Presented at the 2007 ISPA/SCEA Joint Annual International Conference and Workshop - www.iceaaonline.com

Developing Risk Adjusted Results with Limited Data

Presented at SCEA
Gabriela Rohlck, MCR LLC
202-548-5523
grohlck@mcri.com
Co-Author: Marc Rose, MCR LLC

15 June, 2007

"Creating Customer-Focused Success"

The Problem

- Need to give decision makers the best available data
- Risk is important
  - Point estimate alone could provide misleading results
- You want to include risk in your analysis
  - Cost or Benefits
  - Schedule risk possibly also applicable
- Time is limited
- Currently only have a point estimate

What to Do?

© MCR, LLC

"Creating Customer-Focused Success"

Quick Risk Methodology

- Original concept developed by Dr. Steve Book, MCR LLC
- Provides a simple approach for applying risk
- Requires only a few assumptions
  - Point Estimate (value and probability)
  - Overall uncertainty range
  - Most-Likely (Mode) probability (for some distributions)

- <u>NOT</u> a substitute for a complete analysis

© MCR, LLC

"Creating Customer-Focused Success"

Conceptual Overview

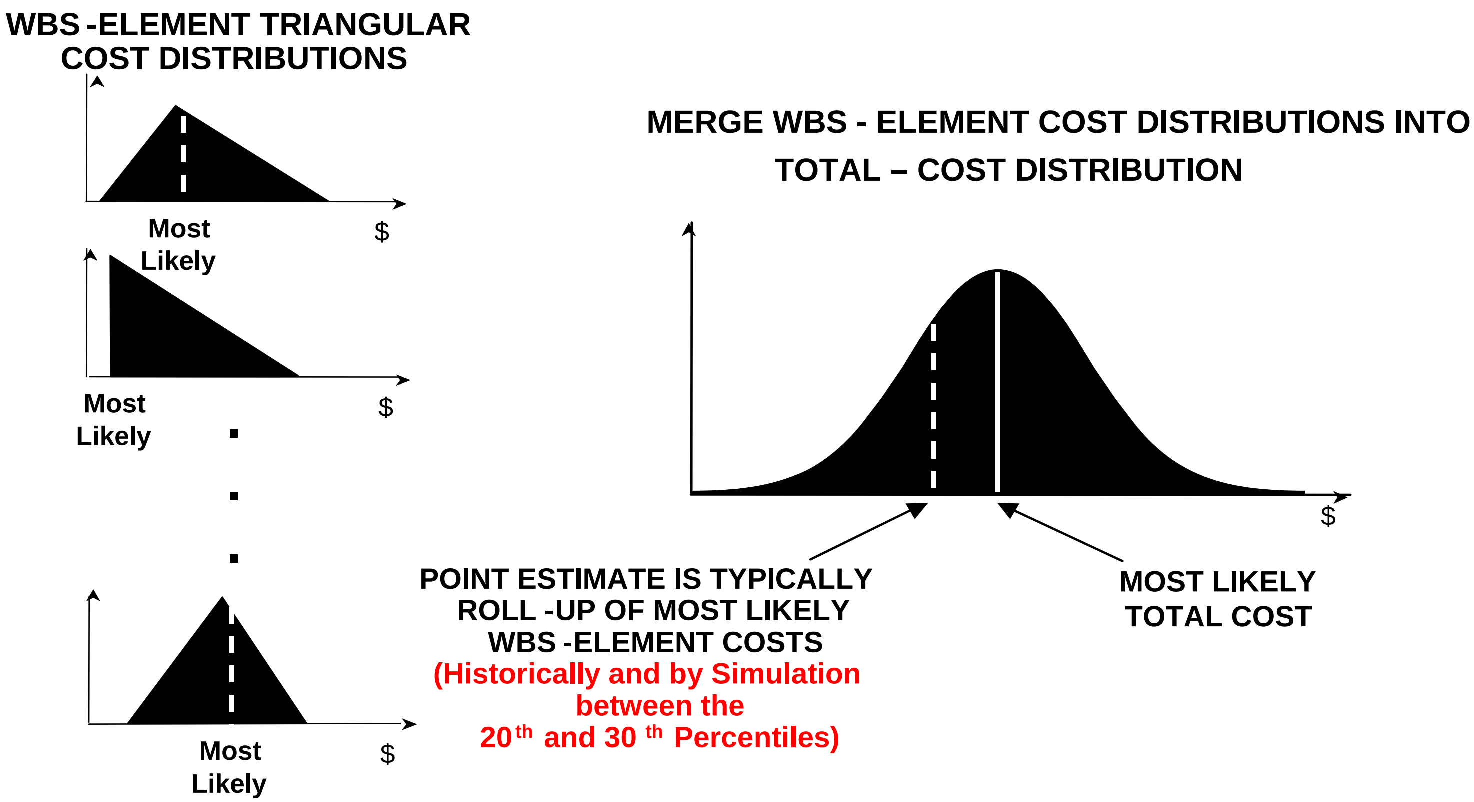
- Risk Analysis is composed of assigning statistical distributions to individual components (e.g., WBS)
- Combine all these distributions into single total distribution

- "Quick Risk" skips these steps and creates the single "final" distribution directly
  - Based on typical results for the point estimate and range of the results
  - Initial development uses a simple Triangular distribution
  - Expanded to include Normal and Log-Normal
- Let's now talk about the "typical" input values

© MCR, LLC

“Creating Customer-Focused Success”

The “Point Estimate”



© MCR, LLC

"Creating Customer-Focused Success"

Uncertainty Range

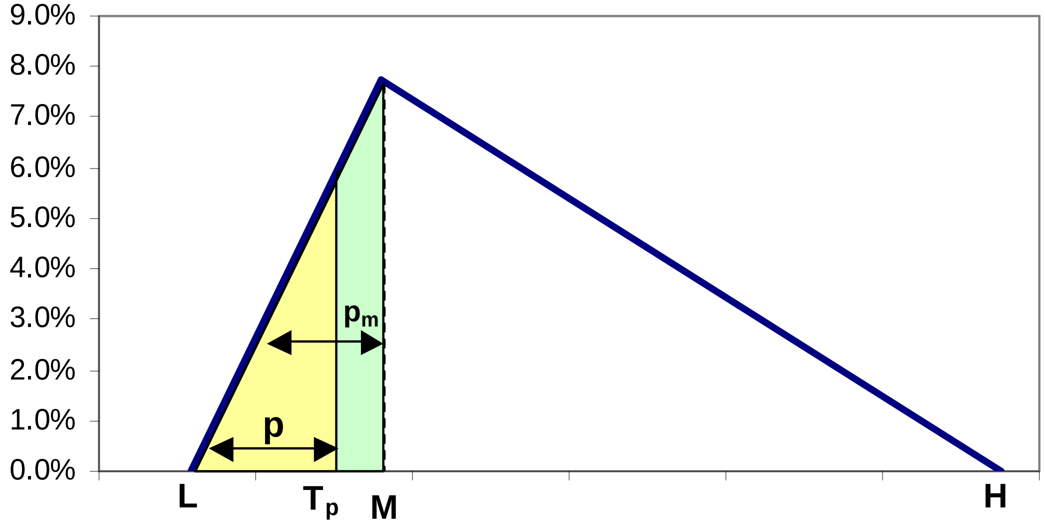
- Basic uncertainty range defined by the ratio of the "high" and the "low" estimates
  - "High" is a reasonable upper bound
  - "Low is a reasonable lower bound
  - Does not use the absolute difference (i.e., High – Low)

  - A Triangular Distribution has a built-in Upper & Lower limits
  - Normal Distribution (and similar) is infinite in extent (not reasonable for limits)

- Let's start with the basic Triangular Distribution

© MCR, LLC

Triangular Distribution



T_p is the point estimate (given)

P is the cumulative probablity at the point estimate (assumed)

k is the ratio of high to low (assumed)

$k = H/L$

$M = (1 - p_m)L + p_m H$

M is the Mode or Most - Likely

p_m is the probability at the mode of the distribution (assumed)

© MCR, LLC

"Creating Customer-Focused Success"

Summary Triangular Equations

- Given the assumed values we can solve for the Triangular distribution parameters (L,M,H)

Two conditions apply:

$$L = T_p \Big/ \left\{ k - (k-1)\sqrt{(1-p)(1-p_m)} \right\} \quad \text{if } p \geq \frac{M-L}{H-L} = p_m$$

$$L = T_p \Big/ \left\{ 1 + (k-1)\sqrt{p \cdot p_m} \right\} \quad \text{if } p \leq \frac{M-L}{H-L}$$

$$H = kL$$

$$M = (1-p_m)L + p_m H$$

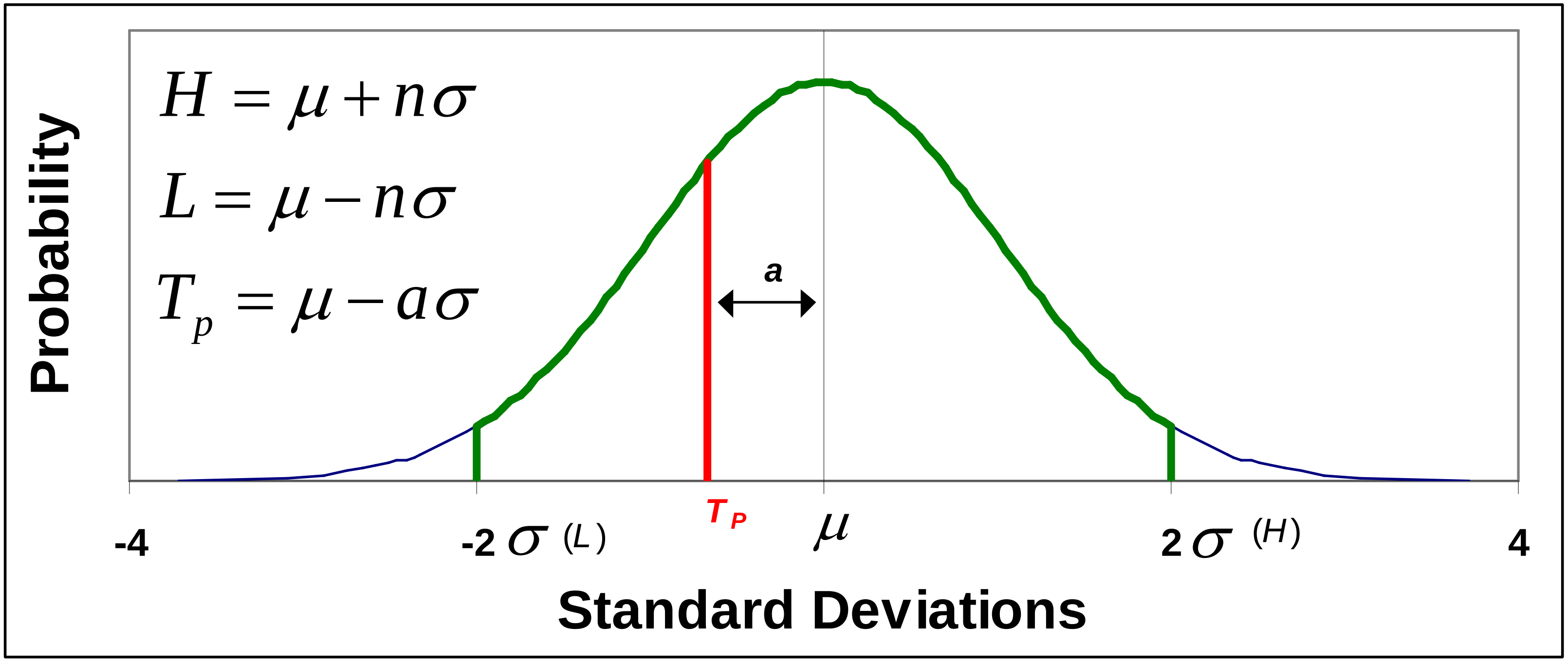
© MCR, LLC

"Creating Customer-Focused Success"

Quick Risk – Normal Distribution

In a similar fashion we develop the method for a Normal Distribution using the concept of high and low values

*** An important distribution due to the Central Limit Theorem**



Note that the point estimate (T_p) is at the 25th percentile as a typical example only. Similarly the *H* and *L* are shown at +/- 2σ

© MCR, LLC

Develop the Quick Risk Equations

- The Assumptions

T_p is the point estimate (given)

n is the number of standard deviations the "High/Low" are off the mean (assumed)

P is the cumulative probablity at the point estimate (assumed) thus a is simply the inverse normal at P

in Excel $a = -\text{norminv}(P,0,1)$

x is the ratio of high to low (assumed)

$x = H/L$

© MCR, LLC

"Creating Customer-Focused Success"

Normal Quick Risk Equations

- Given the previous assumptions we can solve for the unknowns (H, L, μ, σ):

$$L = \frac{2nT_p}{nx - ax + a + n}$$

$$H = xL = x\frac{2nT_p}{nx - ax + a + n}$$

$$\mu = \frac{(H + L)}{2}$$

$$\sigma = \frac{(H - L)}{2n}$$

The full development is provided in the backup along with the log-normal

© MCR, LLC

"Creating Customer-Focused Success"

Application

- Typical Values:
  - p = 25% with n = 2 (standard deviations)
  - for Triangular a p_m = 30%
  - H/L can range from 2 to 8 depending on the program complexity
- Assess, based on the historical record of cost experience, the ratio H/L
  - Hardware: H = 3L
  - Software: H = 8L
  - Testing: H = 2L

© MCR, LLC

“Creating Customer-Focused Success”

Quick Risk Example

As an example here are the inputs for each of the methods developed here:
Normal: n = 2, H/L = 3, Tp = 100, p = 25%
Triangular: H/L = 3, Tp = 100, p = 25%, Pm = 30%
Lognormal: n = 2, H/L = 3, Tp = 100, p = 25%

Results at the 80% confidence level are:
Normal: 145.6
Triangular: 145.5
Lognormal: 151.6

We need a minimum of 46% more funding than at the point estimate.

© MCR, LLC

Presented at the 2007 ISPA/SCEA Joint Annual International Conference and Workshop - www.iceaaonline.com

"Creating Customer-Focused Success"

What if it's a High Risk Project?

- Results at the 80% confidence level for H/L = 8, are :
- Normal: 180
- Triangular: 184.5
- Lognormal: 220

We need a minimum of 80 % more funding than the original estimate

© MCR, LLC

"Creating Customer-Focused Success"

Conclusions

- We developed a quick and "dirty" methodology for performing a "quick-risk" analysis given a simple point estimate and some simple assumptions
- Equations for using three different statistical distributions (Triangular, Normal & Log-Normal) have been developed and provided
- When combined with the assumptions, this methodology allows the decision maker to estimate the high-confidence cost or benefits and make a more informed decision

- Not a substitute for the detailed analysis

© MCR, LLC

Questions?

"Creating Customer-Focused Success"

Backup

Backup - Triangular

BACK UP Notes: "How to Make Your Point Estimate Look Like a Cost-Risk Analysis"

Triangular Distribution

For a quick risk analysis we begin solving below the triangular distribution equations. We assume:

T_p = Point Estimate with p = to the cumulative probability at the point estimate

$H = kL$ where H is the High value in the triangular distribution and L is the Low value

$M = xL + (1 - x)H$ where M is the Mode or Most-Likely

where

$x = 1 - p_m$

and p_m is the cumulative probability at the mode of the distribution

© MCR, LLC

"Creating Customer-Focused Success"

Backup - Triangular

The two general equations for a Triangular Distribution are as follows:

$$T_p = H - \sqrt{(1-p)(H-L)(H-M)} \qquad \text{if } p \geq \frac{M-L}{H-L} \qquad 1)$$

and

$$T_p = L + \sqrt{p(M-L)(H-L)} \qquad \text{if } p \leq \frac{M-L}{H-L} \qquad 2)$$

where T_p is value of the distribution at cumulative probability p

In order to use these two equations we need to solve for one of the variables as a function of the given information. We have chosen to solve for L.

a) Starting with the triangular equation in case 1 we can solve for L:

First substitute for H and M into equation 1:

$$T_p = kL - \sqrt{(1-p)(H-L)(H-M)} \quad = kL - \sqrt{(kL-L)(1-p)\{kL - xL - (1-x)kL\}}$$

Factor the L:

$$= kL - \sqrt{L(k-1)(1-p)L\{k - x - (1-x)k\}} = kL - L\sqrt{(k-1)(1-p)(k - x - k + kx)}$$

Simplify by factoring the (k-1):

$$= kL - L\sqrt{(k-1)(1-p)x(k-1)} = kL - L(k-1)\sqrt{x(1-p)} = L\{k - (k-1)\sqrt{x(1-p)}\}$$

© MCR, LLC

"Creating Customer-Focused Success"

Backup - Triangular

Finally solve for L:

$$L = T_p \Big/ \left\{ k - (k-1)\sqrt{(1-p)(1-p_m)} \right\} \qquad 3)$$

b) Starting with the triangular equation in case 2 we can solve for L:

First substitute for H and M into equation 1:

$$T_p = L + \sqrt{p(M-L)(H-L)} = L + \sqrt{p\{xL + (1-x)kL - L\}(kL - L)}$$

$$= L + \sqrt{pL\{x + (1-x)k - 1\}L(k-1)} = L + L\sqrt{p(x + k - kx - 1)(k-1)}$$

$$= L + L\sqrt{p\{x(1-k) - (1-k)\}(k-1)} = L + L(k-1)\sqrt{p(1-x)} = L + \left\{1 + \left(k - 1\sqrt{p(1 - 1 + p_m}\right)\right\}$$

$$L = \frac{T_p}{\left\{1 + (k-1)\sqrt{p \cdot p_m}\right\}} \qquad 4)$$

© MCR, LLC

"Creating Customer-Focused Success"

Backup - Normal

The more general form of the equations, allowing H & L to be any $n\sigma$ off the mean yields:

$H = xL$ (1)
$H = \mu + n\sigma$ (2)
$L = \mu - n\sigma$ (3)
$T_p = \mu - a\sigma$ (4)
$a = -norminv(T_p, 0, 1)$

In order to use these equations we need to solve for one of the variables as a function of the given information. We have chosen to solve for L. We begin by replacing H into L's equation (3) from above:

$xL = \mu + n\sigma$
$\mu = xL - n\sigma$

Then,

© MCR, LLC

Backup - Normal

$$L = \frac{(\mu + n\sigma)}{x} = T_p + a\sigma - n\sigma$$

From equation (2) we know:

$\mu = L + n\sigma$

It follows that:

$\mu = xT_p + xa\sigma - n\sigma x - n\sigma$

$L = xT_p + xa\sigma - n\sigma x - 2n\sigma = T_p + a\sigma - n\sigma$

$xT_p + xa\sigma - n\sigma x - 2n\sigma - a\sigma + n\sigma = T_p - xT_p$

$a\sigma(x-1) - n\sigma x - n\sigma = T_p - xT_p$

$a\sigma(x-1) - n\sigma x - n\sigma = T_p - xT_p$

$$\sigma = \frac{(T_p - xT_p)}{a(x-1) - nx - n}$$

$$L = T_p + a\frac{(T_p - xT_p)}{a(x-1) - nx - n} - n\frac{(T_p - xT_p)}{a(x-1) - nx - n} = \frac{2n\,T_p}{nx - ax + a + n}$$

$$L = \frac{2n\,T_p}{nx - ax + a + n}$$

© MCR, LLC

"Creating Customer-Focused Success"

Backup - Normal

$$H = x\frac{2nT_p}{nx - ax + a + n}$$

Summing the following equations from above:

$H = \mu + n\sigma$
$L = \mu - n\sigma$

We determine:

$$\mu = \frac{(H + L)}{2}$$

Therefore,

$$\sigma = \frac{(H - \mu)}{n} = \frac{H}{n} - \frac{H + L}{2n} = \frac{H - L}{2n}$$

$$\sigma = \frac{(H - L)}{2n}$$

© MCR, LLC

Quick-Risk LogNormal Distribution

- The Assumptions:

$$z = \text{Ratio of } H / L$$

$$H = zL$$

$$T_P : \text{Point Estimate @ c = Selected Percentile (e.g., 25\%)}$$

$$and\ n = \text{implicit confidence level}$$

let

$$\alpha = \text{mean in } Normal \text{ Space}$$

$$\beta = \text{Standard Deviation in } Normal \text{ Space}$$

then

$$\ln(H) = \alpha + n\beta \quad \text{in } Normal \text{ Space}$$

$$\ln(L) = \alpha - n\beta \quad \text{in } Normal \text{ Space}$$

$$\ln(H) - \ln(L) = \ln(z)$$

$$a = -\text{norminv}(c,0,1)$$

$$\ln(T_P) = \alpha - a\beta$$

© MCR, LLC

Presented at the 2007 ISPA/SCEA Joint Annual International Conference and Workshop - www.iceaaonline.com

"Creating Customer-Focused Success"

Summary Quick-Risk LogNormal Equations

Solving :

$$\beta = \ln(z)/2n$$

$$\alpha = \ln(T_P) + a\ln(z)/2n$$

$$\mu = \exp\left(\alpha + \beta^2 / 2\right)$$

$$\sigma^2 = \exp\left(2\alpha + \beta^2\right)\left(\exp(\beta^2) - 1\right)$$

© MCR, LLC

"Creating Customer-Focused Success"

Backup - LogNormal

Extending the method to a Log-Normal distribution is only a little more complicated than the Normal distribution.

The first step is to convert to "Normal" space by taking the natural logarithm of the values. Thus if your point estimate is 100 then in "Normal" space value is Ln(100) or 4.60517.

Given:

$z = \text{Ratio of } H/L$

$H = zL$

$P: \text{Point Estimate @ c} =$

$\text{Selected Percentile (e.g., 25\%)}$

$\textit{and } n = \text{implicict confidence level}$

let

$\alpha = \text{mean in } \textit{Normal} \text{ Space}$

$\beta = \text{Standard Deviation in } \textit{Normal} \text{ Space}$

then

$\ln(H) = \alpha + n\beta \quad \text{in } \textit{Normal} \text{ Space}$

$\ln(L) = \alpha - n\beta \quad \text{in } \textit{Normal} \text{ Space}$

$\ln(H) - \ln(L) = \ln(z)$

$a = -\text{norminv}(c,0,1)$

$\ln(P) = \alpha - a\beta$

solving:

$\ln(H) - \ln(L) = 2n\beta = \ln(z) \Rightarrow \beta = \ln(z)/2n$

$\ln(P) + a\beta = \alpha \Rightarrow \alpha = \ln(P) + a\ln(z)/2n$

$$\mu = \exp\left(\alpha + \beta^2/2\right)$$

$$\sigma^2 = \exp(2\alpha + \beta^2)(\exp(\beta^2) - 1)$$

© MCR, LLC