$a =$ minimum expected value
$c =$ most likely value
$b =$ maximum expected value
triangular distribution simple trigonometry

Height = \( \frac{2}{b-a} \)

\( y = \text{intercept} + \text{slope} \cdot x \)

Calculate the formula for the line defining the right edge of the triangle. This is needed to calculate \( x' \)

\( y = \text{intercept} + \text{slope} \cdot x \)

We can get a formula for the slope of the line because we know \((c, \frac{2}{b-a})\) and \((b,0)\) are two points on the line.

Slope = \( \frac{\text{rise}}{\text{run}} = \frac{-2/(b-a)}{(b-c)} = \frac{-2}{(b-a)(b-c)} \)

and we can calculate the intercept by plugging in \((b,0)\)

0 = \( \text{intercept} - 2/[(b-a)(b-c)] \cdot b \)

Intercept = \( 2b/[(b-a)(b-c)] \)

\( y = 2b/[(b-a)(c-a)] - 2x/[(b-a)(b-c)] \)

\( y = 2(b-x)/[(b-a)(b-c)] \)

a = minimum expected value

b = maximum expected value

c = most likely value
Triangular distribution simple trigonometry

Height = \( \frac{2}{b-a} \)

For a random \( r' \), which is large [\( > \frac{(c-a)}{(b-a)} \)] solve for \( x' \)

\( r' \) represents an area in the triangular distribution

Define: \( s' = 1 - r' \)

\( s' = \frac{1}{2} \) Base \cdot Height

\( s' = \frac{1}{2} (b-x') \cdot y' \)

\( s' = \frac{1}{2} (b-x') \cdot 2(b-x')/[(b-a)(b-c)] \)

\( s' = (b-x')^2/[(b-a)(b-c)] \)

\( (b-x')^2 = s'(b-a)(b-c) \)

\( b-x' = \text{SQRT}[s'(b-a)(b-c)] \)

\( -x' = \text{SQRT}[s'(b-a)(b-c)]-b \)

\( x' = b-\text{SQRT}[s'(b-a)(b-c)] \)

\( x' = b-\text{SQRT}[(1-r')(b-a)(b-c)] \)

a = minimum expected value

b = maximum expected value

c = most likely value
Triangular distribution simple trigonometry

Now when you generate any random $r'$, you can calculate the corresponding $x'$ value

If $r' \leq (c-a)/(b-a)$
$$x' = \sqrt{r'(b-a)(c-a)} + a$$

If $r' > (c-a)/(b-a)$
$$x' = b - \sqrt{(1-r')(b-a)(b-c)}$$

$\text{Height} = 2/(b-a)$

$a = \text{minimum expected value}$
$c = \text{most likely value}$
$b = \text{maximum expected value}$
What is the centroid of triangle, if a and b are absolute min and max?

- $A = (a, 0)$  
  a is the absolute minimum

- $B = (b, 0)$  
  b is the absolute maximum

- $H = (c, 2/(b-a))$  
  H is the height at the mode c (most likely)

- The Centroid is the center of balance

\[
\begin{bmatrix}
\frac{X_1 + X_2 + X_3}{3} & \frac{Y_1 + Y_2 + Y_3}{3} \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\frac{(a+b+c)}{3} & \frac{2}{3 (b-a)} \\
\end{bmatrix}
\]

- The X-component is the risk adjusted mean

The Centroid
The Risk Adjusted Mean
What if a and b are not absolute, but are the 15% and 85% percentiles?

- Example: $a = 0.9 = \text{the 15}^{th}\ \text{percentile}$
  $b = 1.2 = \text{the 85}^{th}\ \text{percentile}$
  “Original” Risk Adjusted Mean = 1.033

- The adjusted distribution is:
  
  $A = (0.74, 0)$  
  $B = (1.4, 0)$  
  $H = (1.0, 3.03)\ \text{height} = 2/0.66$
  “New” Risk Adjusted Mean = 1.047

- In many defense acquisition situations, it is customary to apply the 15$^{th}$ and 85$^{th}$ percentiles to the min and max estimates of subject matter experts.

The “original” Risk Adjusted Mean
The “new” Risk Adjusted Mean
Using triangular distributions in Monte Carlo simulations

- Define the triangular parameters (min, most likely, max) for each line (or each WBS or group) of your estimate.
  - Min and Max are typically defined as percentages (of the most likely value)
- Do a Monte Carlo Simulation by running thousands of iterations (applying random r’ values) to the defined distribution for each line.
- According to central limit theorem, the resulting distribution of total combined cost estimates will approximate a normal distribution.
Example in Excel

Monte Carlo Simulation, Triangular Distributions

Enter estimates in blue cells only.

<table>
<thead>
<tr>
<th>WBS</th>
<th>Point Est</th>
<th>Triangular Risk Range</th>
<th>Triangular Cost Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Most L</td>
<td>Min</td>
<td>ML</td>
</tr>
<tr>
<td>0</td>
<td>$907,500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$280,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>$185,000</td>
<td>80%</td>
<td>100%</td>
</tr>
<tr>
<td>1.2</td>
<td>$74,000</td>
<td>90%</td>
<td>100%</td>
</tr>
<tr>
<td>1.3</td>
<td>$21,000</td>
<td>80%</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>$240,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>$65,000</td>
<td>90%</td>
<td>100%</td>
</tr>
<tr>
<td>2.2</td>
<td>$99,000</td>
<td>80%</td>
<td>100%</td>
</tr>
<tr>
<td>2.3</td>
<td>$76,000</td>
<td>70%</td>
<td>100%</td>
</tr>
<tr>
<td>3</td>
<td>$387,500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>$1,500</td>
<td>80%</td>
<td>100%</td>
</tr>
<tr>
<td>3.2</td>
<td>$153,000</td>
<td>90%</td>
<td>100%</td>
</tr>
<tr>
<td>3.3</td>
<td>$233,000</td>
<td>85%</td>
<td>100%</td>
</tr>
</tbody>
</table>


where:
- column L contains a random number, RAND()
- column H contains the minimum value of the triangular distribution, a
- column I contains the most likely value of the triangular distribution, c
- column J contains the maximum value of the triangular distribution, b
Example: Triangular distribution of one of the cost elements
Example: normal distribution of total combined cost estimates

Monte Carlo
1000 iterations
Hit F9 to re-run.

Point
Estimate
$ 907,500

Uncertainty
Estimate
~ $ 926,500

Normal Distribution
Example: S-curve

Monte Carlo
1000 iterations
Hit F9 to re-run.

Point Estimate
$907,500

Uncertainty Estimate
~ $926,500

Monte Carlo
1000 iterations
Hit F9 to re-run.
What is Correlation?

- Correlation is a measure of the relationship between two variables.
- There is correlation when a relation between variables exists which tend to vary in a way not expected on the basis of chance alone.
- Due to some underlying relationship between two variables, X and Y, knowing the outcome of one provides additional information as to the likely outcome of the other may be.

**No Correlation**

Knowing the result of X does not provide any additional information as to what the likely outcome of Y may be.

**Correlation = .50**

Knowing the result of X provides additional information as to the likely outcome of Y.
Pearson vs. Spearman’s Rank Correlation

- There are various ways to measure correlation
- Pearson Correlation is the most widely used
- It is a measurement of the **linear** relationship between two variables
- Crystal Ball uses Spearman’s Rank Correlation
- Spearman’s Rank Correlation assumes a **monotonic** relationship rather than a linear relationship
Pearson vs. Spearman’s Rank Correlation

- The Image on the left is perfectly monotone, meaning that each successive point is greater than the previous point, but it is not linear. This shows that the Spearman correlation will show perfect correlation while the Pearson correlation will not.

- The second image shows another example of the differences between Spearman and Pearson correlation.
Calculating Spearman’s Rank Correlation with Excel

In Excel:
1. Start with two data sets (Columns A and B)
2. Rank each data set using the Rank function in Excel. For column, C the formula would be Rank(A2,A2:A15)
3. Using the Correlation function in Excel, you can find the correlation of your data set using Columns C and D. Use the formula Correl(C2:C15,D2:D15)

<table>
<thead>
<tr>
<th>Data 1</th>
<th>Data 2</th>
<th>Rank 1</th>
<th>Rank 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>85</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>41</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>25</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>26</td>
<td>16</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>95</td>
<td>8</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>46</td>
<td>9</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>35</td>
<td>41</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>19</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>14</td>
<td>23</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>74</td>
<td>76</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Correlation
-0.032184496
Evaluating Correlation Results

- In the Excel example, the Correlation was approximately -0.03.
- This shows that the two data sets are slightly negatively correlated. This correlation is so small that it is barely recognizable.
- Below shows in photos linear relationships for correlation.
- The -0.03 would look as if there were no correlation.
Defining Correlation for a PLCCE model

- When applying correlation to assumption variables in Crystal Ball, there are certain guidelines to consider.
- First, only correlate variables that are similar.
- For instance, if you put a risk assumption on separate labor rates, those rates can be correlated. But, if you put risk on labor rates and hardware unit cost, these are probably not correlated.
- The below table are default correlation factors from Page 46 of the Joint Cost Schedule Risk and Uncertainty Handbook.

<table>
<thead>
<tr>
<th>Strength</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Weak</td>
<td>0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>Medium</td>
<td>0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>Strong</td>
<td>0.9</td>
<td>-0.9</td>
</tr>
<tr>
<td>Perfect</td>
<td>1.0</td>
<td>-1.0</td>
</tr>
</tbody>
</table>
Defining Correlation in Crystal Ball

1. More Tools > Define Correlations
2. Add your assumptions (add all assumptions for complete matrix). You can build the correlation matrix in Excel and link it to the Crystal Ball simulation.
3. Add correlation to your assumptions
   - Weak Correlation = 0.3 (use as default)
   - Medium Correlation = 0.5
   - Strong Correlation = 0.9
Example Correlation Matrix

- There are many possible combinations for Crystal Ball assumptions and correlation
  - This is only an example of how a correlation matrix may look
- All PLCCEs are unique

<table>
<thead>
<tr>
<th></th>
<th>Labor Rate 1</th>
<th>Labor Rate 2</th>
<th>HW Unit Cost 1</th>
<th>HW Unit Cost 2</th>
<th>HW Unit Cost 3</th>
<th>SW License Cost</th>
<th>Reuse Code Growth</th>
<th>New Code Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor Rate 1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor Rate 2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HW Unit Cost 1</td>
<td></td>
<td></td>
<td>1</td>
<td>0.9</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HW Unit Cost 2</td>
<td></td>
<td></td>
<td>1</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HW Unit Cost 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW License Cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reuse Code Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>New Code Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Questions?

Thank you

For more information, contact

Jennifer Lampe
jennifer.lampe@engilitycorp.com
724-875-9396

Jeffrey Platten
jeffrey.platten@engilitycorp.com
310-892-9150