Probability Distributions for Risk Analysis

Cost Estimating Advanced (CEA) 07

ICEAA Annual Conference

Peter J. Braxton, Technomics, Inc.

© 2002-2013 ICEAA. All rights reserved.

What This Session Is Not

- What probability distributions are
  - CEB 08 Probability and Statistics Basic
    - Thursday, June 11th, 1430-1615, Seabreeze 1&2
- How to use probability distributions in a simulation
  - CEA 06 Monte Carlo Simulations
    - Thursday, June 11th, 0930-115, Executive Center 2AB
- Distributions for CER risk
  - PT 06 CER Risk and S-Curves
    - Thursday, June 11th, 1215-1400, Executive Center 3AB
- Fitting distributions to data
  - CEA 06 Probability and Statistics Advanced
    - Friday, June 12th, 0900-1030, Marina 6

© 2002-2013 ICEAA. All rights reserved.

What This Session Is Not

- How to do Inputs Risk
  - CEB 07 Risk Analysis Basic
    - Thursday, June 11th, 1215-1400, Seabreeze 1&2
- How to analyze historical program risk data
  - CEA 05 Risk Analysis Advanced
    - Thursday, June 11th, 1530-1815, Spinnaker 1&2
- Joint probability distributions
  - INT 04 Joint Cost and Schedule Risk
    - Thursday, June 11th, 0930-1115, Executive Center 1
- Interaction of risk distributions and contract types
  - INT 05 Contracts Risk
    - Thursday, June 11th, 1215-1400, Executive Center 1

© 2002-2013 ICEAA. All rights reserved.

Outline

- Probability Prerequisites
  - Algebra and Calculus
- Types of Risk Distributions
- Common Distributions
  - Normal, Lognormal, Triangular
- Risk Distribution Applications
- Backup
  - Algebra Refresher
  - Distribution Quad Charts (for reference)

© 2002-2013 ICEAA. All rights reserved.

Calculus for Probability

- Calculus was invented (Newton, Leibniz, et al.) to support **mechanics** (branch of physics)
  - In the modern world, and certainly for cost estimators, **probability** is the primary motivation for the use of calculus
- Differential calculus deals with slopes and rates of change
  - Mechanics: velocity (position), acceleration (velocity)
  - Probability: pdf (cdf), mode
- Integral calculus deals with areas under curves and cumulative values
  - Mechanics: position (velocity), velocity (acceleration)
  - Probability: cdf (pdf), mean, variance, median, percentiles
- Derivatives and integrals are generally inverse operations
  - Fundamental Theorem of Calculus!
- Following slides offer a brief refresher of a few key results needed for probability derivations

© 2002-2013 ICEAA. All rights reserved.

NEW!

Calculus Refresher - Derivatives

- Product Rule
  - Derivative of a product
  - "Derivative of the first, times the second, plus the first times the derivative of the second"

$$\frac{d}{dx}[f(x)\cdot g(x)]=\left[\frac{d}{dx}f(x)\right]\cdot g(x)+f(x)\cdot\left[\frac{d}{dx}g(x)\right]=f'(x)\cdot g(x)+f(x)\cdot g'(x)$$

- Chain Rule
  - Derivative of a compound function
  - "Derivative of the outside, times derivative of the inside"

$$y=g(x)$$

$$\frac{d}{dx}f(g(x))=\frac{d}{dx}f(y)=\frac{d}{dy}f(y)\cdot\frac{dy}{dx}=f'(y)\cdot\left[\frac{d}{dx}g(x)\right]=f'(g(x))\cdot g'(x)$$

© 2002-2013 ICEAA. All rights reserved.

NEW!

Calculus Refresher - Integrals

- Substitution
  - Technique for evaluating (definite) integrals
  - Replace functions, differentials, and limits consistently

$$y = g(x) \qquad \frac{dy}{dx} = \frac{d}{dx} g(x) = g'(x) \Rightarrow dy = g'(x)dx$$

  - Essentially chain rule in reverse

$$\int_a^b f(g(x)) \cdot g'(x)dx = \int_{g(a)}^{g(b)} f(y)dy = F(y)\Big|_{g(a)}^{g(b)} = F(g(b)) - F(g(a))$$

© 2002-2013 ICEAA. All rights reserved.

Calculus Refresher - Series

- Taylor Series
  - Representation of a function as an infinite series (sum)
  - Centered around a point of interest
  - Converges within a certain radius
  - Handy for computation, including approximations (finite number of terms)
  - In cost estimating, tends to be used for:
    - Quantities near 0, e.g., Coefficient of Variation (CV); or
    - Quantities near 1, e.g., Learning Curve Slope (LCS)
- Some handy Taylor Series
  - Log ($0 \le x < 1$)
  - Square root ($-1 \le x \le 1$)
  - Reciprocals ($-1 < x < 1$)

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots \approx -x$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots \approx x$$

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots \approx 1 + x$$

$$\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \ldots \approx 1 + \frac{x}{2}$$

© 2002-2013 ICEAA. All rights reserved.

Probability Notation

- Roman letters are generally used to represent random variables or their associated distribution functions
  - X, Y, Z for variables, x, y, z for realization of those variables
  - F, G for cumulative distribution functions (cdfs)
    - f, g for corresponding probability density functions (pdfs)
    - Subscripts for random variable in case of ambiguity
- Lowercase Greek letters are generally used to represent parameters of probability distributions
  - μ (mu) = mean, a location parameter
  - σ (sigma) = standard deviation, a scale parameter
  - α (alpha), β (beta), θ (theta), and λ (lambda) are often shape, scale, or rate parameters
- Capital Greek letters are generally used to represent particular special functions or mathematical operations
  - Γ (Gamma) = gamma function, Φ (Phi) = standard normal cdf

© 2002-2013 ICEAA. All rights reserved.

Program Risk Distributions

1. Input distributions
   - Characterize uncertainty of inputs to the cost estimating process, such as cost-driver parameters (weight, power, SLOC, etc.)
   - Commonly include **Normal**, **Lognormal**, and **Triangular**
2. Intermediate output distributions
   - Characterize uncertainty about estimates for individual cost elements
     - Prediction interval (PI) associated with a cost-estimating relationship (CER)
       - Commonly include **t** or **log t**
     - Risk ranges provided by a subject matter expert (SME)
       - Commonly include **Triangular**

Taking the Next Step: Turning OLS CER-Based Estimates into Risk Distributions. C.M. Kanick, E.R. Druker, R.L. Coleman, M.M. Cain, P.J. Braxton, SCEA 2008.

3. Final output distributions
   - Characterize the uncertainty of an overall cost estimate
   - Commonly include **Normal** and **Lognormal**

Normality of Work Breakdown Structures, M. Dameron, J. Summerville, R. Coleman, N.St. Louis, Joint ISPA/SCEA Conference, June 2001.

© 2002-2013 ICEAA. All rights reserved.

Portfolio Risk Distributions

4. Cross-program risk distributions
   - Characterize range of cost growth factors (CGFs) associated with historical programs
   - Commonly include **Lognormal**, **Triangular**, or other skew-right distributions (including heavy-tailed distributions)

- A classic problem in risk analysis is how to make inferences about #3 from #4

© 2002-2013 ICEAA. All rights reserved.

NEW!

Befriending Your Distributions

- Normal Demonstration
- Triangular Derivations
- Lognormal Derivations and Demonstration

© 2002-2013 ICEAA. All rights reserved.

Befriending Your Distributions

- Probability distributions are to risk analysts what words are to poets
- Look at distributions in multiple ways:
  - **G**raphical – PDF and CDF graphs
  - **N**umerical – Excel functions
  - **A**lgebraic – formulae, parameters
- Build a "toy problem" and play with it
- Understand properties of distributions, when they "arise," how they are related to other distributions

© 2002-2013 ICEAA. All rights reserved.

NEW!

Excel 2010 Distributions

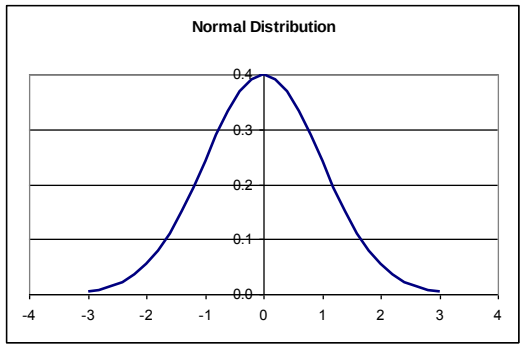
- New and improved format for statistical functions in Excel
  - More consistent syntax
  - Excel 2007 versions maintained for backwards compatibility
- Same set of distributions
  - Continuous: BETA, CHISQ, EXPON, F, GAMMA, GAMMALN, **LOGNORM**, **NORM**, T, WEIBULL
  - Discrete: BINOM, HYPGEOM, NEGBINOM, POISSON
- Suffixes denote different variants
  - .DIST = cdf (and sometimes pdf)
  - .S = standard (normal only)
  - .INV = inverse cdf
  - .2T = two-tail
  - .RT = right tail (left tail is default)
- Examples:
  - =NORM.S.INV(RAND()) will generate a standard normal
  - =T.INV.2T(0.05,30) will give the (positive) critical value for a t-test at alpha = 0.05, 30 degrees of freedom

© 2002-2013 ICEAA. All rights reserved.

NEW!

Normal Distribution Overview

- Distribution



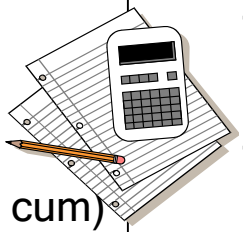
$x \in (\infty, -\infty)$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$F(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

- Parameters and Statistics
  - Mean = μ
  - Variance = σ^2
  - Skewness = 0
  - 2.5th percentile = μ - 1.96σ
  - 97.5th percentile = μ + 1.96σ

- Key Facts
  - If X ~ N(μ, σ^2), then $\frac{(X-\mu)}{\sigma}$ ~ N(0,1) ("standard" normal)
  - Central Limit Theorem holds for n ≥ 30
  - 68.3/95.5/99.7 Rule
  - Limiting case of t distribution
  - Exponential of normal is lognormal
  - Dist: NORMDIST(x, mean, stddev, cum)
    - cum = TRUE for cdf and FALSE for pdf
  - Inv cdf: NORMINV(prob, mean, stddev)

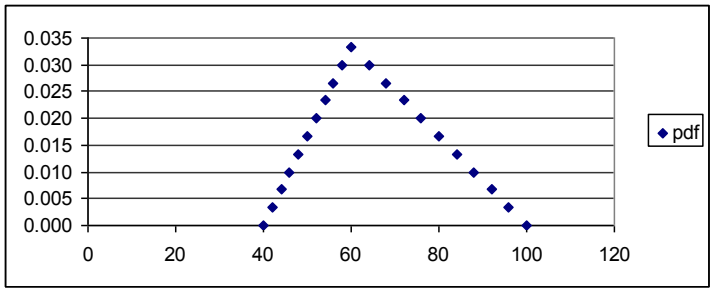


- Applications
  - Central Limit Theorem
    - Approximation of distributions
  - Regression Analysis
    - Assumed error term
  - Distribution of cost
    - Default distribution
  - Distribution of risk
    - Symmetric risks and uncertainties

NEW!
© 2002-2013 ICEAA. All rights reserved.

Triangular Distribution Overview

- Distribution



$$p(x)=\begin{cases}2(x-a)/[(b-a)(c-a)] & a\le x\le c\\ 2(b-x)/[(b-a)(b-c)] & c\le x\le b\end{cases}$$

$$F(x)=\begin{cases}\dfrac{(x-a)^2}{(b-a)(c-a)} & a\le x\le c\\ 1-\dfrac{(b-x)^2}{(b-a)(b-c)} & c\le x\le b\end{cases}$$

- Parameters and Statistics
  - Min = a
  - Max = b
  - Mode = c ($a \le c \le b$)
  - Mean = $\dfrac{a+b+c}{3}$
  - Variance = $\dfrac{a^2+b^2+c^2-ab-ac-bc}{18}$

19

- Key Facts
  - Excel
    - pdf and cdf calculations can be handled as they are listed in the above formulas
  - A symmetrical triangle approximates a normal when

$$a=\mu-\sqrt{6}\sigma, c=\mu, b=\mu+\sqrt{6}\sigma$$

- Applications
  - Risk Analysis
    - SME Input

© 2002-2013 ICEAA. All rights reserved.

Triangular Distribution – PDF and Mean

- For Triangle(L,M,H) , denote L=a, H=b, M=c by T(a,c,b)
- Since the area of the triangle must be 1 (100%), the height is twice the reciprocal of the base
  - We can then derive the PDF by using similar triangles

$$p(x)=\begin{cases}\frac{2}{b-a}\frac{x-a}{c-a} & a\le x\le c\\ \frac{2}{b-a}\frac{b-x}{b-c} & c\le x\le b\end{cases}$$

$$\mu=E[X]=\int_a^b xp(x)dx=\int_a^c \frac{2x}{b-a}\frac{x-a}{c-a}dx+\int_c^b \frac{2x}{b-a}\frac{b-x}{b-c}dx$$

$$=\frac{1}{b-a}\left[\frac{\frac{2}{3}x^3-x^2a}{c-a}\Bigg|_a^c+\frac{x^2b-\frac{2}{3}x^3}{b-c}\Bigg|_c^b\right]=\frac{1}{b-a}\left[\frac{2}{3}c^2+\frac{2}{3}ac+\frac{2}{3}a^2-ac-a^2+b^2+bc-\frac{2}{3}b^2-\frac{2}{3}bc-\frac{2}{3}c^2\right]$$

$$=\frac{1}{b-a}\left[\frac{bc-ac}{3}+\frac{b^2-a^2}{3}\right]=\frac{a+b+c}{3}$$

"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

© 2002-2013 ICEAA. All rights reserved.

Triangular Distribution – Variance

$$\sigma^2 = E\left((X-\mu)^2\right) = E\left(X^2\right) - \mu^2$$

$$E\left(X^2\right) = \int_a^b x^2 p(x)dx = \int_a^c \frac{2x^2}{b-a}\frac{x-a}{c-a}dx + \int_c^b \frac{2x^2}{b-a}\frac{b-x}{b-c}dx = \frac{1}{b-a}\left[\left.\frac{\frac{1}{2}x^4 - \frac{2}{3}x^3 a}{c-a}\right|_a^c + \left.\frac{\frac{2}{3}x^3 b - \frac{1}{2}x^4}{b-c}\right|_c^b\right]$$

$$= \frac{1}{b-a}\left[\frac{1}{2}\left(c^3+ac^2+a^2c+a^3\right) - \frac{2}{3}\left(c^2a+a^2c+a^3\right) + \frac{2}{3}\left(b^3+b^2c+bc^2\right) - \frac{1}{2}\left(b^3+b^2c+bc^2+c^3\right)\right]$$

$$= \frac{2}{3}\left(c^2+bc+ac+b^2+ab+a^2\right) - \frac{1}{2}\left(c^2+bc+ac+b^2+ab+a^2\right) = \frac{a^2+b^2+c^2+ab+ac+bc}{6}$$

$$\mu^2 = \left(\frac{a+b+c}{3}\right)^2 = \frac{a^2+b^2+c^2+2ab+2ac+2bc}{9}$$

$$E\left(X^2\right) - \mu^2 = \frac{3a^2+3b^2+3c^2+3ab+3ac+3bc}{18} - \frac{2a^2+2b^2+2c^2+4ab+4ac+4bc}{18}$$

$$= \frac{a^2+b^2+c^2-ab-ac-bc}{18} = \frac{b^2-2ab+a^2+c^2+ab-ac-bc}{18} = \frac{(b-a)^2-(c-a)(b-c)}{18}$$

$$= \frac{c^2-2ac+a^2+b^2-2bc+c^2+bc-ab-c^2+ac}{18} = \frac{(c-a)^2+(b-c)^2+(c-a)(b-c)}{18}$$

Square of the base minus product of the half-bases!

Sum of squares of half-bases and product of half-bases!

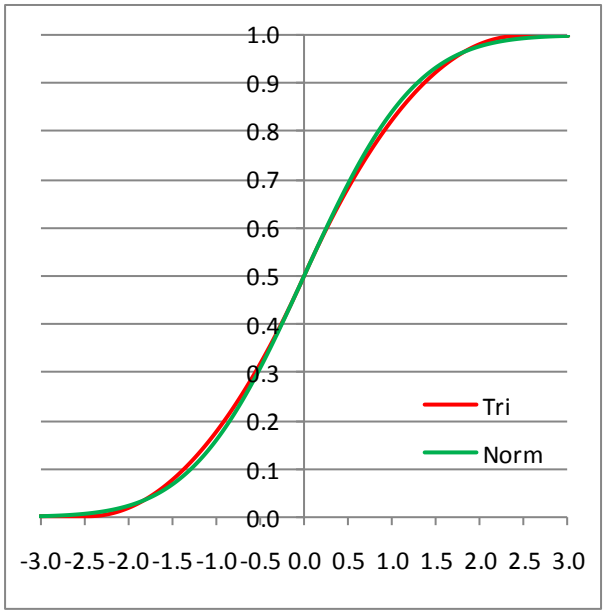
"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

NEW!

© 2002-2013 ICEAA. All rights reserved.

Substituting a Triangular for a Normal: The √6 Factor

- For a symmetric triangle, let ML = m, L = m-w, H = m+w, where w is the half-base
  - Then the mean is m, and the variance is $w^2/6$
- It follows that the half-base is greater than the standard deviation by a factor of √6
- To approximate a normal, N(μ, σ) the factor of √6 is multiplied by the standard deviation of the normal to be emulated to produce the half-base
  - By this means, end points are found that will produce a triangular distribution that emulates the underlying normal in mean and standard deviation
  - This triangular distribution, Tri(μ -√6σ, μ, μ +√6σ) differs from the underlying normal in all other moments, and at all percentiles other than the median and two "cross-over" points, but the difference is minor

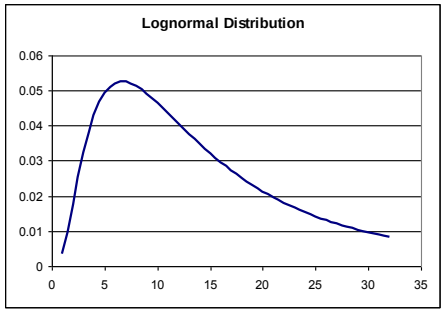


"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

© 2002-2013 ICEAA. All rights reserved.

Lognormal Distribution Overview

- Distribution



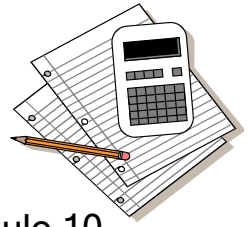
$x \in [0, \infty)$

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-[\ln(x/\mu)]^2}{2\sigma^2}\right)$$

- Parameters and Statistics
  - Median = e^{μ}
  - Std Deviation of ln X = σ
  - Mean = $e^{\mu + \sigma^2/2}$
  - Variance = $e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$

- Key Facts
  - If X has a lognormal distribution, then ln(X) has a normal distribution
  - For small standard deviations, the normal approximates the lognormal distribution
    - For CVs < 25%, this holds
  - Excel
    - Cdf = LOGNORMDIST(x, mean, stddev)
    - Inv cdf = LOGINV(prob, mean, stddev)

- Applications
  - Risk Analysis



© 2002-2013 ICEAA. All rights reserved.

Derivation of Lognormal pdf

- Relationship of Lognormal (X) and corresponding Normal (Y)

$$X = e^{Y} \sim LogN(\mu, \sigma) \Leftrightarrow Y = \ln X \sim N(\mu, \sigma)$$

Both parametrized in terms of mean and standard deviation of the *related normal*

- Lognormal cdf in terms of related Normal cdf

$$F_X(x) = P(X \le x) = P(Y \le \ln x) = F_Y(\ln x)$$

- Lognormal pdf is derivative of cdf

$$f_X(x) = F_X'(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} F_Y(\ln x)$$

  - Apply chain rule

$$= \frac{d}{dy} F_Y(\ln x) \cdot \frac{d}{dx} \ln x = \frac{1}{x} f_Y(\ln x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

© 2002-2013 ICEAA. All rights reserved.

Derivation of Lognormal Median

- Definition of median

$$\int_0^m f_X(x)dx = \frac{1}{2}$$

- Lognormal pdf (see previous slide)
- Substitution

$$\int_0^m \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} dx = \frac{1}{2}$$

$$\int_{-\infty}^{\ln m} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = \frac{1}{2}$$

$$y = \ln x \Rightarrow dy = \frac{1}{x}dx$$

- This is just the pdf of the related normal!

$$= 0.5 \Leftrightarrow \ln m = \mu \Leftrightarrow m = e^{\mu}$$

- Median is preserved by transformation

© 2002-2013 ICEAA. All rights reserved.

NEW!

Derivation of Lognormal Mode

- Definition of mode, apply product rule and chain rule!

$$f_X'(x) = \frac{d}{dx}\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}\left(-\frac{1}{x^2}-\frac{\ln x-\mu}{\sigma^2 x^2}\right)=0$$

- First term is strictly positive, get common denominator

$$\frac{\sigma^2+\ln x-\mu}{\sigma^2 x^2}=0 \Rightarrow \ln x=\mu-\sigma^2 \Rightarrow x=e^{\mu-\sigma^2}=\frac{e^{\mu}}{1+CV^2}\approx e^{\mu}\left(1-CV^2\right)$$

Note mode shift (left)

See later CV derivation

- pdf is concave down at peak
  - We'll spare you the derivative!

© 2002-2013 ICEAA. All rights reserved.

NEW!

Derivation of Lognormal Mean

- Definition of expected value, lognormal pdf (see earlier slide), substitution

$$x = e^y \Rightarrow dx = e^y dy$$

$$E(X) = \int_0^\infty x f_X(x) dx = \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} dx = \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} e^y dy$$

- Completing the square!

$$= \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left(y^2 - 2\mu y + \mu^2\right) + 2\sigma^2 y}{2\sigma^2}} dy = \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left(y - \left(\mu + \sigma^2\right)\right)^2 + 2\mu\sigma^2 + \sigma^4}{2\sigma^2}} dy$$

- Normal pdf, area under the curve = 1!

$$= e^{\mu + \frac{\sigma^2}{2}} \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left(y - \left(\mu + \sigma^2\right)\right)^2}{2\sigma^2}} dy = e^{\mu + \frac{\sigma^2}{2}} = e^\mu \sqrt{1 + CV^2} \approx e^\mu \left(1 + \frac{CV^2}{2}\right)$$

Note mean shift (right)

See later CV derivation

© 2002-2013 ICEAA. All rights reserved.

Derivation of Lognormal Variance

- Same approach, substitution $x = e^y \Rightarrow dx = e^y dy$

$$E\left(X^2\right) = \int_0^\infty x^2 f_X(x)dx = \int_0^\infty \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} e^{2y} dy$$

$$= \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left(y^2-2\mu y+\mu^2\right)+4\sigma^2 y}{2\sigma^2}} dy = \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left(y-\left(\mu+2\sigma^2\right)\right)^2+4\mu\sigma^2+4\sigma^4}{2\sigma^2}} dy$$

$$= e^{2\mu+2\sigma^2} \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left(y-\left(\mu+2\sigma^2\right)\right)^2}{2\sigma^2}} dy = e^{2\mu+2\sigma^2}$$

- Recall "easy-to-compute" variance

$$Var(X) = E\left(X^2\right) - [E(X)]^2 = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$$

© 2002-2013 ICEAA. All rights reserved.

Derivation of Lognormal CV

- Standard deviation is square root of variance

$$e^{\mu+\frac{\sigma^2}{2}}\sqrt{e^{\sigma^2}-1}$$

- First factor is the mean!

$$CV=\sqrt{e^{\sigma^2}-1}$$

- Note that the CV is entirely a function of the variance (*not* CV) of the related normal
  - Pythagorean relationship $1+CV^2=e^{\sigma^2}$
  - Normal standard deviation as function of lognormal CV

$$\sigma^2=\ln\left(1+CV^2\right)\Rightarrow\sigma=\sqrt{\ln\left(1+CV^2\right)}\approx CV$$

std dev (norm)	CV (lognorm)
0.100	10%
0.198	20%
0.294	30%
0.385	40%
0.472	50%

NEW!

© 2002-2013 ICEAA. All rights reserved.

Derivation of Related Normal

- Manipulate first and second moment to solve for normal parameters
- Variance of related normal as function of variance and mean of lognormal

Agrees with previous CV result

$$\frac{E\left(X^2\right)}{\left[E(X)\right]^2} = \frac{Var(X)+\left[E(X)\right]^2}{\left[E(X)\right]^2} = \frac{e^{2\mu+2\sigma^2}}{e^{2\mu+\sigma^2}} = e^{\sigma^2} \Rightarrow \sigma^2 = \ln\left(1+\frac{Var(X)}{\left[E(X)\right]^2}\right)$$

- Mean of related normal as function of variance and mean of lognormal
  - Divide top and bottom by mean of lognormal

Note mean shift (left)

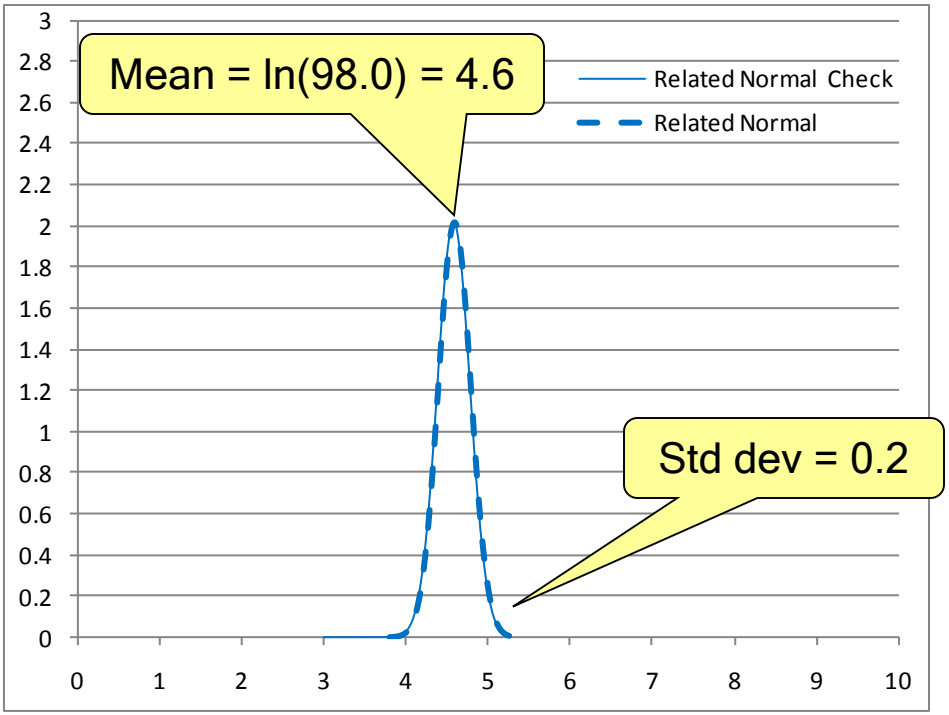
$$\frac{\left[E(X)\right]^2}{\sqrt{E\left(X^2\right)}} = \frac{\left[E(X)\right]^2}{\sqrt{Var(X)+\left[E(X)\right]^2}} = \frac{e^{2\mu+\sigma^2}}{e^{\mu+\sigma^2}} = e^{\mu} \Rightarrow \mu = \ln\frac{E(X)}{\sqrt{1+\frac{Var(X)}{\left[E(X)\right]^2}}} = \ln\left[E(X)\right] - \frac{1}{2}\ln\left(1+\frac{Var(X)}{\left[E(X)\right]^2}\right)$$

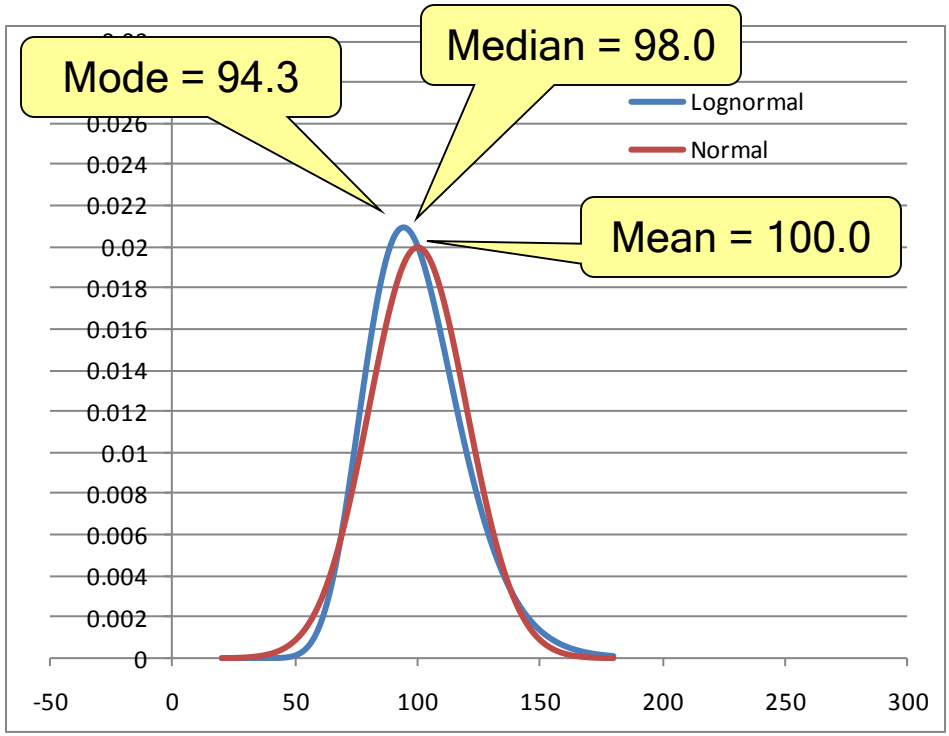
© 2002-2013 ICEAA. All rights reserved.

Related Normal Example

- Mean = 100, CV = 20%
  - Mode shift = -3.8% (relative to median)
  - Mean shift = +2.0% (relative to median)

CV (lognorm)	mode shift factor	mean shift factor	percentile of mean
10%	0.990	1.005	52.0%
20%	0.962	1.020	53.9%
30%	0.917	1.044	55.8%
40%	0.862	1.077	57.6%
50%	0.800	1.118	59.3%





© 2002-2013 ICEAA. All rights reserved.

Selecting From a Distribution

- Random Number Generation
- Inverse CDF Technique

© 2002-2013 ICEAA. All rights reserved.

The Inverse CDF Technique

- Monte Carlo Simulation requires a means to generate probability distributions
  - All events in a simulation must be assigned a value from some sort of probability distribution
  - The Inverse CDF Technique is the easiest and most common

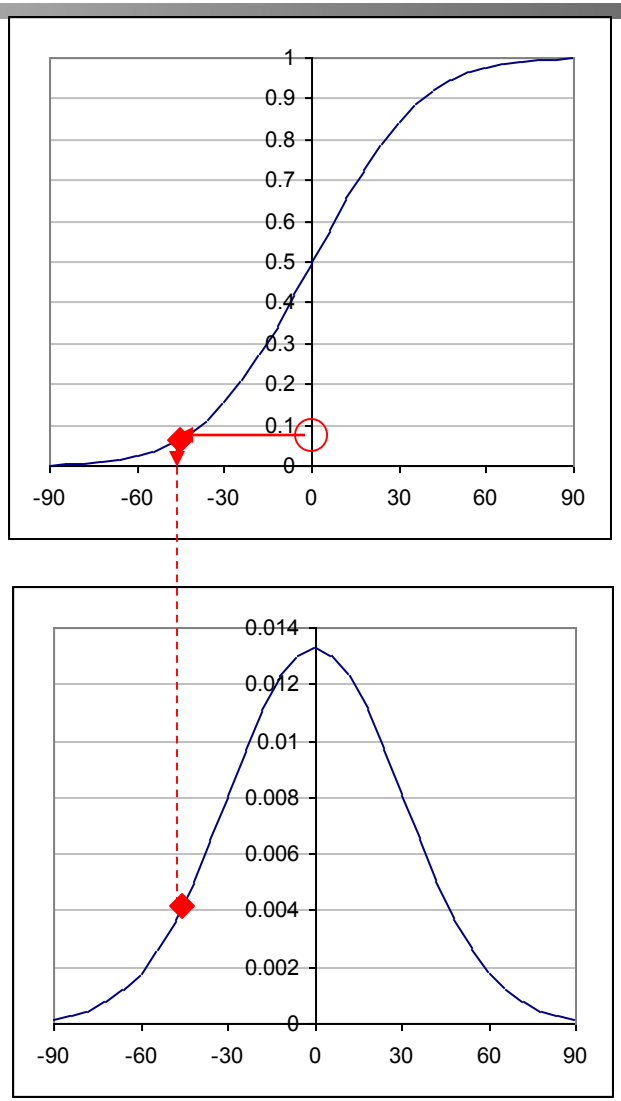
- The CDF of a distribution maps a value (x) to the probability that the random variables takes on a value less an or equal to x
  - Therefore, the inverse of the CDF maps a probability (between 0 and 1) to a value from the distribution
  - By generating a random uniform(0,1) random number, we can produce a value from any distribution with an invertible CDF
- Every simulation must contain some sort of Uniform (0,1) random number generator
  - In Excel this function is "=RAND()", although in terms of "randomness" it is not sufficient

There is an entire area of computer science dedicated to the production of the "most random" numbers (of particular interest to cryptographers)

© 2002-2013 ICEAA. All rights reserved.

Inverse CDF Normal Example

- The RAND() function returns a value of 0.0639
- This is plugged into the inverse CDF
- The value -45.677 is returned
  - 1.523 standard deviations below the mean
- This is at the appropriate percentile of the distribution



© 2002-2013 ICEAA. All rights reserved.

NEW!

Applications of the Triangular Distribution

- Subject Matter Expert (SME) Understatement of Uncertainty
- Conflating SME Inputs

© 2002-2013 ICEAA. All rights reserved.

The Geometry of Symmetric Triangles

- For a symmetric Triangle(L, M, H), where M-L = H-M
- Find points l and h such that l and h are the p^{th} and 1-p^{th} percentiles

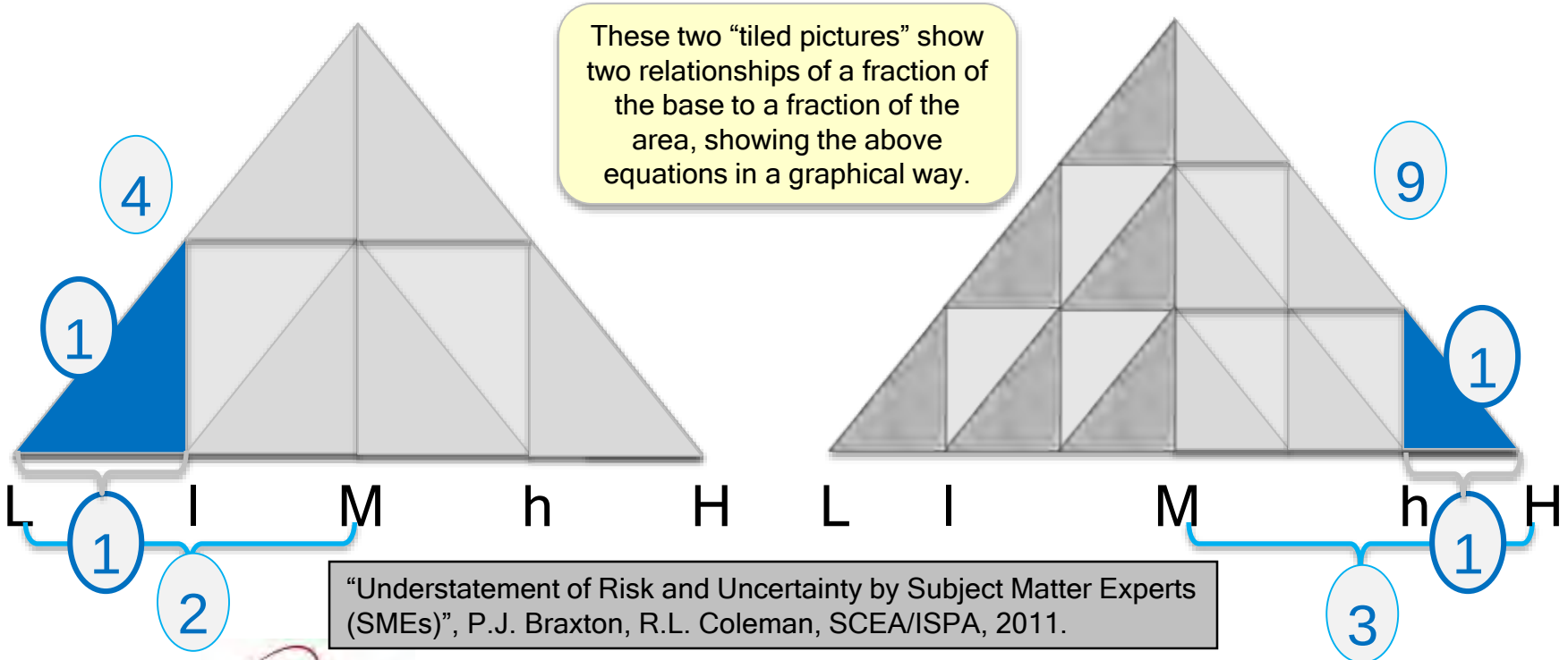
  If l-L = 1/2*(M-L), H-h = 1/2*(H-M), then p = $1/(2*2^2)$ = 1/8 = 12.5%

  If l-L = 1/3*(M-L), H-h = 1/3*(H-M), then p = $1/(2*3^2)$ = 1/18 = 5.6%

  p^{th} percentile → √(p/2) base fraction → √(2p) half-base fraction

  So, the 20^{th} percentile -> 1/5 occurs at point √(1/10) = 0.3162 base fraction

These two "tiled pictures" show two relationships of a fraction of the base to a fraction of the area, showing the above equations in a graphical way.



"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

© 2002-2013 ICEAA. All rights reserved.

Triangles With Related Areas

- We wish to know how to draw triangular distributions that are related to one another
- Constant area: $$A = \frac{1}{2}bh = \frac{1}{2}(bk)\left(\frac{h}{k}\right)$$
  - Used in **expansion of experts** (correcting understated variance)
  - For area to remain constant, in this case A = 1, as the base increases by a factor, the height must be multiplied by the reciprocal of that factor
- Reduction in area:
  - For area to be reduced by a factor, the dimensions of a <u>similar</u> triangle must be reduced by the square root of that factor

$$A_2 = \frac{1}{k}A_1 = \frac{1}{2k}b_1h_1 = \frac{1}{2}\left(\frac{b_1}{\sqrt{k}}\right)\left(\frac{h_1}{\sqrt{k}}\right)$$

  - For area to be reduced by a factor, the height must be reduced by that factor if the base is to remain constant
    - Used in **sampling of experts**

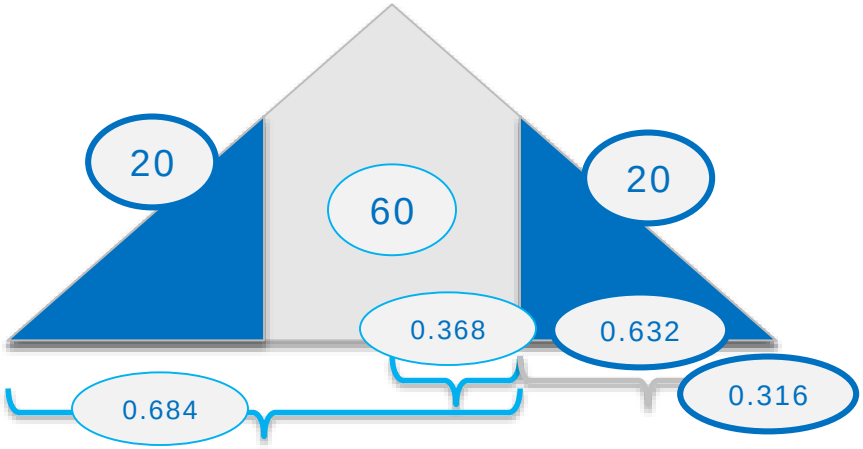
$$A_2 = \frac{1}{k}A_1 = \frac{1}{2k}b_1h_1 = \frac{1}{2}(b_1)\left(\frac{h_1}{k}\right)$$

"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

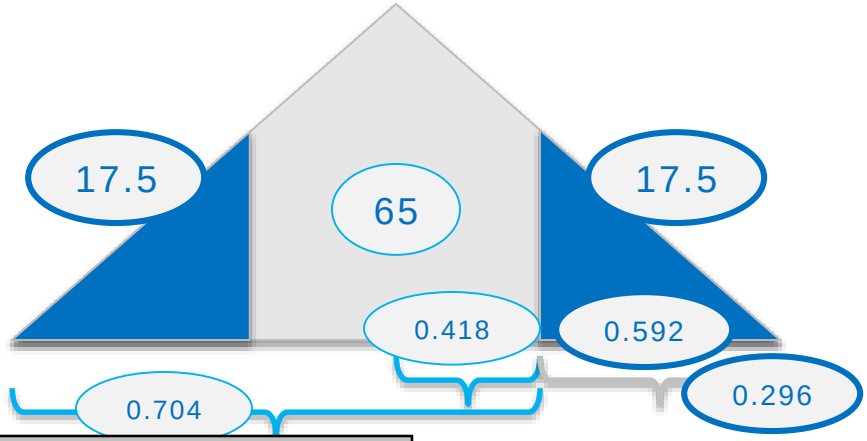
© 2002-2013 ICEAA. All rights reserved.

Correction of Understated Variance for Triangles

- For symmetric triangles
  - To expand from 20-80 to Min-Max, multiply by 2.72 = 1/0.368
  - √(1/10) = 0.3162 base fraction
  - √(2/5) = 0.6325 half-base fraction



  - To expand from plus-or-minus-one-sigma to Min-Max, multiply by 2.45 (√6)
  - (√6-1)/2√6 = 0.2959 base fraction
  - (√6-1)/√6 = 0.5918 half-base fraction
  - Compare with 68.3% within one sigma rule of thumb for Normal distribution

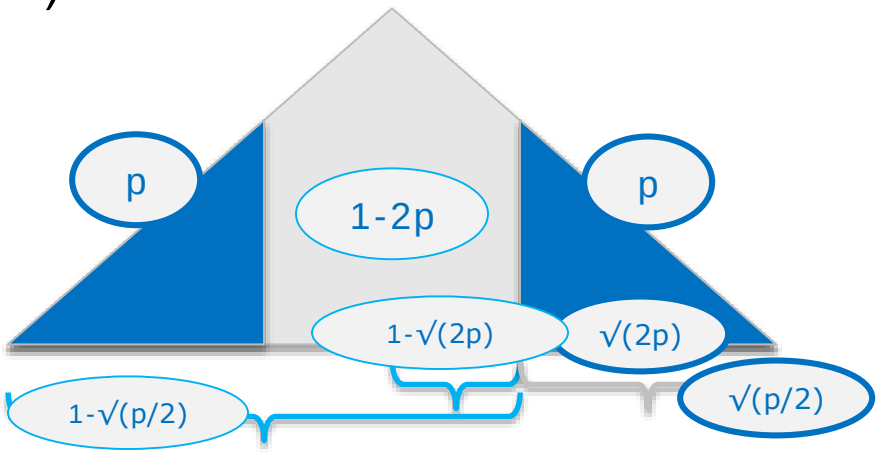


"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

© 2002-2013 ICEAA. All rights reserved.

Correction of Understated Variance for Triangles

- For symmetric triangles
  - General case
  - To expand from p^{th}-$(1-p)^{th}$ to Min-Max, multiply by $1/(1-\sqrt{(2p)})$
  - If $2p = \alpha$, then multiply by $1/(1-\sqrt{\alpha})$
  - To expand from $(\alpha_1/2)^{th}$-$(1-\alpha_1/2)^{th}$ to $(\alpha_2/2)^{th}$-$(1-\alpha_2/2)^{th}$ $[\alpha_1 > \alpha_2]$, multiply by $(1-\sqrt{\alpha_2})/(1-\sqrt{\alpha_1})$
  - For example, to expand from 33-67 to 20-80, multiply by $(1-\sqrt{(2/5)})/(1-\sqrt{(2/3)}) \approx 2.0$



"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

© 2002-2013 ICEAA. All rights reserved.

Variance of Hybrid Distributions – A Pythagorean Relationship

- Suppose k distributions with pdf $p_i(x_i)$, mean μ_i, and standard deviation σ_i are sampled
- Then the pdf of the hybrid distribution is the "average" of the pdfs

$$p(x) = \frac{1}{k}\sum_{i=1}^{k} p_i(x_i)$$

- The mean of the hybrid distribution is the average of the means

$$\mu = E(X) = \frac{1}{k}\sum_{i=1}^{k}\int x_i\, p_i(x_i)\, dx_i = \frac{\sum_{i=1}^{k}\mu_i}{k}$$

- The variance of the hybrid distribution is the average of the variances plus the variance of the means taken as a discrete probability distribution!
  - See next slide for derivation

"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

© 2002-2013 ICEAA. All rights reserved.

Variance of Hybrid Distributions – A Pythagorean Relationship

$$E\left(X^2\right) = \frac{1}{k}\sum_{i=1}^{k}\int x_i^2\, p_i(x_i)\,dx_i = \frac{\sum_{i=1}^{k}\left(\sigma_i^2 + \mu_i^2\right)}{k}$$

$$\sigma^2 = E\left(X^2\right) - \mu^2 = \frac{\sum_{i=1}^{k}\left(\sigma_i^2 + \mu_i^2\right)}{k} - \left(\frac{1}{k}\sum_{i=1}^{k}\mu_i\right)^2$$

$$= \frac{\sum_{i=1}^{k}\sigma_i^2}{k} + \left[\frac{\sum_{i=1}^{k}\mu_i^2}{k} - \left(\frac{1}{k}\sum_{i=1}^{k}\mu_i\right)^2\right]$$

- In the special case of two congruent distributions with centers at m-d and m+d, the variance is

$$= \sigma^2 + \left[\frac{(m-d)^2 + (m+d)^2}{2} - m^2\right] = \sigma^2 + d^2$$

"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

© 2002-2013 ICEAA. All rights reserved.

Equivalence of Averaging Distributions and Averaging Parameters for Symmetric Triangles

- In the case of symmetric triangles, averaging the individual triangles (with perfect rank correlation) can be shown to be equivalent to averaging the parameters
  - We will prove it in the case of two triangles, but the proof can easily be extended to more
- As previously shown, the p^{th} percentile ($p<0.5$) for a symmetric triangle is at the $\sqrt{(2p)}$ half-base fraction
  - So the p^{th} percentiles of the two triangles and their average are:

$$a_1+\sqrt{2p}(c_1-a_1) \quad a_2+\sqrt{2p}(c_2-a_2) \Rightarrow \frac{a_1+a_2}{2}+\sqrt{2p}\,\frac{(c_1-a_1)+(c_2-a_2)}{2}$$

  - But this is clearly just the p^{th} percentile of the average distribution

$$\left(\frac{a_1+a_2}{2}\right)+\sqrt{2p}\left|\left(\frac{c_1+c_2}{2}\right)-\left(\frac{a_1+a_2}{2}\right)\right|$$

  - A similar proof works for $p>0.5$
  - Since all percentiles are equal, the resulting distributions are identical
- Monte Carlo simulation could be used to explore the difference between the two methods for asymmetric triangles, but it is not expected to be large

"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

© 2002-2013 ICEAA. All rights reserved.

Equivalence of Averaging Means and Averaging Modes for Triangles

- If we average parameters, as long as we average mins and maxes, it doesn't matter whether we average means or modes
  - Algebraically equivalent
  - Any number of triangles, symmetry not required
- Let the k^{th} triangle be $T(a_i, c_i, b_i)$, and parameter-averaged triangle be T(A, C, B), where

$$A = \frac{\sum_{i=1}^{k} a_i}{k} \quad C = \frac{\sum_{i=1}^{k} c_i}{k} \quad B = \frac{\sum_{i=1}^{k} b_i}{k}$$

- This is averaging the modes; the resulting mean is

$$\frac{A+B+C}{3} = \frac{\sum_{i=1}^{k} a_i + \sum_{i=1}^{k} b_i + \sum_{i=1}^{k} c_i}{3k} = \frac{\sum_{i=1}^{k} \left(\frac{a_i + b_i + c_i}{3} \right)}{k}$$

  which is just the average of the means!
- Reversing the flow, averaging the means can be shown to produce a mode which is the average of the modes

"Understatement of Risk and Uncertainty by Subject Matter Experts (SMEs)", P.J. Braxton, R.L. Coleman, SCEA/ISPA, 2011.

© 2002-2013 ICEAA. All rights reserved.

Cost Growth Factor and Percentiles

- A result for Final Output Distributions

© 2002-2013 ICEAA. All rights reserved.

Implied Percentile or CGF

- For a given CV, Cost Growth Factor (CGF) and percentile (of the point estimate) are related
  - Think of CGF as the factor it take to bring the point estimate up to the true mean
  - Different answers for normal of lognormal
- Given a CGF and CV, we might ask:
  - At what percentile must the point estimate be?
- Given a percentile and CV, we might ask:
  - What is the CGF between that value and the mean?

© 2002-2013 ICEAA. All rights reserved.

NEW!

Implied Percentile – Normal

- Without loss of generality (w.l.o.g.), assume mean of one (1.0)
  - Then the standard deviation is equal to the CV!
- We are looking for the percentile of the reciprocal of CGF

$$X \sim N(\mu,\sigma) \Leftrightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

$$p = \Phi\left(\frac{(1/CGF)-\mu}{\sigma}\right) = \Phi\left(\frac{(1/CGF)-1}{CV}\right)$$

© 2002-2013 ICEAA. All rights reserved.

NEW!

Implied Percentile – Lognormal

- W.l.o.g., assume lognormal mean of one (1.0)
- Then the mean and standard deviation of the related normal are:

$$\mu = \ln\left(\frac{1}{\sqrt{1+CV^2}}\right) \qquad \sigma = \sqrt{\ln\left(1+CV^2\right)}$$

- We are looking for the percentile of the reciprocal of CGF (unit space)

$$X = e^Y \sim LogN(\mu,\sigma) \Leftrightarrow Y = \ln X \sim N(\mu,\sigma) \Leftrightarrow Z = \frac{Y-\mu}{\sigma} \sim N(0,1)$$

$$p = \Phi\left(\frac{\ln(1/CGF)-\mu}{\sigma}\right) = \Phi\left(\frac{\ln\left(\frac{\sqrt{1+CV^2}}{CGF}\right)}{\sqrt{\ln\left(1+CV^2\right)}}\right)$$

© 2002-2013 ICEAA. All rights reserved.

Percentile-to-CGF Conversion

- This table shows the implied CGF for a given percentile of the Point Estimate for various CVs of a Normal distribution

NORMAL	Coefficient of Variation (CV)									
p	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0.05	1.09	1.20	1.33	1.49	1.49	1.49	2.36	2.92	3.85	5.63
0.10	1.07	1.15	1.24	1.34	1.34	1.34	1.81	2.05	2.36	2.78
0.15	1.05	1.12	1.18	1.26	1.26	1.26	1.57	1.71	1.87	2.08
0.20	1.04	1.09	1.14	1.20	1.20	1.20	1.42	1.51	1.61	1.73
0.25	1.03	1.07	1.11	1.16	1.16	1.16	1.31	1.37	1.44	1.51
0.30	1.03	1.06	1.09	1.12	1.12	1.12	1.22	1.27	1.31	1.36
0.35	1.02	1.04	1.06	1.08	1.08	1.08	1.16	1.18	1.21	1.24
0.40	1.01	1.03	1.04	1.05	1.05	1.05	1.10	1.11	1.13	1.15
0.45	1.01	1.01	1.02	1.03	1.03	1.03	1.05	1.05	1.06	1.07
0.50	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
0.55	0.99	0.99	0.98	0.98	0.98	0.98	0.96	0.95	0.95	0.94
0.60	0.99	0.98	0.96	0.95	0.95	0.95	0.92	0.91	0.90	0.89
0.65	0.98	0.96	0.95	0.93	0.93	0.93	0.88	0.87	0.85	0.84
0.70	0.97	0.95	0.93	0.91	0.91	0.91	0.84	0.83	0.81	0.79
0.75	0.97	0.94	0.91	0.88	0.88	0.88	0.81	0.79	0.77	0.75
0.80	0.96	0.92	0.89	0.86	0.86	0.86	0.77	0.75	0.73	0.70
0.85	0.95	0.91	0.87	0.83	0.83	0.83	0.73	0.71	0.68	0.66
0.90	0.94	0.89	0.84	0.80	0.80	0.80	0.69	0.66	0.63	0.61
0.95	0.92	0.86	0.80	0.75	0.75	0.75	0.63	0.60	0.57	0.55

© 2002-2013 ICEAA. All rights reserved.

Percentile-to-CGF Conversion

- This table shows the implied CGF for a given percentile of the Point Estimate for various CVs of a Lognormal distribution

LOGNORMAL	Coefficient of Variation (CV)									
p	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0.05	1.09	1.18	1.29	1.41	1.55	1.69	1.85	2.03	2.22	2.43
0.10	1.07	1.14	1.22	1.31	1.41	1.52	1.64	1.76	1.90	2.05
0.15	1.05	1.11	1.18	1.25	1.33	1.42	1.51	1.61	1.71	1.82
0.20	1.04	1.09	1.15	1.20	1.27	1.34	1.41	1.49	1.57	1.66
0.25	1.04	1.07	1.12	1.17	1.22	1.27	1.33	1.40	1.46	1.54
0.30	1.03	1.06	1.09	1.13	1.17	1.22	1.27	1.32	1.37	1.43
0.35	1.02	1.04	1.07	1.10	1.13	1.17	1.21	1.25	1.29	1.34
0.40	1.01	1.03	1.05	1.07	1.10	1.12	1.15	1.19	1.22	1.26
0.45	1.01	1.02	1.03	1.05	1.06	1.08	1.11	1.13	1.16	1.19
0.50	**1.00**	**1.00**	**1.01**	**1.02**	1.03	1.04	1.06	1.08	1.10	1.12
0.55	0.99	0.99	**0.99**	**0.99**	**1.00**	**1.01**	**1.02**	**1.03**	**1.04**	**1.05**
0.60	0.99	0.98	0.97	0.97	0.97	**0.97**	**0.97**	**0.98**	**0.98**	**0.99**
0.65	0.98	0.97	0.95	0.94	0.94	0.93	0.93	0.93	0.93	0.93
0.70	0.98	0.95	0.94	0.92	0.91	0.90	0.89	0.88	0.88	0.87
0.75	0.97	0.94	0.91	0.89	0.87	0.86	0.84	0.83	0.82	0.81
0.80	0.96	0.92	0.89	0.86	0.84	0.82	0.80	0.78	0.76	0.75
0.85	0.95	0.91	0.87	0.83	0.80	0.77	0.74	0.72	0.70	0.69
0.90	0.94	0.88	0.84	0.79	0.75	0.72	0.69	0.66	0.63	0.61
0.95	0.92	0.85	0.79	0.74	0.69	0.64	0.61	0.57	0.54	0.51

© 2002-2013 ICEAA. All rights reserved.

Alternate Specifications

- Normal
- Lognormal
- Two Lognormals: Ambiguity from Square Roots

© 2002-2013 ICEAA. All rights reserved.

Alternate Specification of Normal

Given	Mean	Std Dev	CV
Mean, CV	μ	$CV \cdot \mu$	CV
Mean, Percentile (X_p, p) $Z_p = \Phi^{-1}(p)$	μ	$\frac{X_p - \mu}{Z_p}$	$\frac{\left(\frac{X_p}{\mu} - 1\right)}{Z_p}$
CV, Percentile (X_p, p)	$\frac{X_p}{1 + Z_p \cdot CV}$	$\frac{X_p \cdot CV}{1 + Z_p \cdot CV}$	CV
Two Percentiles $Z_i = \Phi^{-1}(p_i)$ $i \in \{1,2\}$	$\frac{Z_2 X_1 - Z_1 X_2}{Z_2 - Z_1}$	$\frac{X_2 - X_1}{Z_2 - Z_1}$	$\frac{\sigma}{\mu}$

© 2002-2013 ICEAA. All rights reserved.

Alternate Specification of Lognormal

Given	Mean	Std Dev	CV	μ	σ
Mean, CV	$E(X)$	$CV \cdot E(X)$	CV	$\ln\left(\frac{E(X)}{\sqrt{1+CV^2}}\right)$	$\sqrt{\ln(1+CV^2)}$
Mean, Percentile (X_p, p) $Z_p = \Phi^{-1}(p)$	$E(X)$	$CV \cdot E(X)$	$\sqrt{e^{\sigma^2}-1}$	$\ln\left(\frac{E(X)}{\sqrt{1+CV^2}}\right)$	$Z_p \pm \sqrt{Z_p^2 - 2\ln\left(\frac{X_p}{E(X)}\right)}$
CV, Percentile (X_p, p)	$e^{\mu+\frac{\sigma^2}{2}}$	$CV \cdot E(X)$	CV	$\ln X_p - Z_p\sqrt{\ln(1+CV^2)}$	$\sqrt{\ln(1+CV^2)}$
Two Per-centiles $Z_i = \Phi^{-1}(p_i)$ $i \in \{1,2\}$	$e^{\mu+\frac{\sigma^2}{2}}$	$CV \cdot E(X)$	$\sqrt{e^{\sigma^2}-1}$	$\frac{Z_2 \ln X_1 - Z_1 \ln X_2}{Z_2 - Z_1}$	$\frac{\ln\left(\frac{X_2}{X_1}\right)}{Z_2 - Z_1}$

© 2002-2013 ICEAA. All rights reserved.

Alternate Specification of Lognormal

Given	Mean	Std Dev	CV	μ	σ
Mean, Median (m)	$E(X)$	$CV \cdot E(X)$	$\sqrt{e^{\sigma^2}-1}$	$\ln(m)$	$\sqrt{2\ln\left(\frac{E(X)}{m}\right)}$
Median (m), Mode (M)	$e^{\mu+\frac{\sigma^2}{2}}$	$CV \cdot E(X)$	$\sqrt{e^{\sigma^2}-1}$	$\ln(m)$	$\sqrt{\ln\left(\frac{m}{M}\right)}$
Mean, Mode (M)	$E(X)$	$CV \cdot E(X)$	$\sqrt{e^{\sigma^2}-1}$	$\frac{1}{3}\ln\left([E(X)]^2 M\right)$	$\sqrt{\frac{2}{3}\ln\left(\frac{E(X)}{M}\right)}$

© 2002-2013 ICEAA. All rights reserved.

NEW!

Alternate Specification of Lognormal

Given	Mean	Std Dev	CV	μ	σ
Median (m), Percentile (X_p, p)	$e^{\mu+\frac{\sigma^2}{2}}$	$CV \cdot E(X)$	$\sqrt{e^{\sigma^2}-1}$	$\ln(m)$	$\frac{\ln\left(\frac{X_p}{m}\right)}{Z_p}$
Mode (M), Percentile (X_p, p)	$e^{\mu+\frac{\sigma^2}{2}}$	$CV \cdot E(X)$	$\sqrt{e^{\sigma^2}-1}$	$\ln\left[M\left(1+CV^2\right)\right]$	$\frac{-Z_p/2 \pm \sqrt{Z_p^2+4\ln\left(\frac{X_p}{M}\right)}}{2}$

$Z_p = \Phi^{-1}(p)$

© 2002-2013 ICEAA. All rights reserved.

NEW!

Two Lognormals

from the quadratic formula!

- The plus-or-minus in the two previous orange cells gives rise to the possibility of two solutions to a given specification
- Mean and a Percentile
  - Percentile > Mean: two tenable solutions
  - Percentile = Mean: one tenable solution and one singular solution (σ = 0)
  - Percentile < Mean: one tenable solution and one untenable solution (σ < 0)

$$\sigma = Z_p \pm \sqrt{Z_p^2 - 2\ln\left(\frac{X_p}{E(X)}\right)}$$

- Mean and a Percentile
  - Percentile < Mode: two tenable solutions
  - Percentile = Mode: one tenable solution and one singular solution (σ = 0)
  - Percentile > Mode: one tenable solution and one untenable solution (σ < 0)

$$\sigma = -\frac{Z_p}{2} \pm \frac{1}{2}\sqrt{Z_p^2 + 4\ln\left(\frac{X_p}{M}\right)}$$

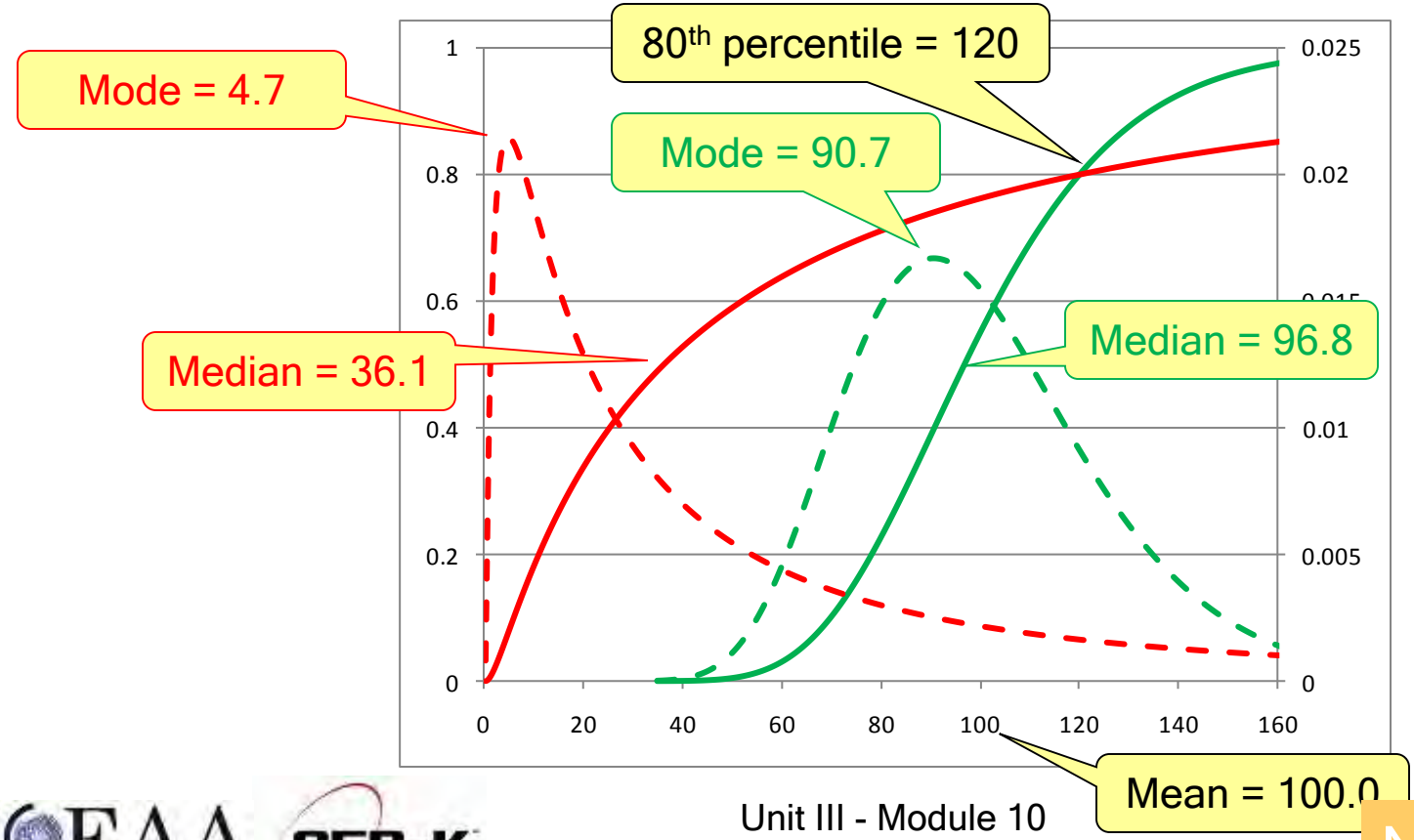
third-class lever!

- Note that the mischief occurs when the percentile (load) is on the other side of the mean/mode (effort) from the median (fulcrum)

© 2002-2013 ICEAA. All rights reserved.

Two Lognormals Example

- Mean = 100, 80th percentile = 120
  - "Regular" solution with 26% CV
  - "Extreme" solution with 258% CV!



NEW!

© 2002-2013 ICEAA. All rights reserved.

Probability Distributions for Risk Analysis

Backup

© 2002-2013 ICEAA. All rights reserved.

Algebra Refresher – Quadratics

- Quadratic = polynomial of degree two $ax^2 + bx + c$
- Completing the Square
  - Allows any quadratic to be written as a perfect square plus a constant
  - Graphically identifies the vertex of the parabola!
  - One way of deriving the quadratic formula

$$a\left[\left(x+\frac{b}{2a}\right)^2 - \frac{b^2-4ac}{4a^2}\right]$$

- Quadratic formula
  - Square root piece = discriminant
  - Gives the two roots of a quadratic equation
    1. Two real roots, branches of parabola intersect x-axis
    2. One real root (duplicate), vertex of parabola tangent to x-axis
    3. Two complex roots (square root of a negative = imaginary), parabola does not intersect x-axis

$$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$$

© 2002-2013 ICEAA. All rights reserved.

NEW!

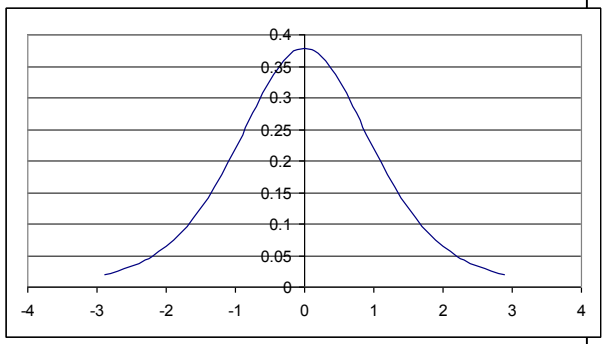
Continuous Distributions

- Normal and t
- Lognormal
- Triangular
- Uniform
- Other Continuous

© 2002-2013 ICEAA. All rights reserved.

t Distribution Overview

- Distribution

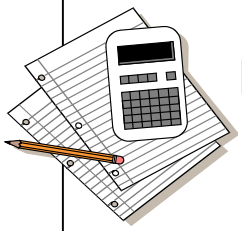


$x \in (-\infty, \infty)$

$$p(x) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\,\Gamma(n/2)[1+(x^2/n)]^{(n+1)/2}}$$

- Parameters and Statistics
  - Degrees of freedom = n
  - Mean = 0
  - Variance = $\frac{n}{n-2}$
  - Skewness = 0

- Key Facts
  - As n approaches infinity, the t distribution approaches Normal
  - The t is distributed as $t = \frac{N(0,1)}{\chi(n)/\sqrt{n}}$
  - Excel
    - cdf = TDIST(x, n, tails)
      - Tails = 1 or 2 depending on whether you want to include the probability in the left-hand tail
    - Inv cdf = TINV(prob, n)



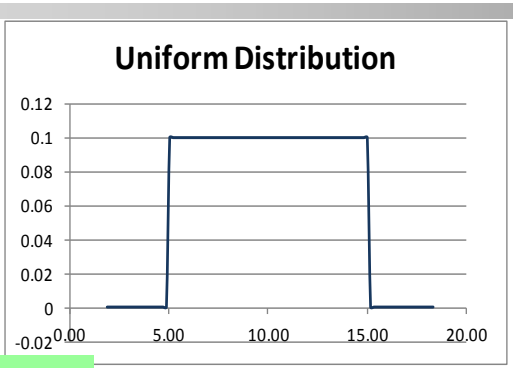
- Applications
  - Confidence Intervals
    - Mean of Normal variates
  - Regression Analysis
    - Significance of individual coefficients
  - Hypothesis Testing

© 2002-2013 ICEAA. All rights reserved.

Uniform Distribution Overview

- Distribution

$$F(x)=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\le x\le b\\ 1 & x>b\end{cases}$$



$$p(x)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & x<a \text{ or } x>b\end{cases}$$

- Parameters and Statistics
  - Min = a
  - Max = b
  - Mode = any value [a,b]
  - Mean = $\dfrac{a+b}{2}$
  - Variance = $\dfrac{(b-a)^2}{12}$

19

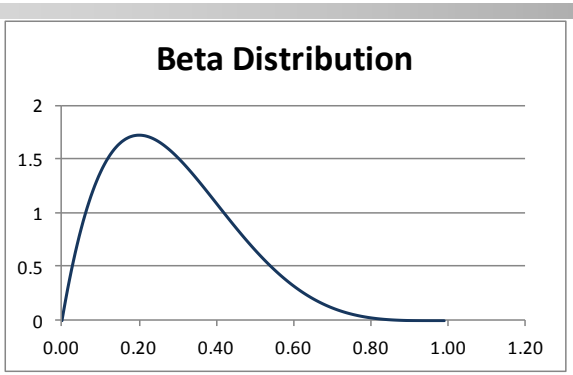
- Key Facts
  - Excel
    - pdf and cdf calculations can be handled as they are listed in the above formulas

- Applications
  - Risk Analysis
    - SME Input
  - Sampling from arbitrary distributions
    - Example: Rejection Sampling

© 2002-2013 ICEAA. All rights reserved.

Beta Distribution Overview

- Distribution



$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

$$F(x)=I_x(\alpha,\beta)$$

- Parameters and Statistics
  - Min = 0
  - Max = 1
  - Mode = $\frac{\alpha-1}{\alpha+\beta-2}$ for $\alpha>1, \beta>1$
  - Mean = $\frac{\alpha}{\alpha+\beta}$
  - Variance = $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

19

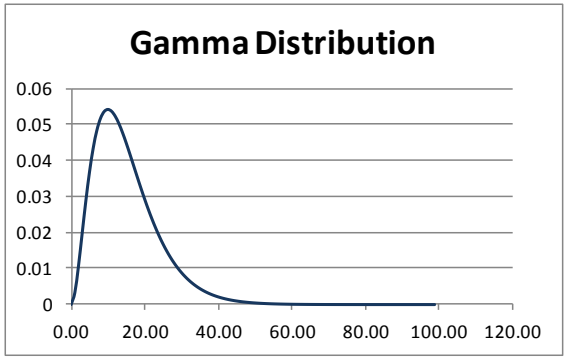
- Key Facts
  - Excel
    - pdf and cdf calculations can be handled as they are listed in the above formulas

- Applications
  - Risk Analysis
  - Order Statistics
  - Rule of Succession
  - Bayesian Inference
  - Task Duration

© 2002-2013 ICEAA. All rights reserved.

Gamma Distribution Overview

- Distribution



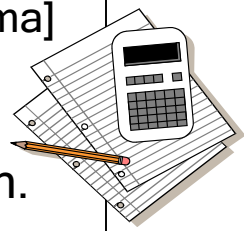
$$F(x)=\frac{1}{\Gamma(\alpha)}\gamma(\alpha,\beta x)$$

$$p(x)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$$

- Parameters and Statistics
  - Min = 0
  - Max = ∞
  - Mode = $\frac{\alpha-1}{\beta}$ for $\alpha \geq 1$
  - Mean = $\frac{\alpha}{\beta}$
  - Variance = $\frac{\alpha}{\beta^2}$

19

- Key Facts
  - Excel
    - GAMMADIST, GAMMAINV, and GAMMALN [natural log of gamma]
  - It is the conjugate prior for the precision (i.e. inverse of the variance) of a normal distribution.

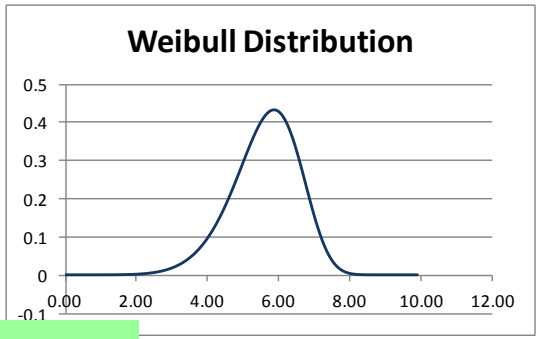


- Applications
  - Risk Analysis
    - Due to shape and scale parameters
  - Modeling
    - Size of insurance claims and rainfall

© 2002-2013 ICEAA. All rights reserved.

Weibull Distribution Overview

- Distribution



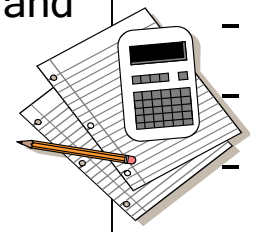
$$F(x) = 1 - e^{-(x/\lambda)^k}$$

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- Parameters and Statistics
  - Min = 0
  - Max = ∞
  - Mode = $\begin{cases} \lambda\left(\frac{k-1}{k}\right)^{\frac{1}{k}} & k > 1 \\ 0 & k = 1 \end{cases}$
  - Mean = $\lambda\Gamma\left(1 + 1/k\right)$
  - Variance = $\lambda^2\Gamma\left(1 + 2/k\right) - \mu^2$

19

- Key Facts
  - Excel
    - WEIBULL function where α = k and β = λ, as shown above
  - Interpolates between the exponential distribution with intensity 1/λ when k =1 and a Rayleigh Distribution of when k = 2.

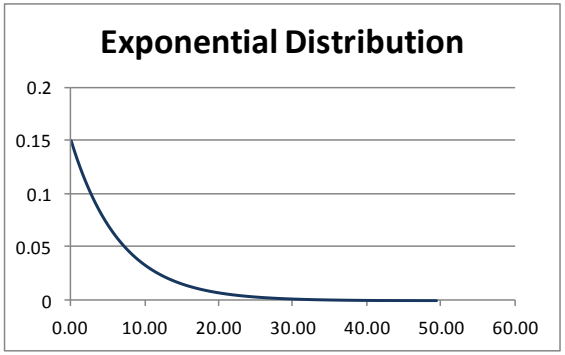


- Applications
  - Risk Analysis
  - Reliability Engineering
  - Failure Analysis
  - Delivery Times
  - RF Dispersion

© 2002-2013 ICEAA. All rights reserved.

Exponential Distribution Overview

- Distribution



$$F(x)=\begin{cases}1-e^{-\lambda x}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

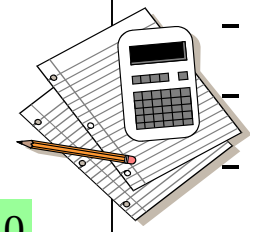
$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0,\\ 0, & x<0.\end{cases}$$

- Parameters and Statistics
  - Min = 0
  - Max = ∞
  - Mode = 0
  - Mean = λ^{-1}
  - Variance = λ^{-2}

19

- Key Facts
  - Excel
    - EXPONDIST
  - Exponential distribution exhibits infinite divisibility
  - Memoryless

$$\Pr(T>s+t|T>s)=\Pr(T>t)\text{ for all }s,t\geq 0$$

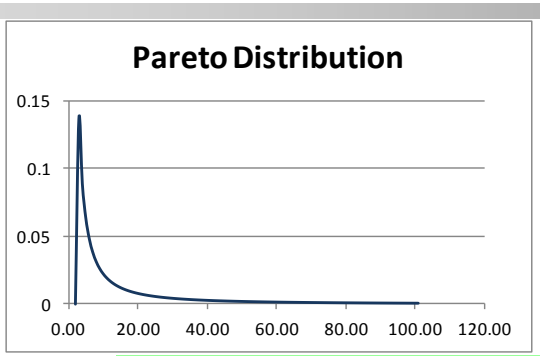


- Applications
  - Risk Analysis
  - Inter-arrival times (Poisson)
  - Time between events
  - Reliability engineering
  - Hazard Rate
  - Bathtub Curve

© 2002-2013 ICEAA. All rights reserved.

Pareto Distribution Overview

- Distribution



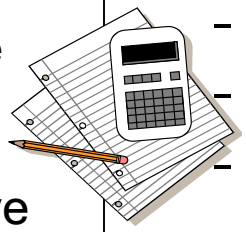
$$f(x)=\begin{cases}\alpha\dfrac{x_m^{\alpha}}{x^{\alpha-1}} & \text{for } x>x_m\\ 0 & \text{for } x<x_m\end{cases}$$

$$F(x)=\begin{cases}1-\left(\dfrac{x_m}{x}\right)^{\alpha} & \text{for } x>x_m\\ 0 & \text{for } x<x_m\end{cases}$$

- Parameters and Statistics
  - Min = x_m
  - Max = +∞
  - Mode = x_m
  - Mean = $\dfrac{\alpha x_m^{\alpha}}{x^{\alpha+1}}$ for $x \geq x_m$
  - Variance = $\dfrac{x_m^2\alpha}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$

19

- Key Facts
  - Excel
    - pdf and cdf calculations can be handled as they are listed in the above formulas
  - The Pareto distribution and log-normal distribution are alternative distributions for describing the same types of quantities.



- Applications
  - Risk Analysis
  - Population sizes
  - File size and Internet traffic
  - Hard disk error rates
  - Distribution of Income

© 2002-2013 ICEAA. All rights reserved.

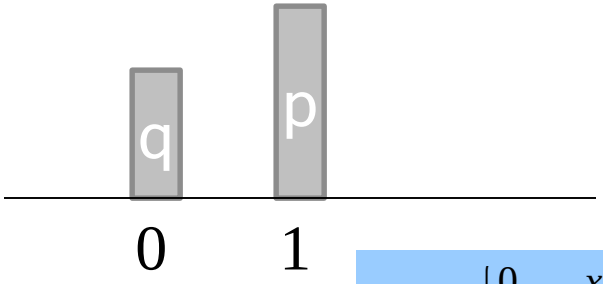
Discrete Distributions

- Bernoulli
- Binomial
- Other Discrete

© 2002-2013 ICEAA. All rights reserved.

Bernoulli Distribution Overview

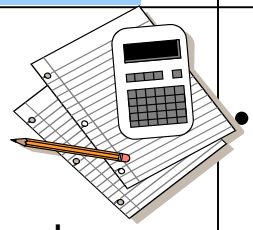
- Distribution



$$p(x)=\begin{cases} q=1-p & x=0 \\ p & x=1 \end{cases}$$

$$F(x)=\begin{cases} 0 & x<0 \\ q & 0\le x<1 \\ 1 & x\ge 1 \end{cases}$$

- Parameters and Statistics
  - Min = 0
  - Max = 1
  - Mean = p
  - Variance = pq



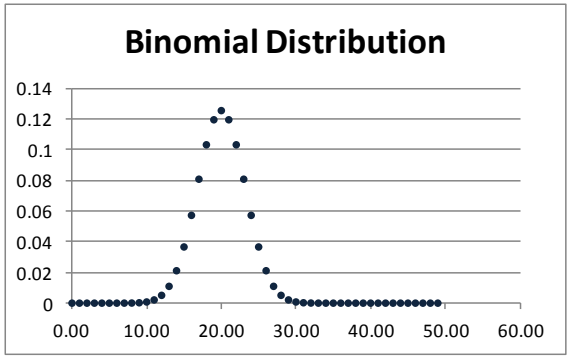
- Key Facts
  - Excel
    - pdf and cdf calculations can be handled as they are listed in the above formulas
  - The sum of n Bernoullis is Binomial (n,p)

- Applications
  - Risk Analysis
    - Discrete Risks
    - X = Cf * Bernoulli
    - p = Pf

© 2002-2013 ICEAA. All rights reserved.

Binomial Distribution Overview

- Distribution

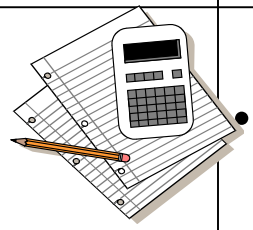


$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$F(x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}$$

- Parameters and Statistics
  - Min = 0
  - Max = n
  - Mean = np
  - Variance = np(1-p)

- Key Facts
  - Excel
    - BINOMDIST
  - The number of "successes" in a sequence of n independent experiments
  - Beta Distribution is the conjugate prior

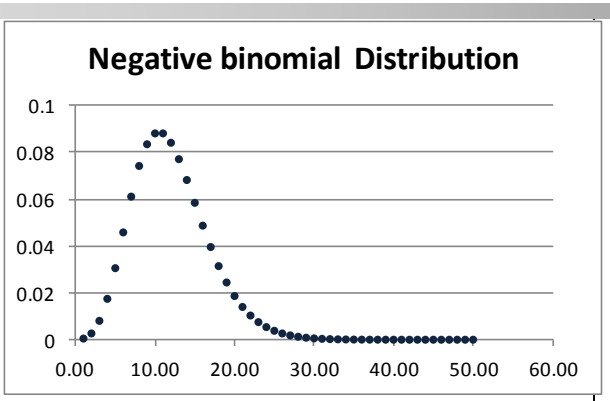


- Applications
  - Risk Analysis
  - Models for repeated processes or experiments
  - May be used for batch processing failures

© 2002-2013 ICEAA. All rights reserved.

Negative Binomial Distribution Overview

- Distribution

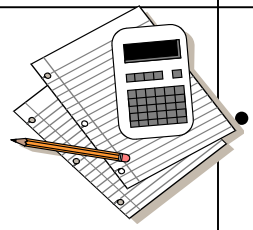


$$p(k) = \binom{k+r-1}{k}(1-p)^r p^k$$

$$\Pr(X \le k) = 1 - I_p(k+1, r)$$

- Parameters and Statistics
  - Min = 0
  - Max = ∞
  - Mean = $\frac{pr}{1-p}$
  - Variance = $\frac{pr}{(1-p)^2}$

- Key Facts
  - Excel
    - NEGBINOMDIST



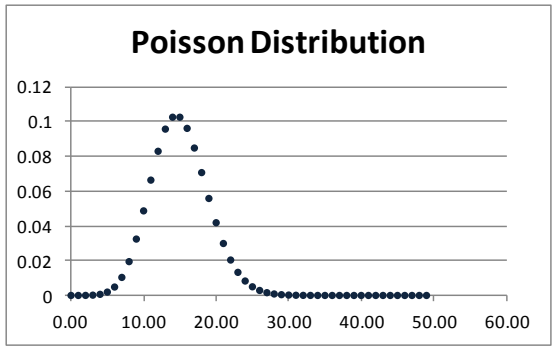
- Applications
  - Risk Analysis
    - Discrete Risks
  - Process Analysis

© 2002-2013 ICEAA. All rights reserved.

NEW!

Poisson Distribution Overview

- Distribution

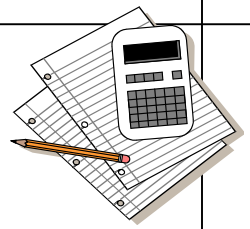


$$p(x) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$F(x) = e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}$$

- Parameters and Statistics
  - Min = 0
  - Max = ∞
  - Mean = λ
  - Variance = λ

- Key Facts
  - Excel
    - POISSON
  - Ladislaus Bortkiewicz, 1898, used to investigate the number of Prussian soldiers killed by horse kick.
  - Law of Small Numbers or Law of rare events

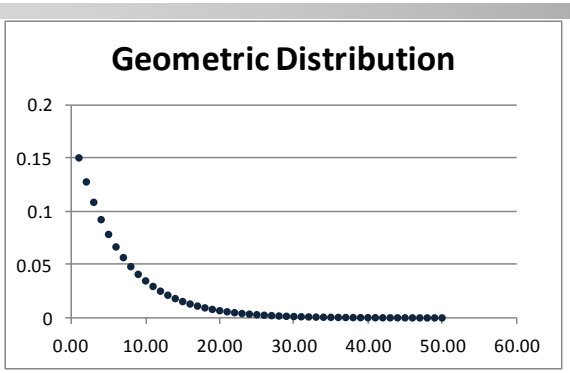


- Applications
  - Risk Analysis
  - Reliability Engineering
  - Customer Service
  - Civil Engineering
  - Astrology

© 2002-2013 ICEAA. All rights reserved.

Geometric Distribution Overview

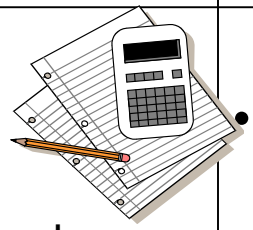
- Distribution



$$p(x) = (1-p)^{k-1} p$$

$$F(x) = 1-(1-p)^{k}$$

- Parameters and Statistics
  - Min = 1
  - Max = ∞
  - Mean = 1/p
  - Variance = $\frac{1-p}{p^2}$



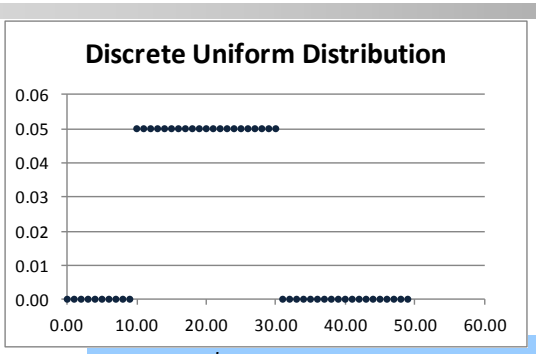
- Key Facts
  - Excel
    - pdf and cdf calculations can be handled as they are listed in the above formulas
  - Continuous analogue is the exponential distribution

- Applications
  - Risk Analysis
    - Discrete Risks

© 2002-2013 ICEAA. All rights reserved.

Discrete Uniform Distribution Overview

- Distribution

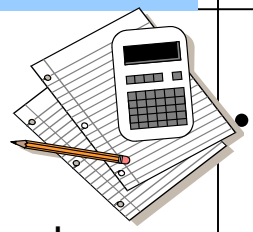


$$p(x)=\begin{cases}\dfrac{1}{(b-a+1)} & a\le k\le b\\ 0 & \text{Otherwise}\end{cases}$$

$$F(x)=\begin{cases}0 & k<a\\ \dfrac{k-a+1}{b-a+1} & a\le k\le b\\ 1 & k>b\end{cases}$$

- Parameters and Statistics
  - Min = a
  - Max = b
  - Mean = $\dfrac{a+b}{2}$
  - Variance = $\dfrac{(b-a+1)^2-1}{12}$

- Key Facts
  - Excel
    - pdf and cdf calculations can be handled as they are listed in the above formulas
  - German Tank Problem



- Applications
  - Risk Analysis
    - Discrete Risks
    - Quantity selections

© 2002-2013 ICEAA. All rights reserved.

NEW!