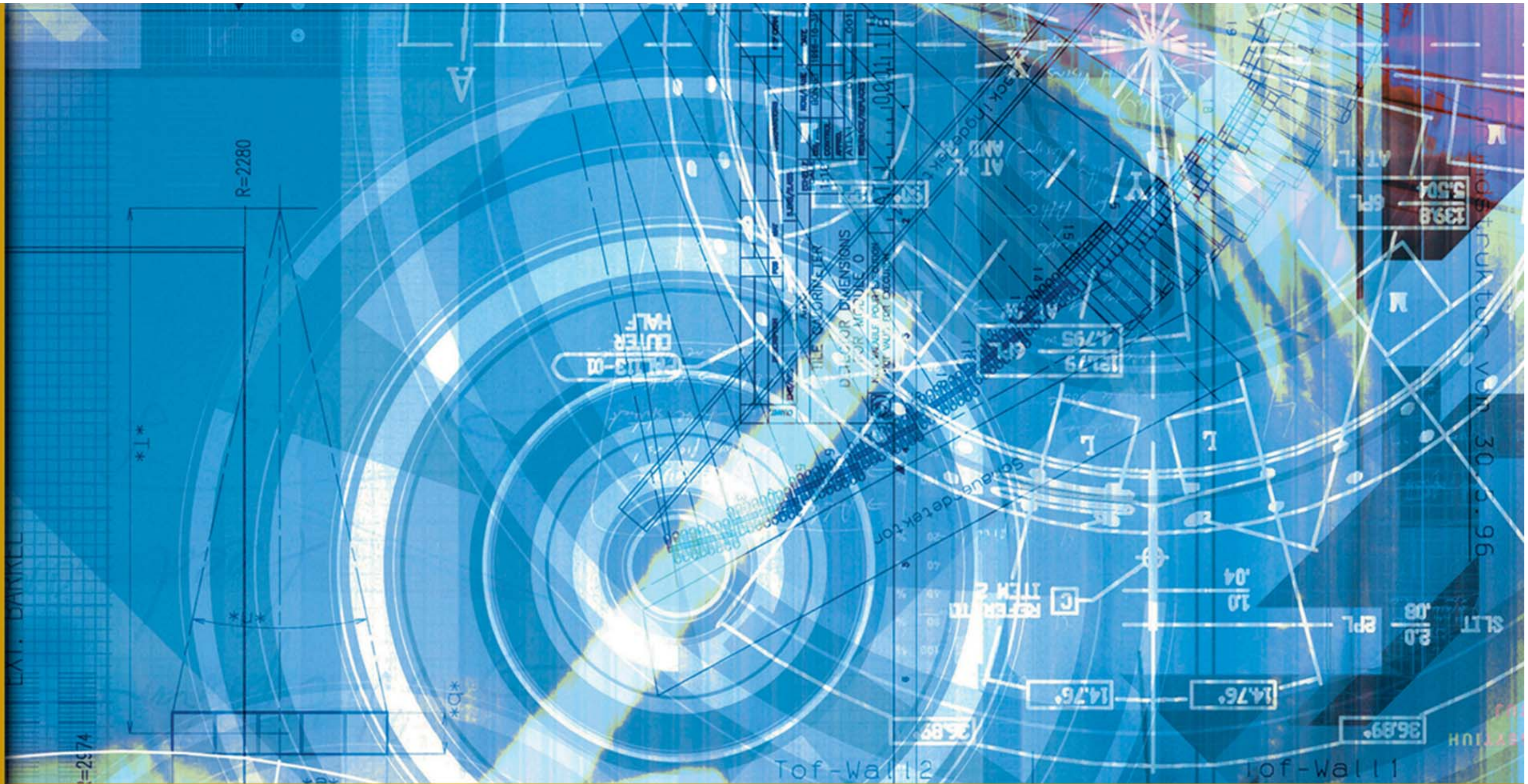


Fat Tailed Distributions For Cost And Schedule Risks

presented by:

John Neatrour jneatrour@mcri.com

SCEA: January 19, 2011

MCR

CRITICAL THINKING. SOLUTIONS DELIVERED.

Introduction to a Problem

- Risk distributions are informally characterized as fat-tailed if the variance is infinite or does not exist, otherwise they are thin-tailed
- In cost and schedule risk analysis it is almost always implicitly assumed that the risks are represented mathematically by thin-tailed probability distributions
- Why not test this assumption scientifically?
  - Get a data set
  - See if fat-tailed distributions are rejectable at say 95% confidence (a probability value of 0.05)
- Our experience is that cost and schedule are never under run. Perhaps that is because people “game the system” (act in self interest, not in project team’s interest) or the true underlying distribution for the random process is fat-tailed?
- Consider: in situations where people “gaming the system” is correct behavior (markets) it is already known that only fat-tailed models fit the data while thin-tailed models fail

Introductory Definitions 1/2

- We define risk as uncertainty in outcomes in random variables which may or may not be actualized
- Risks are generally represented mathematically by probability distributions
- For "continuous" outcomes (such as cost and schedule) the n^{th} moment of random variable X with respect to the cumulative probability distribution F or the probability density distribution f is given by the formula

$$\mu_n \equiv E(X^n) \equiv \int_{-\infty}^{\infty} x^n dF(x) = \int_{-\infty}^{\infty} x^n f(x) dx$$

- Thin-tailed distributions have moments for any possible value of n:
  - n = 0 gives the normalization, usually 1.
  - n = 1 gives the mean, $\mu = \mu_1$
  - n = 2 gives μ_2 in terms of the variance $\mu_2 = \sigma^2 + \mu^2$
  - Etc.

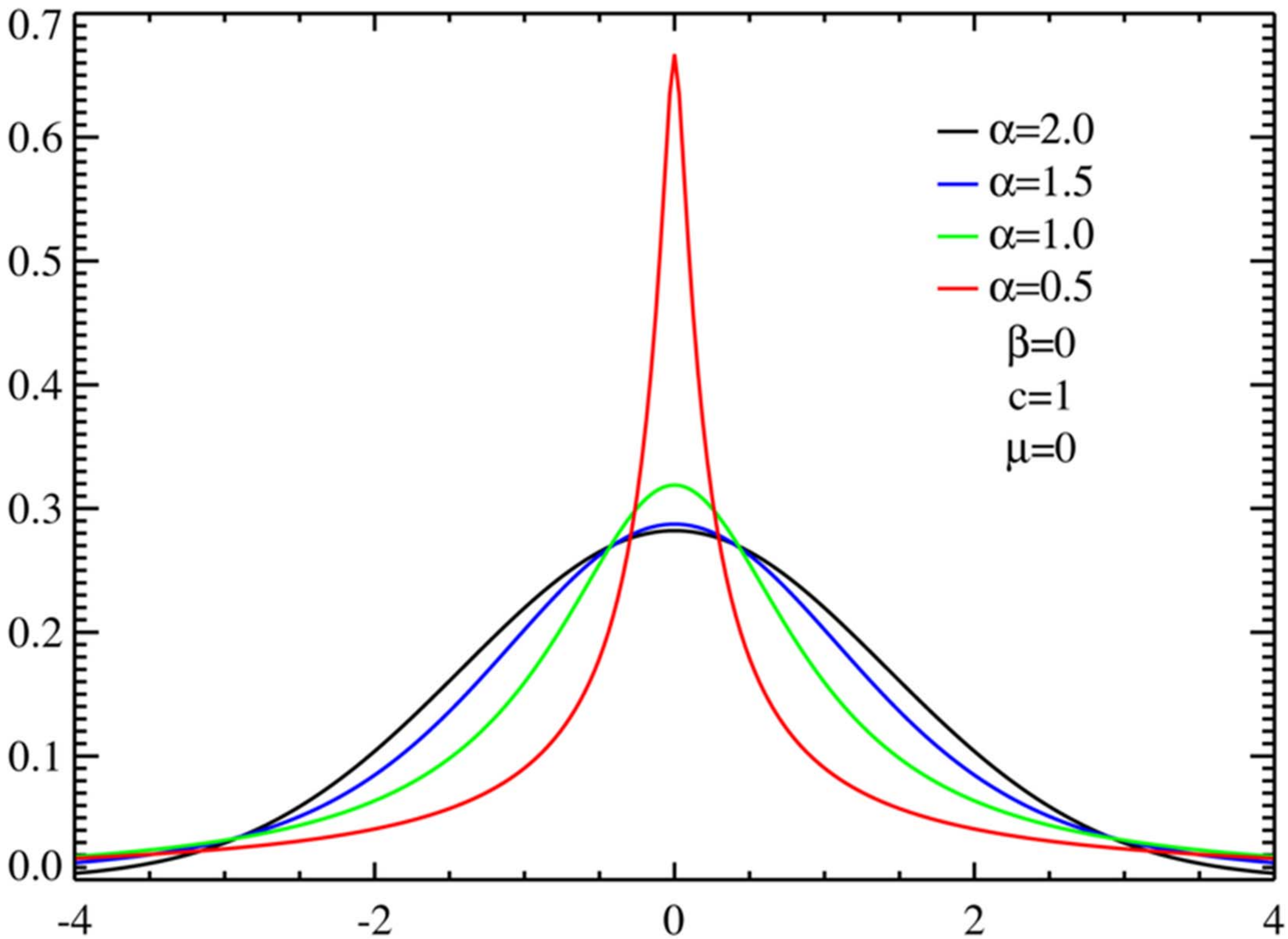
Introductory Definitions 2/2

- Fat-tailed distributions DO NOT have moments for $n \geq 2$, the integral becomes infinite, so it is impossible to calculate a variance for them
- Note: when the $n = 1$ moment also fails to exist then it is impossible to calculate the mean. So there is no expected outcome, a very bad risk!
  - Any portfolio of risks for which the mean outcome fails to exist begs to be dumped, restructured or dealt with severely!
- A popular probability distribution for risk analysis is the Lévy skew alpha-stable distribution which has four parameters:
  - A shape parameter that measures how fat the tail is, α
  - A skew parameter that measures the distribution asymmetry, β
  - A scale parameter that measures the width of the distribution peak, c
  - A shift parameter that locates the distribution's peak position, μ

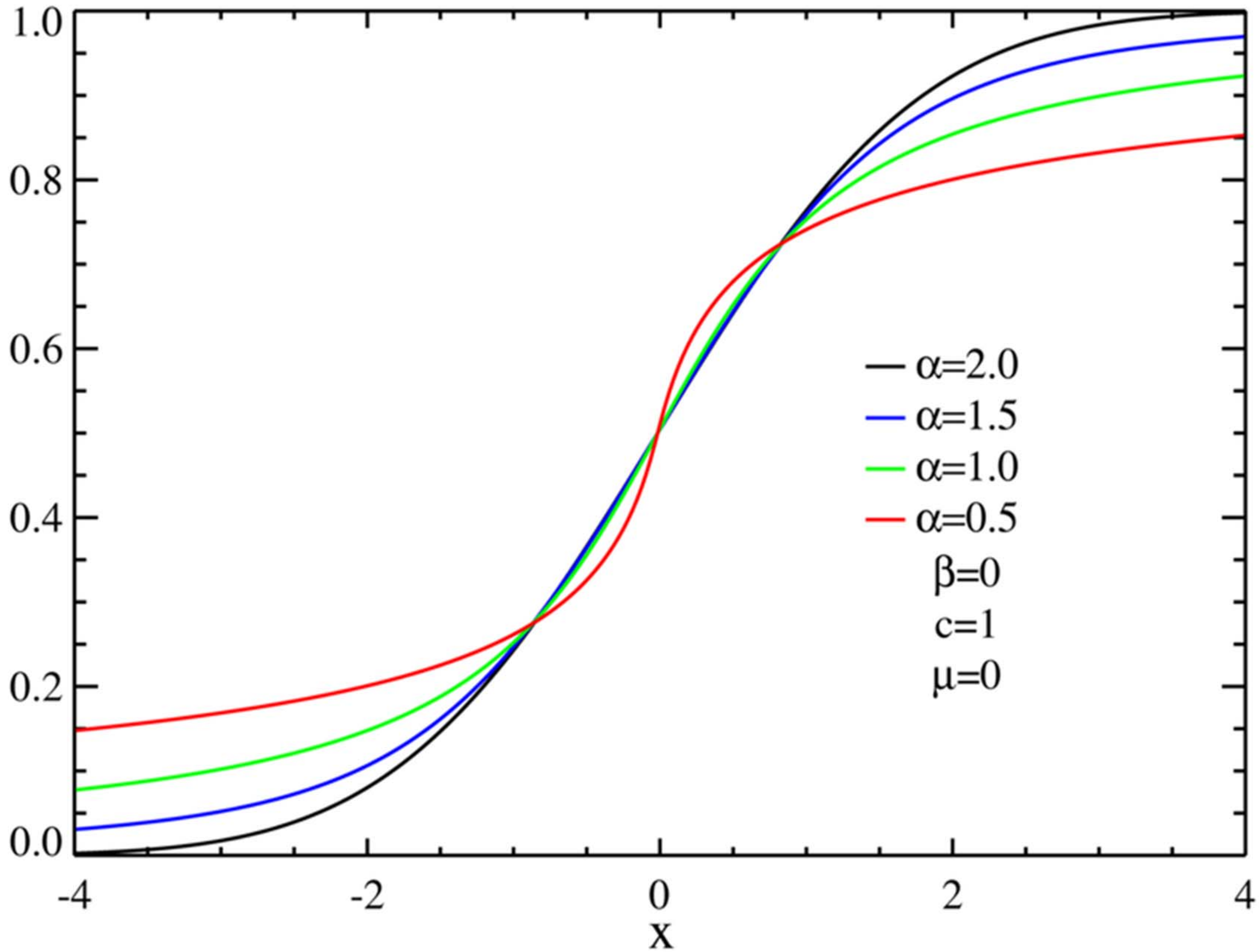
Lévy Skew Alpha-Stable Distributions

- For fat-tailed risk-portfolio analysis, we need a distribution that is stable: sums of distribution type must remain of that type
  - Sums of normal distributions are normal
    - OK for stability but not for thick tails
  - Sums of triangular distributions are normal
    - No good either for stability or for thick tails
- Most general stable fat-tailed distribution is the Lévy skew alpha-stable
  - For $\alpha = 2$, the distribution reduces to Gaussian with variance $2c^2$, mean μ, and the skewness parameter β has no effect
  - For $\alpha = 1$ and $\beta = 0$, the distribution reduces to Cauchy with scale parameter c and shift parameter μ
  - For $\alpha = 1/2$ and $\beta = 1$, the distribution reduces to Lévy with scale parameter c and shift parameter μ

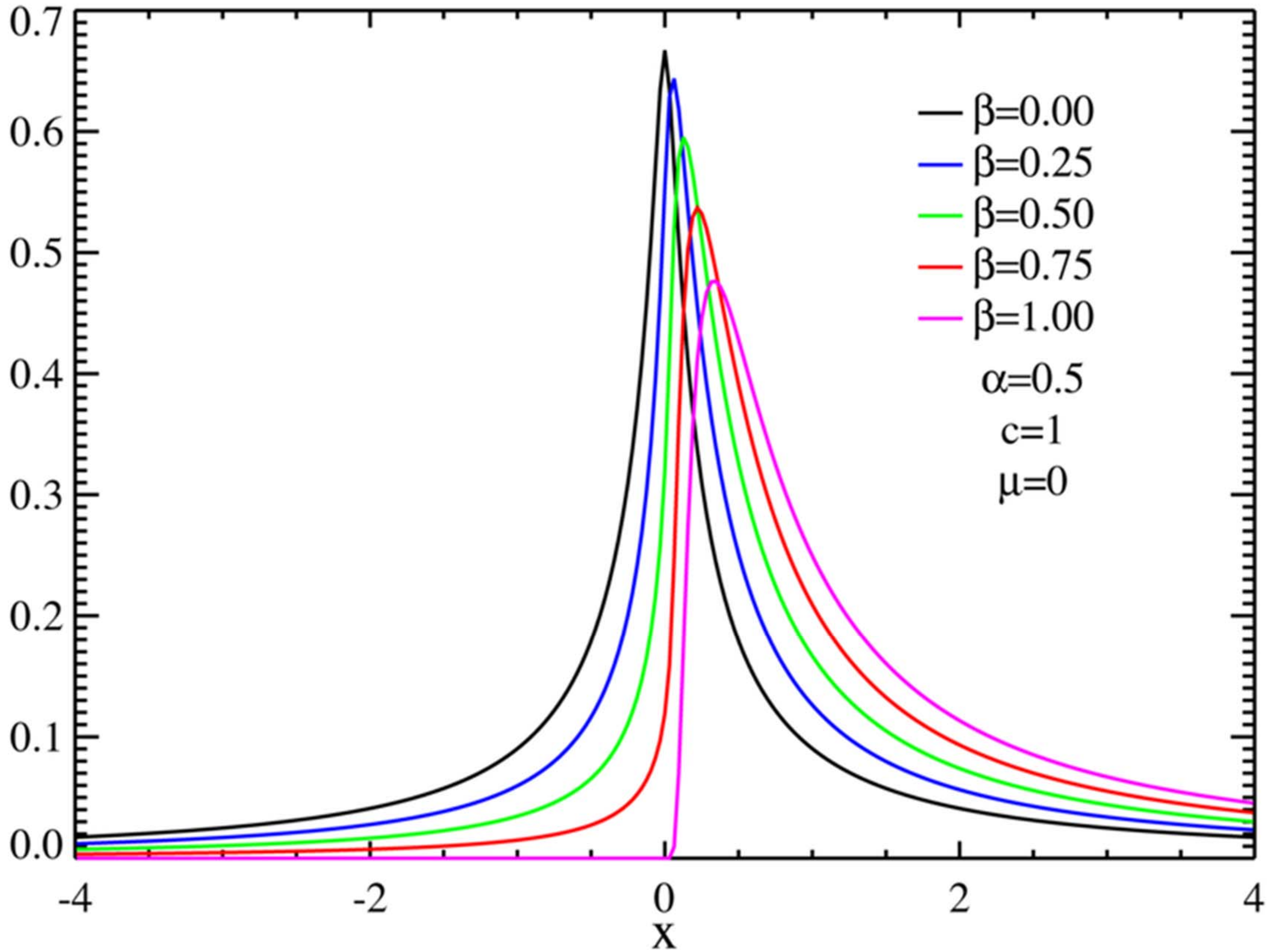
Symmetric Lévy Skew Alpha Stable Distribution Densities



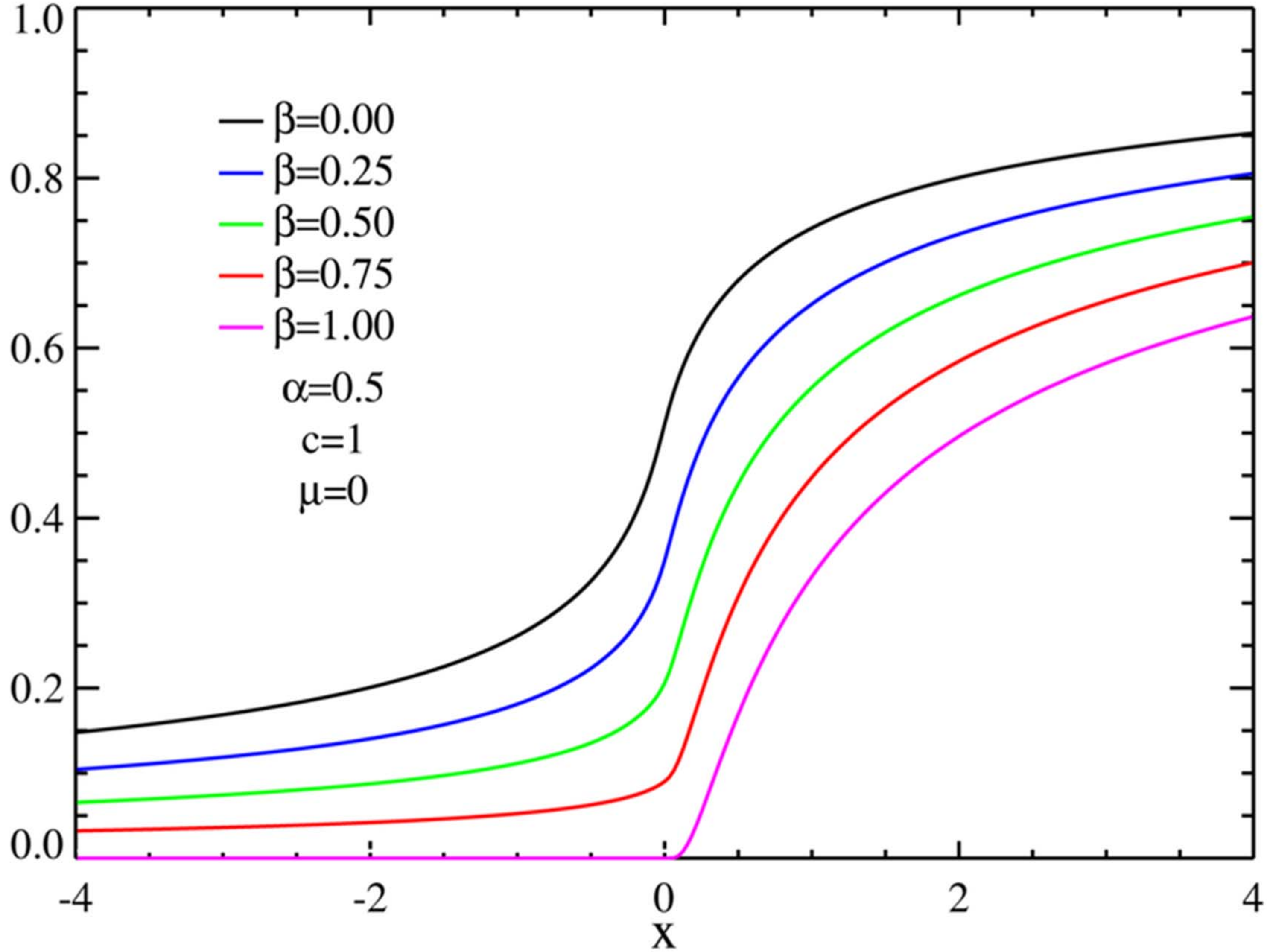
Symmetric Lévy Skew Alpha Stable Cumulative Distributions



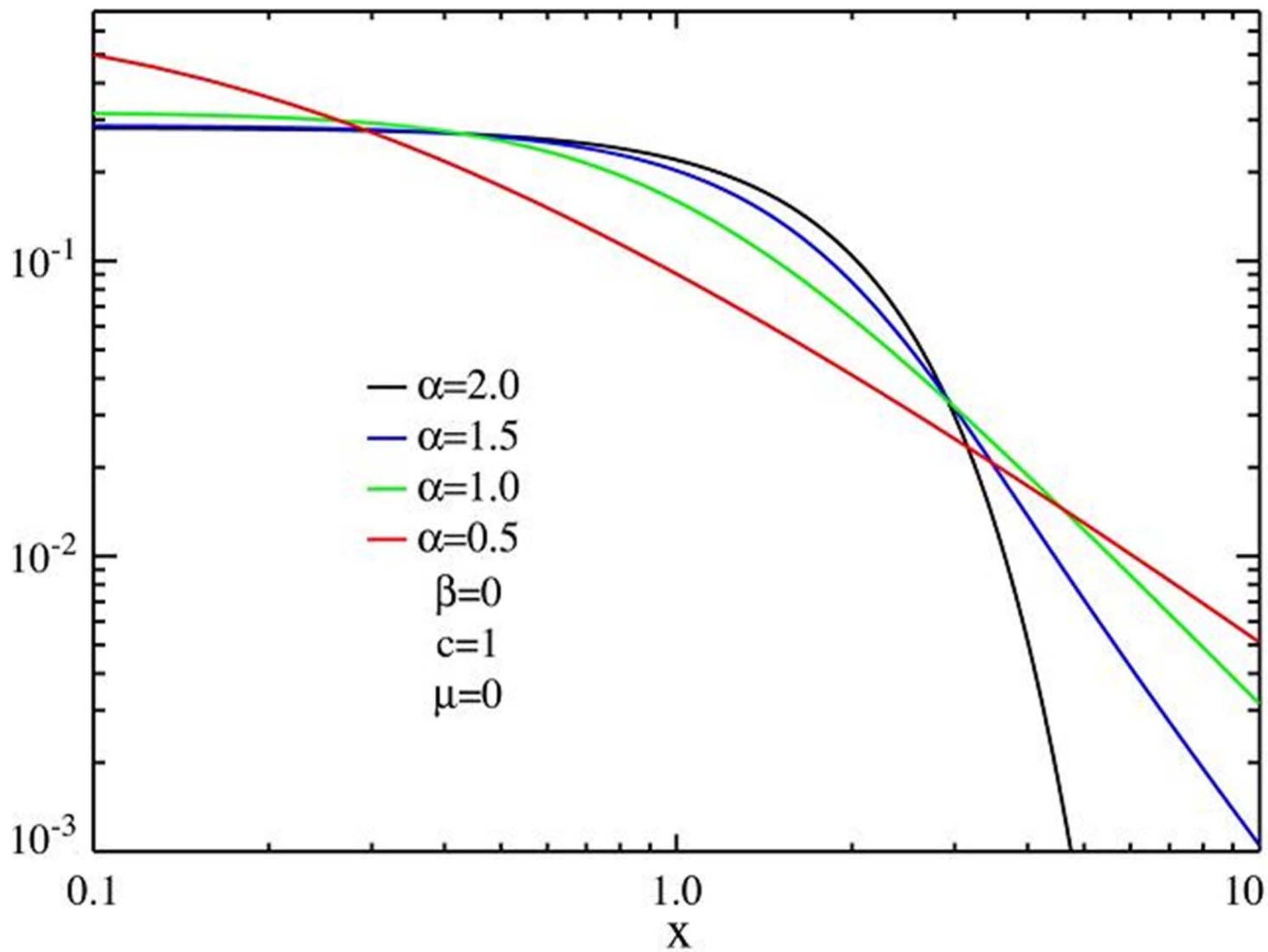
Non-Symmetric Lévy Skew Alpha Stable Distribution Densities



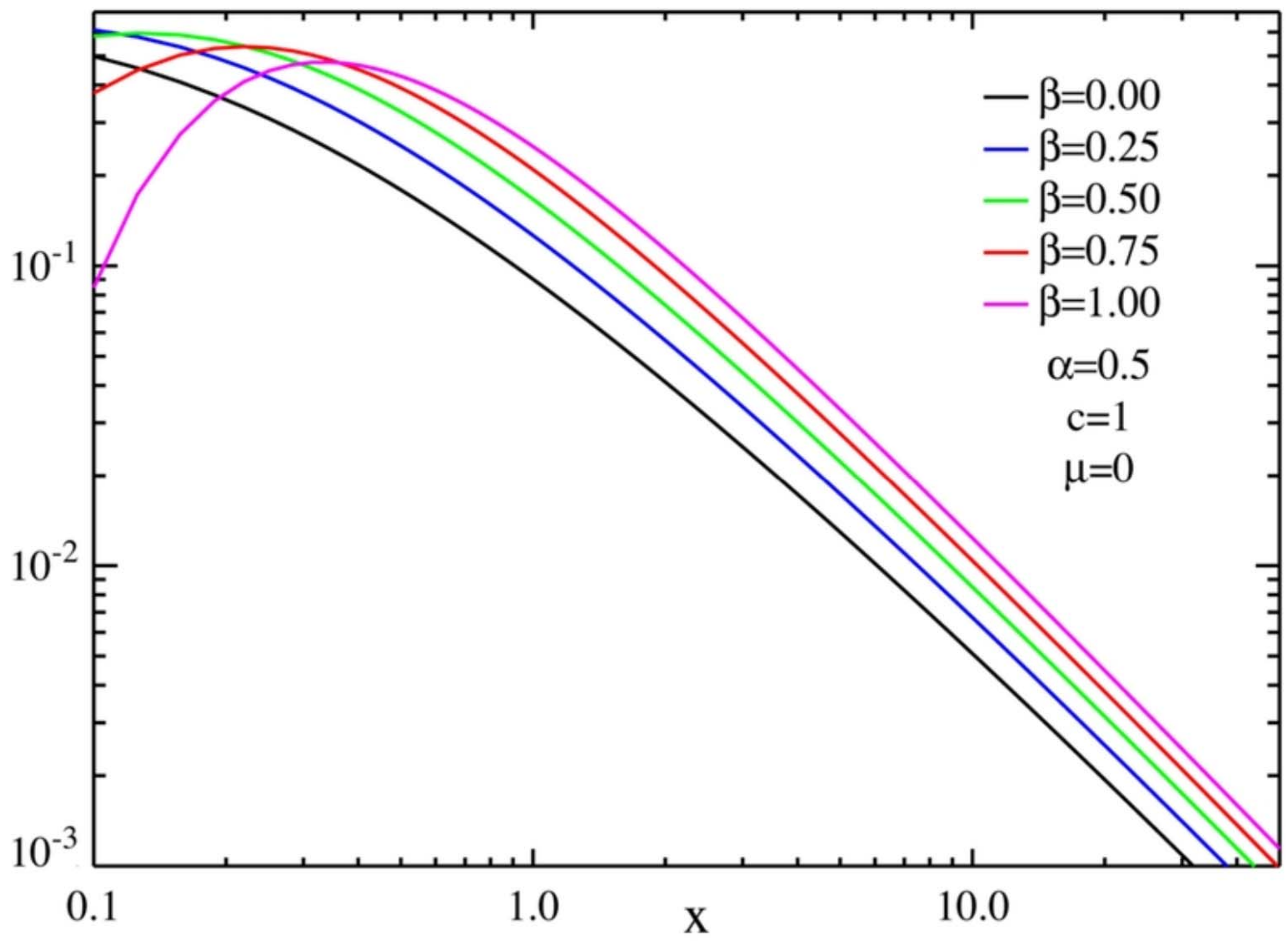
Non-Symmetric Lévy Skew Alpha Stable Cumulative Distribution



Symmetric Lévy Skew Alpha Stable Log-Log or Zipf Plot



Non-Symmetric Lévy Skew Alpha Stable Log-Log or Zipf Plot



The Problem in More Detail

- **Despite more sophisticated statistical methods, cost and schedule estimates, even at high confidence levels, tend to fall short of actual experience**
- **Consider:**
  - **Projects/programs and markets are treatable as random processes**
  - **Methods to estimate most likely outcomes (cost and schedule) are typically based on data from completed programs (data bias)**
  - **Cost and schedule risk analyses typically use thin-tailed distributions for each element which are summed statistically to produce the total program cost and schedule distribution**
  - **Financial and market modeling is moving away from thin- to fat-tailed distributions because thin tails have proven to be a poor fit to data**
  - **Insurance and Reinsurance probabilities of ruin models use fat-tailed distribution tools**
- **This experience suggests that Cost/Schedule Risk Analysis should adopt similar methods, letting project data (including failures) be the guide but also using knowledge of the nature of the process**

Solution Heuristics: 1 of 2

- Distinction: modeling random data is not the same as modeling random processes
- Data modeling assumes convenient functional forms and makes best fits to historical data
  - Functional forms might be arbitrarily chosen
  - Functional forms may have built-in bias
  - Goodness of fit is the only criterion (and is not falsifiable)
  - No theoretical justification is derived from the nature of the process
- Data modeling considers only project outcomes; process modeling considers how we get to the outcomes and provides testable ideas
  - Improve predictability and understanding by using knowledge of the nature of the process to guide data modeling
- Cost and Schedule Risk Analysis do not have a theoretical foundation so data fitting may seem to work (representing historical data for successes)
  - Astronomical example: Ptolemy, Tycho, and Copernicus versus Kepler and Newton
    - Agreement of data with an idea for fitting data does not prove the idea, it is really scientifically uninformative

Solution Heuristics: 2 of 2

- The point is that a good fit does not mean you have modeled reality
  - Ptolemy, Tycho, and Copernicus models all fit the data equally well
  - Kepler knew they were mutually contradictory and couldn't all be right so Kepler's model replaced them all by fitting the data more simply and paved the way for . . .
  - Newton and fundamental laws
- Only contradiction is scientifically informative: A. Einstein, "No amount of experimentation can ever prove me right; a single experiment can prove me wrong."
  - We can never expect improvements in cost estimating practice until we scientifically examine our assumptions
- In our context all previous cost-risk data fitting ***assumes*** that distributions are thin-tailed
  - So success in fitting is a circular argument, which is not valid
- We propose to see if fat-tailed distributions can be ruled out (we'll find they can't!)

CRITICAL THINKING. SOLUTIONS DELIVERED.

Project Process

- Consider that a project starts with no budget expended, no time passed and no requirements satisfied
- Treat Cost and Schedule as random variables
- Execution of the project develops a history of expending actual budget (in time and money) to obtain values (requirements satisfied) that are measurable (earned values)
- The nature of project execution limits the types of processes that are suitable as models
- In turn this rules in/out certain statistical models and techniques as compatible/incompatible with the processes.
- There are discontinuities in the realization of values against the prices paid. (Completions of subsystems and tasks, changes in requirements, failures in tests)
- Project behavior is analogous to the behavior of markets and other open human behavioral models in all the above aspects [Refs. 1-4]

Random Processes

- Distinguish two types of random (stochastic) processes by how the random variables evolve in time [Ref 5]
- Continuous random processes / Diffusion processes / Drift-Diffusion
  - Characterized by smooth changes that vary in time
  - Example: concentration at a point that then spreads or spreads and drifts
  - Modeled by thin-tailed distributions (e.g. normal)
- Jump processes / Jump-Diffusion processes
  - Characterized by jumps of random variables in time
  - Modeled by fat-tailed distributions (e.g., Lévy)
- Projects/programs…
  - Characterized by jumps in value delivered versus cost
  - Exhibit jumps due to changes in requirements
  - Have risks that are often best represented by discrete events (e.g., test success/failure, external political events)
- Therefore projects/programs should not be modeled with thin-tailed distributions

Jump-Diffusion Processes, Chart 1 of 2

MCR CRITICAL THINKING. SOLUTIONS DELIVERED.

- **Jump-Diffusion Processes or Lévy Processes are modeled by Lévy skew alpha-stable distributions and three component types of random changes:**
  - **Drift: a steady motion corresponding to both level-of-effort tasks and continuous progress in schedule (no "time off") and cost expenditures with regard to real physical time**
  - **Diffusion: a random deviation from the drift due to risk that can be represented as continuous variables such as prices, productivities and cost/schedule cost drivers with continuous values**
  - **Jumps: discontinuous changes due to duty cycles (scheduled "time off"), discrete risks (launch windows, test failures), discrete external events (changes in requirements and budget).**

Jump-Diffusion Processes, Chart 2 of 2

- Distributions have four parameters: position, spread, asymmetry, and shape (tail thickness).
- Except for special cases, the variance, skewness and kurtosis don't exist
- More flexible in modeling jump-diffusion processes than two-parameter distributions: the ratio of the probabilities to the right and to the left of the mode is adjustable
- Intuition suggests that for small data sets two-parameter lognormal distributions will work as well as four-parameter Lévy skew alpha-stable distributions
- We expect that sufficiently large data sets will show inconsistencies with lognormal modeling – incompatible moments will appear
- Large data sets from financial markets show modeling with Lévy skew alpha-stable distributions to be superior to lognormal distributions (data fitting and bailouts)

Data Description

- Analysis based on the cost growth of programs from the initial estimate at System Requirements Review or Critical Design Review to the actual cost at launch
- Majority of data gathered from two sources
  - May 2004 GAO Report
  - NASA CADRes
- NASA Phase E costs not included
  - Assumed this was the case where level-two data not available
- Costs were converted to FY08$ from TY$ using 2008 NASA New Start inflation indices

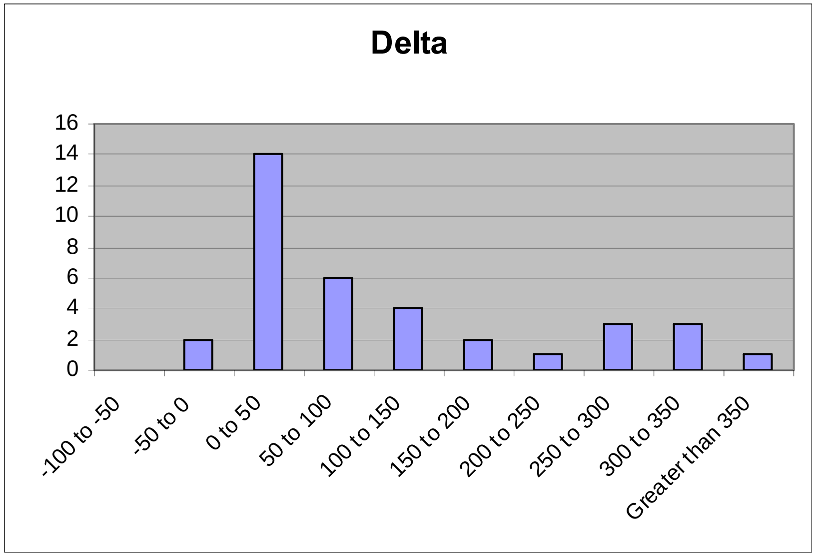
Data Problems

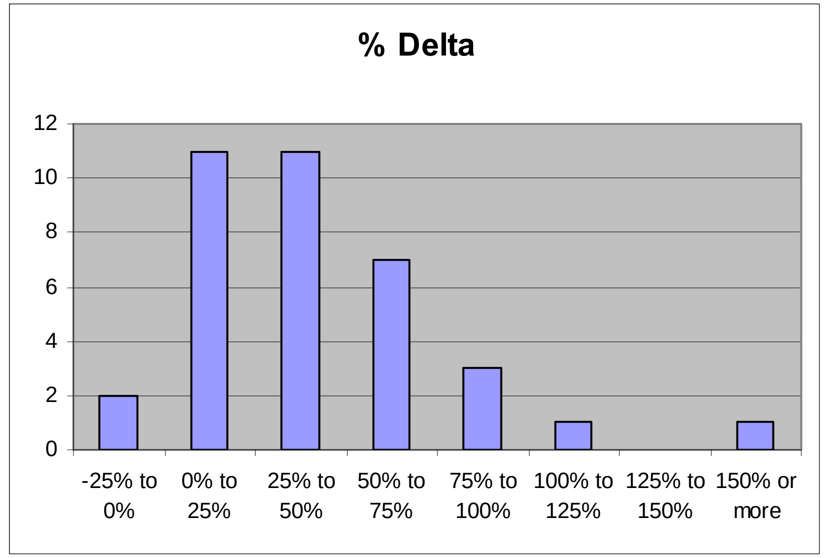
- Data are biased towards successful or completed projects
  - Methodology should include effects of an unbiased sample that contains failed or canceled projects
  - EV techniques can be applied to cost and schedule data to derive estimates to complete
- Relatively small sample size – need more data including …
  - More from NASA space segment
  - Non-NASA space
  - Non-space DoD
- Confidence intervals (CIs) for moments expected to scale with inverse square root of sample size [Ref 6]:
  - CI(Skewness) ~ z*sqrt(6/N)
  - CI(Kurtosis) ~ z*sqrt(24/N)

A Promising First Analysis

- **An example of an empirically derived fat-tailed distribution is displayed on the next charts**
  - **Based on 36 NASA missions**
  - **Shows an exponential relationship indicating fat-tailed distributions.**
- **For project outcome Deltas**
  - **Sample skewness is 3.3, 95% confidence [2.5, 4.1]**
  - **Sample kurtosis is 14.2, 95% confidence [12.6, 15.8]**
- **For project outcome %Deltas**
  - **Sample skewness is 2.1, 95% confidence [1.3, 2.9]**
  - **Sample kurtosis is 7.0, 95% confidence [5.4, 8.6]**

Histogram for NASA Missions





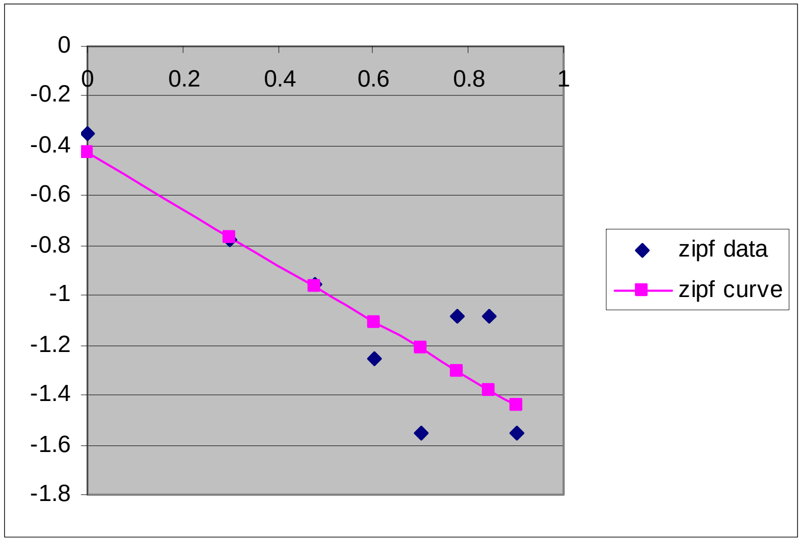
- **Histograms indicate frequencies of project outcomes in terms of absolute TY$ (2008) and in percent deviation from the expected outcome (CDR estimate)**
- **The data are highly skewed with a heavy right tail suggested visually**

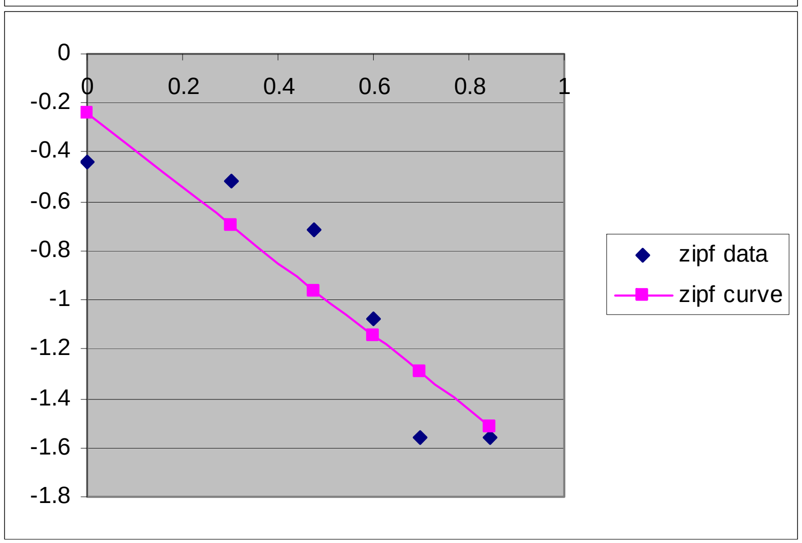
CRITICAL THINKING. SOLUTIONS DELIVERED.

Testing the Data

- A variety of techniques was used to help determine whether the data could be classified as "fat-tailed"
  - Plots
    - Quantile-Quantile (Q-Q) plots – scatterplots of the actual quantiles of the data against expected quantiles, given a particular distribution (normal, lognormal, etc)
      - Deviation from a 45° angle straight line indicates that the assumption that the data following this distribution may be incorrect
    - Zipf plots – the drop-off in probability as data points get further from the mean
      - A steep drop-off indicates a thin-tailed distribution, while a more moderate drop-off indicates a fat-tailed distribution
  - Statistical Tests
    - Pearson's Chi-Sq Test – compares actual and expected frequencies in user-defined bins
    - Kolmogorov-Smirnov Test – based on the Q-Q plot and uses the maximum deviation between actual and expected quantile as the test statistic

Zipf Plot for NASA Missions

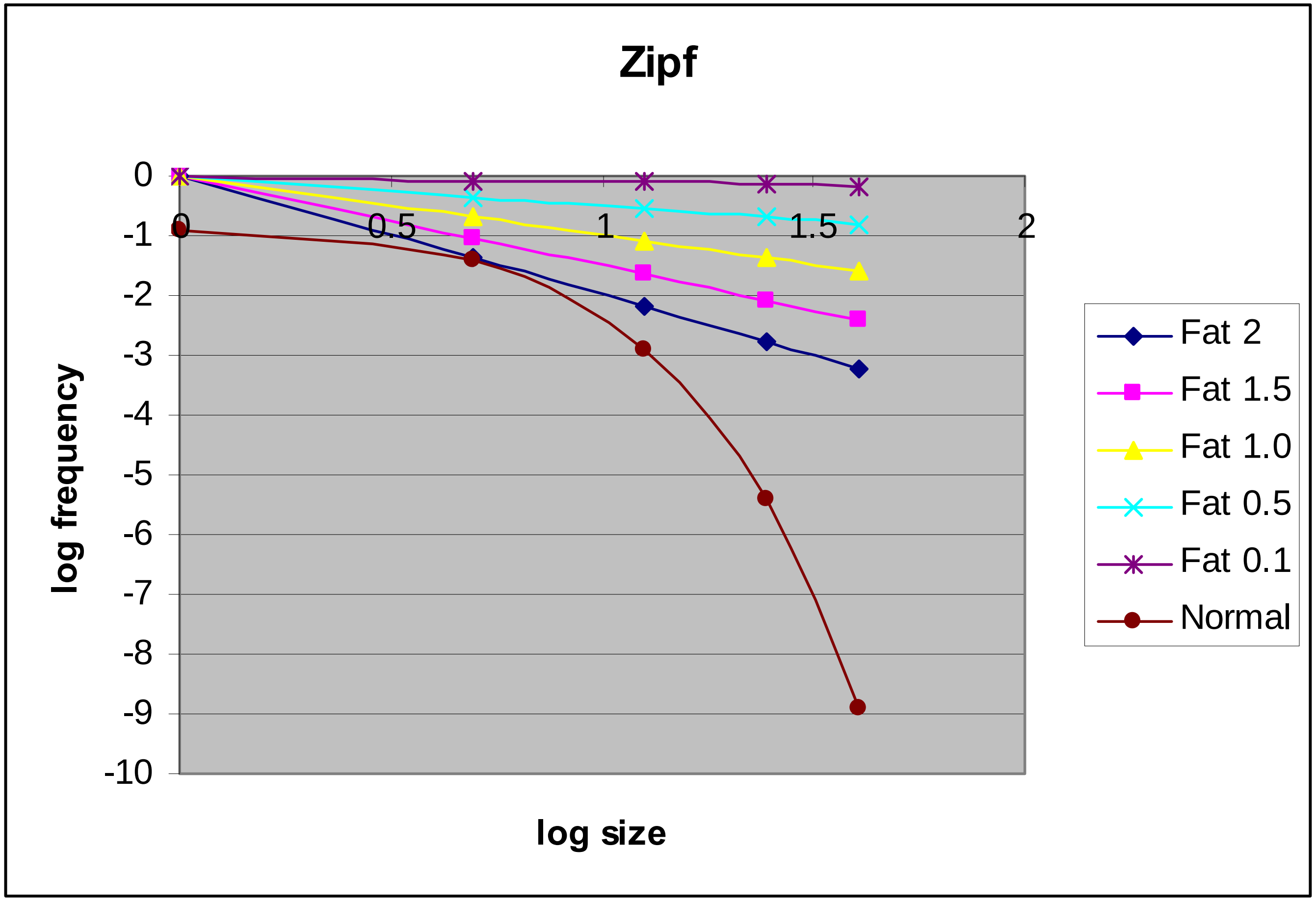




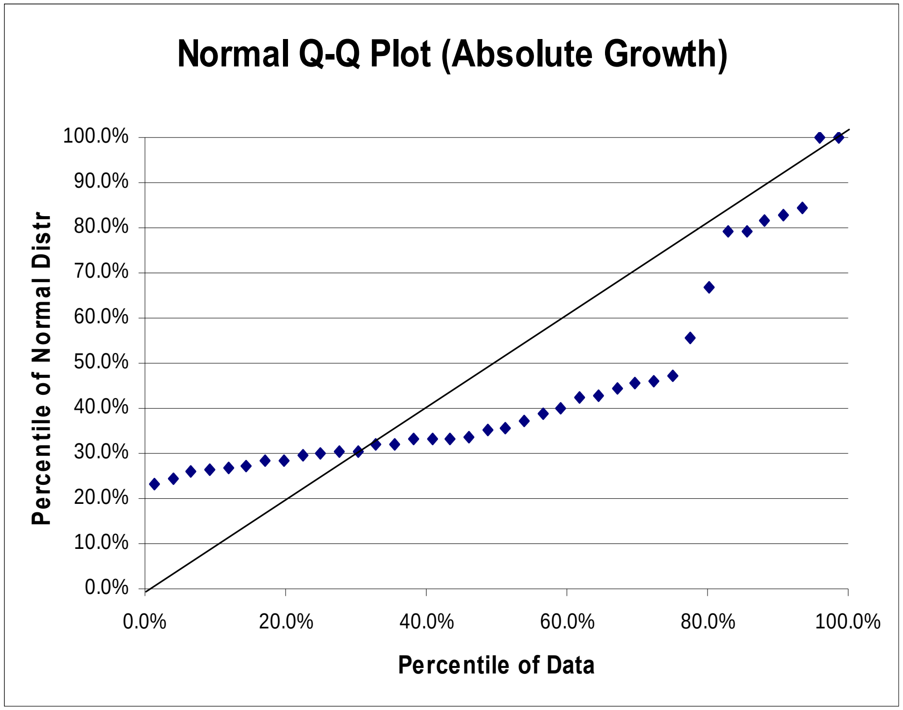
- Linear trend contradicts thin-tailed assumptions
  - Vertical axis is log frequency
  - Horizontal is excursion log magnitude
- Slope of curve implies a shape parameter alpha = 1.12 for Delta and alpha = 1.50 for %Delta
- Best linear fit is drawn through data (Solver Excel add-in used).

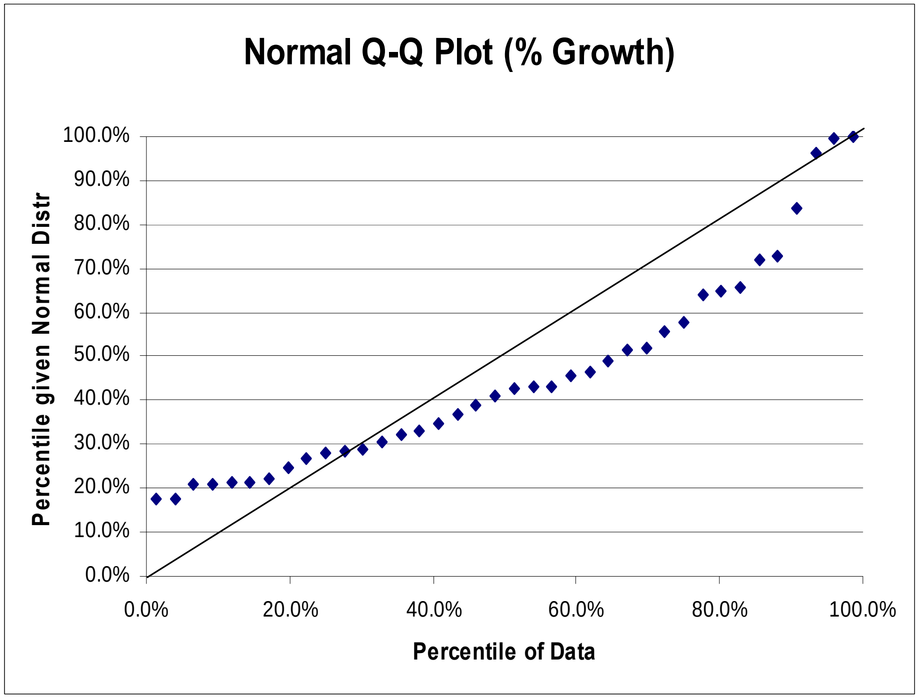
Zipf Plot for Normal Distribution

Normal distributions have a completely different Zipf plot.



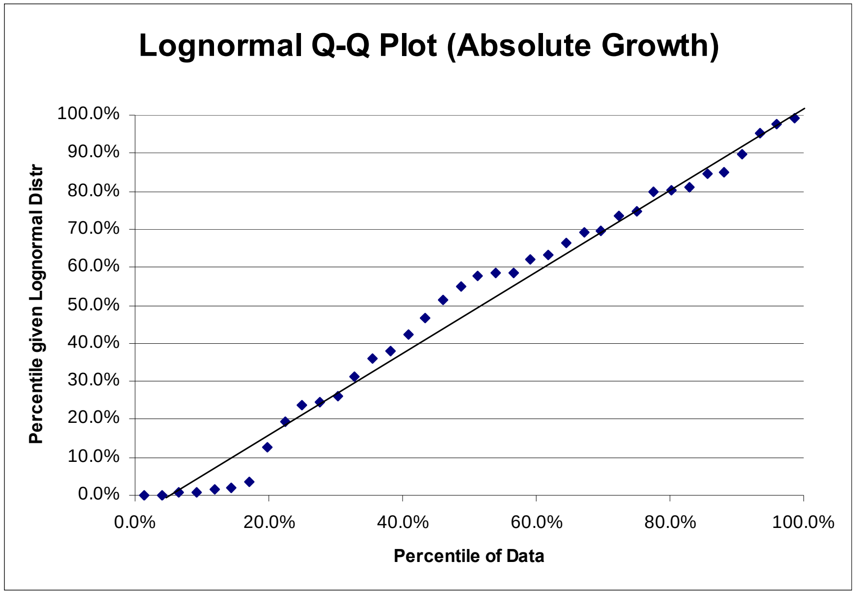
Normalized Data and a Thin-Tailed Distribution for NASA Missions

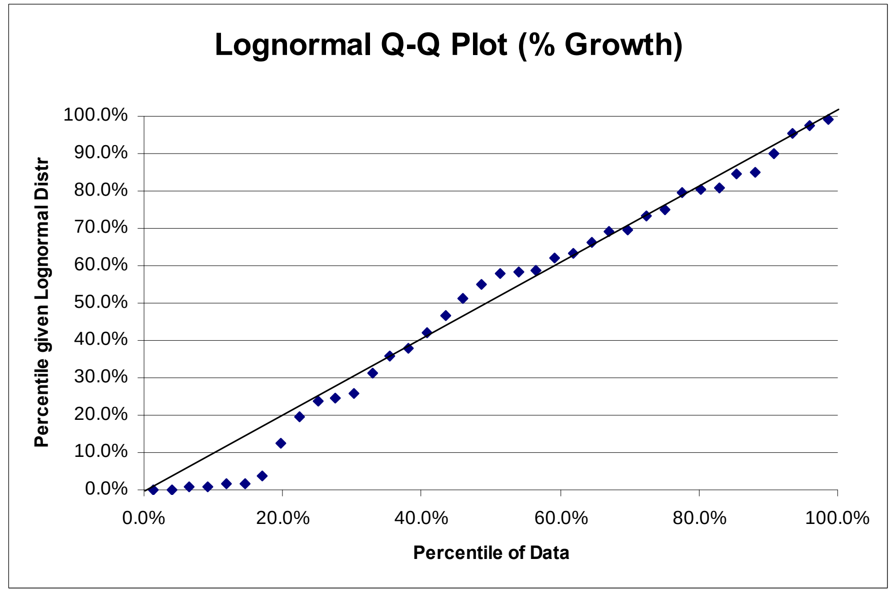




Q-Q plots show normal distribution fits poorly when mean and variance of the distribution are estimated from the sample

Normalized Data and a Fat-Tailed Distribution for NASA Missions





- **Q-Q plots show data fit much better by a lognormal**
- **Delta lognormal has skewness = 5.7, kurtosis = 94.7**
- **%Delta lognormal has skewness = 3.1, kurtosis = 23.7**
- **In both cases the consistent moments are outside the confidence intervals for sample estimates – Is this a problem?**

Pearson Chi-Square Test

- Tests whether two binned distributions have the same underlying distribution function by testing against a null hypothesis that the user-proposed distribution is correct
- Weakness: depends on binning scheme
- Chi-Square Test Statistic is as follows:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i},$$

  - X^2 = the test statistic that asymptotically approaches a χ^2 distribution
  - O_i = an observed frequency
  - E_i = an expected (theoretical) frequency, asserted by the null hypothesis
  - n = the number of possible outcomes of each event
- For all our tests we require 95% confidence or a probability level of 0.05 to reject the null hypothesis that the distribution is successfully fit

Pearson's Chi-Square Test for Delta and Lognormal Fits

- Lognormal Distribution not applicable to negative variables - must artificially shift the Deltas to be positive
- Test statistic is 4.83 with 5 degrees of freedom (8 bins – 2 estimated parameters – 1)
- Probability of data as extreme as observed given a lognormal distribution is 0.086
- Null hypothesis that shifted lognormal is correct distribution <u>can</u> be rejected at the 10% likelihood level but <u>cannot</u> reject at 5%
- We prefer to require 5% likelihood or 95% confidence level

Delta	Count	Lognormal Expected	Chi Sq
0 to 50	12	10	0.19
50 to 100	8	10	0.18
100 to 150	6	6	0.01
150 to 200	2	3	0.34
200 to 250	1	2	0.39
250 to 300	1	1	0.06
300 to 350	5	1	2.78
> 350	1	3	0.88
SUM	36	36	**4.83**

Pearson's Chi-Square Test for %-Delta

- **Lognormal Distribution not applicable to negative variables - must artificially shift the %Deltas to be positive**
- **Test statistic is 3.06 with 4 degrees of freedom (7 bins – 2 estimated parameters – 1)**
- **Probability of data as extreme as observed given a lognormal distribution is 0.190**
- **Null hypothesis that lognormal is correct distribution <u>cannot</u> be rejected**

% Delta	Count	Lognormal Expected	Chi Sq
0% to 25%	11	6	1.71
25% to 50%	13	16	0.35
50% to 75%	7	9	0.16
75% to 100%	3	3	0.02
100% to 125%	1	1	0.04
125% to 150%	0	1	0.51
150% or more	1	0	0.27
SUM	36	36	**3.06**

Kolmogorov-Smirnov Test

- Tests whether two distributions have the same underlying distribution function by testing the null hypothesis that the proposed distribution is correct
- Strength: independent of binning since no binning is required – uses a maximum distance measure
- Test whether lognormal distributions underlie the Delta and %-Delta sample data
- For all our tests we require 95% confidence or a probability level of 0.05 to reject the null hypothesis that the distribution is successfully fit

K-S Test against Lognormal Distributions

- Delta
  - K-S Statistic = 0.101
  - Critical Value (α = 0.05) = 0.321
  - Null hypothesis that lognormal is correct distribution cannot be rejected
- %-Delta
  - K-S Statistic = 0.113
  - Critical Value (α = 0.05) = 0.321
  - Null hypothesis that lognormal is correct distribution cannot be rejected

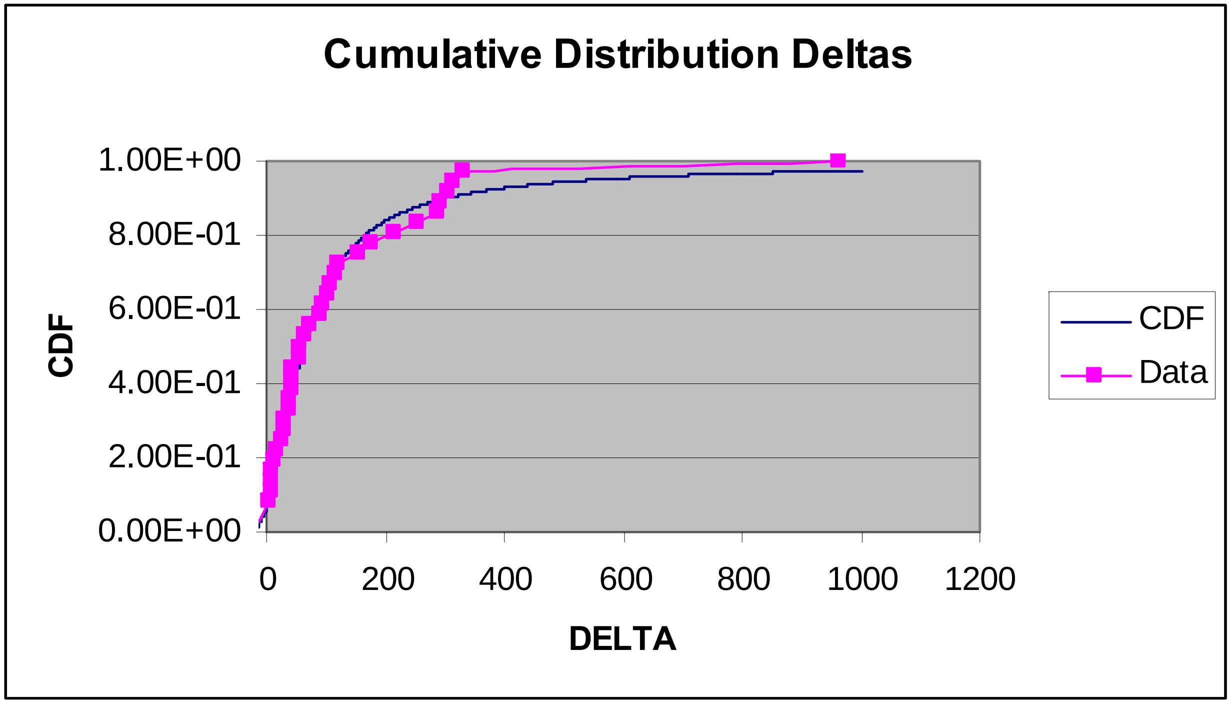
Analysis With Lévy Distributions

- **Adopted** John P Nolan's program "stable": [Ref 7]
  - http://academic2.american.edu/~jpnolan/stable/stable.html

- Issues:
  - Licensing
    - Not for profit use only
    - Free
  - Technical
    - Peer reviewed
    - Has some difficulties in performance for some values of parameters, but not fatal
    - Uses a parametrization which has nice numerical properties

Mission Delta Data

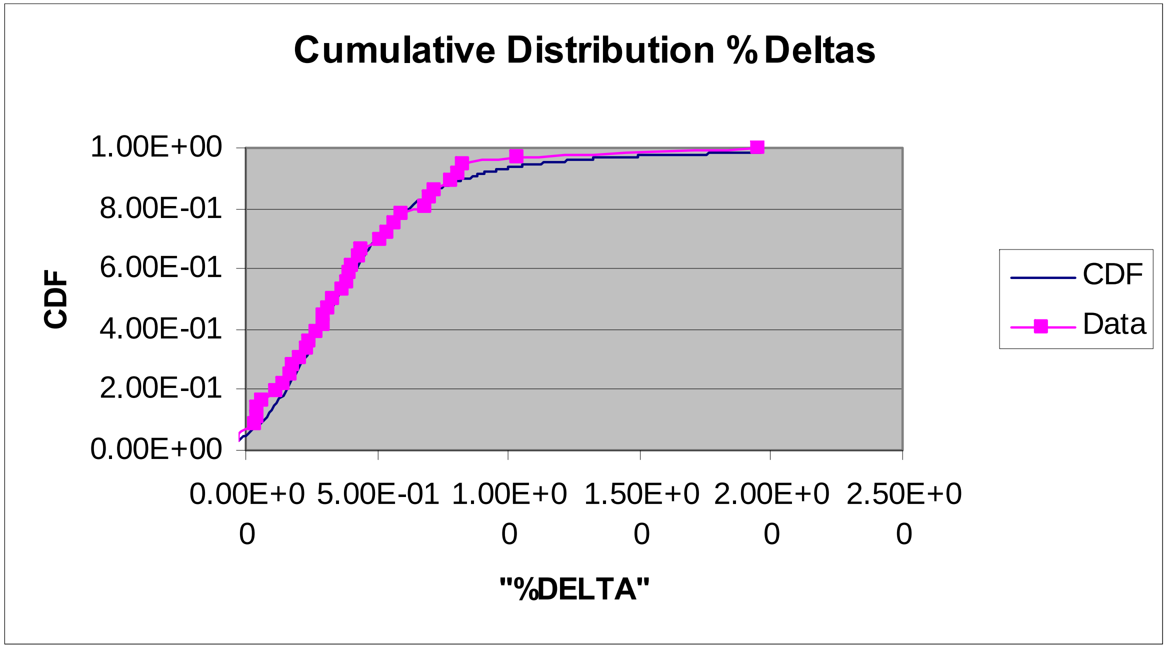
- No need to shift data arbitrarily to avoid underruns
- Good news: this procedure produces a good fit for the CDF
- Bad news: the shape parameter alpha has a value that has a 50% confidence that the portfolio is improperly formed
- Tentative Conclusion: reject modeling (and policies) based on mission Deltas
- Caveat: Program cannot estimate confidence intervals for beta when beta is near 1



Parameter	Estimated Value	Conf Int 95%
α	1	0.3306
β	1	0
c	36.6199	13.5643
μ	42.2318	20.6158

Mission %Delta Data

- No need to shift data arbitrarily to avoid underruns
- Good news: this procedure produces a good fit for the CDF
- Better news: the shape parameter alpha has a value that has a 95% confidence that the portfolio is properly formed
- Tentative Conclusion: modeling (and policies) based on mission %Deltas is reasonable
- Caveat: Program cannot estimate confidence intervals for beta when beta is near 1



Parameter	Estimated Value	Conf Int 95%
α	1.5	0.4178
β	1	0
c	0.173567	0.051029
μ	0.297631	0.099506

Pearson's Chi-Square Test for Delta and Lévy Fit

- **Need not artificially shift Deltas to be positive, the fat-tailed distribution can handle underruns gracefully**
- **Test statistic is 3.63: 4 degrees of freedom (9 bins – 4 estimated parameters – 1)**
- **Probability of data as extreme as observed given a Lévy distribution is 0.304**
- **Null hypothesis that Lévy is correct distribution cannot be rejected since the likelihood of data this extreme is 30.4%, so confidence is only 69.6%.**

Delta	Count	Lévy Expected Count	Chi Sq
0 to 50	16	15	0.03
50 to 100	6	9	0.60
100 to 150	4	4	0.00
150 to 200	2	2	0.00
200 to 250	1	1	0.00
250 to 300	3	1	1.00
300 to 350	3	1	1.00
Greater than 350	1	3	1.00
Sum	36	36	3.63

Pearson's Chi-Square Test for %-Delta and Levy Fit

- Lognormal Distribution not applicable to negative variables – must artificially shift the Deltas to be positive
- Test statistic is 0.97: 2 degrees of freedom (7 bins – 4 estimated parameters – 1)
- Probability of data as extreme as observed given a Levy distribution is 0.617
- Null hypothesis that Lévy is correct distribution cannot be rejected because of the high likelihood of fit at 61.7%, the confidence level would be a mere 38.2%

% Delta	Count	Lévy Expected Count	Chi Sq
0% to 25%	13	12	0.04
25% to 50%	11	13	0.17
50% to 75%	7	6	0.08
75% to 100%	3	2	0.20
100% to 125%	1	1	0.00
125% to 150%	0	1	0.00
150% or more	1	1	0.00
Sum	36	36	0.48

K-S Test Against Lévy Distributions

- **Delta**
  - K-S Statistic = 0.060
  - Critical Value (α = 0.05) = 0.321
  - Null hypothesis that Levy is correct distribution cannot be rejected

Parameter	Estimated Value
α	1
β	1
c	36.6199
μ	42.2318

- **%-Delta**
  - K-S Statistic = 0.056
  - Critical Value (α = 0.05) = 0.321
  - Null hypothesis that Levy is correct distribution cannot be rejected

Parameter	Estimated Value
α	1.5
β	1
c	0.173567
μ	0.297631

Conclusions, 1/3

- Pearson Chi-Square test <u>does not</u> reject hypothesis that lognormal distributions represent the data at 95% confidence or a probability of 0.05
- Kolmogorov-Smirnov test <u>does not</u> reject hypothesis that lognormal distributions represent the data at 95% confidence or a probability of 0.05
- Pearson Chi-Square test <u>does not</u> reject hypothesis that Lévy distributions represent the data at 95% confidence or a probability of 0.05
- Kolmogorov-Smirnov test <u>does not</u> reject hypothesis that Lévy distributions represent the data at 95% confidence or a probability of 0.05
- We prefer Kolmogorov-Smirnov to Pearson's Chi-square for the following technical reasons:
  - Neither are the number of bins >> 1, nor is the number of instances in each bin >>1, so the conditions for Chi-square to be good are not satisfied
  - So we take Kolmogorov-Smirnov as our prime criterion and use Chi-square only as a cross check

Conclusions, 2/3

- **Therefore: Fat-tailed distributions CANNOT be rejected as modeling project outcomes on the basis of this data sample**
- **The assumption that project outcomes are thin-tailed is therefore not grounded in these data**
- **Observation: a study of NASA data that rejected fat-tailed distributions [Ref 8] . . .**
  - **Did not consider Lévy distributions**
  - **Did not use techniques to allow for undersampling the tails**
  - **So that study is not relevant in discussing the fat-tail hypothesis**
- **Note: managing/modeling to Delta outcomes may be a poor idea since the tails may be so fat that a proper risk portfolio doesn't exist**
  - **i.e. there is no expectation value**
  - **This is a consequence of the shape parameter α being possibly < 1, the mean is infinite and does not exist**
  - **With not even an expectation value for the outcome the risk is very high!**

Conclusions, 3/3

- Note that while our sample is too small to reject the thin-tailed or fat-tailed hypothesis with 95% confidence, the fat-tailed distributions perform better than lognormal distributions under both statistical tests:

Pearson's chi-squared			
Delta		lognormal	Lévy
	probability	0.086	0.304
	confidence	91%	70%
%Delta			
	probability	0.190	0.617
	confidence	81%	38%

Kolmogorov-Smirnov (lower score is better)			
Delta		lognormal	Lévy
	K-S statistic	0.101	0.060
	95% confidence	0.321	0.321
%Delta			
	K-S statistic	0.113	0.056
	95% confidence	0.321	0.321

Next Steps

- Collect and analyze more data
  - We need to get a large enough data set to falsify either the thin-tailed or fat-tailed hypothesis
  - Collect data from non NASA sources (we expect they will look similar)
- Improve automated Lévy distribution analysis tools

References

1. Mandelbrot, Benoit B., *Fractals and Scaling in Finance,* (1997), New York, NY: Springer.
2. Rachev, Svetlozar, T., C. Menn, and F. Fabozzi, (2005), *Fat-Tailed and Skewed Asset Return Distributions*, Hoboken NJ: John Wiley & Sons.
3. Resnick, Sidney I., (2007), *Heavy-Tail Phenomena,* New York, NY: Springer
4. Embrechts, P., C. Klüppelberg, and T. Mikosch (1997) *Modelling Extremal Events,* Berlin: Springer-Verlag
5. http://en.wikipedia.org/wiki/Jump_process
6. Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). New York: Harper Collins.
7. http://academic2.american.edu/~jpnolan/stable/stable.html
8. Smart, Christian, "The Portfolio Effect and The Free Lunch," 2010 SCEA/ISPA Joint Annual Conference.