2010-04-30

The Correct Use of Subject Matter Experts in Cost Risk Analysis

Braxton, Peter J.; Druker, Eric R.; Coleman, Richard L.

Monterey, California. Naval Postgraduate School

https://hdl.handle.net/10945/33464

NPS-AM-10-031

EXCERPT FROM THE
PROCEEDINGS

OF THE

SEVENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM
WEDNESDAY SESSIONS
VOLUME I

Acquisition Research

Creating Synergy for Informed Change

May 12 - 13, 2010

Published: 30 April 2010

Prepared for: Naval Postgraduate School, Monterey, California 93943

The research presented at the symposium was supported by the Acquisition Chair of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request Defense Acquisition Research or to become a research sponsor, please contact:

NPS Acquisition Research Program
Attn: James B. Greene, RADM, USN, (Ret.)
Acquisition Chair
Graduate School of Business and Public Policy
Naval Postgraduate School
555 Dyer Road, Room 332
Monterey, CA 93943-5103
Tel: (831) 656-2092
Fax: (831) 656-2253
E-mail: jbgreene@nps.edu

Copies of the Acquisition Sponsored Research Reports may be printed from our website www.acquisitionresearch.net

The Correct Use of Subject Matter Experts in Cost Risk Analysis

Richard L. Coleman—Richard L. Coleman is a Naval Academy graduate, received an MS with Distinction in Operations Research from the US Naval Postgraduate School and retired from active duty as a Captain, USN, in 1993. His service included tours as Commanding Officer of *USS Dewey (DDG 45)*, and as Director, Naval Center for Cost Analysis. At TASC and Northrop Grumman, he worked extensively in cost, CAIV, and risk for the Missile Defence Agency (MDA), Navy ARO, the intelligence community, NAVAIR, and DD(X). He has supported numerous ship programs, including the DDG 1000 class, DDG 51 class, Deepwater, LHD 8 and LHA 6, LPD 17 class, *Virginia* class SSNs, CVN 77, and CVN 78. He was the Director of Independent Cost Estimation for Northrop Grumman Information Systems and conducted Independent Cost Evaluations on Northrop Grumman programs. He is currently at TASC, Inc., as a PM for cost and risk. He has more than 75 professional papers to his credit, including five ISPA/SCEA Best Paper Awards and two ADoDCAS Outstanding Contributed Papers. He was a senior reviewer for Cost Programmed Review Of Fundamentals (CostPROF) and Cost Estimating Body of Knowledge (CEBoK) and lead author of the Risk Module. He has been Regional and National VP of the Society of Cost Estimating & Analysis (SCEA) and is currently a Board Member. He received the SCEA *Lifetime Achievement Award* in 2008.

Richard L. Coleman (TASC)
6211 Elati Ct, Alexandria, VA, 22310
richard.lewis.coleman@gmail.com
703.615.4482

Peter J. Braxton—Peter J. Braxton holds an AB in Mathematics from Princeton University and an MS in Applied Science (Operations Research) from the College of William and Mary. He currently serves as a senior cost and risk analyst for TASC, Inc. He has worked in cost estimating and risk analysis, Cost As an Independent Variable (CAIV), and Target Costing on behalf of the Navy Acquisition Reform Office (ARO), the DD(X) development program, and other ship and intelligence community programs. He has reviewed several high-profile proposals as part of the Independent Cost Estimation (ICE) function for Northrop Grumman. He has over 12 professional papers to his credit, including SCEA/ISPA International Conference award-winners in CAIV (1999) and Management (2005). He served as managing editor for the development of both the acclaimed CostPROF body of knowledge and training course materials and its successor CEBoK. He serves as SCEA's Director of CEBoK and an elected Director, and was named SCEA's 2007 *Estimator of the Year for contributions in Education*. He serves as a TASC Technical Fellow, and received a TASC *President's Award* and a Northrop Grumman *Corporate Contracts, Pricing, & Supply Chain Award* in 2008.

Peter J. Braxton (TASC)
9937 Wood Grouse Ct., Burke, VA, 22015
peter.braxton@tasc.com
703. 449.3623

Eric R. Druker— Eric R. Druker CCE/A graduated from the College of William and Mary with a BS in Applied Mathematics in 2005, concentrating in both Operations Research and Probability & Statistics with a minor in Economics. He is employed by Booz Allen Hamilton as an Associate. He currently is President of the St. Louis SCEA Chapter. In 2009, he was named SCEA's *National Estimator/Analyst of the Year for Technical Achievement*. He currently supports NASA, performing joint cost & schedule risk analysis across a variety of projects. In his previous position as a Technical/Research Lead at Northrop Grumman, he developed the risk analysis process/methodology used to assess financial risk to the company during Independent Cost Evaluations (ICEs). As a part of this, he participated in or led over 30 ICEs on Northrop Grumman proposals. He has been an invited presenter at SCEA, the Naval Postgraduate School's Acquisition Research Symposium, SCEA, DoDCAS and the NASA PM Challenge. Prior to coming to Booz Allen, he served as lead author of the Regression and Cost/Schedule Risk modules for the Cost Estimating Body of Knowledge (CEBoK) update.

Eric R. Druker (BAH)
4584 Emerald View Ct, Eureka, MO, 63025
druker_eric@bah.com,
314.368.5850

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 correct solutions in case (1), in which 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 as well as errors committed by the risk analyst.

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.

Overview

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. These parameters are for the distribution of various types of risk such as (typically, but not exclusively) triangles for cost risk, Bernoullis for technical risk and occasional normals.

Single or multiple experts may offer estimates (expert testimony) of a particular risk via some form of parameterization.

This paper will discuss two topics in correction of expert testimony: 1) The “best” approach to converting extrema and quartiles from expert opinion into risk distributions, and 2) 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. Problems with inconsistent percentiles and problems with unusual characterizations will both be discussed.

This topic was addressed in general in a prior paper by Coleman, Braxton, Druker, Cullis, and Kanick (2009) under the rubric “Omission of Elements of Variability.” A paper by St. Louis, Blackburn, and Coleman (1998) espoused a form of combination of expert testimony that this paper now recommends against.

The “Best” Approach to Converting Extrema and Quartiles from Expert Opinion into Risk Distributions

Correcting Extrema and Quartiles for Truncation

The Problem. Our estimated distributions tend to be “too tight,” as shown by Brown (1973) and Alpert and Raifa (1982). Without feedback, we provide extreme values near the 20th percentile and 80th percentile when we are asked Min and Max. This can be improved, with feedback to the 10th and 90th percentile. This can be improved by asking for more-extreme values. For example, “astonishingly-low-probability outcomes” equate to the 0.1th percentile and 99.9th percentile. Without feedback, we give 25th and 75th quartiles that actually contain only 33% of the outcomes versus the expected 50%. This can be improved with feedback to 43% versus the expected 50%. Independent investigations of this over-tightness are remarkably consistent in the degree to which it occurs, as shown by Brown (1973) and Alpert and Raifa (1982). Our ability to probabilistically characterize the past or future or to estimate our certainty on general-knowledge facts are all about comparable, as noted by Lichtenstein, Fischhoff, and Phillips (1982).

Correcting Extrema and Quartiles. For extrema, 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 0.316 of the real base (see Appendix A). 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. For quartiles, 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 inter-quartile 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.

Errors of Extrema—Pictorially. The 20th percentile occurs at a point that is 0.316 of the base, so the understatement of experts is on the order of 1/3. Pictorially, then, we are experiencing a reduction in distribution on the order of the blue (claimed) to the white (actual) portrayed in Figure 1. For a tutorial on computing percentiles, see Appendix A.

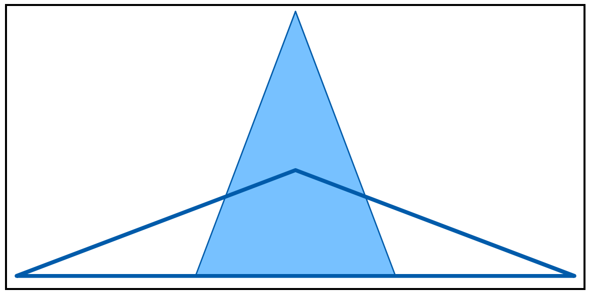

Figure 1. Visualization of Expert Truncation of Dispersion

The "Best" Approaches to Conflating Multiple Views of a Distribution

Conflation

By definition, 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 conflation. Ad hoc conflations are often used for k experts each giving estimates for the same risk or WBS element. For example:

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 (as done by St. Louis, Blackburn, and Coleman (1998)).
    b. Make k random draws from the k different distributions with correlation and average them.
        1. 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, as demonstrated by Brown (1973), or the means of the respective end points for triangles [Average the Parameters].
    b. Make a new distribution with the mean of the k experts and the lowest low and the highest high as end points.
        2. Sampling has been endorsed by Brown (1973). Sampling is done as follows: for each run of the Monte Carlo, pick the answer from a randomly selected expert who provided testimony.

We will only examine ad hoc methods 1a, 2a and sampling. The others can be inferred from those. Also, note that 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.

The First Question

The first question in conflation is to determine what we believe to be the underlying model. No single conflation method will work for the two possible scenarios that can confront the estimator, namely single or multiple realities.

"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. Examples: What is your estimate for the GNP of Brazil for 2009? How big is a brown bear? What is the range of technical risk for the cost of the engine?

"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. Examples: How far away is your favorite planet? (There could be up to 9 answers, depending on the inclusion of Pluto and Earth!) 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) 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 realities 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 if you lead the witness.

Desiderata for Single and Multiple Realities. Cases dictate different characteristics for the conflation technique. Single reality requires the best estimate for the mean, the best estimate for the dispersion and the best estimate for the distribution. Multiple realities dictate the best portrayal of the multiple choices we are confronted with. We will discuss each in turn.

We will describe the apparent preferred solution for each method after asserting them. For single reality, average the parameters and correct for the understatement of extrema (using method 1b or 2a from above). For multiple realities, sample from the experts after correcting each for understatement of the extrema. If we cannot discern whether we are in single or multiple reality, then 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 Conference Best Paper by St. Louis et al. (1998). To see why, we will show its characteristics and indicate why it is probably unsuitable.

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, as shown below, 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. 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 $s/\sqrt{k}$. 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 $\sqrt{k}$, 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.

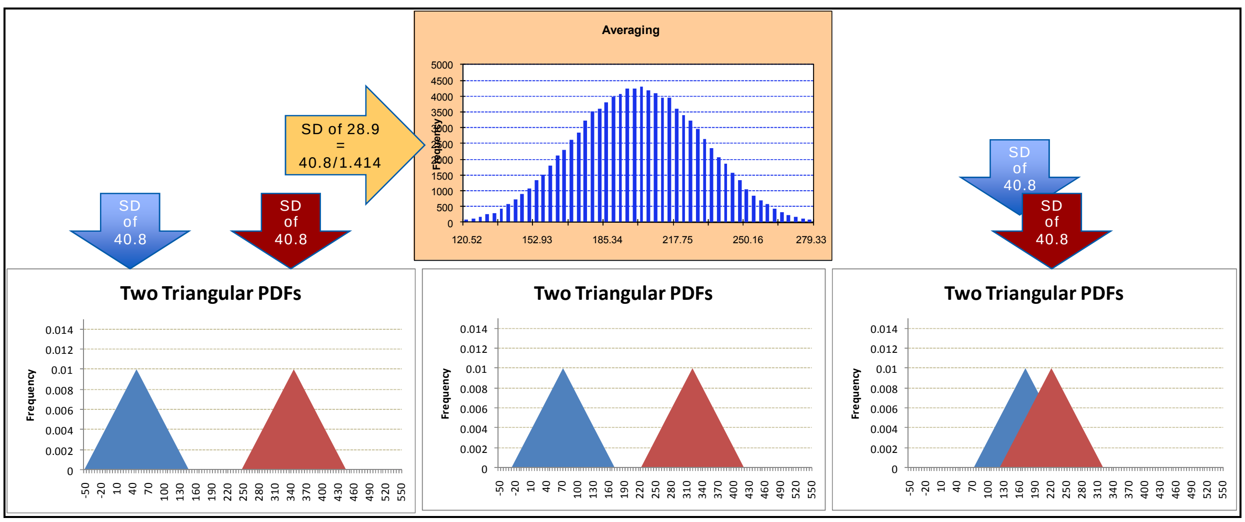

Figure 2. Conflation by Averaging on Each Iteration

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. Averaging parameters provides simple results: Three very different sets of testimony by two experts will produce 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 s. The standard deviation of the conflation will not vary with k. Correction can be achieved by a spreading with $\sqrt{k}$, but this is likely to be done wrong or omitted altogether and, at best, would require row-by-row corrections.

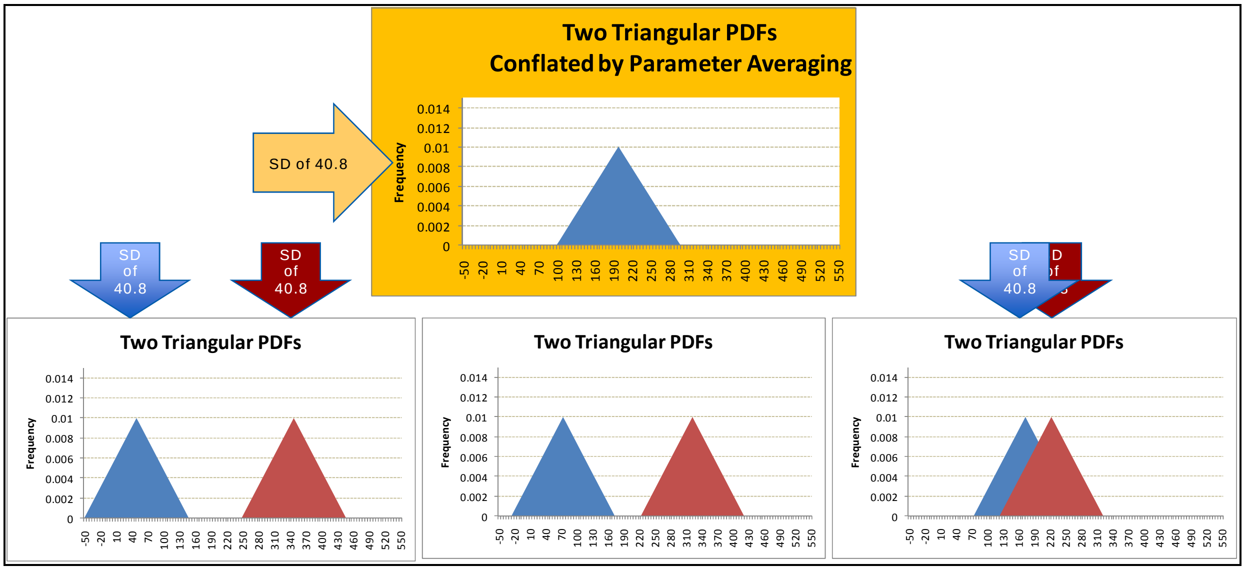

Figure 3. Conflation by Averaging Parameters

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 "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. Put all the distributions in the mix, and scale each by 1/k, creating a (probably) multi-mode custom distribution, as recommended by Brown (1973). We will see this pictorially on the next slide. Alternatively, 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, as used by Flynn et al. (2010). The two methods are mathematically identical. The resulting distribution will have two characteristics: 1) a better estimate of the mean and, generally, better predictive performance than other conflation schemes; 2) 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 conflation 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.

To conflate two triangular distributions, "average" them as illustrated in Figure 4.

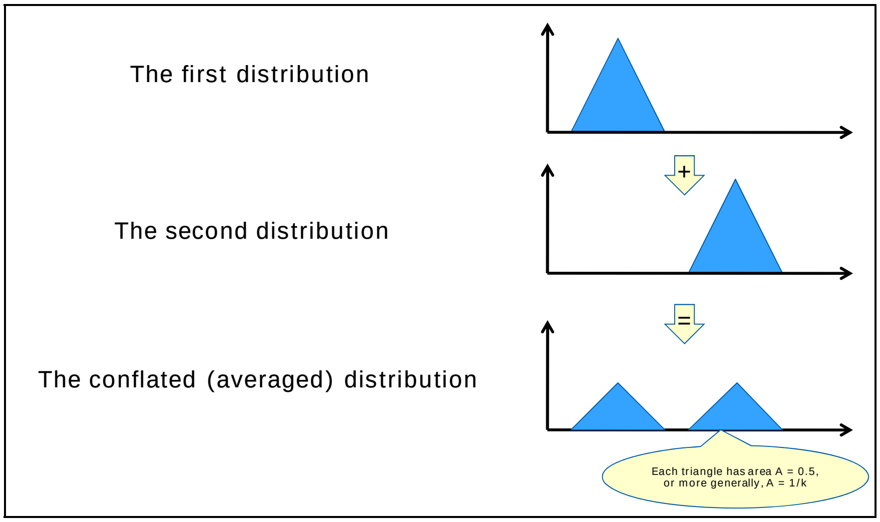

Figure 4. Conflation by Sampling

The charts in Figure 5 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 probability density functions (PDFs), (also described as sampling). The two triangles move closer in such a way that the conflated mean remains constant at 200 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 in Figure 6.

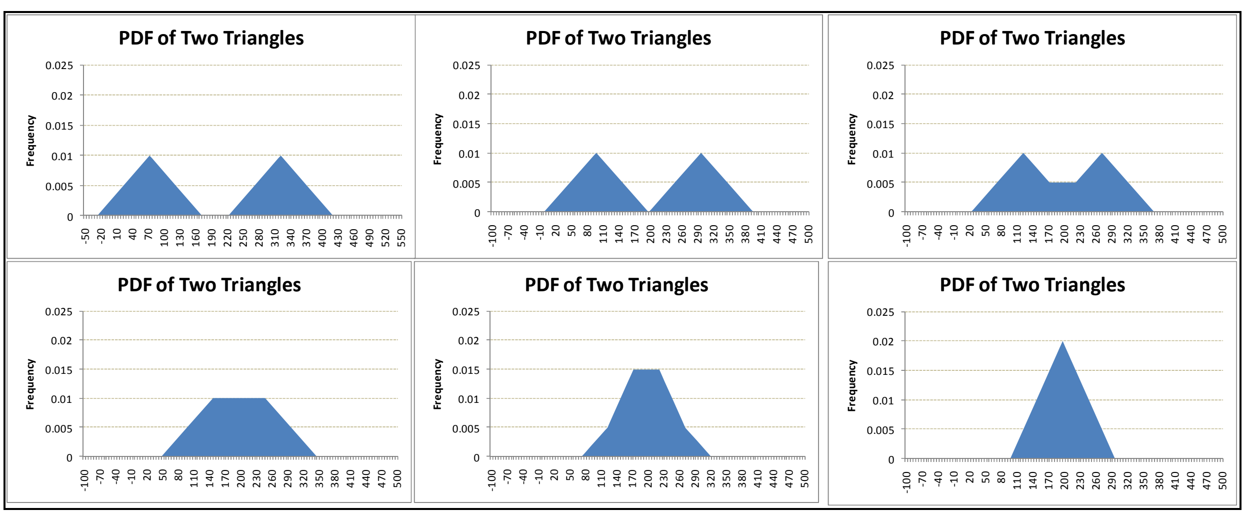

Figure 5. Conflation of Two Triangles by Sampling Maintaining a Constant Mean

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. Note that the variance of two identical triangles separated by distance 2d can be shown to be $\sqrt{(\sigma^2+d^2)}$, which we prove in Appendix A.

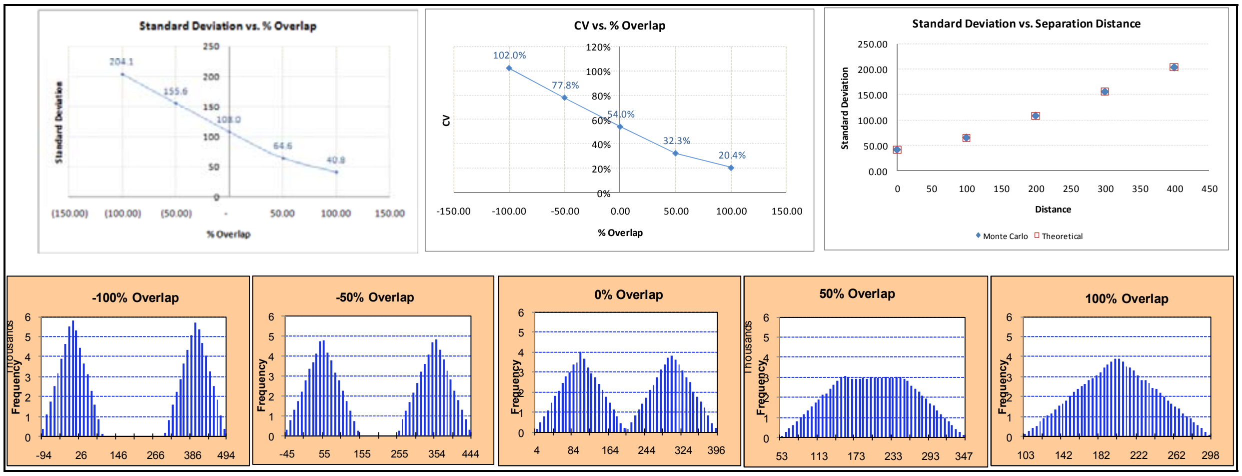

Figure 6. The Standard Deviation and Coefficient Deviation of Two Sampled Triangles as a Function of Their Separation

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, by examination,

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, then 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. We will now speculate on the distribution of the experts themselves, which we have come to call the meta-distribution. Our assumptions are that: 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; and 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.

The implications of the above are that: 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; and 2) Experts will perceive a standard deviation as a variance σ times a chi-square (n) divided by n, at best.

The above thoughts do not yet consider the implications of truncation of the value of σ, but this needs to be incorporated.

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 is that we should <u>not</u> endeavor to both <u>expand and sample</u> expert distributions. If we correct the individual distributions, then we will have the dispersion "about right." If we then sample them, then 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, as recommended by Coleman, Summerville, and Gupta (2002), sampling is the priority. 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 Realities. The mean of the multiple realities case 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. We recommend using sampling. Be sure to correct each expert's testimony before sampling; you cannot easily correct it afterwards—order matters.

Conclusion for the Conflation of Single and Multiple Realities

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. 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," and 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 is that "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). That said, expert testimony in risk analysis should be accorded the same respect that cost data is in cost analysis. We recommend three tenets in correcting apparently erroneous expert testimony. We will list them, and we will apply them in several actual examples of errors the authors have encountered, chosen because they are the most common.

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 that greater priority).

When making corrections, 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.

Example One—Implausible Percentiles. The expert told us that "The 20/50/80 are $0.0M/$0.9M/$3.6M." The difficulty is that no triangle can fit this, and the distribution is very skewed, so simplifying steps were taken. We assumed that the stated "50% percentile" is really the mode. We took the 20 and 80 as "about true," and assume they are $\pm\sigma$. We used the rule that the half-base lengths of a symmetric triangle are $\sqrt{6}*\sigma$. We noted that these triangles are not symmetrical, but we still used it as a factor that probably does a decent job. The results are in the table in Figure 7.

Inputs		Outputs	
20%-ile	0	L	-1.305
50%-ile	0.9	M	0.900
80%-ile	3.6	H	7.514

Figure 7. Table of Inputs and Corrections

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).

Example 2—Unlikely Distributions. The expert gave us three discrete points: 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% percentile =-$2M, Mode = 0M, 80th percentile = +$4M. The expert agreed readily that the precise distribution wasn't what he meant, and the triangle captured the sense of it.

Example 3—Errors of Characterization Induced by the Risk Analyst. Below are three typical errors of characterization introduced by the risk analyst after the expert has given his testimony. They are actual examples, chosen because they are the most common.

Categorical Risk Distributions. Risk models cannot always easily (or, rather, obviously) implement a categorical random variable beyond a Bernoulli. Categorical risk distributions are like Bernoulli's but allow 2 or more values (the Bernoulli is a member of the categorical family.) Many models can handle categoricals, but most analysts don't realize that. For a 3-value categorical, with choices of 0, 1 and 2, many analysts implement it as two independent Bernoulli's with values of 0 or 1 and 0 or 2. This is an error as the results are not the same, the two Bernoulli's can turn out as 1 and 2 at the same time, but the original formulation prohibits that. To fix this problem, either implement it as a categorical or create two Bernoulli'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—even though 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, then there is little harm; but this is often not done right. 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.

Conclusion for Correcting Mischaracterizations. 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 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 for 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.

References

Alpert & Raiffa. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Brown, T.A. (1973). *An experiment in probabilistic forecasting* (R-944-ARPA).

Coleman, R.L., Braxton P.J., Druker, E.R., Cullis, B.L., & Kanick, C.M. (2009). *Are we at the 50th percentile now and can we estimate to the 80th*? SCEA Conference, St. Louis, MO; ADoDCAS Conference, Williamsburg, VA.

Coleman, R.L., Summerville, J.R., Gupta, S.S. (2002). *The manual for intelligence community CAIG independent cost risk estimates*. ADoDCAS, Williamsburg, VA; SCEA Conference, Scottsdale, AZ; ASC/Industry Cost & Schedule Workshop, Seattle, WA.

Flynn, A.J., Nethery, B.J., Thomas, K., Gerstner, A.E., Dickey, B., Kanick, C.M., & Braxton, P.J. (2010). *Determining the cost of the certification and accreditation process using expert opinion and Monte Carlo simulation*. SCEA Conference, San Diego, CA.

Lichtenstein, Fischhoff, & Phillips. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

St. Louis, N., Blackburn, F.K., & Coleman R.L. (1998). *Risk analysis of a major government information production system, expert-opinion-based software cost risk analysis methodology*. SCEA/ISPA Conference, Toronto, Canada; ADoDCAS Conference, Williamsburg, VA; Journal of Parametrics.

Appendix A. Derivations and Proofs

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 pth and 1-pth percentiles (see Figure 8).

If l-L = 1/4*(H-L), H-h = 1/4*(H-L), then p = 2*$(1/4)^2$ = 1/8 = 12.5%

If l-L = 1/9*(H-L), H-h = 1/9*(H-M), then p = 2*$(1/9)^2$ = 1/18 = 5.6%

The pth percentile corresponds to the √(p/2) base fraction, so the 20th percentile, expressed as 1/5, occurs at point √(1/10) = 0.316228 base fraction.

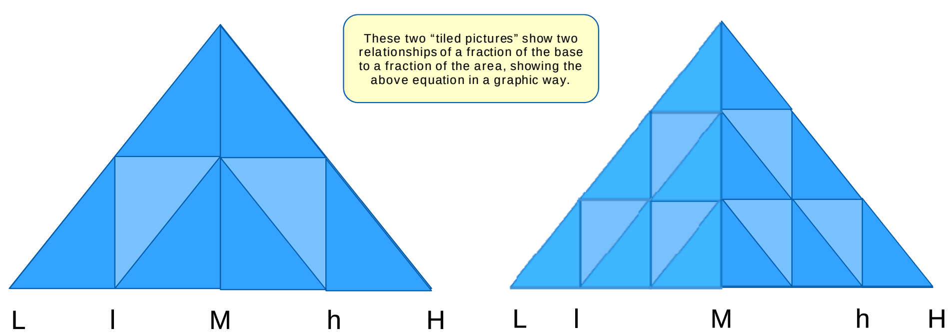

Figure 8. Visual Aid to Demonstrate the Relationship of Percentiles and Base Fraction

Triangles with Related Areas. We wish to know how to draw triangular distributions that are related to one another for our illustrations:

Triangles of Constant Area. 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:

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

Similar Triangles of Reduced Area. The dimensions of a similar triangle must be reduced by the square root of that factor:

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

Reduction of Height to Reduce Area with Constant Base. For area to be reduced by a factor, the height must be reduced by that factor, if the base is to remain constant:

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

Triangular Distribution—PDF and Mean. For a Triangle(L,ML,H), denote L = a, H = b, ML = c denoted 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.

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

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

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

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

Triangular Distribution—Variance

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

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

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

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

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

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

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

Note that the variance is thus the square of the base minus the product of the half-bases.

Substituting a Triangular for a Normal: The √6 Factor. For a symmetric Triangle(L, ML, H), let ML = m, L = m-w, H = m + w, where w is the half-base. Then the mean is m, and the variance is w2/6 (see previous proofs) and the variance is thus w/√6. It follows that the half-base is greater than the standard deviation by a factor of √6. So, to approximate a normal, the factor of √6 is multiplied by the standard deviation of the original normal to be emulated to produce the half-base of the triangle we wish to use in emulation. By this means, end points are found that will produce a triangular distribution to emulate the underlying Normal(μ, σ) in mean and standard deviation. This symmetrical triangular distribution, Triangle(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, as shown in Figure A-2.

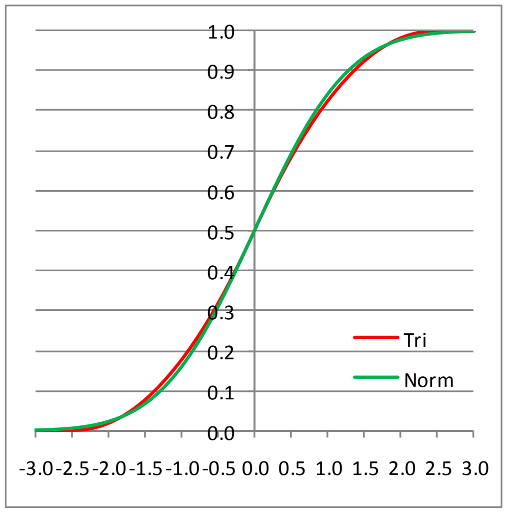

Figure 9. Comparison of Triangle(μ-√6σ, μ, μ+√6σ) and Normal(μ, σ)

Variance of Hybrid Distributions—A Pythagorean Relationship. The Mean 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 the next proof for the derivation of the variance.

Variance of Hybrid Distributions—A Pythagorean Relationship—The Variance.

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

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

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

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

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

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 √(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. 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 expected to be small.

THIS PAGE INTENTIONALLY LEFT BLANK

2003 - 2010 Sponsored Research Topics

Acquisition Management

- Acquiring Combat Capability via Public-Private Partnerships (PPPs)
- BCA: Contractor vs. Organic Growth
- Defense Industry Consolidation
- EU-US Defense Industrial Relationships
- Knowledge Value Added (KVA) + Real Options (RO) Applied to Shipyard Planning Processes
- Managing the Services Supply Chain
- MOSA Contracting Implications
- Portfolio Optimization via KVA + RO
- Private Military Sector
- Software Requirements for OA
- Spiral Development
- Strategy for Defense Acquisition Research
- The Software, Hardware Asset Reuse Enterprise (SHARE) repository

Contract Management

- Commodity Sourcing Strategies
- Contracting Government Procurement Functions
- Contractors in 21st-century Combat Zone
- Joint Contingency Contracting
- Model for Optimizing Contingency Contracting, Planning and Execution
- Navy Contract Writing Guide
- Past Performance in Source Selection
- Strategic Contingency Contracting
- Transforming DoD Contract Closeout
- USAF Energy Savings Performance Contracts
- USAF IT Commodity Council
- USMC Contingency Contracting

Financial Management

- Acquisitions via Leasing: MPS case
- Budget Scoring
- Budgeting for Capabilities-based Planning

- Capital Budgeting for the DoD
- Energy Saving Contracts/DoD Mobile Assets
- Financing DoD Budget via PPPs
- Lessons from Private Sector Capital Budgeting for DoD Acquisition Budgeting Reform
- PPPs and Government Financing
- ROI of Information Warfare Systems
- Special Termination Liability in MDAPs
- Strategic Sourcing
- Transaction Cost Economics (TCE) to Improve Cost Estimates

Human Resources

- Indefinite Reenlistment
- Individual Augmentation
- Learning Management Systems
- Moral Conduct Waivers and First-tem Attrition
- Retention
- The Navy's Selective Reenlistment Bonus (SRB) Management System
- Tuition Assistance

Logistics Management

- Analysis of LAV Depot Maintenance
- Army LOG MOD
- ASDS Product Support Analysis
- Cold-chain Logistics
- Contractors Supporting Military Operations
- Diffusion/Variability on Vendor Performance Evaluation
- Evolutionary Acquisition
- Lean Six Sigma to Reduce Costs and Improve Readiness
- Naval Aviation Maintenance and Process Improvement (2)
- Optimizing CIWS Lifecycle Support (LCS)
- Outsourcing the Pearl Harbor MK-48 Intermediate Maintenance Activity
- Pallet Management System
- PBL (4)
- Privatization-NOSL/NAWCI
- RFID (6)

- Risk Analysis for Performance-based Logistics
- R-TOC AEGIS Microwave Power Tubes
- Sense-and-Respond Logistics Network
- Strategic Sourcing

Program Management

- Building Collaborative Capacity
- Business Process Reengineering (BPR) for LCS Mission Module Acquisition
- Collaborative IT Tools Leveraging Competence
- Contractor vs. Organic Support
- Knowledge, Responsibilities and Decision Rights in MDAPs
- KVA Applied to AEGIS and SSDS
- Managing the Service Supply Chain
- Measuring Uncertainty in Earned Value
- Organizational Modeling and Simulation
- Public-Private Partnership
- Terminating Your Own Program
- Utilizing Collaborative and Three-dimensional Imaging Technology

A complete listing and electronic copies of published research are available on our website: www.acquisitionresearch.org

THIS PAGE INTENTIONALLY LEFT BLANK

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CALIFORNIA 93943

www.acquisitionresearch.org