TASC

The Correct Use of Subject Matter Experts in Cost Risk Analysis

Richard L. Coleman (TASC), Peter J. Braxton (TASC), Eric R. Druker (BAH), Bethia L. Cullis (TASC)

Presented at the Naval Postgraduate School *Acquisition Research Symposium*

12 May 2010

The Correct Use of Subject Matter Experts in Cost Risk Analysis

Richard L. Coleman (TASC), Peter J. Braxton (TASC), Eric R. Druker (BAH), Bethia L. Cullis(TASC)

Presented at the DoE Cost Analysis Symposium

19-20 May 2010

Abstract

- Subject Matter Experts (SMEs) are commonly used in cost risk analysis (and in other fields as well) for values that either are not available in historical data or for which no appropriate analogy can be found. Problems commonly arise in two areas in particular: (1) when multiple experts give opinions on a single effect or entity and the inputs are not identical in distribution (which is almost inevitable); and (2) when a single expert provides distributional information that is intractable or suspiciously unlikely in its form (which is common).

- This paper will put forward a correct solutions in case (1), where the authors' experience shows that practitioners (and even experts) use incorrect solutions. It is important to note that the commonly exercised incorrect solution underestimates the dispersion, and thus the 80th percentile, in some cases by a large margin. The authors believe that their solution is rare and further are unaware of any use of the solution, and will recommend tenets to guide the practitioner. In preparation for the solutions laid out above, the authors will first describe the method of expert-based risk analysis, with the erroneous method for combining SME testimony, and then show the correction. An analytical treatment will quantify the impacts of the erroneous approach. The paper will also explain why the new method of conflating expert assessments is to be preferred to the common Delphi technique, which may fall prey to both anchoring and domination by a vocal minority.

- The paper will also briefly address case (2) by presenting common examples of problematic formulations and proposed resolutions. These include intractable specification of a triangular distribution; specification of a discrete categorical distribution when triangular was intended; and specification of a triangular with low and high values that are not the true extremes.

- In any situation, correct treatment of risk is important. In the current era, with 80th percentiles required for all weapon systems cost estimates by the Weapon Systems Acquisition Reform Act of 2009, and budgeting to the 80th percentile as the default practice, the correct determination of the distribution is more important than ever before.

TASC

Problem Statement

- Expert-based risk methodologies are a common approach to cost risk
- Expert-based risk methodologies are defined for the purposes of this paper as follows:
  - Notwithstanding that the cost estimate may be based on actuals, expert-based risk methods rely on elicitation of the parameters of the risk distribution from expert opinion
    - Typically triangles for cost risk
    - Typically Bernoullis for technical risk
    - May include normals
  - Single or multiple experts may offer estimates of a particular risk via some form of parameterization
- This paper will discuss two topics
  - The “best” approach to converting extrema and quartiles from expert opinion into risk distributions
  - The “best” approaches to conflating multiple views of the parameterization of a single risk
- For completeness, the paper will also discuss some difficult characterizations that they have encountered and the approach that they have evolved for “correcting” them
  - Inconsistent percentiles
  - Unusual characterizations
- This topic was addressed in general in a prior paper[1] under the rubric “Omission Of Elements Of Variability”
- A confession: A prior paper[2] espoused a form of combination of expert testimony that this paper now recommends against

1. *Are We at the 50th Percentile Now and Can We Estimate to the 80th?* Richard L. Coleman (TASC), Peter J. Braxton (TASC) , Eric R. Druker (BAH), Bethia L. Cullis (TASC), Christina M. Kanick (TASC)
2. *Risk Analysis of a Major Government Information Production System, Expert-Opinion-Based Software 2. Cost Risk Analysis Methodology*, N. L. St. Louis, F. K. Blackburn, R. L. Coleman SCEA/ISPA International Conference 1998, ADoDCAS 1998, Journal of Parametrics, June 1998, Awarded DoDCAS Outstanding Contributed Paper and Overall Best Paper Award SCEA/ISPA

The "Best" Approach To Converting Extrema And Quartiles From Expert Opinion Into Risk Distributions

Correcting Extrema and Quartiles for Truncation

- Our estimated distributions tend to be "too tight"[3,4]
  - Extrema
    - Without feedback, we provide values near the 20th %-ile and 80th %-ile when we are asked Min and Max
    - This can be improved, with feedback to the 10th and 90th %-iles
    - This can be improved by asking for more-extreme values:
      - "Astonishingly-low-probability outcomes" equate to the 0.1th %-ile and 99.9th %-ile
  - Quartiles
    - Without feedback, we give 25th and 75th quartiles that actually contain only 33% of the outcomes vs. the expected 50%
    - This can be improved with feedback to 43% vs. the expected 50%
  - Independent investigations of this over-tightness are remarkably consistent in the degree to which it occurs[3,4]
  - Our ability to probabilistically characterize the past or future or to estimate our certainty on general-knowledge facts are all about comparable[5]

3. *Judgment under uncertainty; Heuristics and biases,* Edited by Daniel Kahneman, Paul Slovic and Amos Tversky, Cambridge University Press, 1982, Chapter 21, A progress report on the training of probability assessors, Alpert & Raiffa
4. *An experiment in Probabilistic Forecasting,* Thomas A. Brown, R-944-ARPA, July 1973
5. *Judgment under uncertainty; Heuristics and biases,* Edited by Daniel Kahneman, Paul Slovic and Amos Tversky, Cambridge University Press, 1982, Chapter 22, Calibration of Probabilities: the state of the art to 1980 Lichtenstein, Fischhoff & Phillips

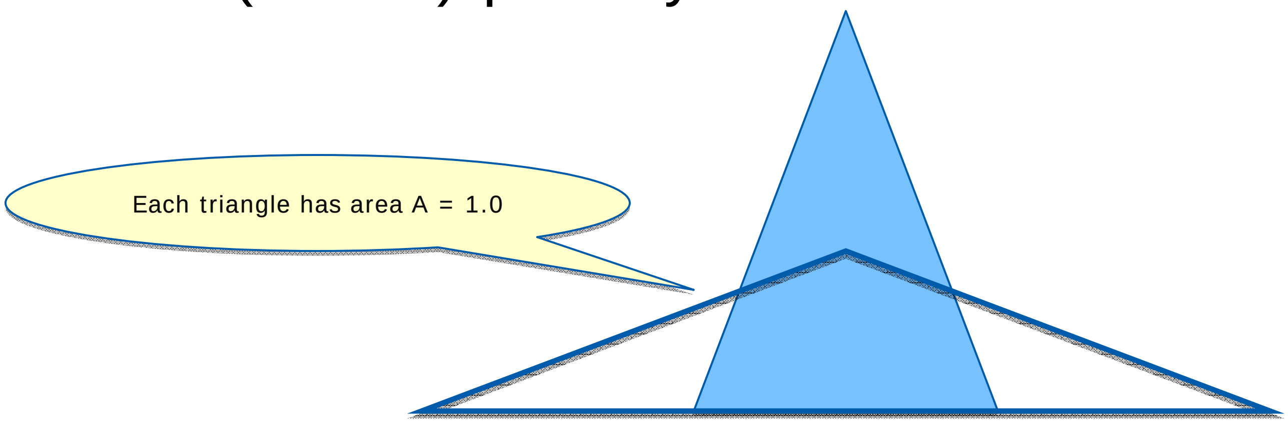
Correcting Extrema and Quartiles – Two Views

- Assume that experts will return 20th and 80th percentiles when asked for the full range
  - In other words, when given highs and lows, assume you are getting something more like standard deviations masquerading as extrema, it's not quite that bad, but it's close, it's about .316 of the real base*
  - This could be presumed to improve to 10th and 90th but only if the experts can be assumed to have gotten specific feedback about their accuracy at this task in the past
    - Note that this is not the same as saying they are very well qualified, it refers specifically to feedback training
    - We believe that practitioners have mistaken expertise for being trained and that this is why many practitioners believe experts provide 10th and 90th percentiles
- Although we don't typically ask for quartiles, we recommend assuming that a claimed 25-75 inter-quartile range is actually a 33-67 percentile range
  - This can be improved to a 28-72 range with specific feedback
- The two distortions above are not strictly coherent, meaning that they yield different corrections
  - The full range case is a greater understatement than the interquartile case
  - *The wider the confidence interval you ask for, the more the witness will understate it*
- When given expert testimony, therefore, it is appropriate to correct the testimony by adjusting the standard deviation or the end points using the two general results above, depending on the form given*

* See the backup

Errors of Extrema - Pictorially

- We note that experts appear to be providing approximately the 20th and 80th percentiles
- We know* that the 20th percentile occurs at a point that is √(1/10) = 0.316 of the base
  - The understatement of variance by experts is on the order of 2.5
- Pictorially, then, we are experiencing a reduction in distribution on the order of the blue (claimed) to the white (actual) portrayed below



* For the geometry of triangles with regard to percentiles and area, see the backup

The "Best" Approaches To Conflating Multiple Views Of A Distribution

Conflation of Expert Information

- Conflation refers to the combining of different (independent) views of a thing to arrive at a single (better, and more complete) view of it

- We seek to conflate expert testimony principally because we will arrive at a better estimate for the mean
  - But, what about the dispersion?

- Conflation is the most difficult problem for expert-based risk methodologies
  - This is not immediately obvious, but it is so
  - Dispersion is in turn the hard part of the conflation

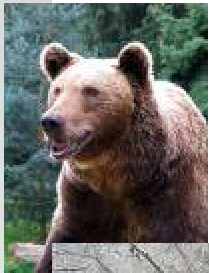
- Ad hoc conflations are often used for k experts each giving estimates for the same risk or WBS element, e.g.:
  1. Use the individual expert testimonies in each run of the Monte Carlo:
     a. Make k random draws from the k different distributions and average them[6]
     b. Make k random draws from the k different distributions with correlation and average them
  2. Derive the parameters of a single distribution from the parameters of the expert testimony and then Monte Carlo
     a. Make a new distribution with i) the mean of the k expert means and ii) the mean of the standard deviations, for normals[7], or the means of the respective end points for triangles [Average the Parameters]
     b. Make a new distribution with the average mode of the k experts and the lowest low and the highest high as end points
     c. Make a new distribution with the average mean of the k experts and the lowest low and the highest high as end points
  3. Sampling has been endorsed in the literature[7]
     - For each run of the Monte Carlo, pick the answer from a randomly selected expert who provided testimony

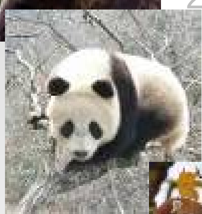
- We will examine each of these methods
  - In backup we prove that 1b and 2a are equivalent for symmetric triangles and we speculate that for asymmetric triangles there is no significant difference, and so there is nothing to separate these beyond ease of implementation

6. *Risk Analysis of a Major Government Information Production System, Expert-Opinion-Based Software Cost Risk Analysis Methodology*, N. L. St. Louis, F. K. Blackburn, R. L. Coleman SCEA/ISPA International Conference 1998, ADoDCAS 1998, Journal of Parametrics, June 1998, Awarded DoDCAS Outstanding Contributed Paper and Overall Best Paper Award SCEA/ISPA
7. *An experiment in Probabilistic Forecasting,* Thomas A. Brown, R-944-ARPA, July 1973

The First Question

- No single conflation method will work for the two possible scenarios that can confront the estimator





  1. “Single Reality”: There is a one (typically uni-modal) distribution, which we do not know, but which experts are presumed to know to some degree of accuracy
     - Example: What is your estimate for the GNP of Brazil for 2009?
     - Example: How big is a brown bear?
     - Example: What is the range of technical risk for the cost of the engine?
  2. “Multiple Realities”: There are k (typically uni-modal) distributions, we generally know neither k nor the individual distributions, but experts are presumed to know at least one each to some degree of accuracy
     - Example: How far away is your favorite planet? [there could be up to 9 answers depending on the inclusion of Pluto and Earth!]
     - Example: How big is a panda? [there is a lesser panda and a greater panda, but we don’t happen to know that and fail to specify]
     - Example: What is the cost risk for the engine on the F-35? [There is a main and an alternate engine, each has a range]

- This problem may seem silly, but it is not, and our choice of conflation methods depends on the case we believe to apply

- We will recommend approaches for both, but first, decide which case applies

- The amount of spread in your expert testimony will give you an idea whether single or multiple reality is more likely
  - We recommend against feedback or “drilling down” until after testimony is gathered because witnesses are notoriously vulnerable to witness leading, anchoring and all other sorts of mischief … you’ll never know

Richard.lewis.coleman@gmail.com, 703-615-4482,

Desiderata for Single and Multiple Reality Cases

- Each case dictates different characteristics for the conflation technique
- Single reality:
  - Best estimate for the mean
  - Best estimate for the dispersion
  - Best estimate for the distribution
- Multiple Realities
  - Best portrayal of the multiple choices we are confronted with
- We will discuss each in turn

The Preferred Methods

- We will describe the apparent preferred solution for each method after asserting them below
- Single Reality:
  - Average the parameters and correct for the understatement of extrema (using method 1b or 2a from an earlier slide)
- Multiple Realities
  - Sample from the experts after correcting each for understatement of the extrema
- If we cannot discern whether we are in Single Reality or Multiple Realities, we recommend sampling
  - Because this is more conservative, meaning it will have wider dispersion
- We reject the use of averaging answers on each iteration despite having used the method in a Best paper Overall[8] in 1998. To see why, we will show its characteristics and indicate why it is probably unsuitable.

8: *Risk Analysis of a Major Government Information Production System, Expert-Opinion-Based Software Cost Risk Analysis Methodology*, N. L. St. Louis, F. K. Blackburn, R. L. Coleman SCEA/ISPA International Conference 1998, ADoDCAS 1998, Journal of Parametrics, June 1998, Awarded DoDCAS Outstanding Contributed Paper and Overall Best Paper Award SCEA/ISPA

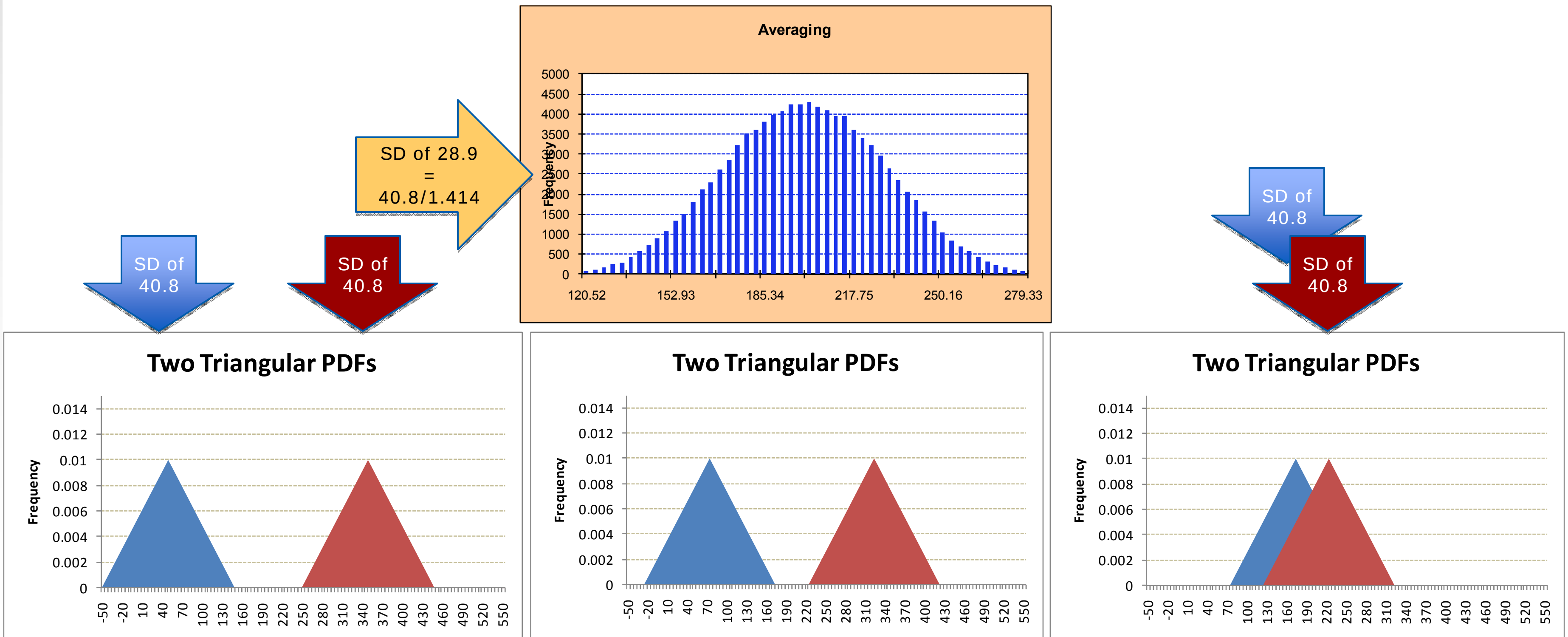
Recommendation - Single Reality

- The mean of the single reality not troublesome, almost any *reasonable* approach will yield the same mean
  - We use the word "reasonable" with trepidation
- The standard deviation presents the problem, since individuals are known to under-report, and some methods are vulnerable to distortions
- We recommend averaging parameters of the expert testimony because it is clear what is happening
- Correct each expert's testimony for truncation of the standard deviation, or correct the average, there is no obvious difference in the order of the operations
  - Techniques for correcting the standard deviation were shown in the first part of the paper

Conflation: Averaging on Each Iteration (1a)

- Averaging on each iteration can have an unexpected result: Three very different sets of testimony by two experts will produce exactly the same picture
  - This is not obvious at first, but it is so

- The standard deviation of k identical but scattered triangles, with SD = s, when iteration-averaged will produce a standard deviation s/√k
  - The SD of the conflation can be thus be arbitrarily small, if k is sufficiently large
  - This does not comport with our desire that the SD be well modeled
  - Correction for k can be achieved by a spreading with √ k but this is likely to be done wrong or omitted altogether, and at best would require row-by-row corrections
  - Correction for expert truncation can be achieved by treating the end points as if they were 20/80 points, this can be done before or after conflation



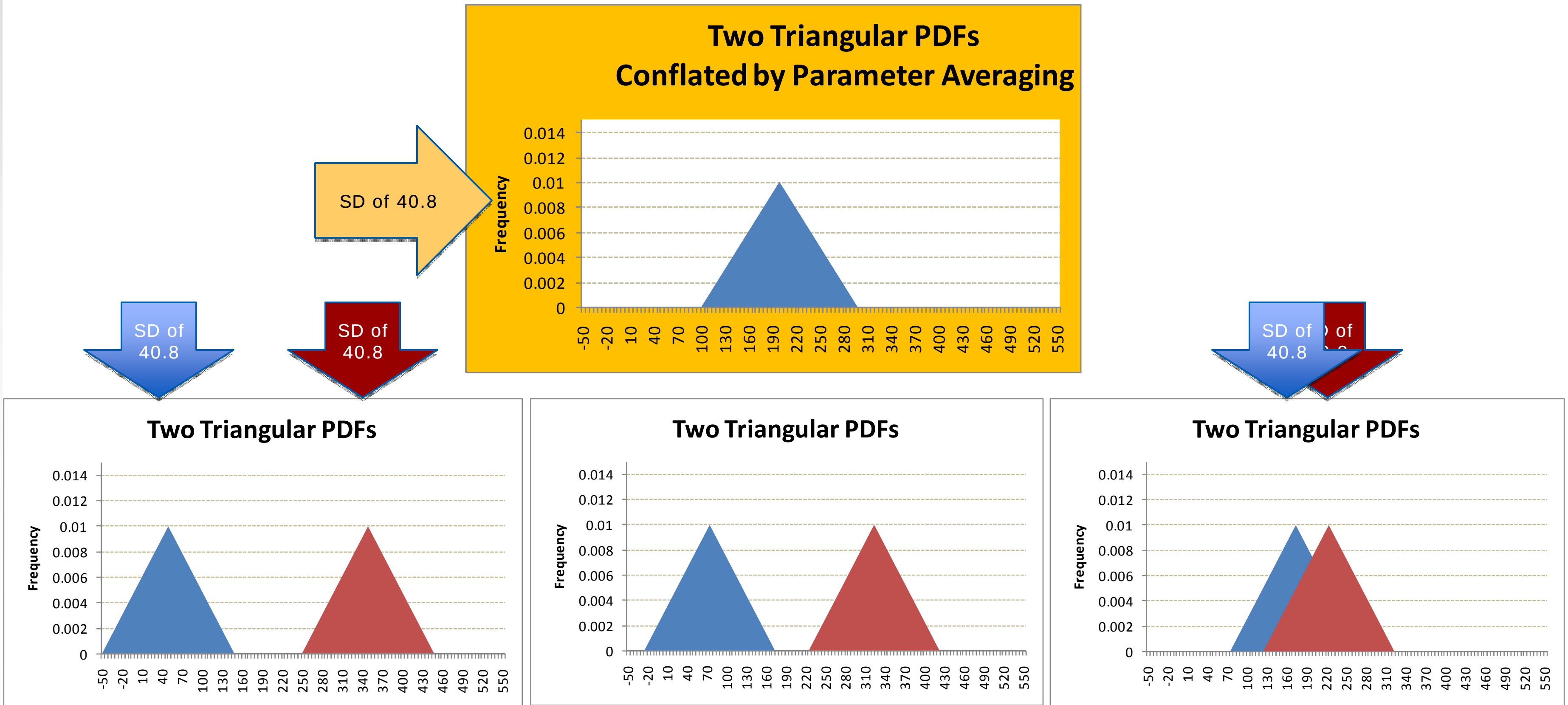
TASC

Conflation: Averaging on Each Iteration (1a)

- We conclude that averaging on each distribution has some good and bad characteristics, but on the whole is not desirable
- It produces a good confidence interval for the mean of the experts, but this is not what we want
  - We already know the mean of the experts, the point estimate is the simple average of the means of each
  - What we really want is the full range of the possible outcomes, but averaging on each iteration does not do this, instead it shrinks the answer
  - By analogy, this is the same problem as the confidence interval for a CER … it bounds the line, but not the data … what we really want is the prediction interval
  - It is only a candidate (and flawed at that) for clear cases of single reality

► Conflation: Averaging Parameters (2a)

- Averaging parameters provides simple results: Three very different sets of testimony by two experts produces exactly the same picture
- The standard deviation of k identical but scattered triangles, with standard deviation of s, when iteration-averaged will produce a standard deviation of s
  - The SD of the conflation will not vary with k
  - Correction can be achieved by a spreading with √k but this is likely to be done wrong or omitted altogether, and at best would require row-by-row corrections



TASC

Conflation: Averaging Parameters (2a)

- We conclude that averaging parameters has some good and bad characteristics, but on the whole is simple and wieldy
  - It produces good estimates of the mean and the standard deviation
  - It is insensitive to scatter of expert testimony, so is only useable in clear cases of single reality
  - Correct the parameters as shown earlier because each expert is likely to truncate
    - The order of the operations does not matter

Conflation: Sampling (3)

- "Average" the probability distributions of the k experts, using one of two schemes, depending on the speed implications and the ease of implementation in your model:
  1. Put all the distributions in the mix, and scale each by 1/k, creating a (probably) multi-mode custom distribution[9]
     - We will see this pictorially on the next slide
  2. Characterize each of the k distributions and choose a first random number to select which expert distribution to use for each run of the Monte Carlo and a second random number to draw from that expert's distribution[10]
  - The two above methods are mathematically identical

- The resulting distribution will have two characteristics:
  - A better estimate of the mean and generally better predictive performance than other conflation schemes[9]
  - A wider (actually, "not narrower") standard deviation for the conflated result than those of the original individual distributions
    - We don't know the degree to which sampling will correct dispersion, although the more experts the wider the dispersion
    - We plan to attempt a study of this
  - We will give a demonstration of this effect with representative data

9. *An experiment in Probabilistic Forecasting,* Thomas A. Brown, R-944-ARPA, July 1973
10. *Determining the Cost of the Certification and Accreditation Process using Expert Opinion and Monte Carlo Simulation*, A. J. Flynn, B. J. Nethery, K. Thomas, A. E. Gerstner, B. Dickey, C. M. Kanick, and P. J. Braxton, SCEA 2010

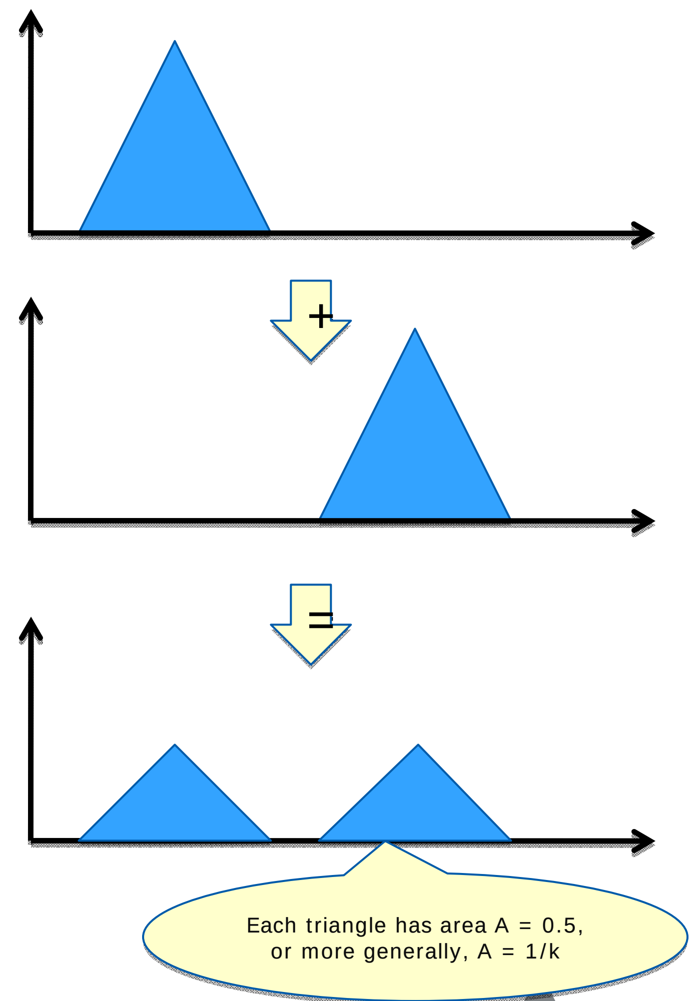
► Conflation: Sampling (3)

- To conflate two triangular distributions, “average” them

The first distribution

The second distribution

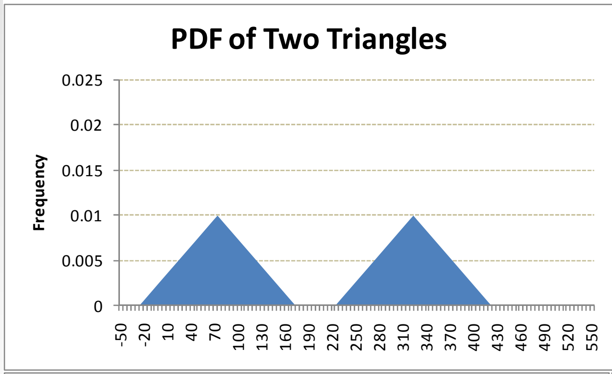
The conflated (averaged) distribution

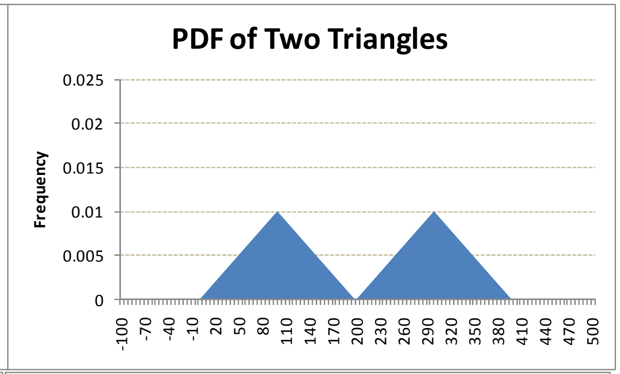


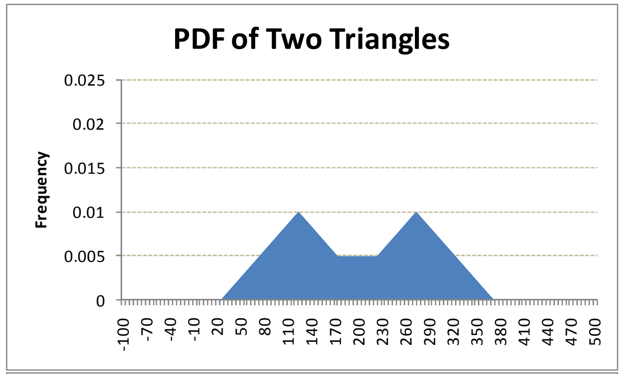
TASC

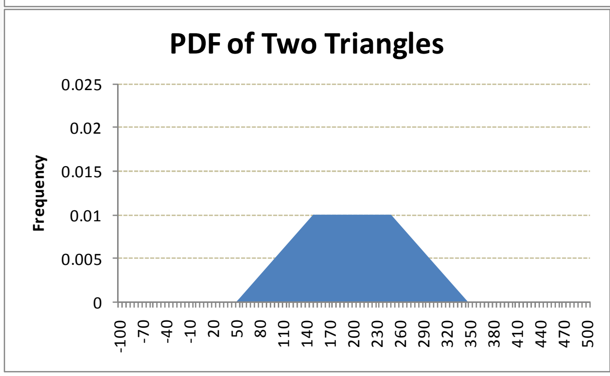
Sampling of Two Triangles - PDF

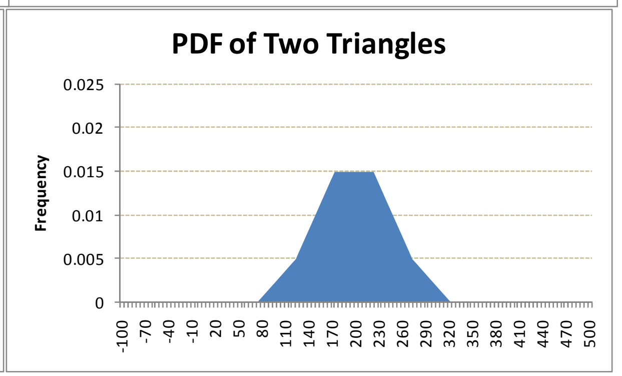
- These charts portray the conflation of two triangles as the respective experts who estimated them come into alignment
  - Each original individual triangle is symmetric, has a base length of 200, and a standard deviation of 40.8
  - Conflation is done by averaging the two PDFs (also described as sampling)
- The two triangles move closer in such a way that the conflated mean remains constant
  - We maintained the same conflated mean of 200
  - We kept the conflated mean constant to allow us to discuss the CV in a meaningful way
  - When the two triangles merge, we get a triangle that has the height and width of each individual triangle before conflation
- The standard deviation of the conflated distribution will be shown on the next graph

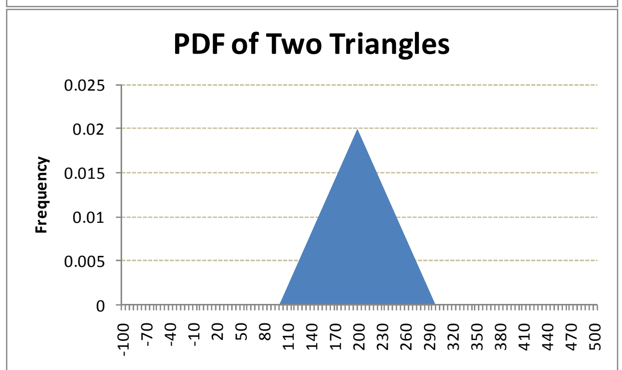










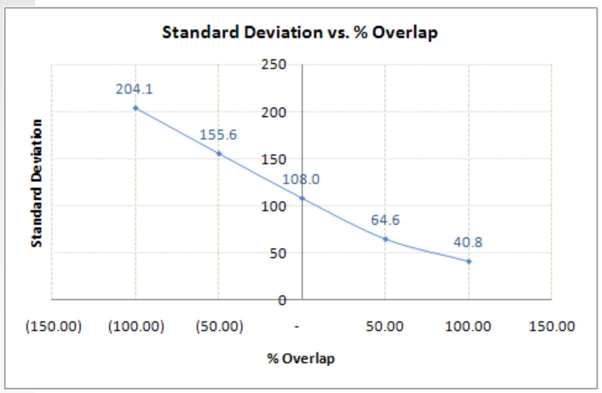


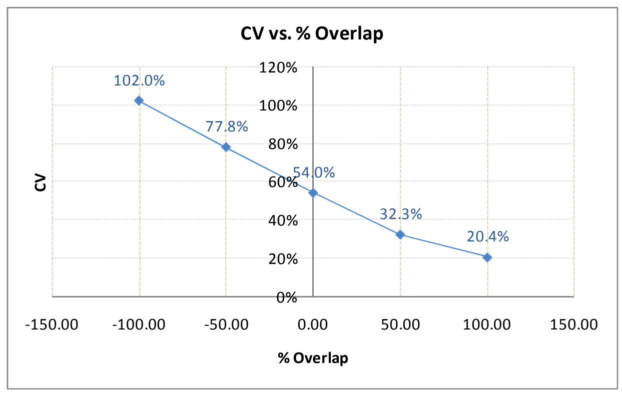
TASC

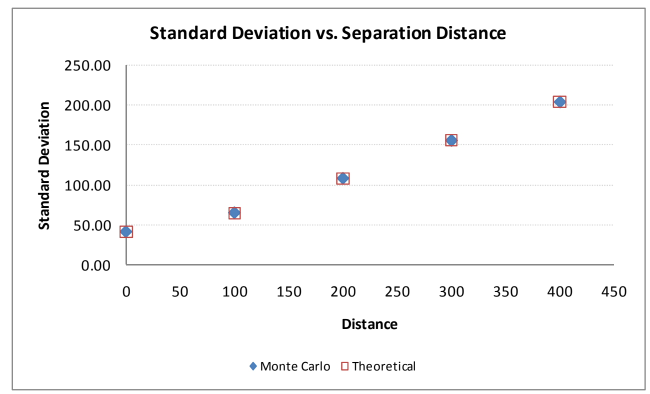
► Conflation of Two Triangles - CV and SD

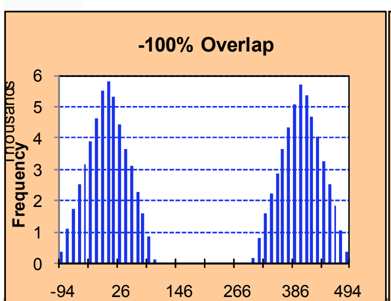
- As two triangular PDFs move closer, the conflated standard deviation and CV drop until the triangles merge and achieve the same standard deviation as that of each triangle
  - Since we chose to maintain the mean of the conflation at 200, the CV drops
- The unsettling conclusion is that the CV of conflated expert opinion can be uncontrollably large, depending on how far apart their triangles
- The standard deviation of two identical triangles separated by distance 2d can be shown* to be $\sqrt{(\sigma^2+d^2)}$

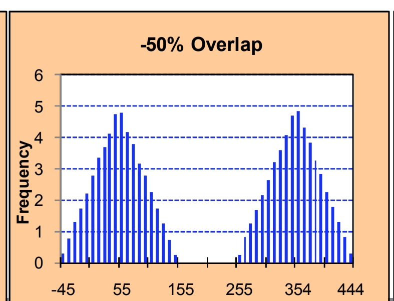
**We aren't saying it's easy … this phrase usually means the Professor is too lazy to show you or too kind to bore you, and the former is by far the more likely! We're the latter, the proof is in backup*

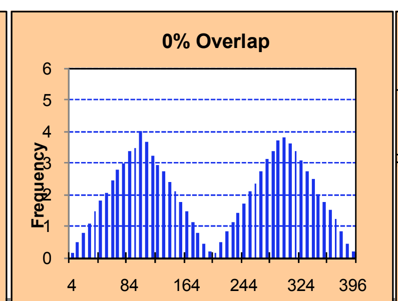


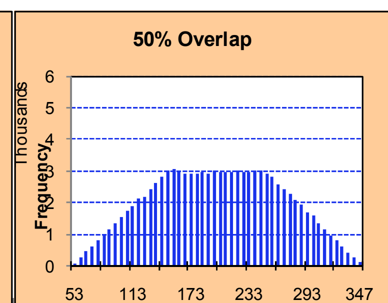


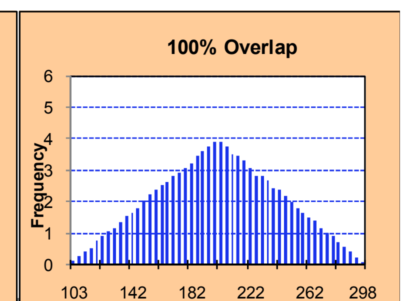












TASC

The Dispersion of Sampled Distributions

- Let:
  σ = SD of the underlying risk
  S_e = SD for the individual experts (we think it is about ½σ)
  S_m = SD for the meta distribution of the experts opinions
  S_c = SD of the conflation
- Then,
  if $S_e = 0$, then $S_c = S_m$
  if $S_m = 0$, then $S_c = S_e$
- And, further
  - $S_c \geq \max(S_e, S_m)$
  - This also implies that if S_e is corrected to σ, S_c exceeds σ
- We have shown, in backup, that once the experts have produced k triangles, then:

  $$S_c = \sqrt{(S_e^2 + S_t^2)}$$

  where S_t is the calculated sum of the squares of the differences of the k triangles from their means. We have yet to prove that:

  $$S_c = \sqrt{(S_e^2 + S_m^2)}$$

  But we believe it to be true

Thoughts on the Distribution of Expert Opinion

- Assumptions:
  1. Experts will not be versed in the distribution of costs, but will be estimating the distribution based on the outcomes they have experienced and perhaps some hearsay
  2. Experts are most likely to be technical people, not cost estimators, so will have experience in a handful of projects and hearsay of somewhat larger number
- Implications
  1. Experts will perceive a mean (and perhaps the mode?) according to Chebyshev's inequality or a confidence interval bounded by $\sigma/(\sqrt{n})$, at best
     - Where n is the number they have observed
  2. Experts will perceive a standard deviation (and thus perhaps the extrema of a triangle?) as a variance σ times a chi-square (n) divided by n, at best
  3. The above do not yet consider the implications of truncation of the value of σ

Combining Corrections for Extrema and Conflation

- We have shown that individual distributions can be corrected for a consistent pattern of understatement
- We have shown that sampling of multiple experts will improve the mean and widen the spread
  - But we don't have a good sense of how much the spread will be improved
- The implication of the two above statements is that we should <u>not</u> endeavor to both <u>expand and sample</u> expert distributions
  - If we correct the individual distributions, we will have the dispersion "about right", if we then sample them, we will have a dispersion that exceeds "about right"
- So, for "single reality", do one or the other but not both
  - Expansion of a single distribution focuses on the dispersion
  - Sampling of diverse experts focuses on getting the mean right
  - Since we generally recommend correcting lower order moments first[11], conflation is the priority

11. *The Manual for Intelligence Community CAIG Independent Cost Risk Estimates*, R. L. Coleman, TASC, Inc., J. R. Summerville, TASC, Inc., S. S. Gupta, Director, IC CAIG, 35th DoDCAS & SCEA 2002, ASC/Industry Cost and Schedule Workshop, Oct 2002 [see tenets]

Conflation: Sampling

- Sampling of each distribution has excellent characteristics
  - It replicates what the experts told us exactly
- It has a problem in use for a single reality situation because the standard deviation is not easily correctible for scatter nor is it useable without correction
  - We can easily correct each expert's testimony for truncation
  - But we cannot undo the growth caused by expert scatter, which is theoretically unbounded … the adjustment would be a function of k, the number of experts and has yet to be ascertained
- We conclude that, despite its popularity in the literature, the sampling technique is too tricky in a single reality case and should not be used

Recommendation - Multiple Reality

- The mean of the multiple reality is not troublesome, almost any *reasonable* approach will yield the same mean
  - Again that dangerous word "reasonable"!
- The standard deviation does not present as much of a problem in a multiple reality case because we believe each expert, like the six blind men, sees a piece of the truth
- Use sampling
- Correct each expert's testimony before sampling, you cannot easily correct it afterwards, order matters

An Actual Case Study

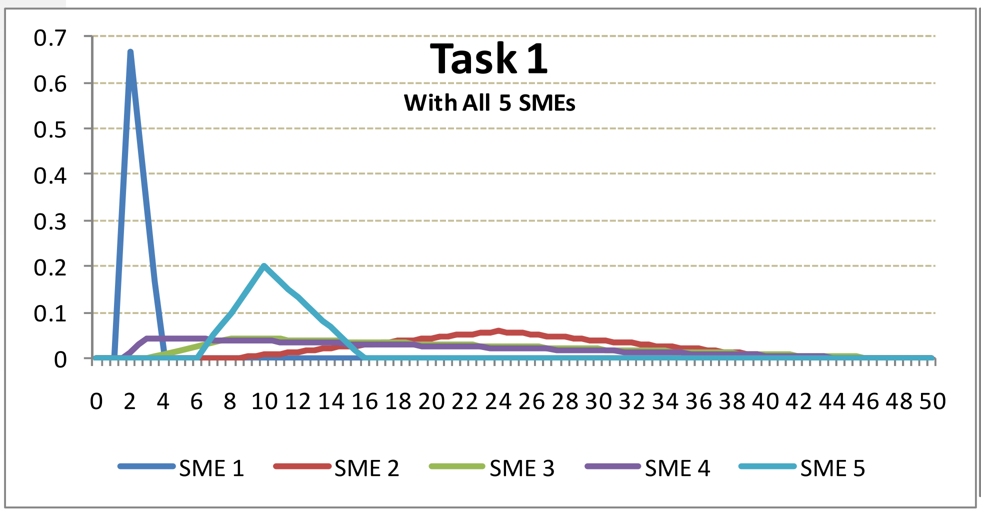
TASC

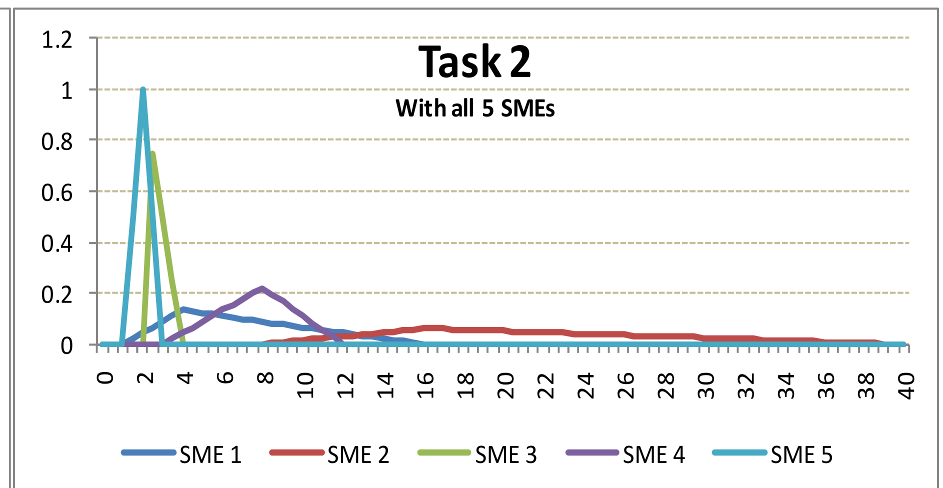
The SME Data

- Actual SME data was collected on a number of subtasks
- Each SME was providing estimates of the same tasks without collaboration
- The data, while not strictly pathological, was sufficiently different to provide a good test of our findings
- Our paper was written for this study, but our methodology development was divorced from the data until the end
- The data source is sufficiently obscured, by a single linear transformation, to prevent traceback

► The Original Data

- The transformed source data shows a dispersion of opinion
- It was unclear whether this was a case of multiple reality
  - The study authors concluded that it might be, so they chose sampling
- We will compute the results from all the methods we examined and plot the results of the two methods we selected

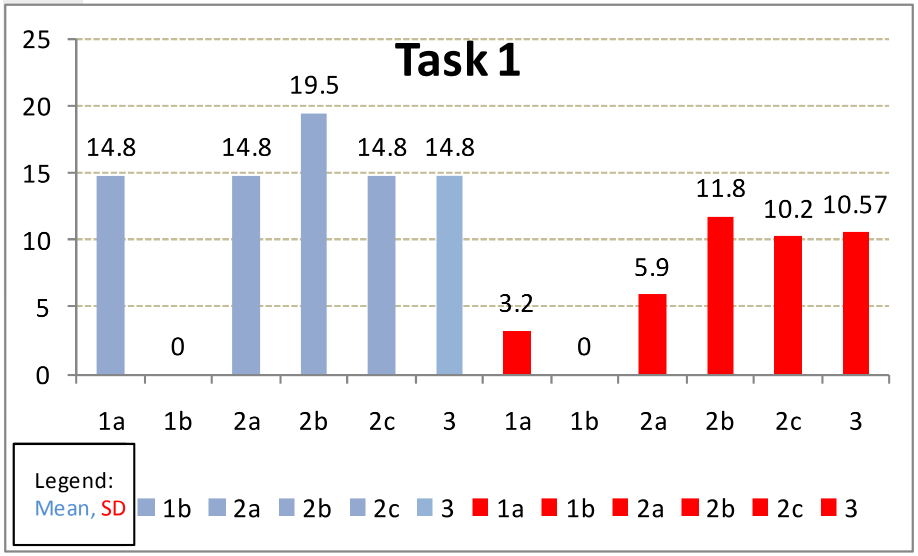


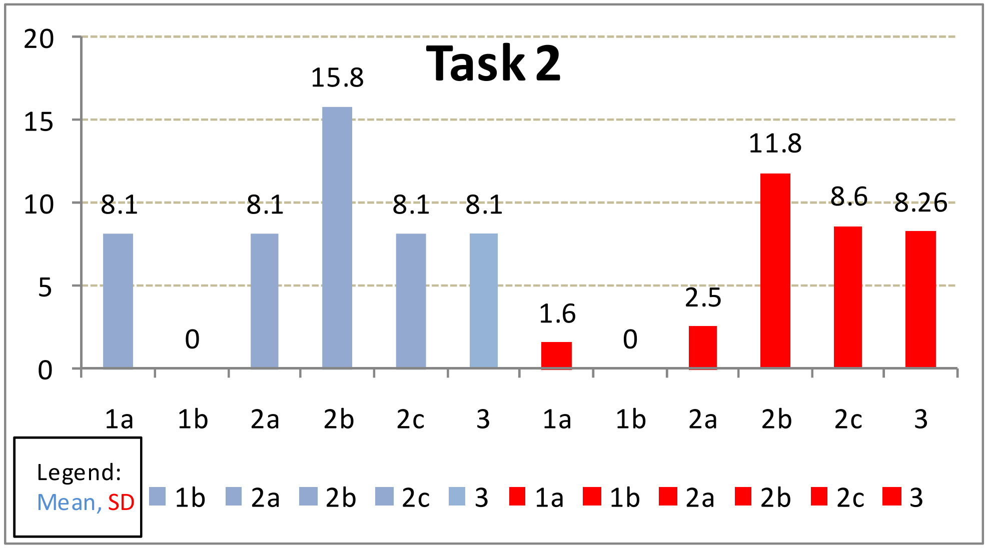


TASC

Moments of the Postulated Methods

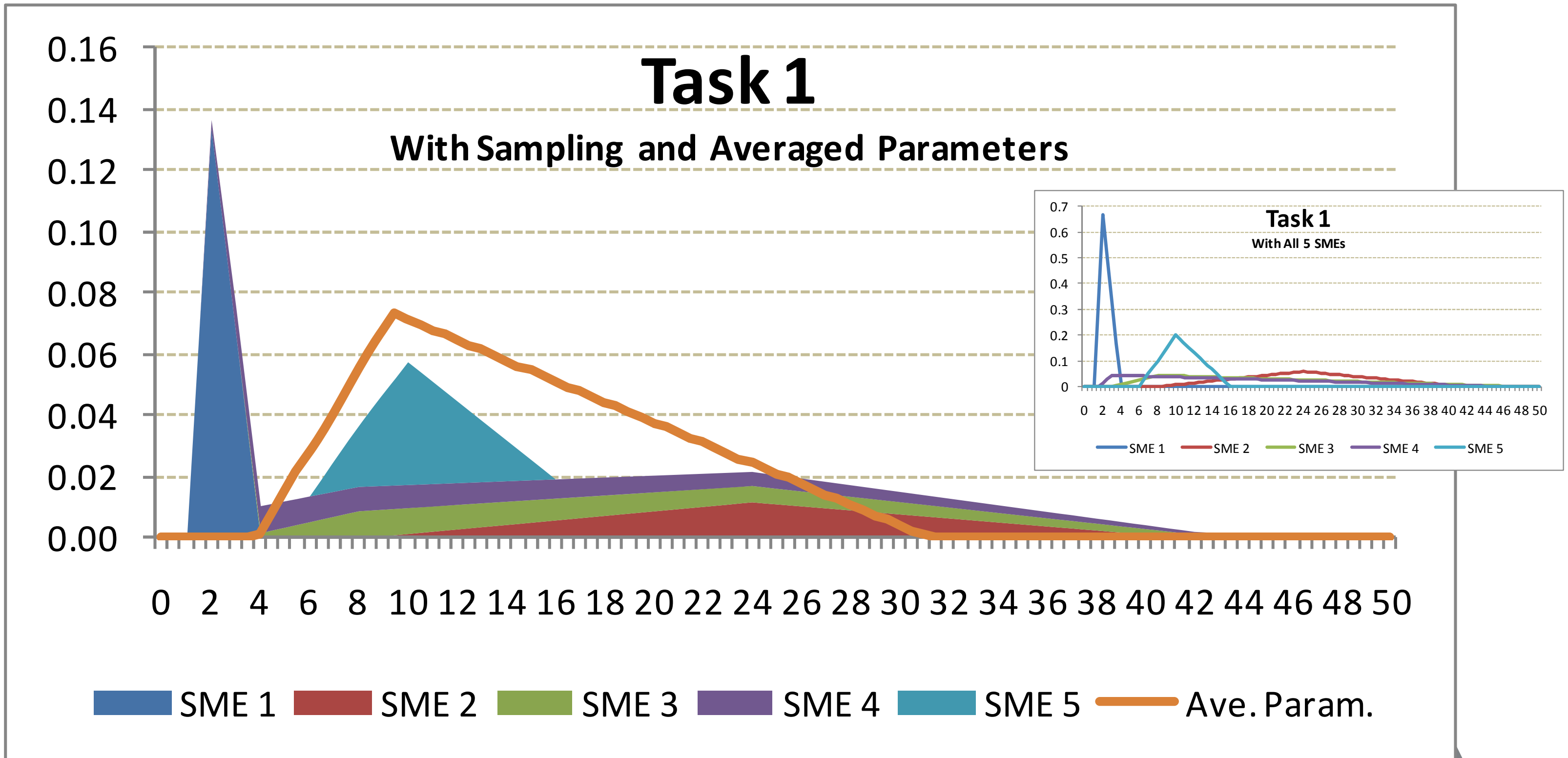
- Methods recap
  - 1a — Average the results of each SME on each run
  - 1b — Same as 1a with correlation = 1.0 (same as 2a below for symmetric, a bit different for skewed)
  - 2a — Average the parameters of the SMEs (use the average of the means *or* the average of the modes)
  - 2b — Min of the mins, average of the modes and max of the maxes
  - 2c — Min of the mins, average of the means and max of the maxes
  - 3 — Sampling (equivalent to averaging PDFs)
- As we expected, the means are all almost all the same
  - Method 2b used averaged modes, so the mean is not preserved
  - Method 2c, an attempt to salvage 2b, used average means but routinely returned modes below the min so was unusable
- As we expected, the standard deviation is the parameter that responds to our choices
  - SD of 1a was "too small"
  - SD of 2b, the rejected 2c and 3 were "too big"
  - The SD of 2a was "Goldilocks"





Graphs of the Two Recommended Methods

- The "Sand Chart" shows sampling, the preferred method for multiple realities
  - It retains all the information told to us by the SMEs equally
  - It suggests, in this example, that there may be three different modes, representing 3 different possibilities
- The "Line Chart" shows averaged parameters, the preferred method for single reality
  - It responds to all SMEs, but produces a uni-modal, less dispersed solution
  - It suggests, in this example that SME 1 was too low while SMEs 3 and 4 were a bit pessimistic on the high end
- Both methods are credible and both do a decent job of synthesizing



Conclusion for the Conflation of Experts

- As asserted, we have illustrated that the averaging of parameters for k triangles, is equivalent to averaging of draws from those k triangles with a single draw of a random number used to simulate expert's draw, and then averaging the draws
- We have demonstrated why those two equivalent methods give the simplest and clearest result for Single Reality and seem the best representation of what the k experts seem to have meant
- We have shown why Sampling of k experts gives the best representation of what the k experts seem to have meant in the case of Multiple Realities
- We presented a case study with actuals that shows that the two recommended approaches do a decent job of synthesizing what the SMEs told us
- The issue of deciding between Single and Multiple Realities remains the most difficult issue
  - Sometimes it will be as simple as learning that each expert has in mind "a different engine"
  - Sometimes it will be a concession to the wide dispersion and the recognition that there "must be a reason."
- We will now move to a different topic, that of correcting mischaracterization of distributions, without which this paper would seem incomplete

Correcting the (Mis)characterization of Distributions

The Problem

- "Experts" who may know a lot about the technical issues, and maybe even the cost of them, will not necessarily be well versed in probability
  - Consequently, the characterizations they will produce will not be easily used and will sometimes be incoherent (meaning, internally contradictory)
- Expert testimony in risk analysis should be accorded the same respect that cost data is in cost analysis
  - Tenet 1: "Do no harm" meaning preserve as much of what the expert said as is possible in achieving coherence
  - Tenet 2: Preserve lower order moments above higher order moments
  - Tenet 3: If particular aspects are more important than others, preserve those aspects (e.g., if the variability or upper percentiles are the focus, accord those greater priority)
- It is preferable to make the corrections with direct feedback to the expert, but this feedback should be done under the same precepts as the corrections
  - Meaning, follow the tenets in your persuasions and probing

TASC

Implausible Percentiles

- "The 20/50/80 are $0.0M/$0.9M/$3.6M"
- No triangle can fit this, and the distribution is wildly skewed, so simplifying steps were taken:

1. Assume that the stated "50%-ile" is really the mode
2. Take the 20 and 80 as "about true", and assume they are ±σ. Use the rule that the half-base lengths of a symmetric triangle are √6*σ. Note that these triangles are not symmetrical, but use it as a factor that probably does a decent job
3. Results:

Input		Output	
20%-ile	0	L	-1.305
50%-ile	0.9	M	0.900
80%-ile	3.6	H	7.514

- Note that the correction *may* be distorting the central tendency
  - But, this distribution is clearly intended to be skewed, and the mean is therefore above the median
  - We cannot actually compute the mean with the information given
  - We also knew that in this analysis, the ROS at the 80th percentile was a particular focus, so we felt that preservation of that point should take priority (Tenet 3)

Unlikely distributions

- Risk values:
  - 20% probability of -$2M
  - 40% probability of $0
  - 20% probability of +$4M
- Suspecting that this was a just clumsy way to characterize a triangle, we asked if a triangle with the below characteristics was along the lines of what the expert meant:
  - 20%-ile -$2M
  - Mode 0M
  - 80^{th} %-ile+$4M

  … the expert agreed readily that the precise distribution wasn't what he meant, and the triangle captured the sense of it.

Errors Of Characterization Induced by the Risk Analyst

- Categorical* risk distributions
  - Many risk models cannot easily (or rather obviously) implement a categorical random variable beyond a Bernoulli
    - Many can do it, most analysts don't realize they can
  - For a 3-value categorical, with choices of 0, a and b, many analysts implement it as two independent Bernoullis with values of 0 or a and 0 or b
  - This results in an error as the results are not the same … the two Bernoullis can turn out as a and b at the same time, but the original formulation prohibits that
  - Either implement it as a categorical or create two Bernouli's with the right characteristics
- Triangular risk distributions
  - Sometimes the end points are set at the standard deviation of the formulation
  - Sometimes triangles are used instead of normals, even when the normal was proposed, out of aversion to negative outcomes
    - In practice, negative outcomes are harmless in Monte Carlo
    - Negative outcomes ought to be fairly rare anyway
- Normals
  - Sometimes triangles are substituted incorrectly (see above)
    - If the mean and standard deviation are captured correctly there is little harm
  - Sometimes the negative portion of the normal is truncated despite that this causes a shift of the formulated mean and a reduction in the standard deviation

* Categorical risk distributions are like Bernoullis but allow 2 or more values (the Bernoulli is a member of the family)

Conclusion for Correcting Mischaracterization of Distributions

- We have presented tenets by which apparent errors of characterization may be corrected and have listed the most common Risk-Analyst-induced errors
- We finish by reiterating that the testimony of the experts we consult should be handled much as we should handle data
  - We must be careful in not ignoring the symptoms of the testimony, and avoid such elementary errors as causing anchoring* and "leading the witness."
  - We should, nonetheless carefully repair any clear errors caused by the unfamiliarity with probability that can result in unlikely distributions

Final Thoughts

- The conflation of expert testimony has received some attention in the literature, but little to none of the conclusions seem to have permeated the cost risk discipline
- We hope that we have provided a reasonably thorough paper by which risk Analysts might be guided
- We also hope that we have provided a few good tenets fr correcting mischaracterization, along with some illustrative (actual) examples.
- We hope to be able to take on the issue of what we call the meta-distribution, the likely distribution of individual expert testimony
  – Without a good model for the meta-distribution, the full demonstration of the best answers will remain incomplete, because the meta-distribution is the unseen ground truth against which these answers can be measured
  – Until we can be satisfied we have the meta-distribution, we are confined to showing the behavior of various methods and deciding if that behavior seems correct

Backup

► The Geometry of Symmetric Triangles

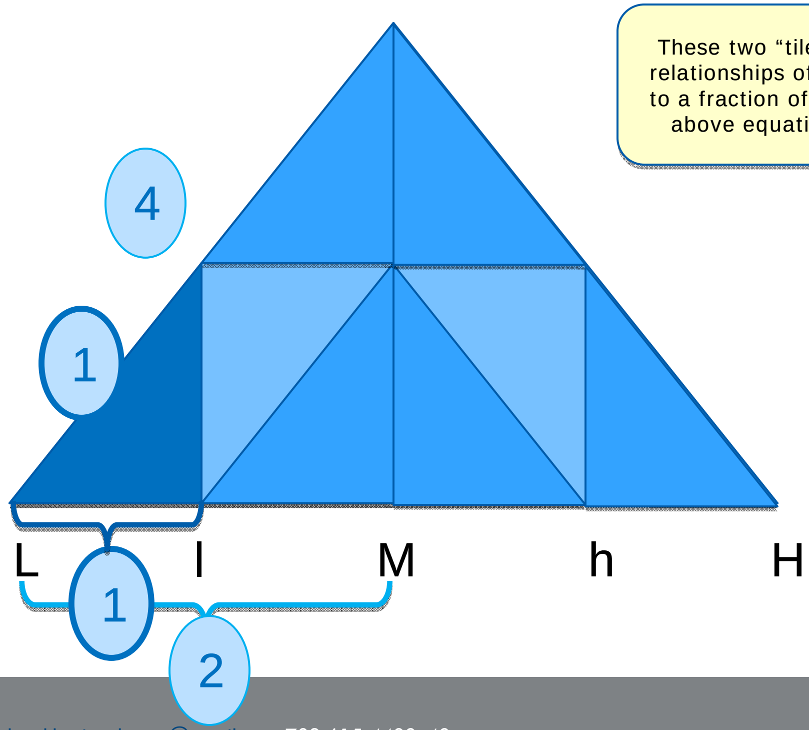
- For a symmetric Triangle(L, M, H), where M-L = H-M
- Find points l and h such that l and h are the p^{th} and $1\text{-}p^{th}$ percentiles

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

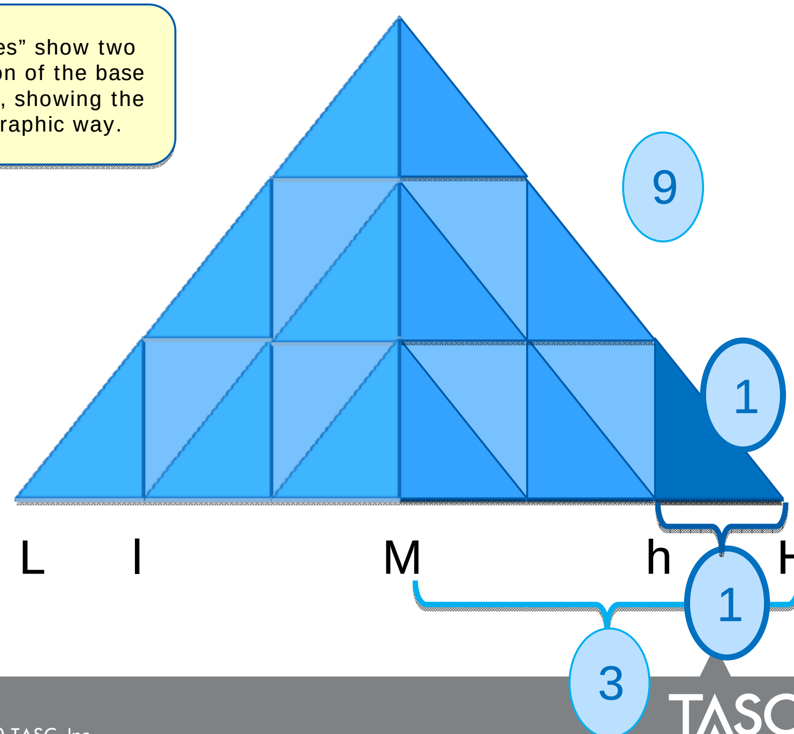
  If l-L = 1/3*(M-L), H-h = 1/3*(H-M), then $p = 1/(2*3^2) = 1/18 = 5.6\%$

  p^{th} percentile -> $\sqrt{(p/2)}$ base fraction -> $\sqrt{(2p)}$ half-base fraction

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



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



TASC

Triangles With Related Areas

- We wish to know how to draw triangular distributions that are related to one another:
- Constant area:
  - 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

$$A = \frac{1}{2}bh = \frac{1}{2}(bk)\left(\frac{h}{k}\right)$$

- Reduction in area:
  - For area to be reduced by a factor, the dimensions of a similar triangle must be reduced by the square root of that factor

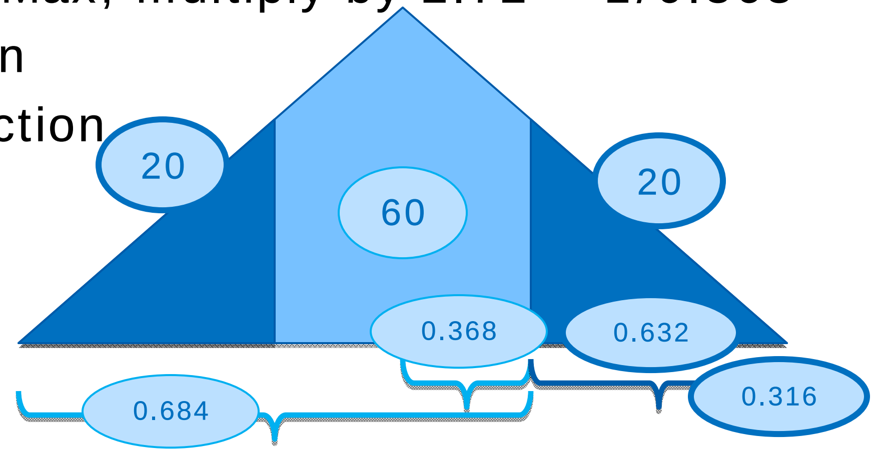
$$A_2 = \frac{1}{k}A_1 = \frac{1}{2k}b_1h_1 = \frac{1}{2}\left(\frac{b_1}{\sqrt{k}}\right)\left(\frac{h_1}{\sqrt{k}}\right)$$

  - 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**

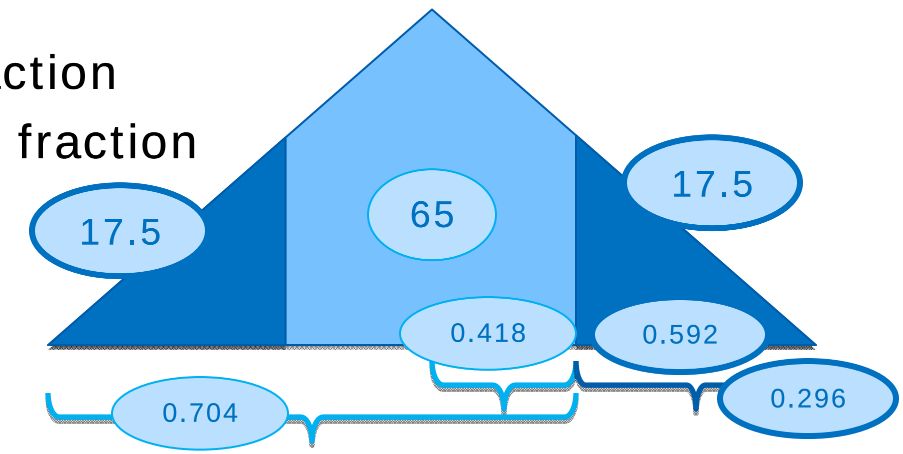
$$A_2 = \frac{1}{k}A_1 = \frac{1}{2k}b_1h_1 = \frac{1}{2}(b_1)\left(\frac{h_1}{k}\right)$$

► 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



TASC

Triangular Distribution – PDF and Mean

- For Triangle(L,M,H) , denote L=a, H=b, ML=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

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

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

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

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

► Triangular Distribution – Variance

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

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

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

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

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

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

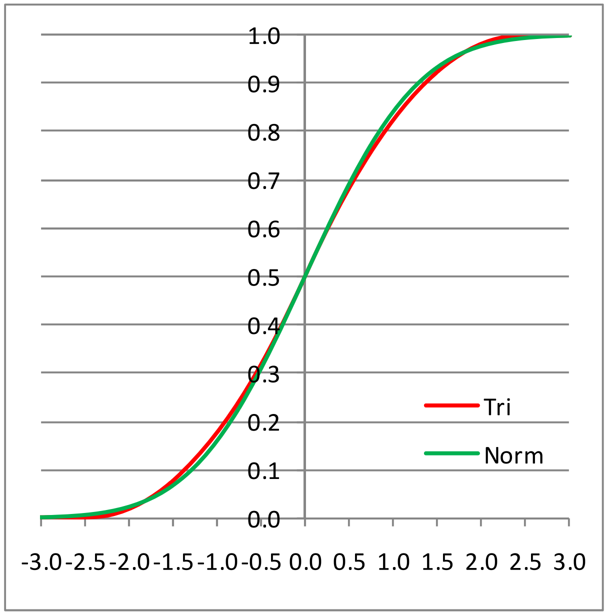
$$= \frac{a^2 + b^2 + c^2 - ab - ac - bc}{18} = \frac{b^2 - 2ab + a^2 + c^2 + ab - ac - bc}{18} = \frac{(b-a)^2 - (b-c)(c-a)}{18}$$

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

TASC

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, Triangular(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



TASC

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

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

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

TASC

Equivalence of Averaging Distributions and Averaging Parameters for Symmetric Triangles

- In the case of symmetric triangles, averaging the individual triangles (with perfect rank correlation) – method **1b** – can be shown to be equivalent to averaging the parameters – method **2a**
  - 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:

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

  - 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

Equivalence of Averaging Means and Averaging Modes for Triangles

- If we average parameters – method **2a** – 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 kth 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

Richard.lewis.coleman@gmail.com, 703-615-4482,