AEROSPACE
Assuring Space Mission Success

Compensating for "Lack of Tails" on Triangular Distributions

Melvin Broder
Developmental Planning and Projects

SPE/System Planning and Architectures
11 March 2009

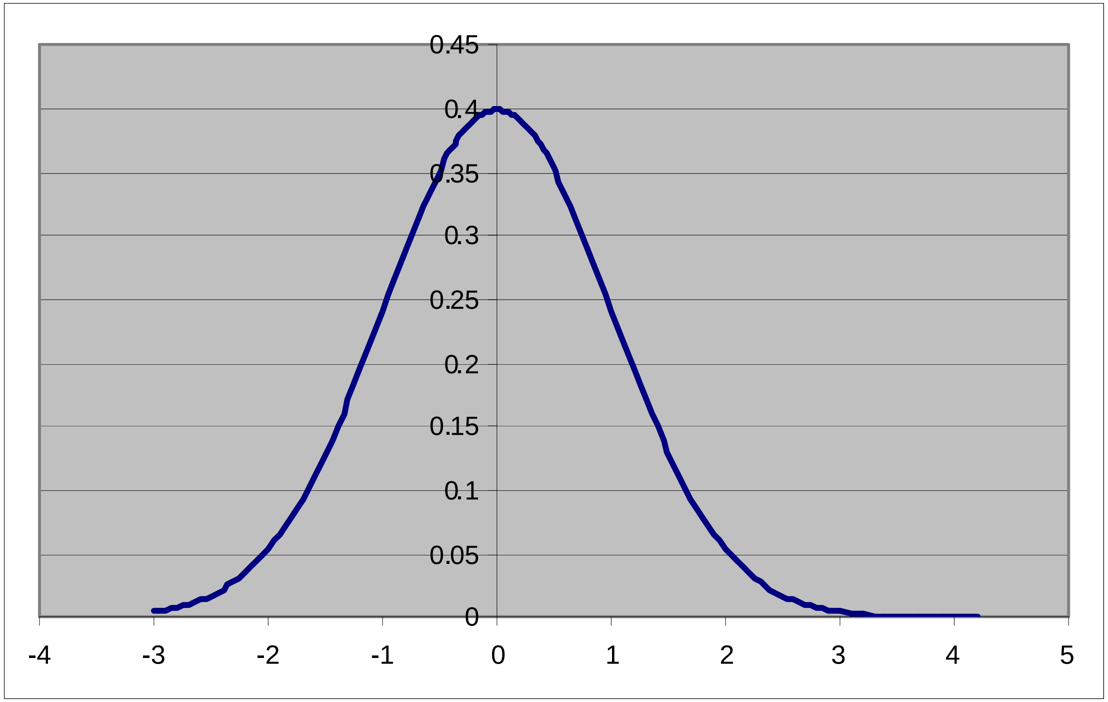
Approved for public release; distribution unlimited

Overview

- Statement of Problem
  - *Triangular distributions have exceptional attributes for the collection of expert information on the uncertainties of cost risk*
  - *Triangles have one major drawback, however, the absolute limitation of risk at the extremes, i.e. values outside the low and high range have 0% probability of occurrence*
  - *Conversely the normal or Gaussian distribution has infinite "tails" but information collection, although requiring only two parameters (mean and sigma) is difficult to elicit from experts*
- This paper uses the two distributions (Triangular and Normal) as representative of two classes of probability distributions to show a solution that offers minimum impact on the underlying cost risk while expanding the probability of high and low values
- Results are also expanded to Lognormal distributions

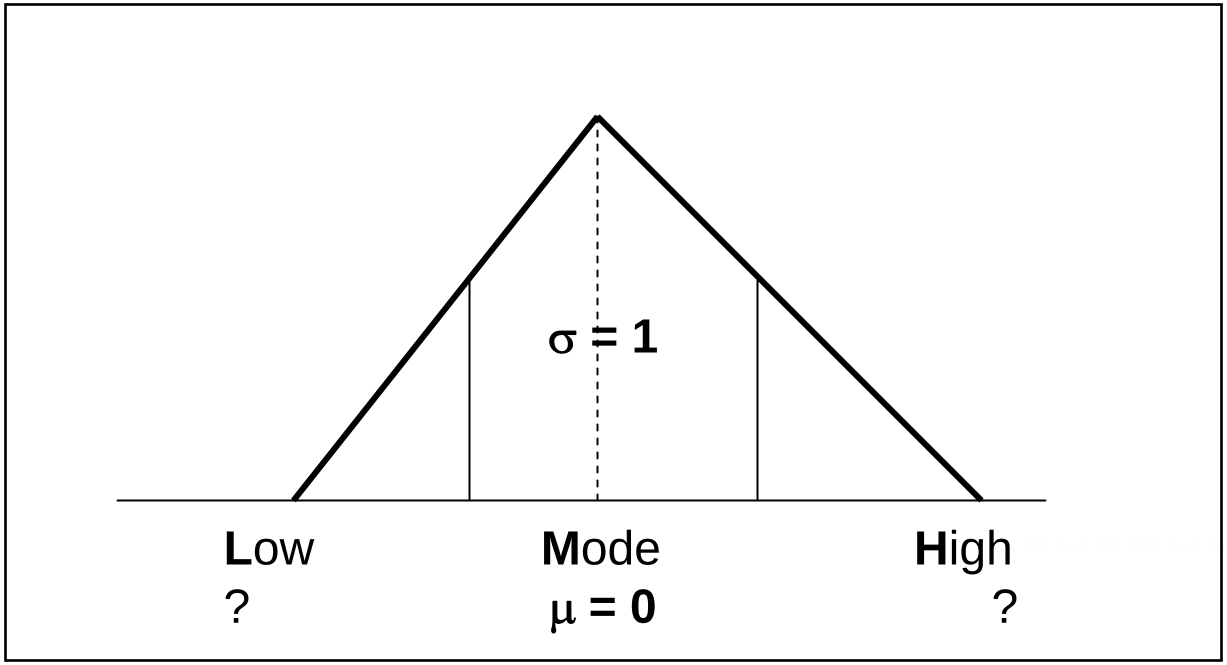
The Standard Normal

- The standard normal distribution is the normal distribution with a mean of zero and a variance of one N (0,1)



“The Standard Triangle”

- Unlike the Normal Distribution N(0,1), there is no generally accepted definition of a “standard triangle;” however we can define it as a triangular distribution one that has the same mean and sigma of the standard Normal, i.e. T(0,1).



Lower and Upper Ends of the Standard Triangle

- We can then calculate the Low (L) and High (H) for a standard triangle given the mean of zero and the a sigma of value of 1

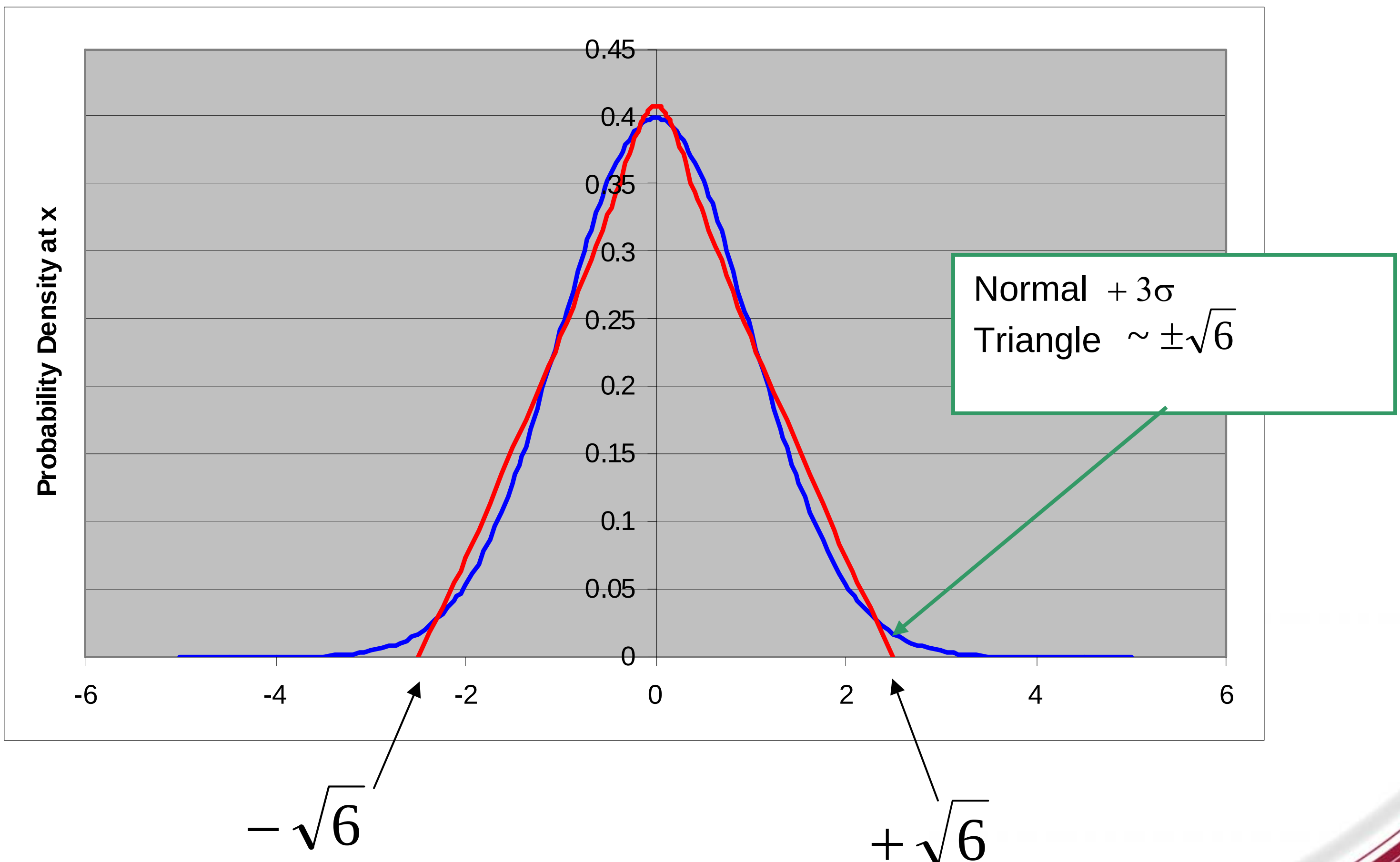
- The mode, M = mean = 0
- The low value L < M < H and H = -L

- Sigma = 1 = $\sqrt{\frac{L^2 + M^2 + H^2 - LM - LH - MH}{18}}$

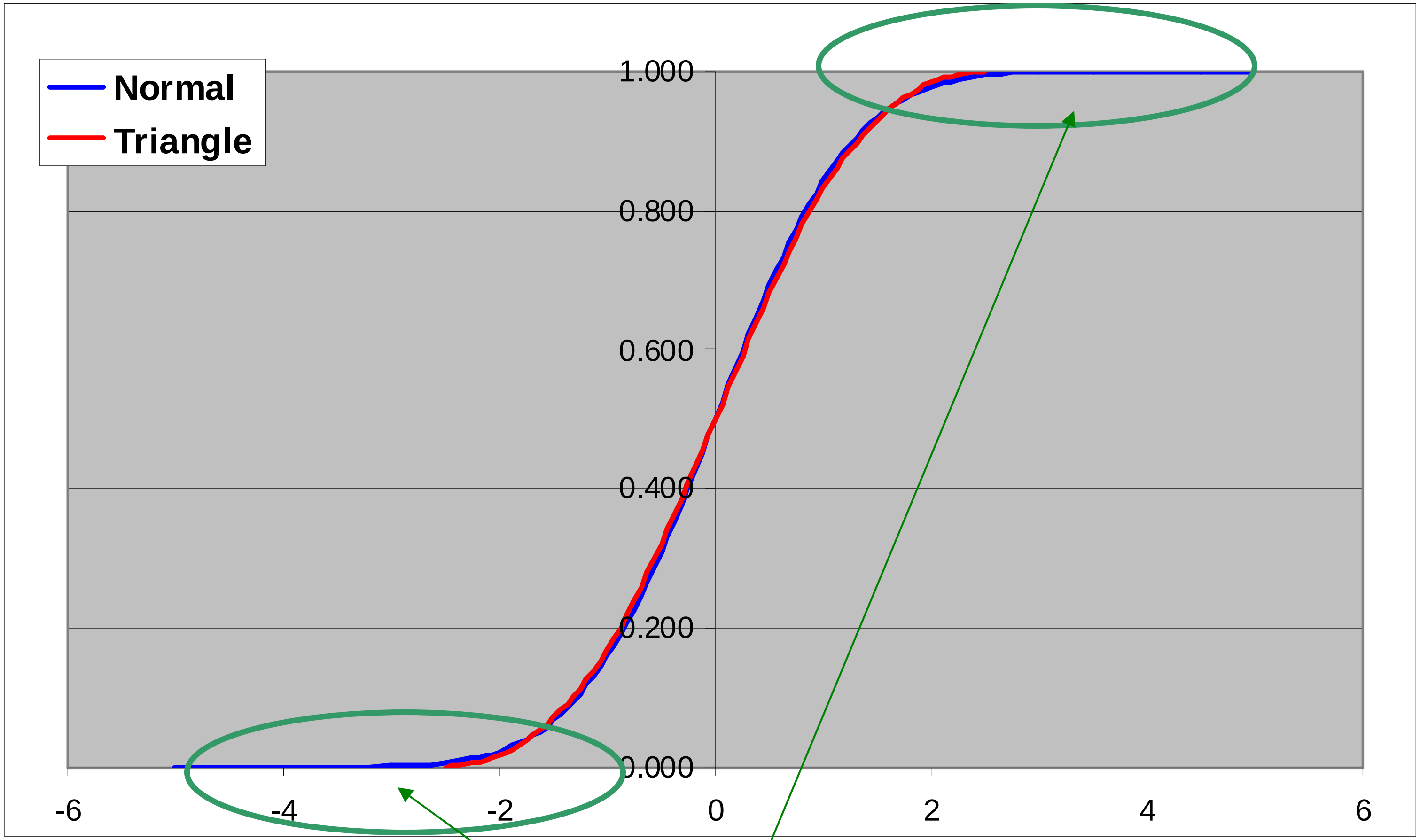
- Because M = 0 and 0-L = H-0, it follows that that L = - $\sqrt{6}$ and H = + $\sqrt{6}$

Normal and Triangle-Fitted Distributions

Normal Distribution at +/- 3σ Approximates Triangle $+/-\sqrt{6}\ \sigma$



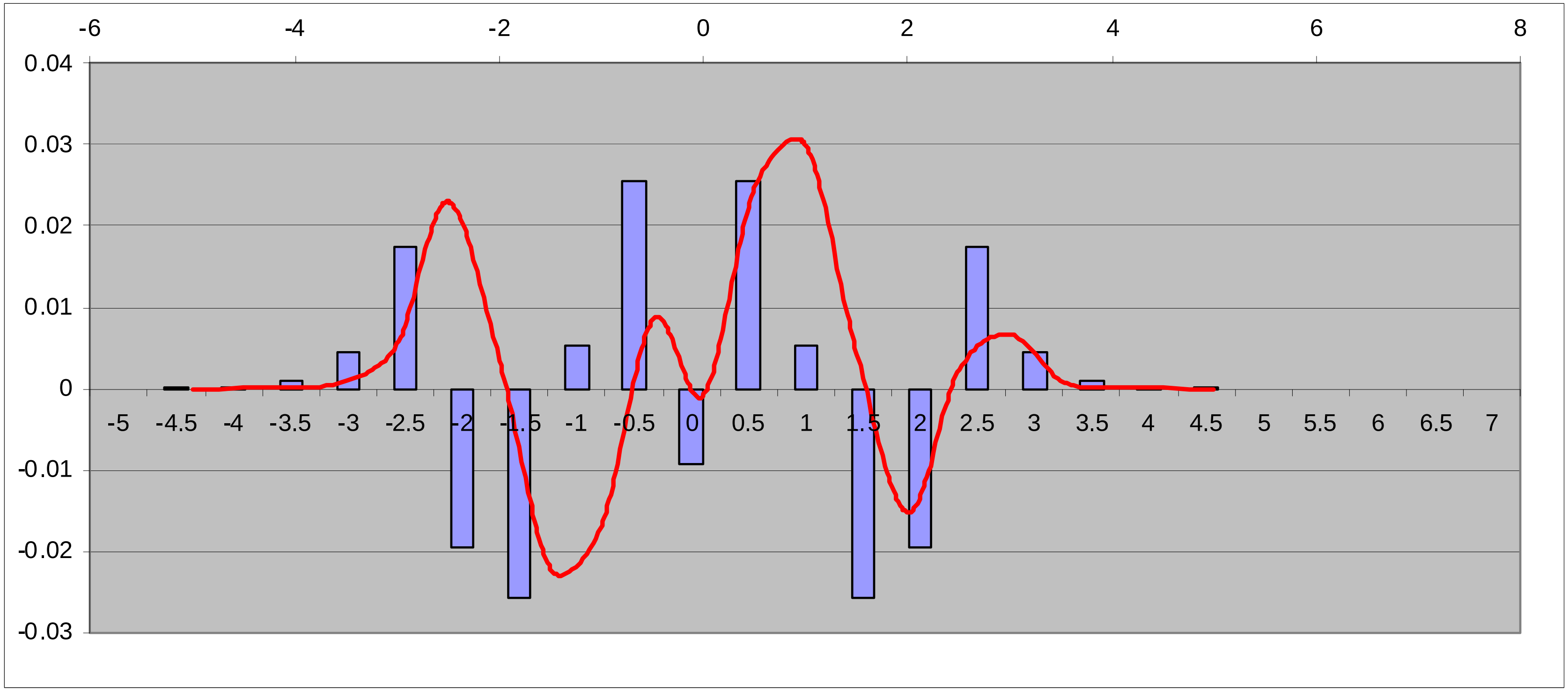
Normal and Triangle-Fitted Cumulative Distributions



What about the “tails”?

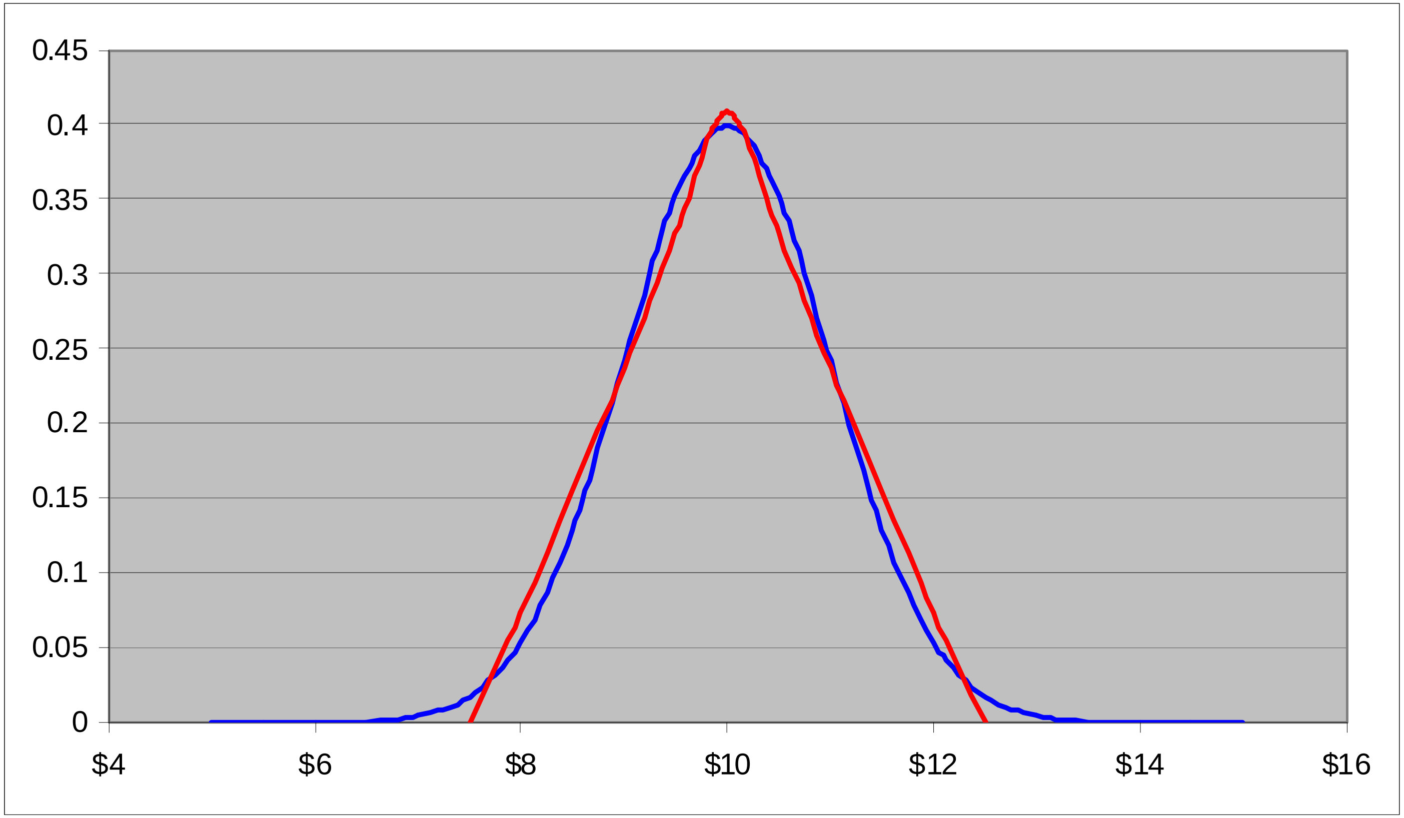
Probability Differences

Normal Minus Triangle-Fitted Distributions



Difference in cumulative (red) probability are off-setting between +/- 3σ

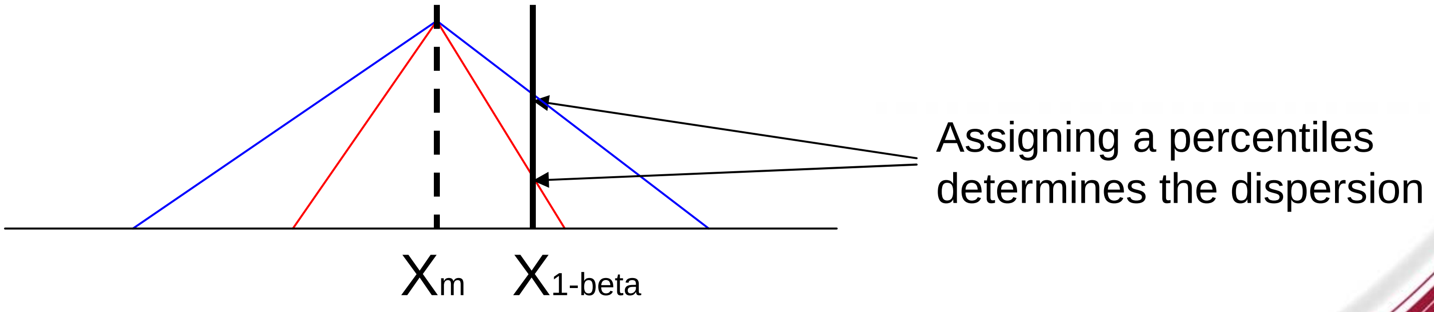
Using with Actual Cost Type Numbers



- Assumes a WBS line item task
  - *Normally distributed with mean of $10M and sigma of 1*
  - *CoV = σ/μ = 1/10 = 10% (note triangle is not equilateral, angle = 10%)*
- Triangle is calculated as L and H = Mean + (Multiplier *($+/-\sqrt{6}$)
  - *Multiplier is the sigma number, in this case = 1*

Eliciting Risk Information Using the Triangular Distribution

- Usual situation is being given a cost range but not knowing the probabilities associated with them
- Assume that you are given a $1.0 M ($\mathbf{X}_m$) estimate by a vendor and that it is accurate +/- $100 K ($\mathbf{X}_{1\text{-beta}} = \mathbf{X}_m + \$1m$)
- Your experts tell you:
  - *"They usually deliver right on time"*
  - *"Their costs can be trusted"*
- You must assign percentiles to that +/- $100 K and calculate a distribution



AFCAA Table was Built to Help with this Problem

- In the example you need to make assumptions:
  - *Is the distribution Normal or is it a triangle*
  - *What is the skewness of the distribution*
- In this table a "Normal has CoV values from 15% to 35%
- In the example a mean of $1.0 M and the assumption of +/-$100 K as a 1 sigma value would give a CoV of 10%

Table 8 Default Bounds (1 of 2) For Subjective Distributions

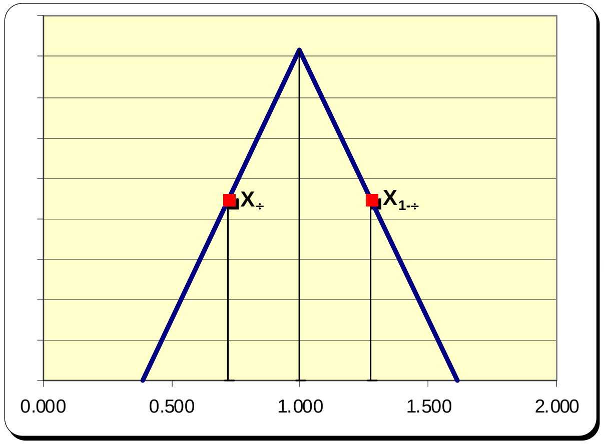
Distribution	Point Estimate Interpretation	Point Estimate and Probability	Mean	CV based on mean	CV Based on PE	15%	85%
Lognormal Low	Median	1.0 (50%)	1.011	0.151	0.153	0.856	1.168
Lognormal Med	Median	1.0 (50%)	1.032	0.254	0.262	0.772	1.296
Lognormal High	Median	1.0 (50%)	1.063	0.361	0.384	0.696	1.437
Normal Low	Mean	1.0 (50%)	1.000	0.150	0.150	0.845	1.155
Normal Med	Mean	1.0 (50%)	1.000	0.250	0.250	0.741	1.259
Normal High	Mean	1.0 (50%)	1.002	0.346	0.347	0.640	1.363
Weibull Low	Mode	1.0 (25%)	1.158	0.179	0.208	0.956	1.370
Weibull Med	Mode	1.0 (20%)	1.393	0.332	0.463	0.956	1.855
Weibull High	Mode	1.0 (15%)	2.104	0.572	1.204	1.000	3.277
Triangle Low Left	Mode	1.0 (75%)	0.878	0.178	0.156	0.695	1.041
Triangle Low	Mode	1.0 (50%)	1.000	0.150	0.150	0.834	1.166
Triangle Low Right	Mode	1.0 (25%)	1.123	0.139	0.156	0.959	1.305
Triangle Med Left	Mode	1.0 (75%)	0.796	0.327	0.260	0.492	1.069
Triangle Med	Mode	1.0 (50%)	1.000	0.250	0.250	0.723	1.277
Triangle Med Right	Mode	1.0 (25%)	1.204	0.216	0.260	0.931	1.508
Triangle High Left	Mode	1.0 (74%)	0.745	0.448	0.334	0.347	1.103
Triangle High	Mode	1.0 (50%)	1.000	0.350	0.350	0.612	1.388
Triangle High Right	Mode	1.0 (25%)	1.286	0.283	0.364	0.903	1.711

Calculating the Triangle by Direct Formula is Complex (Quartic Equation)

Table 8 Default Bounds (1 of 2) For Subjective Distributions

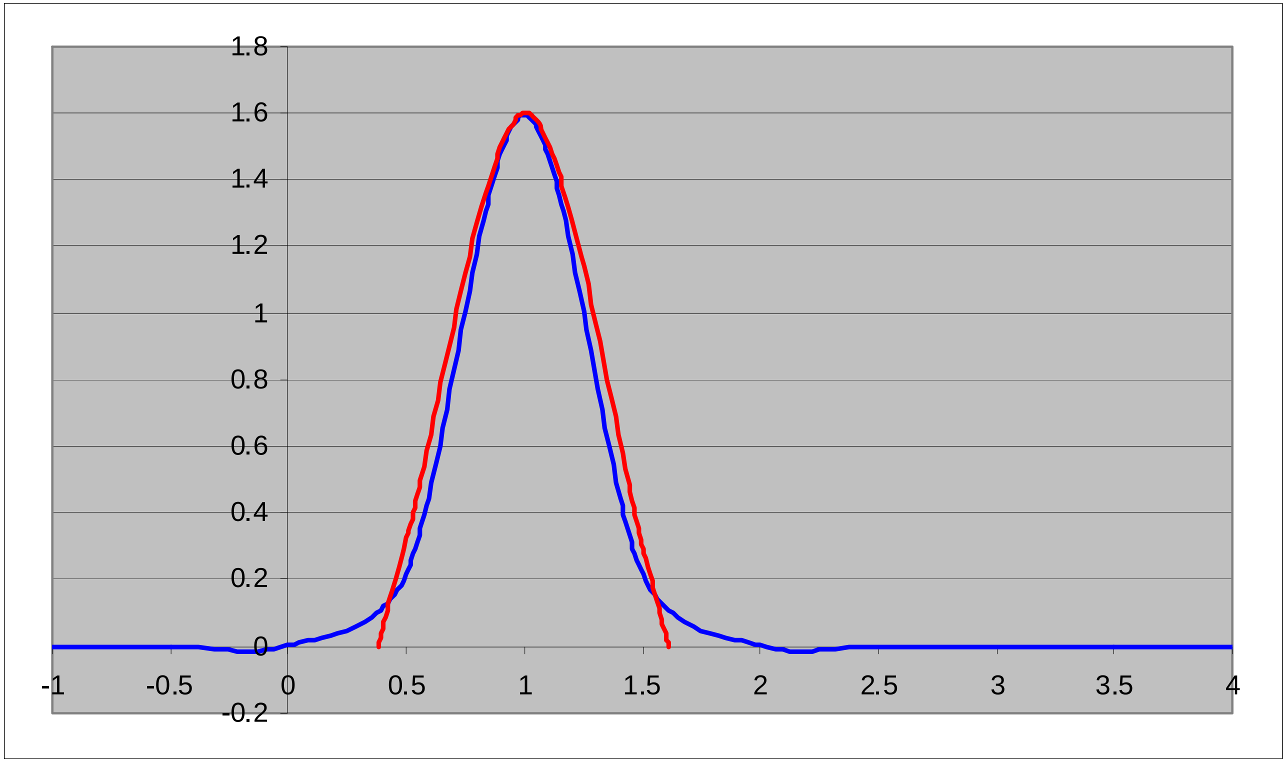
Distribution	Point Estimate Interpretation	Point Estimate and Probability	Mean	CV based on mean	CV Based on PE	15%	85%
Lognormal Low	Median	1.0 (50%)	1.011	0.151	0.153	0.856	1.168
Lognormal Med	Median	1.0 (50%)	1.032	0.254	0.262	0.772	1.296
Lognormal High	Median	1.0 (50%)	1.063	0.361	0.384	0.696	1.437
Normal Low	Mean	1.0 (50%)	1.000	0.150	0.150	0.845	1.155
Normal Med	Mean	1.0 (50%)	1.000	0.250	0.250	0.741	1.259
Normal High	Mean	1.0 (50%)	1.002	0.346	0.347	0.640	1.363
Weibull Low	Mode	1.0 (25%)	1.158	0.179	0.208	0.956	1.370
Weibull Med	Mode	1.0 (20%)	1.393	0.332	0.463	0.956	1.855
Weibull High	Mode	1.0 (15%)	2.104	0.572	1.204	1.000	3.277
Triangle Low Left	Mode	1.0 (75%)	0.878	0.178	0.156	0.695	1.041
Triangle Low	Mode	1.0 (50%)	1.000	0.150	0.150	0.834	1.166
Triangle Low Right	Mode	1.0 (25%)	1.123	0.139	0.156	0.959	1.305
Triangle Med Left	Mode	1.0 (75%)	0.796	0.327	0.260	0.492	1.069
Triangle Med	Mode	1.0 (50%)	1.000	0.250	0.250	0.723	1.277
Triangle Med Right	Mode	1.0 (25%)	1.204	0.216	0.260	0.931	1.508
Triangle High Left	Mode	1.0 (74%)	0.745	0.448	0.334	0.347	1.193
Triangle High	Mode	1.0 (50%)	1.000	0.350	0.350	0.612	1.388
Triangle High Right	Mode	1.0 (25%)	1.286	0.283	0.364	0.903	1.711

Given Mode	Mode	1.000
Given Percentiles	□	15%
	X□	0.723
	1-□	85%
	X1-□	1.277
Calculated Triangle Parameters	L	0.388
	M	1.000
	H	1.612
	Mean	1.000
	Std Dev	0.250
	P(X<Mode)	50.0%



See reference #3 for details

Comparing AFCAA Recommended..

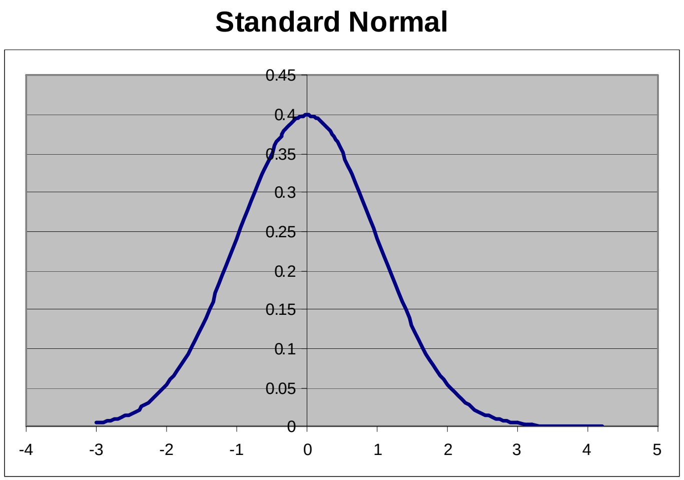


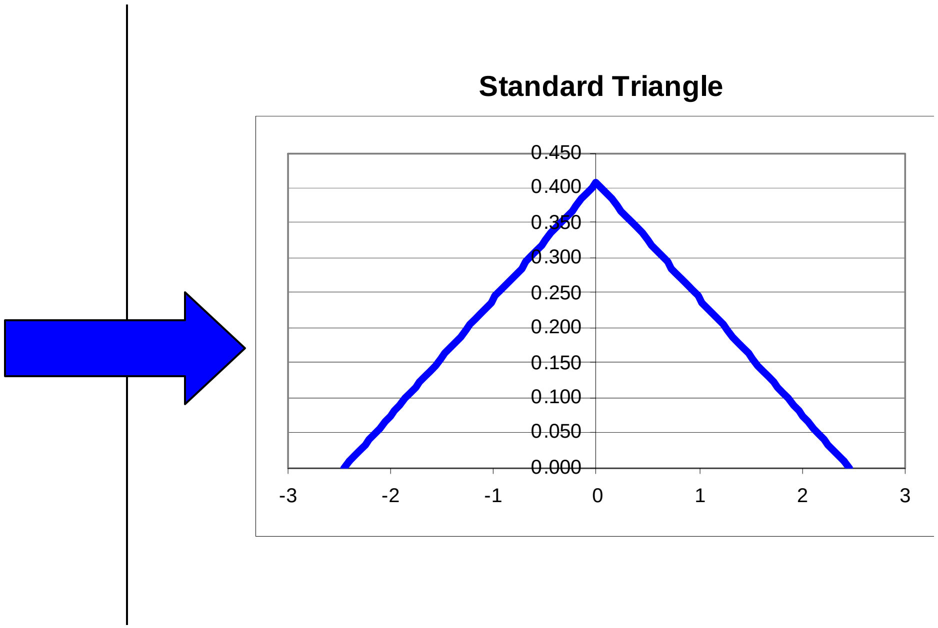
- Triangle is calculated as L and H = Mean + (Multiplier *($+/-\sqrt{6}$)
  - *Multiplier is the sigma number, in this case = .25 (CoV = .25/1 = 25%)*
- Triangle formed is the same as given in AFCAA Table and approximates the Normal N(1, .25)

Summary Comments on Normal

- Using a Multiplier of the assumed sigma value will derive a normal triangular-fitted distribution
- The ratio of the normal triangular-fitted distribution can be used to quickly judge whether a cost-risk triangle created by expert opinion is “normal”, noting that triangle-fitted to normal has base angles = 10%
- No adjustment to the percentile values are necessary for any evaluations within +/- 3 sigma of the mode/mean

What we have so far

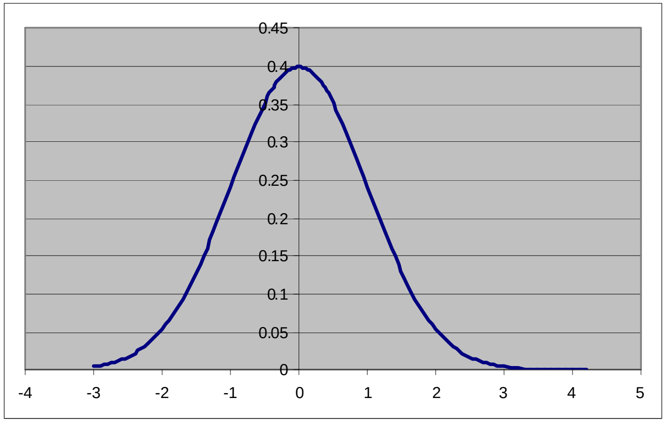


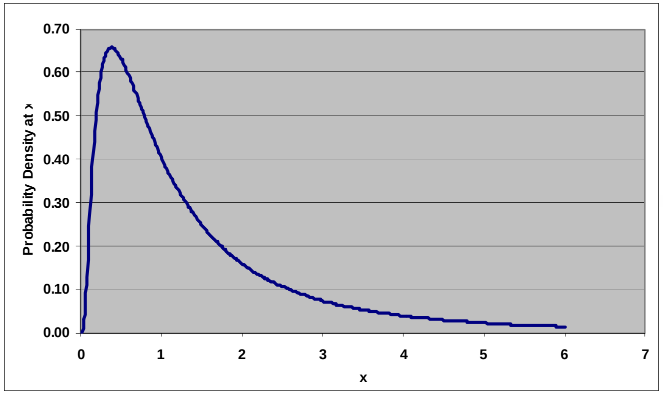


- But this is not typical of eliciting cost-risk input
- On most occasions it is skewed to the right

Non-Standard Distributions

- What if the underlying distribution is not a Normal or symmetrically distributed curve?
- We will use the Lognormal as a second example





Converting Normal to Lognormal

- Using the method of moments conversion* for finding the underlying "Standard" Lognormal based on a Standard Normal N(0,1)
  - *where P is the mean = 0*
  - *and Q is the sigma =1 then:*

- Then Mode is $= e^{P-Q^2} = e^{0-1^2} = 1/e$

- the Median is $= e^{P} = e^{0} = 1$

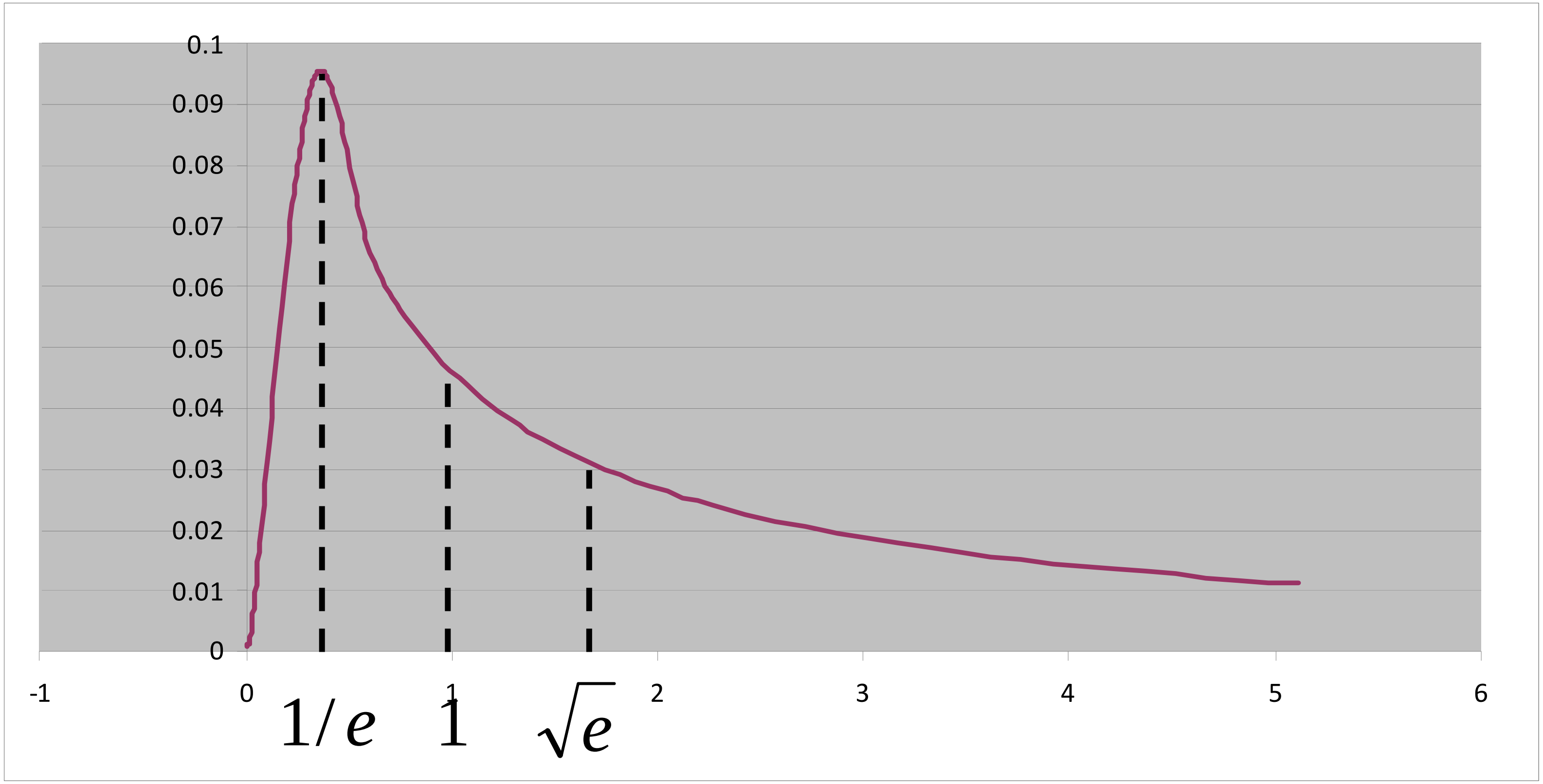
- the Mean is $= e^{P+\frac{1}{2}Q^2} = e^{0+\frac{1}{2}(1)^2} = \sqrt{e}$

- the Sigma is $= e^{P+\frac{1}{2}Q^2}\sqrt{e^{Q^2}-1} = \sqrt{e^2-e}$

See reference #1 for details

 melvin.broder@aero.org

"Standard" Lognormal



- By Inspection it is impossible to fit a straight line through the Mode, Median and Mean points of a lognormal distribution

 melvin.broder@aero.org

Finding the "Standard Lognormal Triangle" (Chart 1 of 2)

- Mathematically given Mode, Median Mean and Sigma then we have four equations with two unknowns, the Low (L) and High (H) values, so that the problems is over-constrained.

- One solution is to assume that the Low value is zero because as the random variable $x \to -\infty, L \to 0$

- Then choice of known parameter equations can be used to solve for the High (H) value

 melvin.broder@aero.org

Finding the Standard Lognormal Triangle (Chart 2 of 2)

- Using formulas for the analytic geometry of a triangle*, we can calculate three answers depending on the choice of parameters from lognormal standard distribution (sigma not used:
  - ***Using Mean and Mode***

$$\frac{L+M+H}{3}=\sqrt{e} \quad \textit{then} \quad H=3\sqrt{e}-\frac{1}{e}$$

  - ***Using Median and Mode***

$$1=H-\sqrt{.5(H-L)(H-M)}=H-\sqrt{.5H\left(H-\frac{1}{e}\right)} \quad \textit{then} \quad H=\frac{-4-\frac{1}{e}\pm\sqrt{8+\frac{8}{e}+\frac{1}{e^2}}}{2}$$

  - ***Using Median and Mean***

$$H=\sqrt{.5(H-L)(H-\{3\sqrt{e}-H-L\})} \quad \textit{then} \quad H=\frac{2}{4-3\sqrt{e}}$$

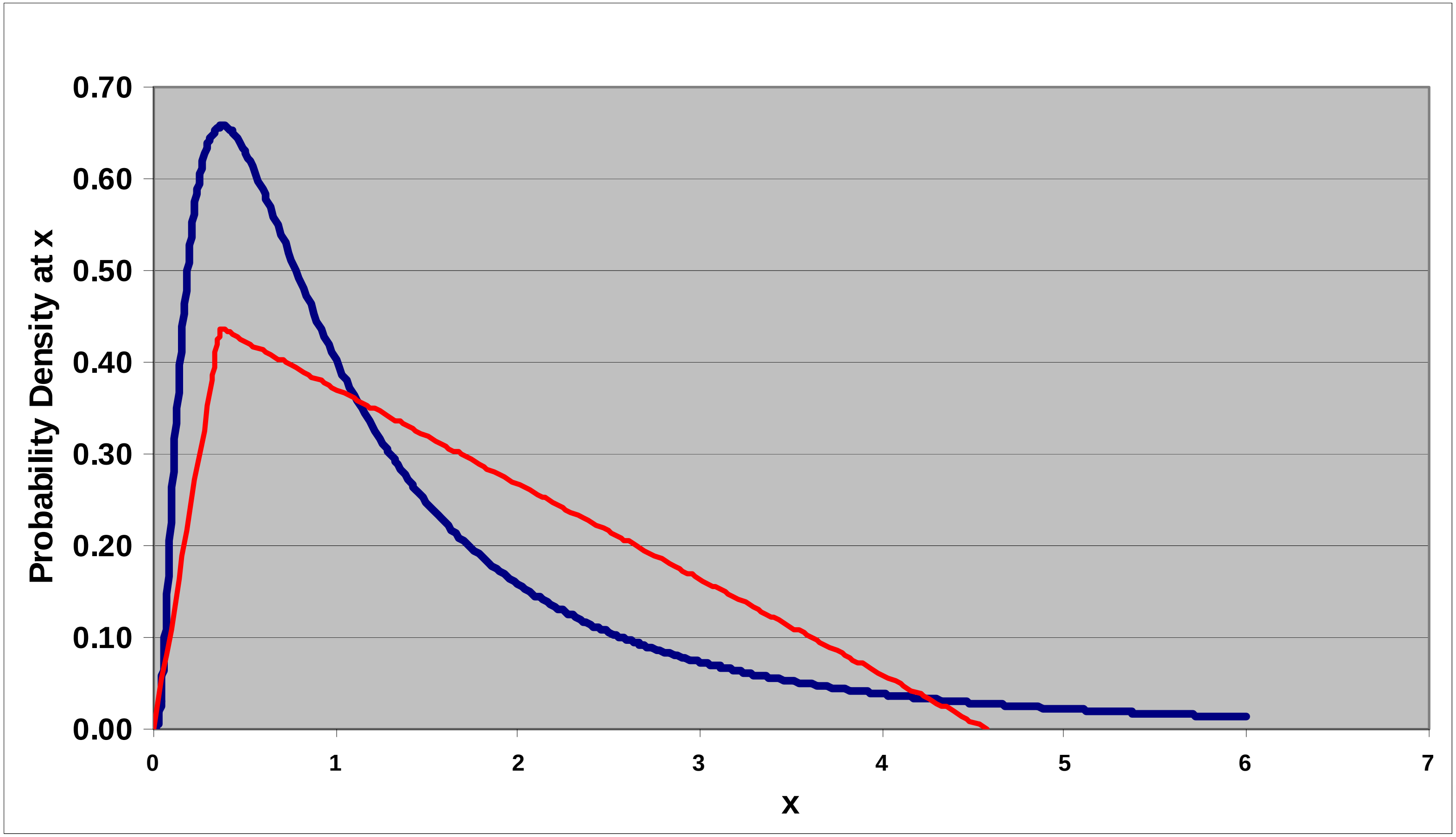
See reference #2 for details

The Three Alternatives Compared Using Numeric Values

	Low	Mode	Median	Mean	High	Sigma
Lognormal	0.00	0.37	1.00	1.65		2.16
Mean & Mode	0.00	0.37	1.47	1.65	4.58	1.04
Median & Mode	0.00	0.37	1.00	-0.05	-0.52	0.18
Median & Mean	0.00	7.06	1.00	1.65	-2.11	1.96

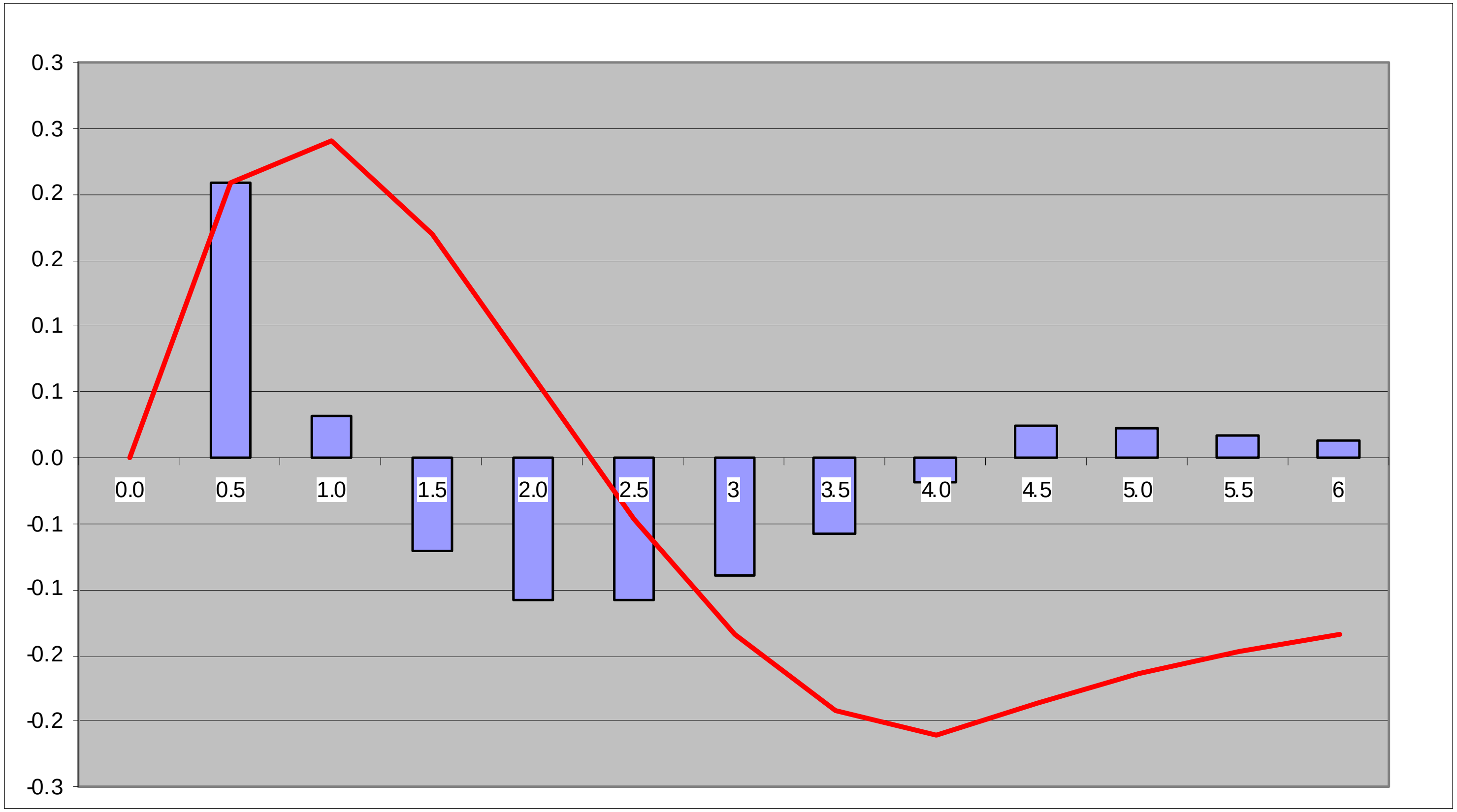
Results for these parameters violate the requirements for triangular distributions

Lognormal and Log-Triangle-Fitted Using Mode and Mean Distributions



- Using the Mode and Mean Triangle-fitted Lognormal

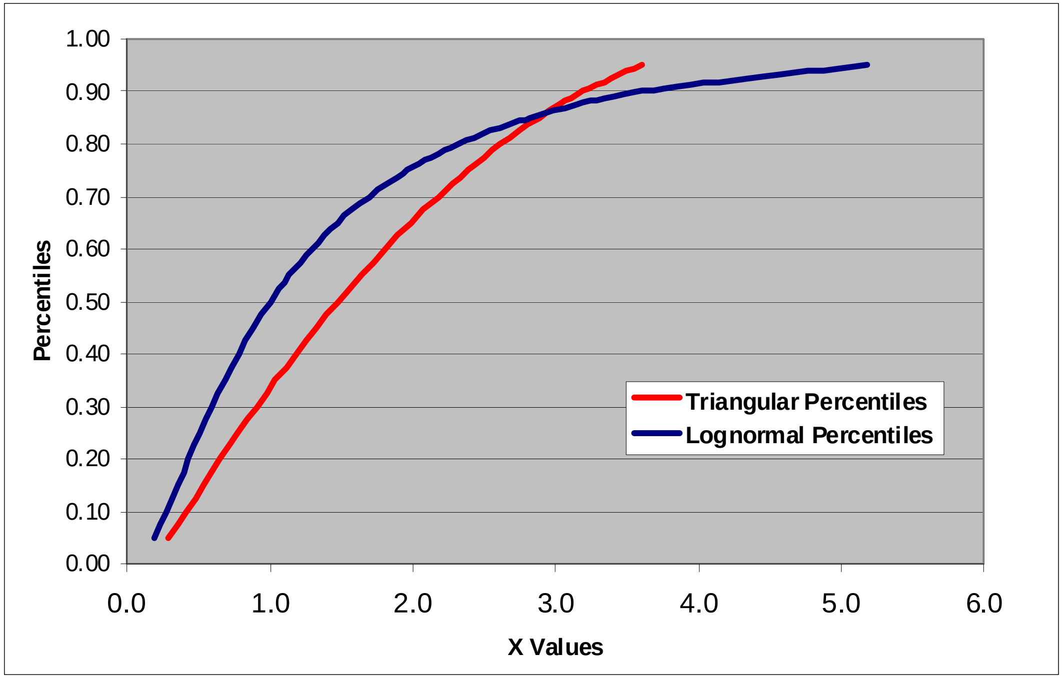
Probability Differences
Lognormal Minus Triangle-Fitted Distributions



Difference in cumulative (red) probability is close only near +1.5 sigma value

 melvin.broder@aero.org

Compensating for Percentile Differences is not in the Tails



Percentiles	Triangle	LogNormal	Delta
5th	0.3	0.2	-0.10
10th	0.4	0.3	-0.14
15th	0.5	0.4	-0.18
20th	0.7	0.4	-0.22
25th	0.8	0.5	-0.27
30th	0.9	0.6	-0.31
35th	1.0	0.7	-0.36
40th	1.2	0.8	-0.40
45th	1.3	0.9	-0.44
50th	1.5	1.0	-0.47
55th	1.6	1.1	-0.50
60th	1.8	1.3	-0.51
65th	2.0	1.5	-0.51
70th	2.2	1.7	-0.48
75th	2.4	2.0	-0.42
80th	2.6	2.3	-0.29
85th	2.9	2.8	-0.06
90th	3.2	3.6	0.41
95th	3.6	5.2	1.58

- The lognormal median value of 1.0 is only the 35 percentile on the triangle
- Although the means of both distributions are the same
- Note this assumes both start with random value, X, at zero

Comparing to SSCAG/AFCAA Ratios for Highly Skewed Triangles

Table 6-4 Risk Levels Determined by Average Risk Factors and Distribution Symmetry

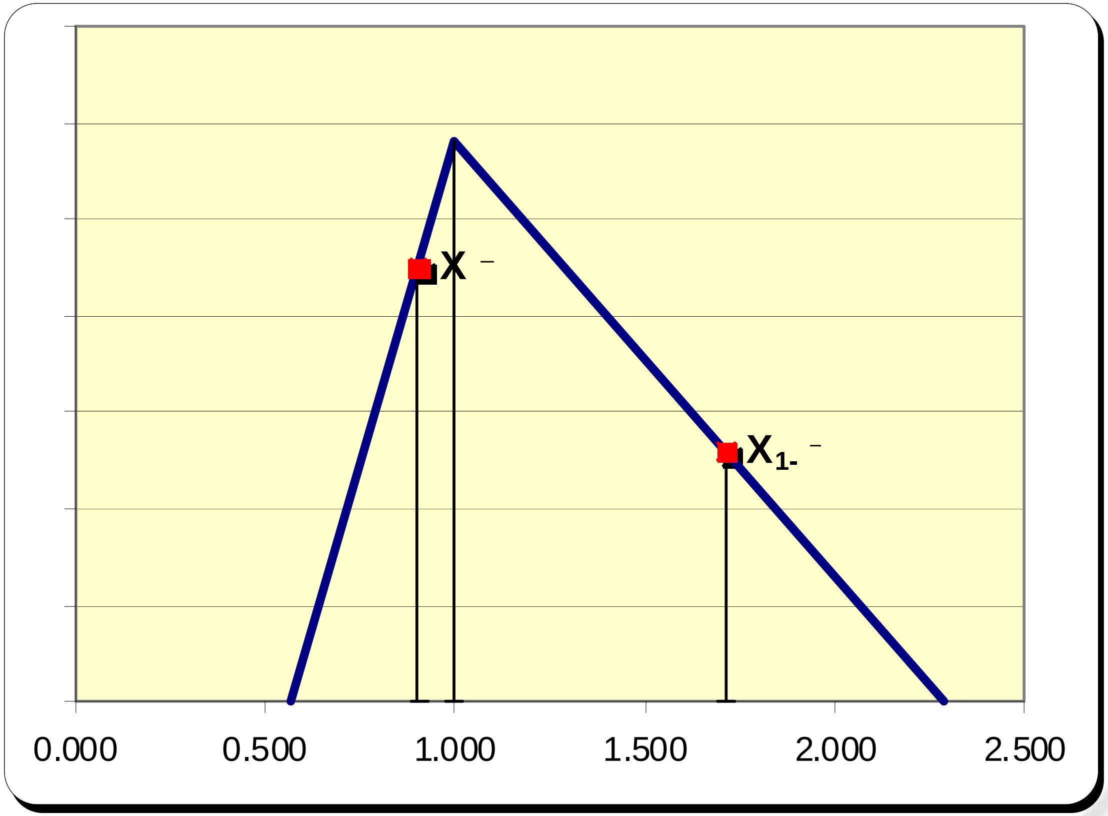
Average Probability Risk Factor Value (Pr)	Risk Levels	Cost Risk Factors					
		Skewed Left SL		Symmetric		Skewed Right SR	
		Low	High	Low	High	Low	High
1.0 < Pr ≤ 1.2	VL	0.96	1.02	0.97	1.03	0.98	1.12
1.2 < Pr ≤ 2.0	L	0.93	1.03	0.95	1.05	0.97	1.21
2.0 < Pr ≤ 2.5	ML	0.90	1.04	0.93	1.07	0.96	1.30
2.5 < Pr ≤ 3.5	M	0.85	1.05	0.90	1.10	0.95	1.45
3.5 < Pr ≤ 4.0	MH	0.80	1.10	0.85	1.15	0.90	1.60
4.0 < Pr ≤ 4.8	H	.070	1.10	0.80	1.20	0.90	1.90
4.8 < Pr ≤ 5.0	VH	0.50	1.10	0.70	1.30	0.90	2.50

Table 8 Default Bounds (1 of 2) For Subjective Distributions

Distribution	Point Estimate Interpretation	Point Estimate and Probability	Mean	CV based on mean	CV Based on PE	15%	85%
Lognormal Low	Median	1.0 (50%)	1.011	0.151	0.153	0.856	1.168
Lognormal Med	Median	1.0 (50%)	1.032	0.254	0.262	0.772	1.296
Lognormal High	Median	1.0 (50%)	1.063	0.361	0.384	0.696	1.437
Normal Low	Mean	1.0 (50%)	1.000	0.150	0.150	0.845	1.155
Normal Med	Mean	1.0 (50%)	1.000	0.250	0.250	0.741	1.259
Normal High	Mean	1.0 (50%)	1.002	0.346	0.347	0.640	1.363
Weibull Low	Mode	1.0 (25%)	1.158	0.179	0.208	0.956	1.370
Weibull Med	Mode	1.0 (20%)	1.393	0.332	0.463	0.956	1.855
Weibull High	Mode	1.0 (15%)	2.104	0.572	1.204	1.000	3.277
Triangle Low Left	Mode	1.0 (75%)	0.878	0.178	0.156	0.695	1.041
Triangle Low	Mode	1.0 (50%)	1.000	0.150	0.150	0.834	1.166
Triangle Low Right	Mode	1.0 (25%)	1.123	0.139	0.156	0.959	1.305
Triangle Med Left	Mode	1.0 (75%)	0.796	0.327	0.260	0.492	1.069
Triangle Med	Mode	1.0 (50%)	1.000	0.250	0.250	0.723	1.277
Triangle Med Right	Mode	1.0 (25%)	1.204	0.216	0.260	0.931	1.508
Triangle High Left	Mode	1.0 (74%)	0.745	0.448	0.334	0.347	1.103
Triangle High	Mode	1.0 (50%)	1.000	0.350	0.350	0.612	1.388
Triangle High Right	Mode	1.0 (25%)	1.286	0.283	0.364	0.903	1.711

Converting AFCAA Percentile Input to Low and High Values

Given Mode	Mode	1.000
Given Percentiles	□	15%
	X□	0.903
	1-□	85%
	X1-□	1.711
Calculated Triangle Parameters	L	0.571
	M	1.000
	H	2.286
	Mean	1.286
	Std Dev	0.365
	P(X<Mode)	25.0%



Comparison of Low to High Ratio

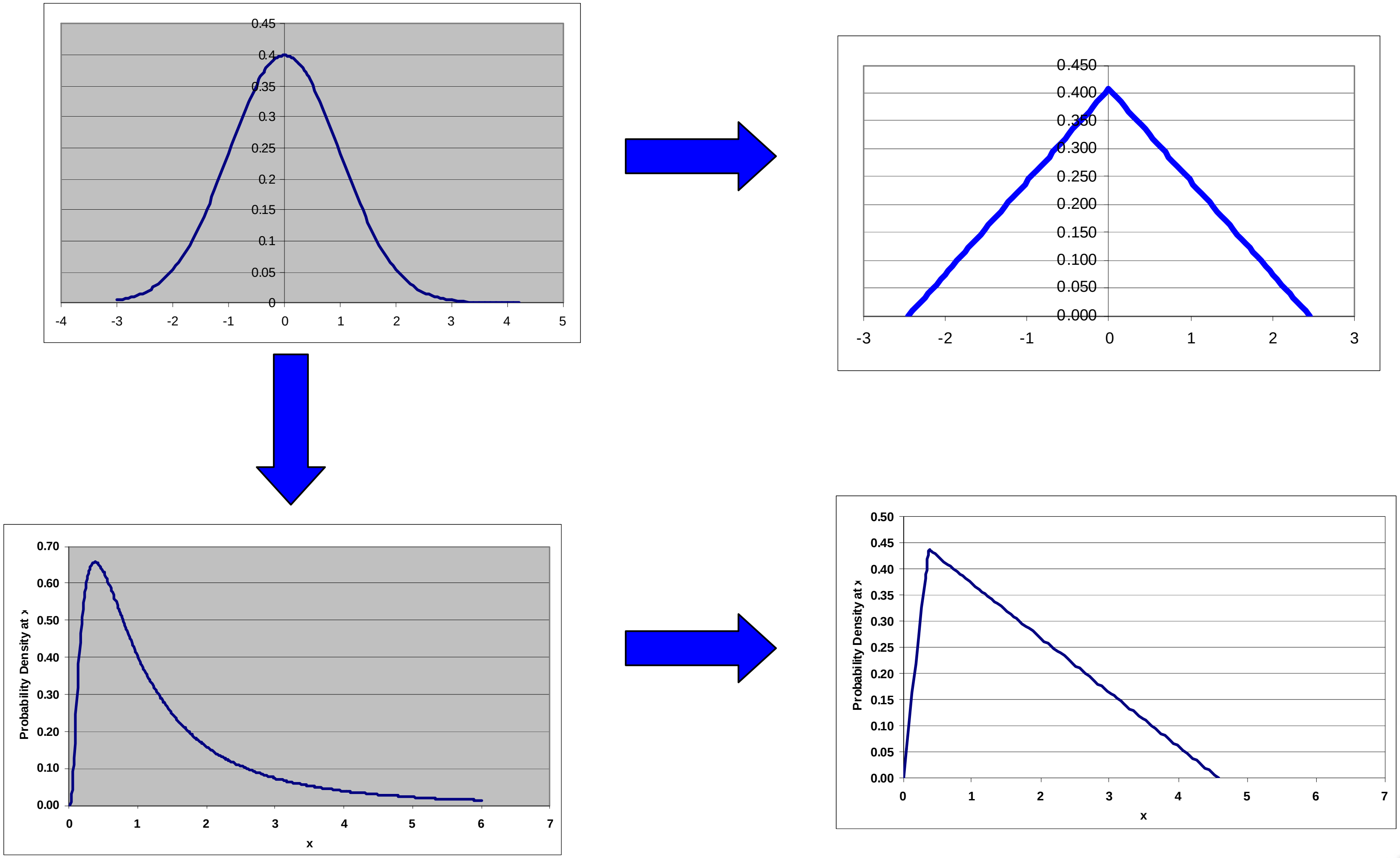
	Low	Mode	High	Ratio
Lognormal	0.00	0.37	6.00	**15.22**
Log-fitted	0.00	0.37	4.58	**11.37**
AFCAA	0.57	1.00	2.29	**3.00**
SSCAG	0.90	1.00	2.50	**15.00**

- Ratio is calculated as (High – Mode) /(Mode - Low)
- Comparison to Lognormal not possible as High is infinite, assumed value of 6 for comparison

Uses of Standard Lognormal Triangle

- The Ratio of High to Low (H / L) can be used as a easy check to see how close a risk input is to a lognormal distribution
- Given a most likely cost (i.e. the mode or M value) the analyst only needs to obtain a High (H) value and assume the distribution is lognormal to build a triangular risk input. In other words the Low (L) can be calculated using the ratio if there is no justification for determining a cost-opportunity
- Conversions
  - *If a lognormal mean and sigma are provided but analyst wants to use triangular distributions then these can be converted into a triangle*
  - *If triangular data is given but user wants to use a lognormal to ensure the percentiles near the median are accurately calculated*

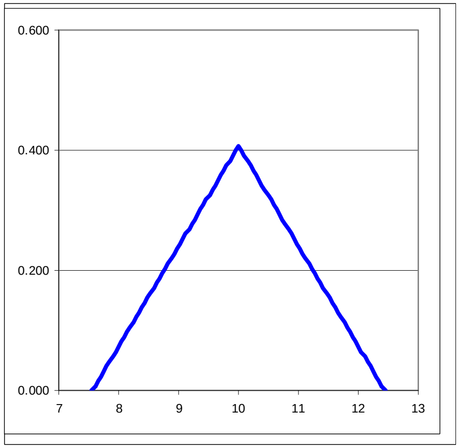
What Have We Modeled

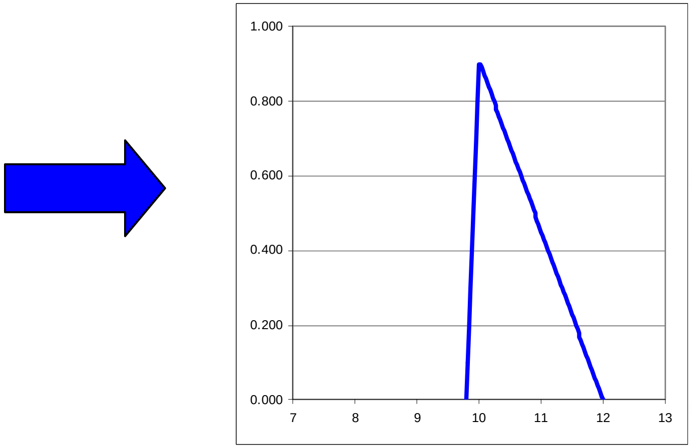


Application Example -1

The Expert's Credibility Scenario

- Your expert tells you:
  - *"I was told to fit my estimate into a $10 M budget, which is reasonable"*
  - *"I think I can meet this cost goal plus or minus $2 M"*
- You want to keep as much expert information as possible but:
  - *Experience with costs on other programs and with expert opinion leads you to believe that the risks are not symmetrical*
- You assume the high value is correct and then use the triangle-fitted to the lognormal ratio to generate the low value.

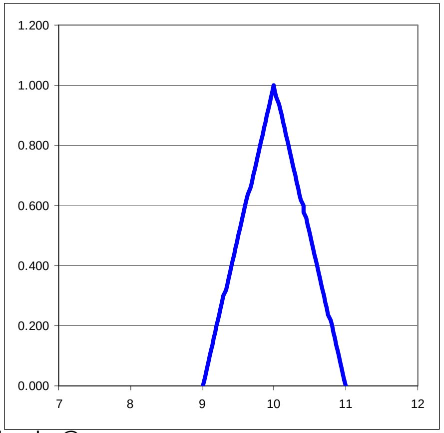


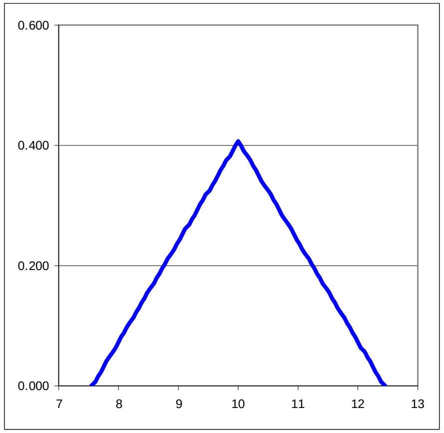


Application Example -2

The Expert's Understanding of True Cost Risk Scenario

- Your expert tells you:
  - *"I did a $10 M budget estimate based on a lot of similar jobs I have worked"*
  - *"I am confident I can meet my estimate plus or minus $1 M"*
- You want to keep as much expert information as possible but:
  - *Experience with costs on other programs and with expert opinion leads you to believe that there is much more uncertainty about the costs*
- You assume the expert's L and H values (CoV of 4%) are one sigma values equal to a higher CoV of 10% (about same as L/H = 15%/85% percentiles)
- You can use a triangle-fitted normal where multiplier is 10% of $10M:
  - *calculated as L and H = Mean + (Multiplier * (* $+/-\sqrt{6}$ *)*





Summary

- Other Methods of Modeling
  - *Use of low and high risks as percentiles (eg. 10% / 90%) for the triangle input values and convert to the Low and High end points of the risk triangle using references #1 and #2*
  - *Use the mean and sigma of the triangle transferred directly to a Normal, Lognormal (or any other) distribution for use in a Monte Carlo sampling analysis using reference*

- Key Take-Away
  - *Normal triangle-fitted value for Low and High* = $+/-\sqrt{6}$
  - *Lognormal triangle-fitted ratio* = *11.4*
  - *Don't worry about the tails but a triangle-fitted to the Lognormal could have significant differences at the median*

References

1. "FRISK - Formal Risk Assessment of System Cost Estimates," AIAA 1992 Aerospace Design Conference, P. H. Young, Dated February 1992

2. "Issues in Specifying a Triangular Cost Distribution", Stephen A. Book, ISPA Conference 2002,The International Society of Parametric Analysts, Dated May 2002

3. "On the Calculation of the Parameters of a Triangular Distribution Using Percentiles", Timothy P. Anderson, Dated August 2008

4. "U.S. Air Force Cost Analysis Agency Cost Risk and Uncertainty Analysis Handbook", Alfred Smith, Jeff McDowell, Dr. Lew Fichter, Table 2-5, Dated April 2007

5. "Space Systems Cost Risk Handbook", Edited by Timothy P. Anderson and Raymond P. Covert, Table 6-4, Dated November 2005

AEROSPACE
Assuring Space Mission Success

For More Information on Cost Risk Models and Handbooks

Melvin Broder
Developmental Planning and Projects
melvin.broder@aero.org
or
melvin.broder@losangeles.af.mil

310.336.2567