SAND2011-2229C

An Initial Comparison of Methods for Representing and Aggregating Experimental Uncertainties involving Sparse Data

Vicente Romero, Laura Swiler, Angel Urbina

Sandia National Laboratories*

Albuquerque, NM

* Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

13th AIAA Non-Deterministic Approaches (NDA) Conference

April 4-7, 2011, Denver, CO

National Nuclear Security Administration

Sandia National Laboratories

Introduction and Motivation

- **This talk discusses characterization and treatment of uncertainty associated with random variability of stochastic materials, systems, and processes when limited samples of the randomness are available.**
  - pervasive and important problem in experimental data UQ and model validation or calibration to experimental data
- **Sparse sampling introduces epistemic uncertainty that undermines accurate characterization of the quantity's aleatory uncertainty (i.e., of its randomness as represented by a probability density function (PDF) or "variability distribution")**
- **The view is therefore here taken that:**
  - One should not endeavor for the impossible (accurate estimation of the underlying variability distribution from which the sparse samples come)
  - Rather, a pragmatic goal is that the uncertainty representation should be conservative so as to bound a specified percentage of the underlying PDF
  - An opposing goal (making this a difficult problem) is that the uncertainty representation not be overly conservative—i.e., should minimally over-estimate the specified variability range of the true PDF.

Four Methods Tried for Sparse-Data Uncertainty Representation

Sandia National Laboratories

- **<u>Method 1</u>: Fit the data with a Normal distribution**
  - An easy and very common approach used in engineering practice.
  - But how well does it work?

- **<u>Method 2</u>: Tolerance Intervals from classical statistics**

- **<u>Method 3</u>: PDF estimation by Pradlwarter-Schueller KDE (kernel density estimation) approach to fitting the data (slide follows)**

- **<u>Method 4</u>: PDF estimation by Sankararaman-Mahadevan "Non-Parametric" PDF fit to the data (slides follow)**

- **Evaluation criteria:**
  - how well the approaches perform on the two stated objectives:
    - conservative ⇔ but not overly conservative
    - <u>assess on illustrative problem</u>:
      - estimate 95% PDF coverage range (central 0.95 percentile) of exact PDFs from which data samples are drawn
  - ease of implementation and use for engineering practice

Classical Tolerance Interval Method for Dealing with Sparse Data

Sandia National Laboratories

- **Approach**
  - calculate the standard deviation σ of the data
  - multiply σ by appropriate factor *f* from statistical tables
  - create interval bars of extent $f\sigma$ about the mean μ of the data: $\mu \pm f\sigma$
  - 0.9/0.95 Tolerance Intervals — in this study, the factors *f* correspond to approximate 90% confidence that the produced tolerance interval encompasses the central 95-percentile range (between the 0.025 and 0.975 percentiles) of the true PDF

# samples	$f_{0.9/0.95}$
2	18.80
4	4.94
8	3.26
12	2.86
20	2.56
30	2.41
40	2.33
∞	1.96

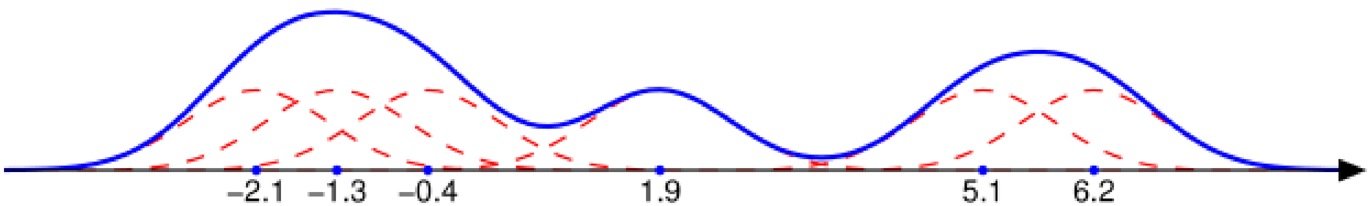
- **Very simple to use in practice**

Pradlwarter-Schueller KDE Method for Dealing with Sparse Data

Sandia National Laboratories

Pradlwarter, H.J., and G.I. Schuëller, "The use of kernel densities and confidence intervals to cope with insufficient data in validation experiments," *Computer Methods in Applied Mechanics and Engineering*. Vol. 197, Issues 29-32, May 2008, pp. 2550-2560.

- **Kernel Density Estimation (KDE) is a technique used to estimate the density of a random variable *X* given *n* independent samples $X_1, \ldots, X_n$ of it.[1]**



$$f_n^{\{KDE\}}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

- **We used a Gaussian Kernel**
- **KDE is very sensitive to the bandwidth, *h*. Bandwidth estimation typically involves optimization (e.g. maximize a cross-validation likelihood)**
- **For small samples, we used the approach in Pradlwarter and Schuëller: find *h* to satisfy a fixed probability:** $\int_{-\infty}^{a} f(x;h)dx + \int_{b}^{+\infty} f(x;h)dx = P(\alpha, n)$

1. Figure taken from http://en.wikipedia.org/wiki/Kernel_density_estimation

Sankararaman-Mahadevan Method for constructing a PDF from Sparse Data

Sandia National Laboratories

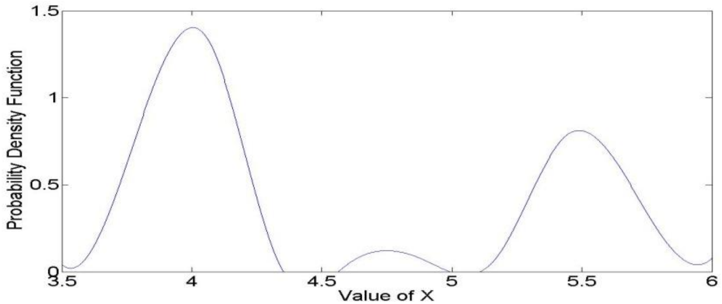
Shankar Sankararaman & Sankaran Mahadevan, "Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data," *Reliability Engineering and System Safety*, doi:10.1016/j.ress.2011.02.003.

- **Discretize the domain of $X \rightarrow \theta_i$, i = 1 to Q**
- **PDF values at each of these Q points**
  - **$f_X(x=\theta_i)=p_i$ for i = 1 to Q**
- **Interpolation technique (cubic splines used here)**
- **Construct likelihood as a function of:**
  - Discretization points θ_i selected, i = 1 to Q
  - Corresponding probability density function values p_i
  - Type of interpolation technique

$$L \propto \left(\prod_{i=1}^{n} \int_{a_i}^{b_i} f_X(x)dx \right) \left(\prod_{i=1}^{m} f_X(x_i) \right)$$

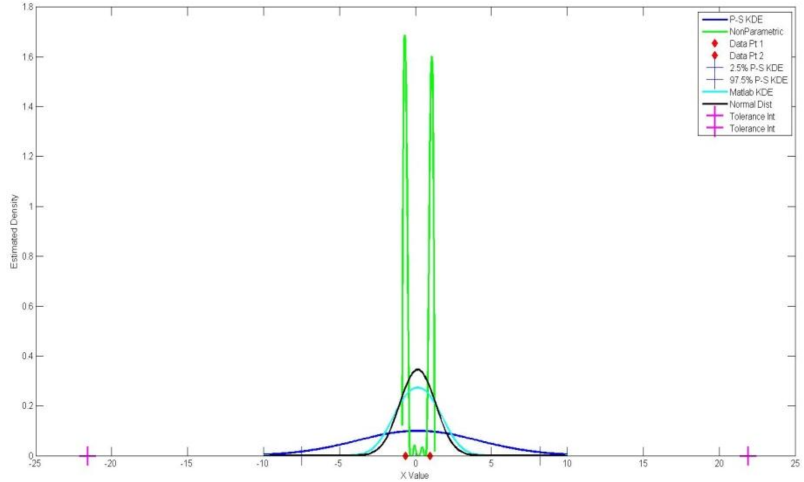
- **Maximize Likelihood $L(p)$ where p = {p_i ; i = 1 to Q) subject to applicable constraints**
  - **Estimate $f(x)$**

Probability Density Function



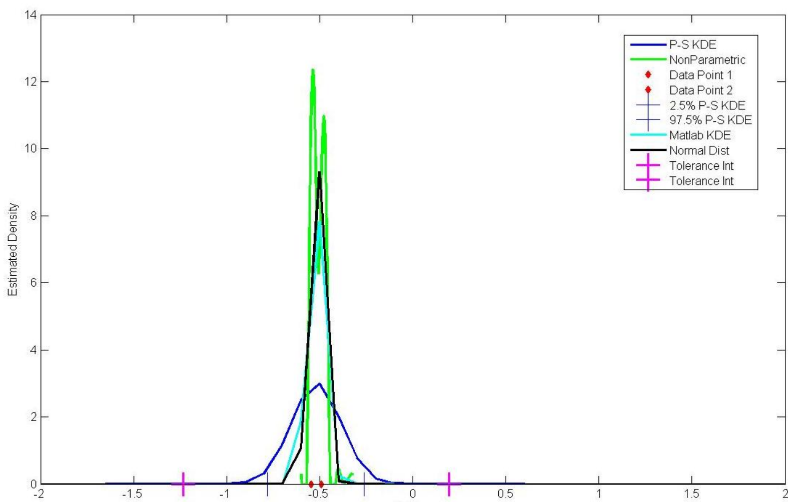
Illustrative PDFs or 95% Intervals from the Four Methods + a Matlab-KDE Method

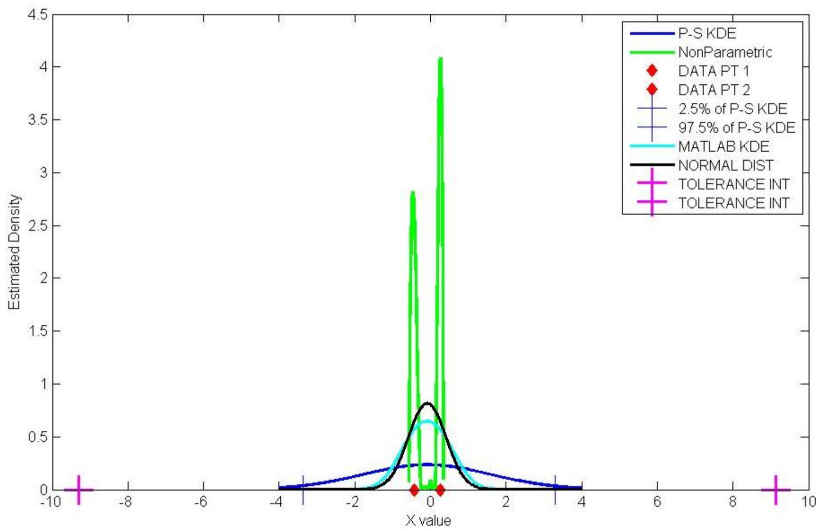
- **Three instances of 2 samples from a Normal PDF**
  - **2 samples "close" together**
  - **2 samples "far" apart**
  - **2 "intermediately" spaced**



far

close

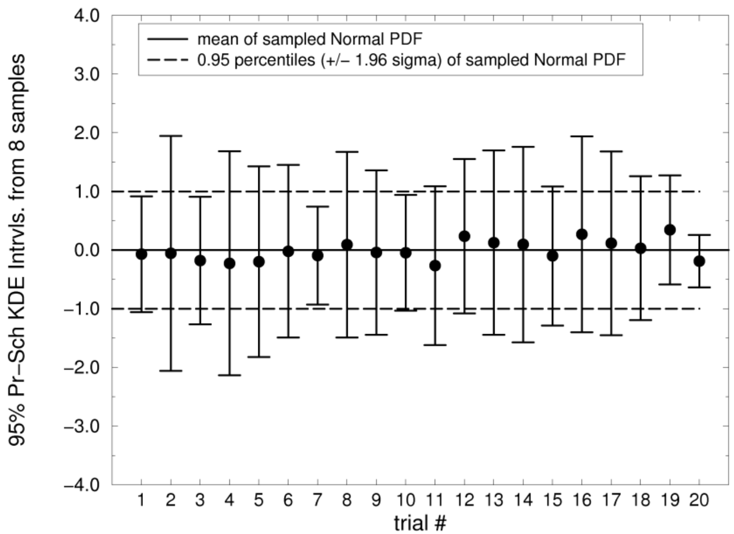


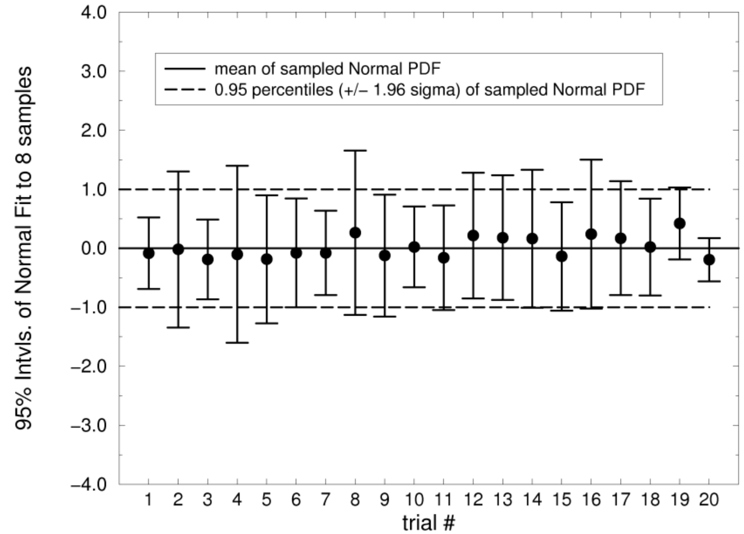


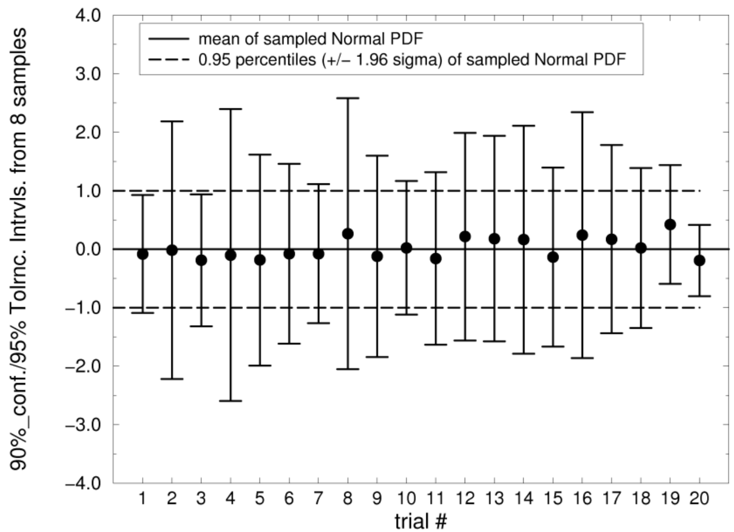
inter med.

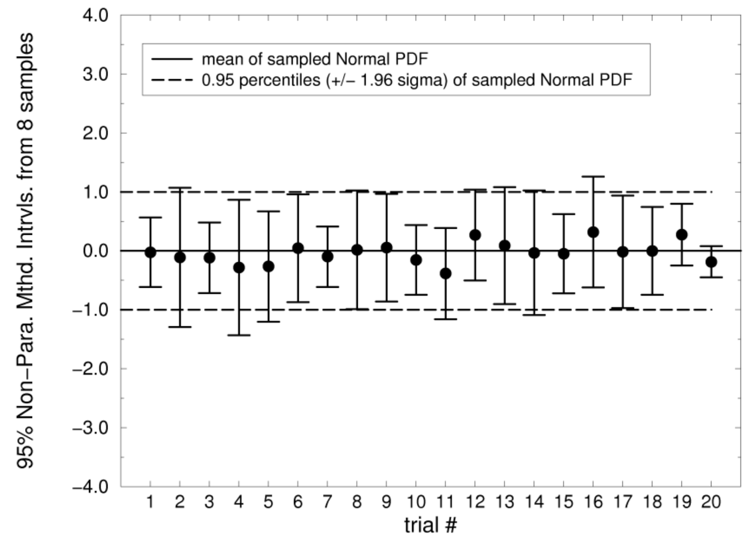
Illustrative 95% Intervals from the Four Methods

– 20 trials per method, 8 random samples of a Normal PDF per trial







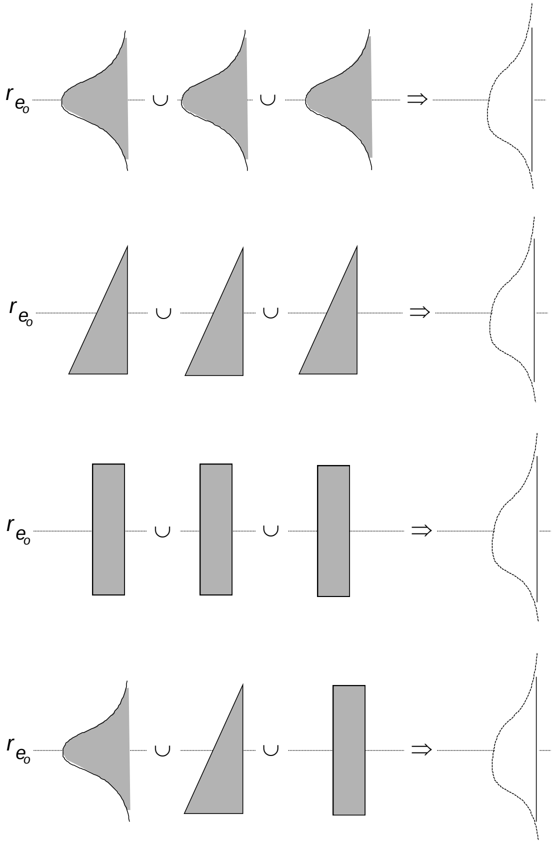


- **Tol. Intvls. most conservative, then Pr-Sch, Normal Fit, and Non-Para.**

The Larger Test Matrix

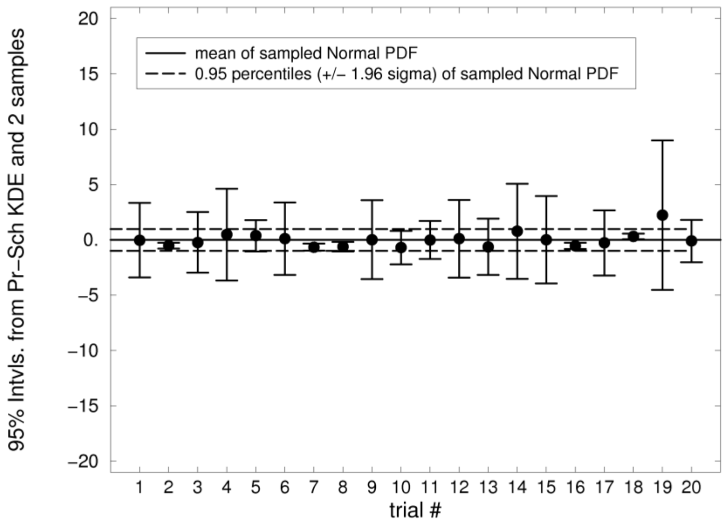
Sandia National Laboratories

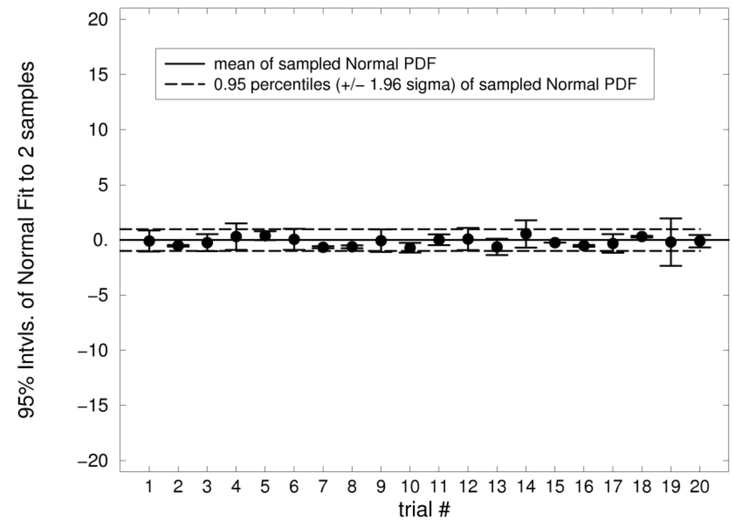
- **1000 trials of each method for fitting sample data from each of:**
  - normal PDF
  - right-triangular PDF
  - uniform PDF
  - Convolutions of these PDF types (figure at right) acting as three equally dominant sources of random uncertainty in a linear system.
    - Answer question: does the presence of multiple sources of uncertainty smooth or mitigate the errors in representing the individual PDFs?
- **Fit the data for sample sizes of *n* = 2, 8, 32 for each PDF**
- **Initial results reported here and in the paper are for Top Row of figure only**

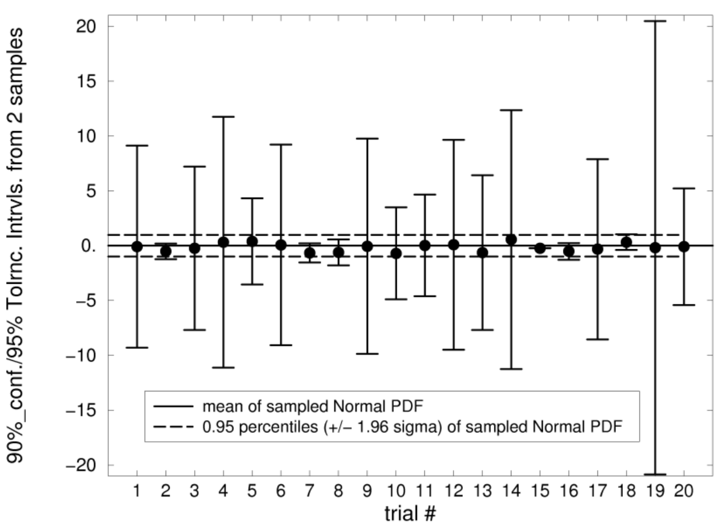


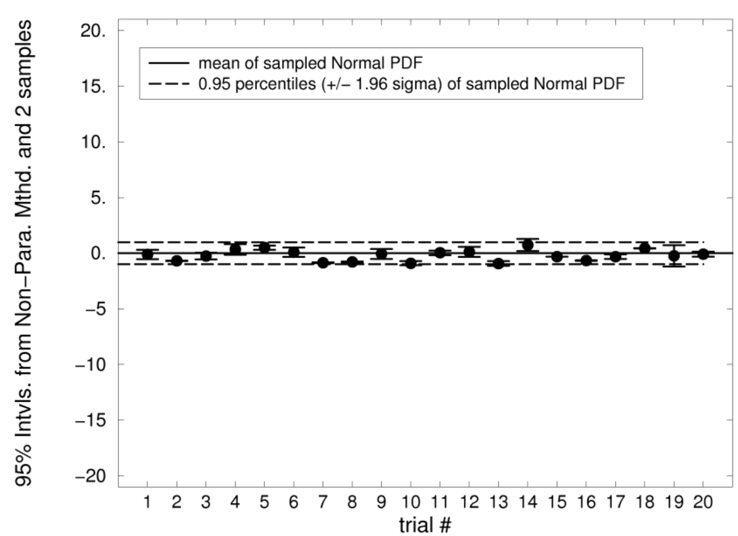
Glimpse of 95% Intervals from the Four Methods

– 20 trials per method, 2 random samples of a Normal PDF per trial





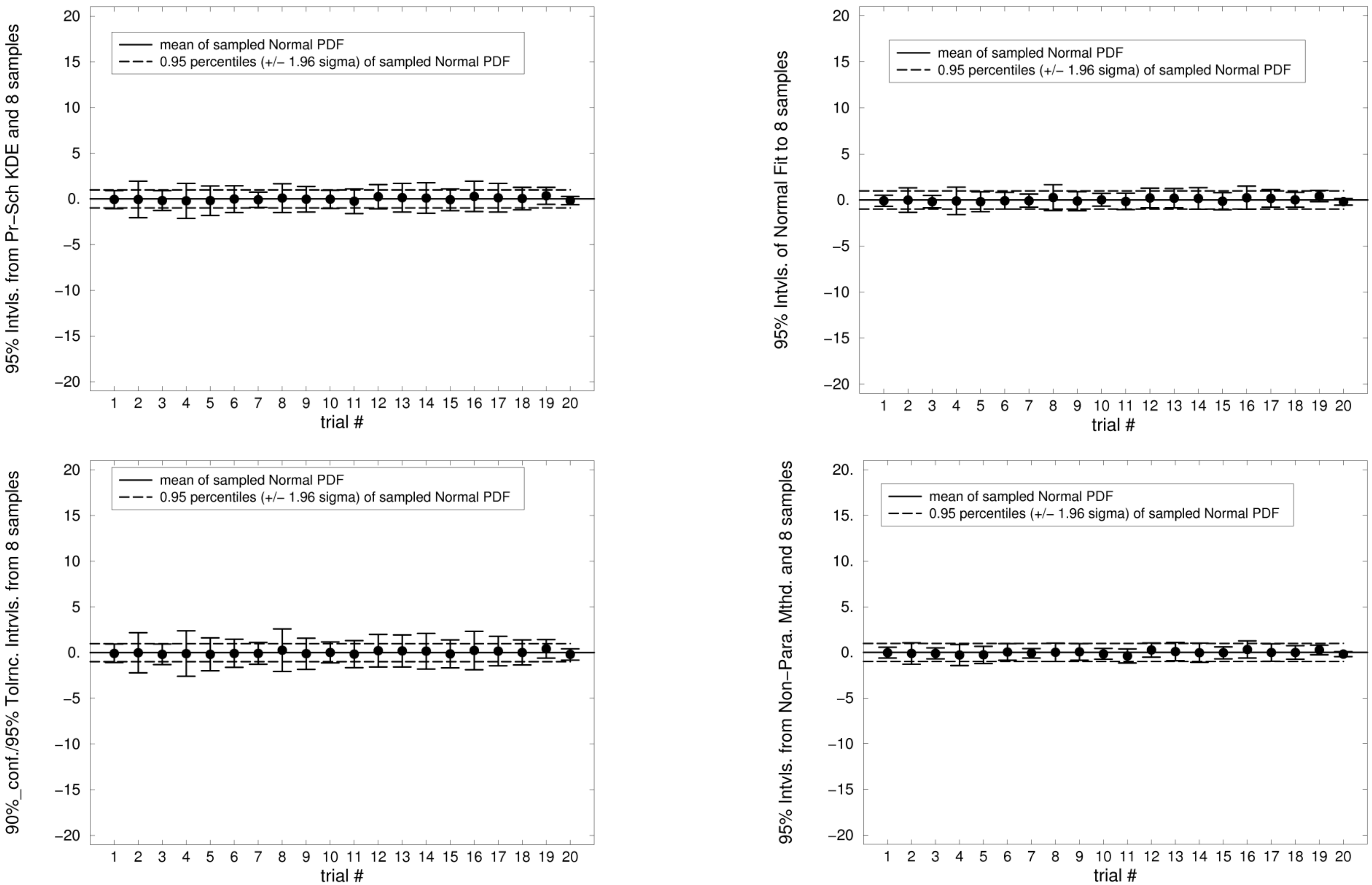




- **Method results differ greatly for low numbers of samples.**

Glimpse of 95% Intervals from the Four Methods

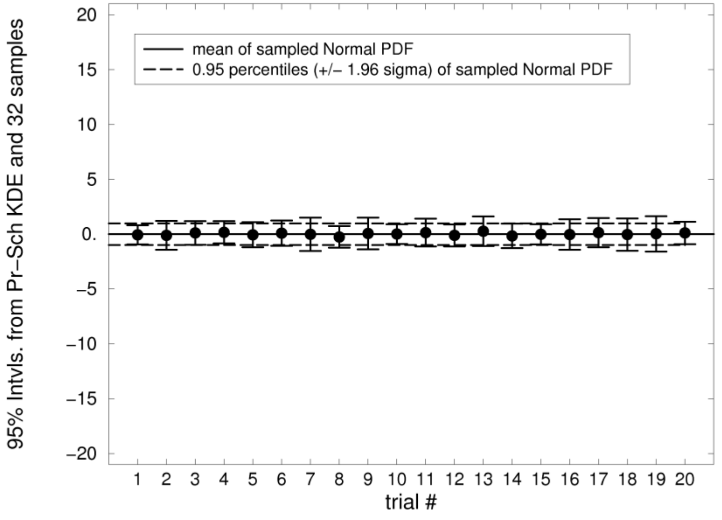
– 20 trials per method, 8 random samples of a Normal PDF per trial

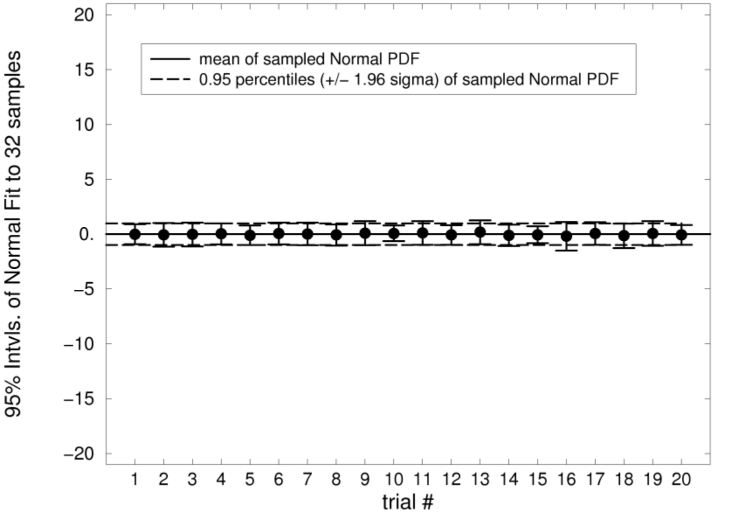


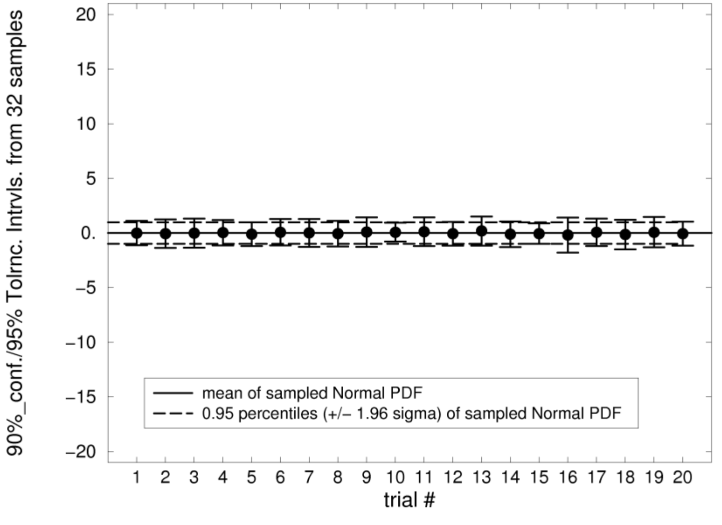
- **Interval estimation improves markedly as # of samples increases to 8.**

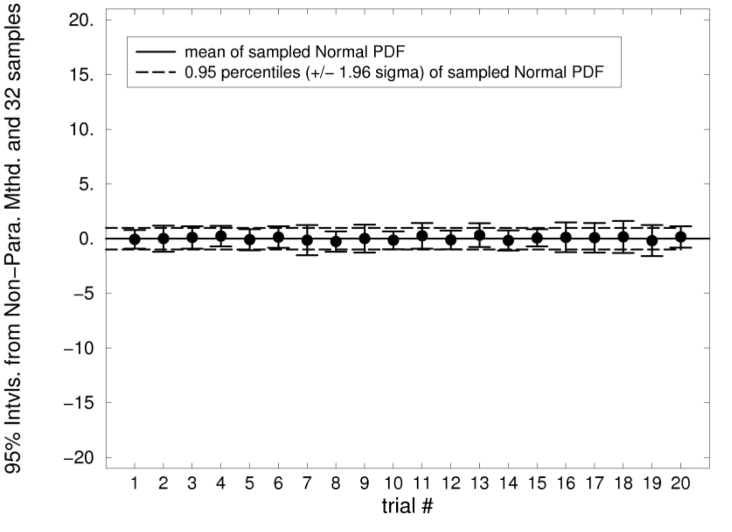
Glimpse of 95% Intervals from the Four Methods

– 20 trials per method, <u>32</u> random samples of a Normal PDF per trial







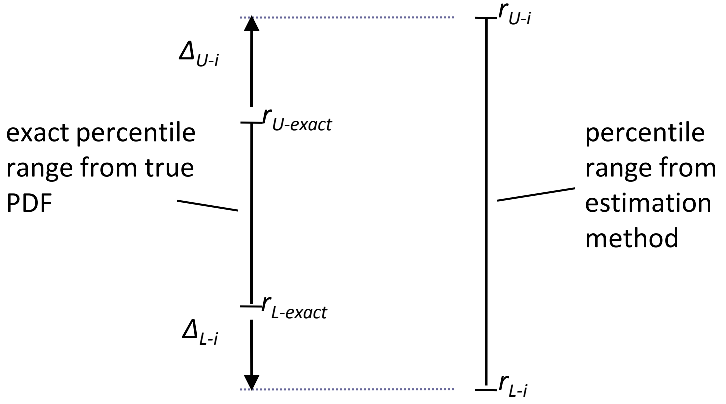


- **Differences in method performance diminish for “large” # of samples.**

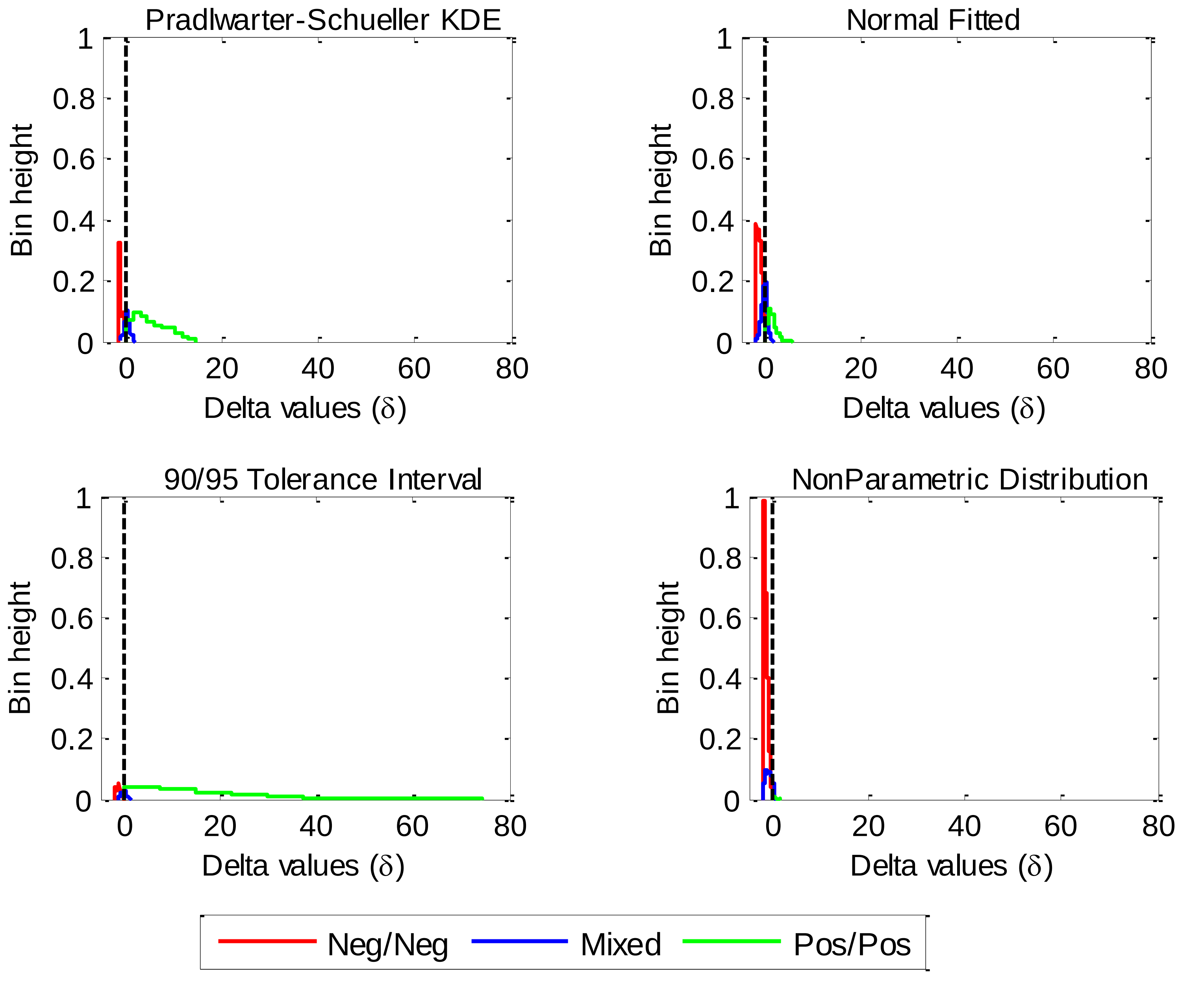
Quantification of Error of 95% Intervals produced by the various methods

Sandia National Laboratories

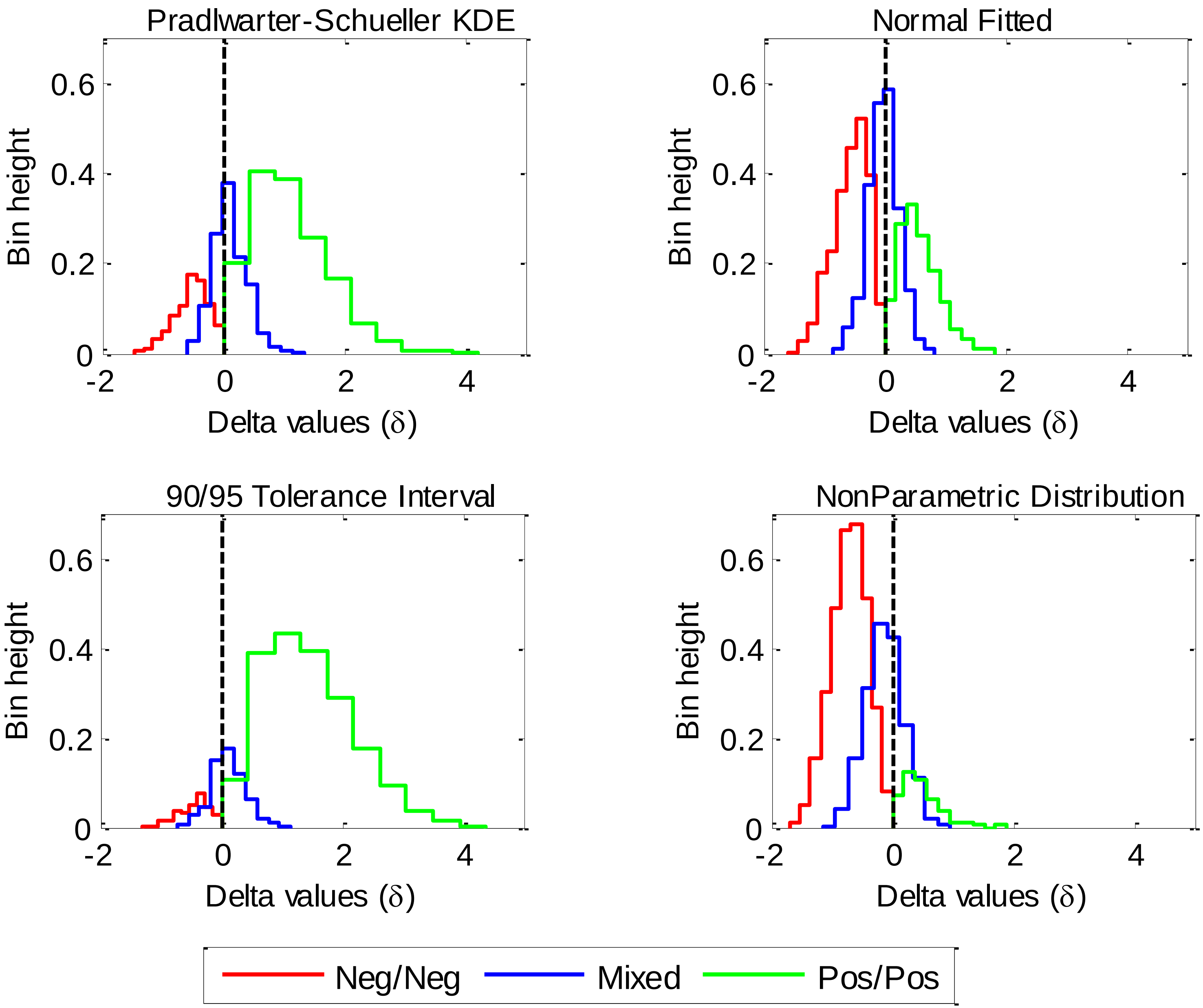
- **Interval estimation error is quantified and put into one of three bins as follows:**
  - **+/+** *or* **pos/pos error (**both Δ_U and Δ_L in schematic at right are positive**)**
    - **estimated interval exceeds true 95% range at both top and bottom (estimated interval encompasses true interval)**
    - **most desirable according to objective**
  - **+/− *or* −/+ "mixed" error (**Δ_U and Δ_L are of opposite sign**)**
    - **estimated interval encompasses true 95% range at one end, but falls short at other end**
    - **less desirable**
  - **−/−** *or* **neg/neg error (**both Δ_U and Δ_L are negative**)**
    - **estimated interval falls short of true 95% range at both top and bottom**
    - **least desirable**



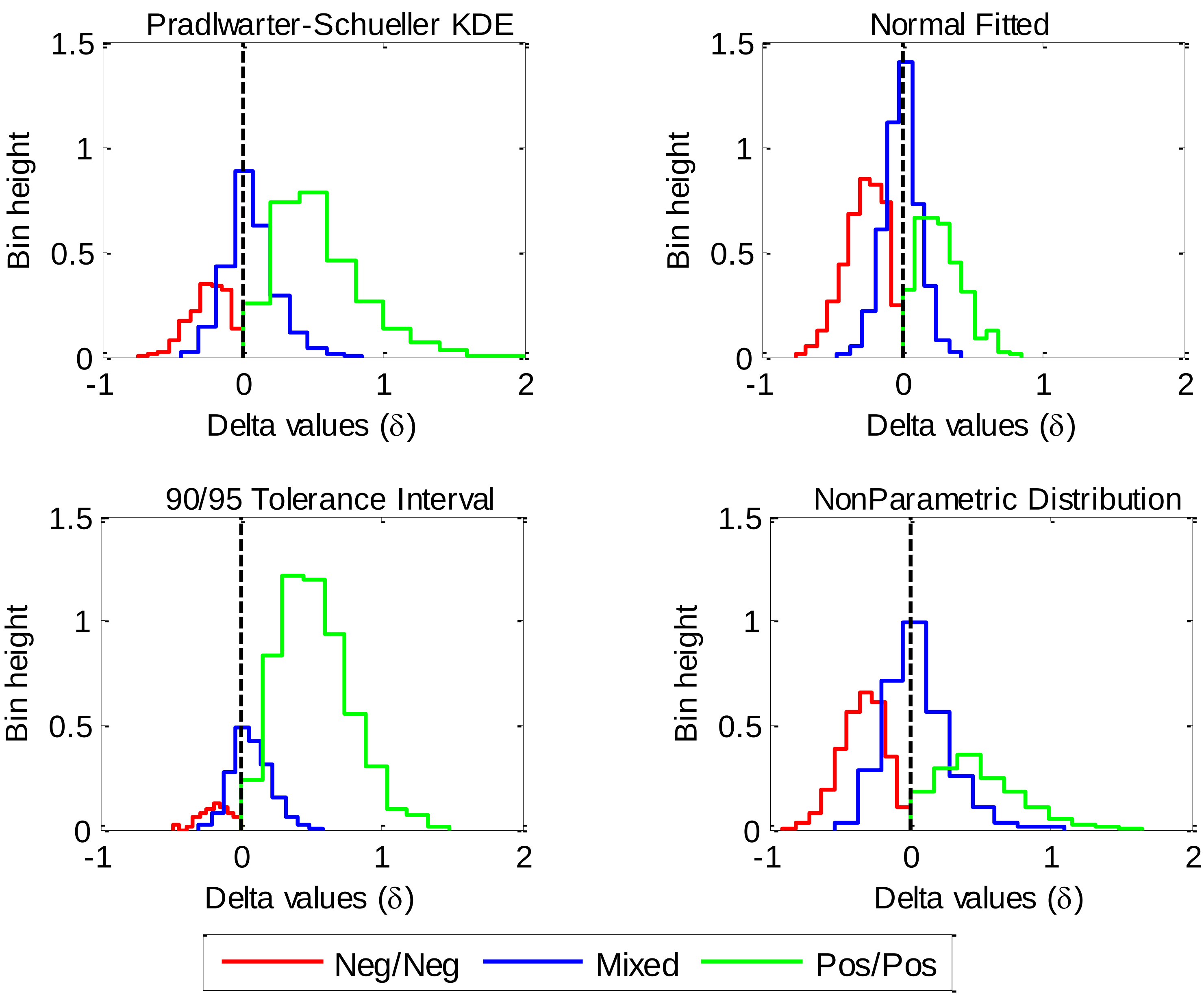
Error histograms — Normal PDF, 2 samples



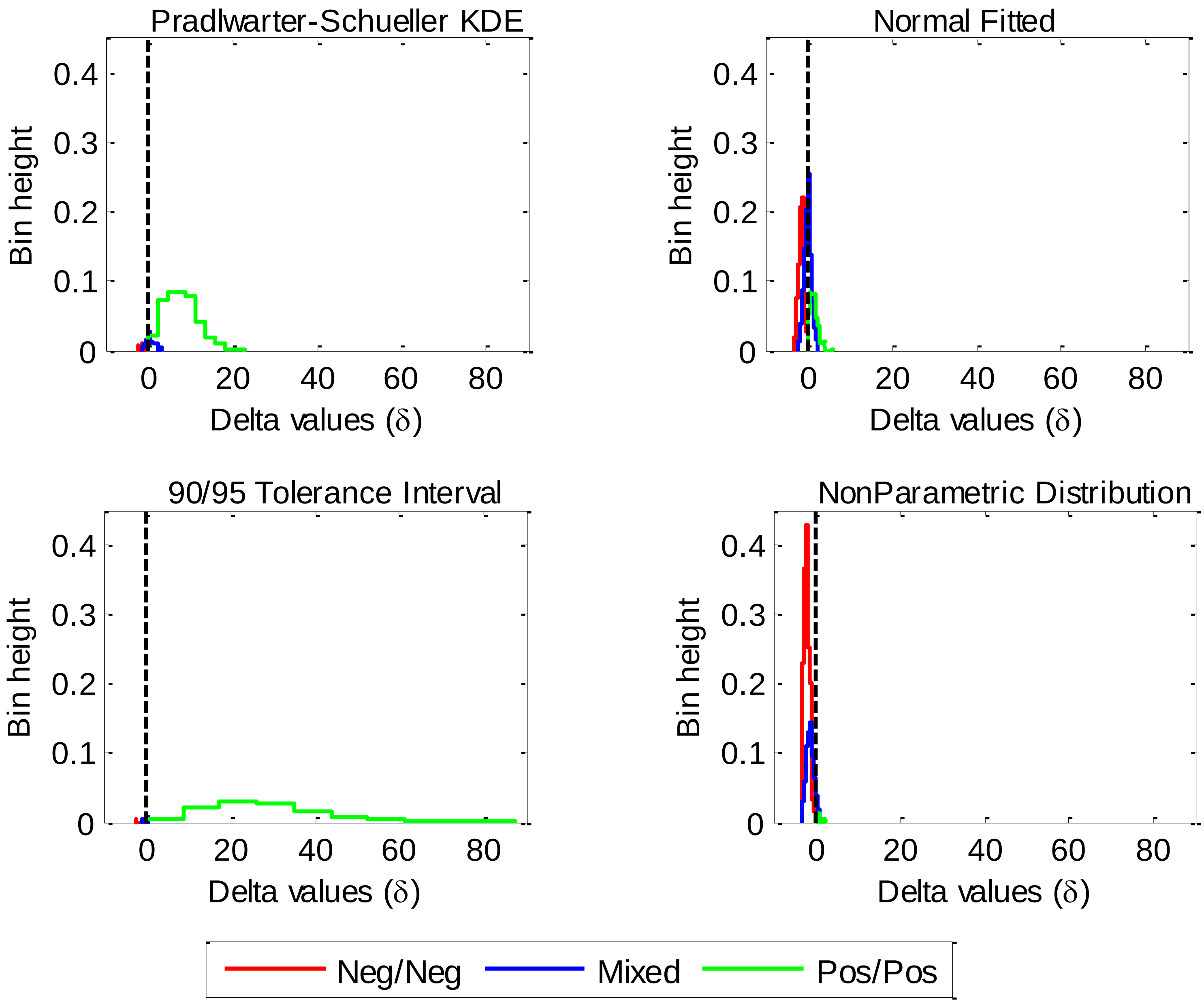
Error histograms — Normal PDF, 8 samples



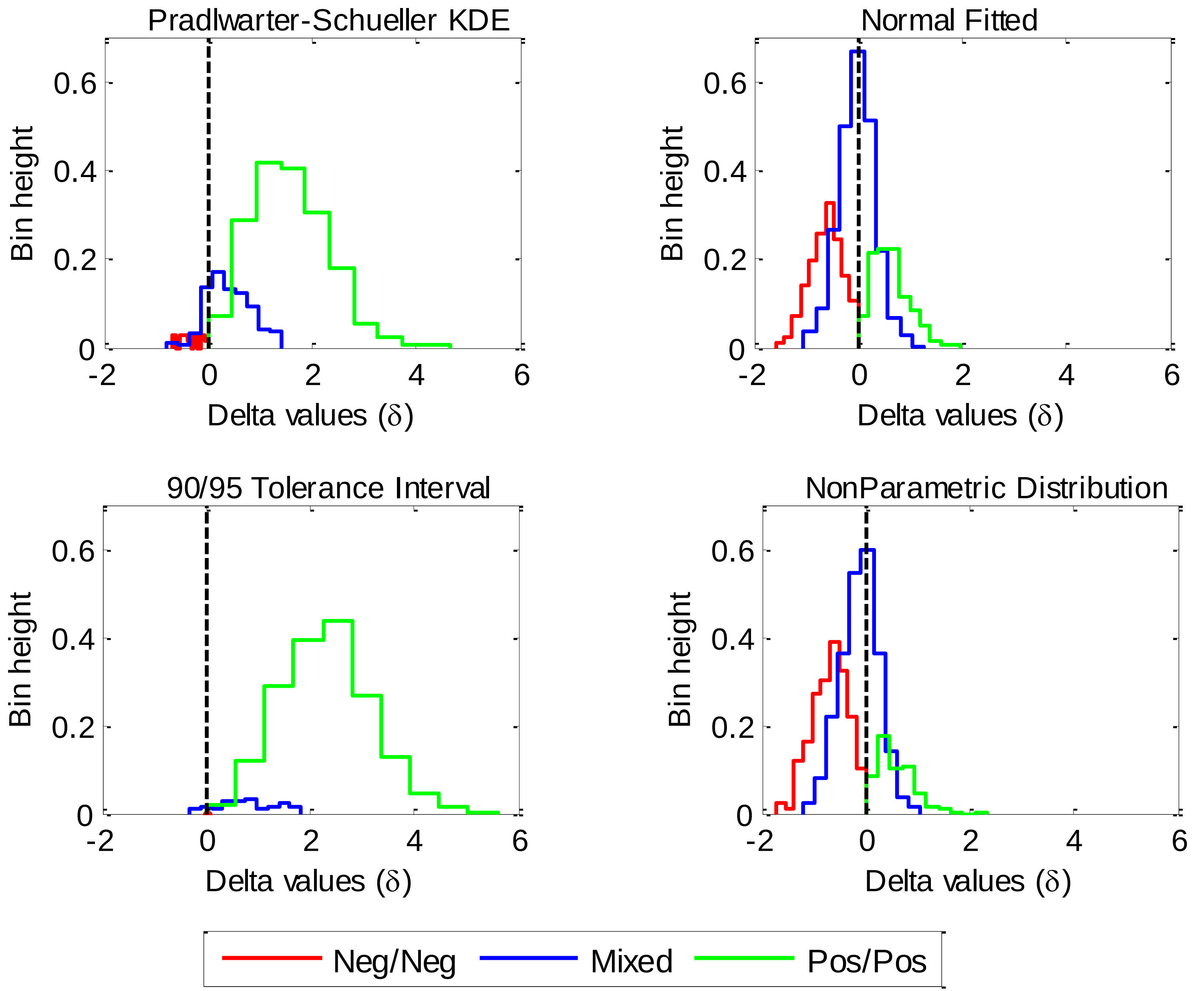
Error histograms — Normal PDF, 32 samples



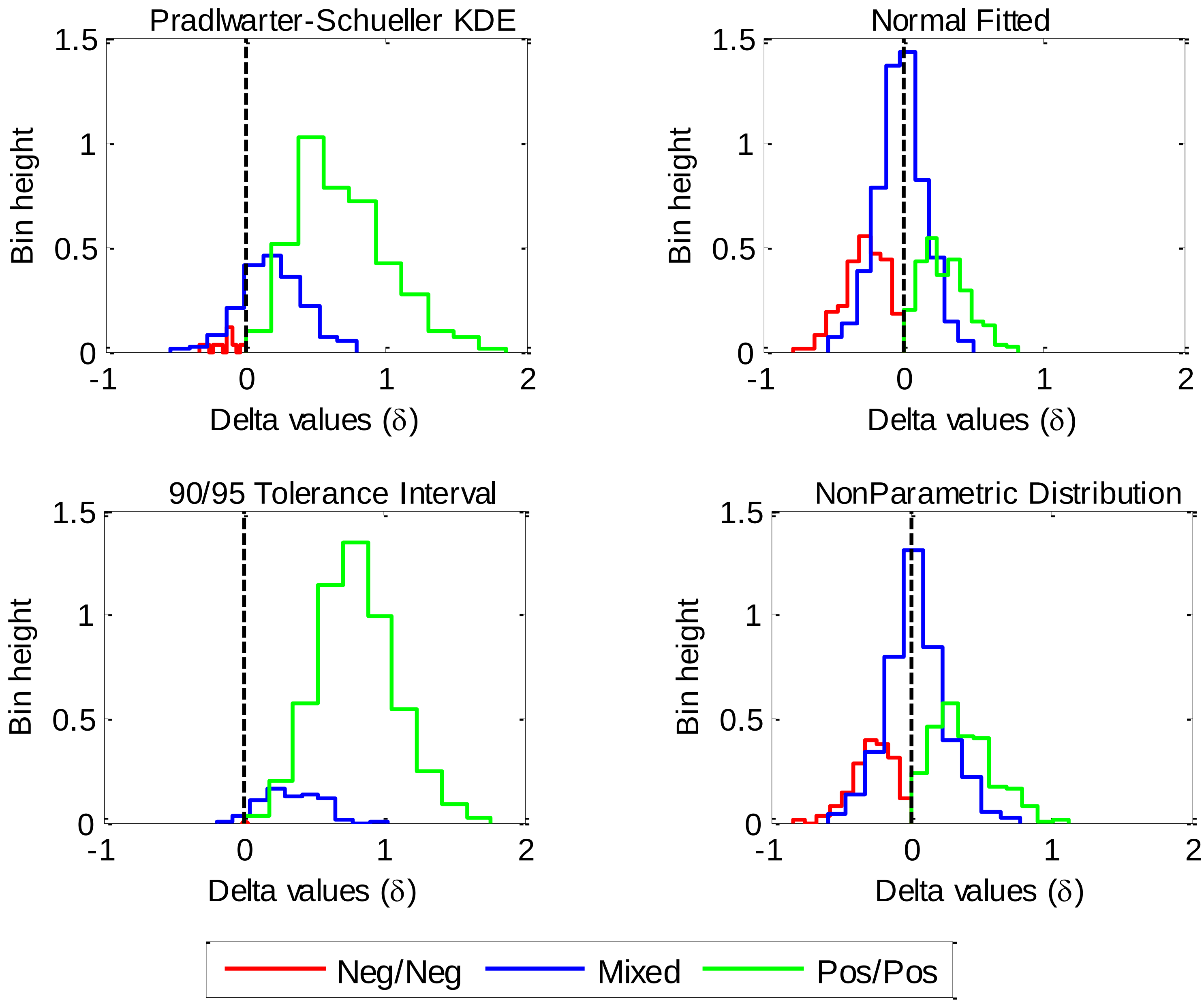
Error histograms — Convolution, 2 samples



Error histograms — Convolution, 8 samples



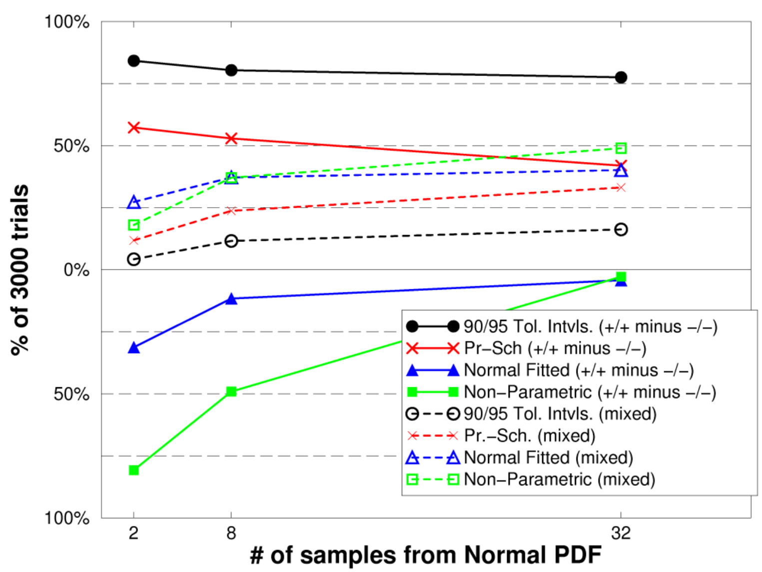
Error histograms — Convolution, 32 samples



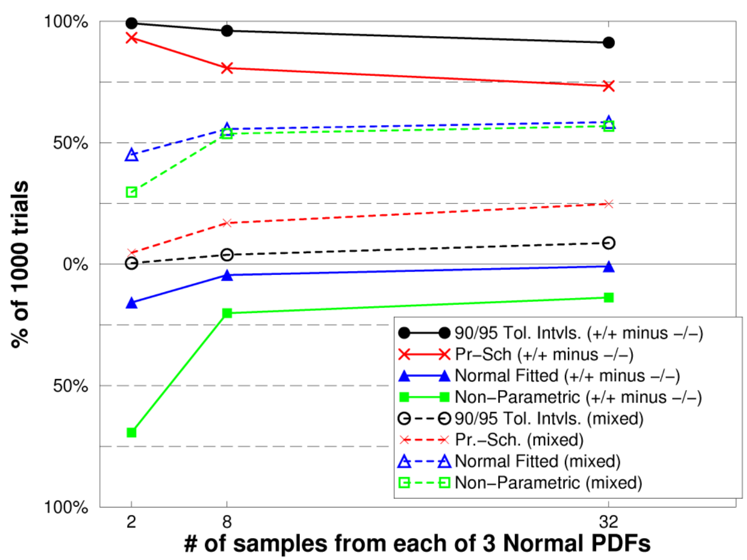
% of Errors in +/+, −/−, and mixed categories for each Method

Sandia National Laboratories

Performance in Fitting Normal PDF



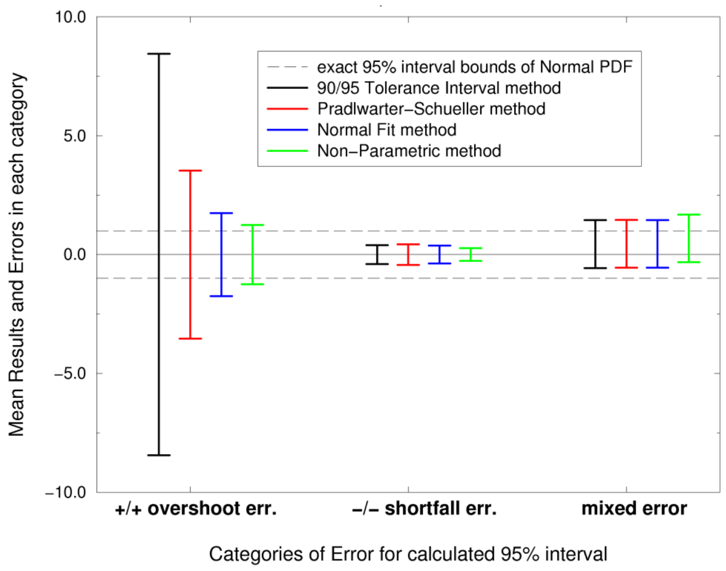
Performance in Convolution



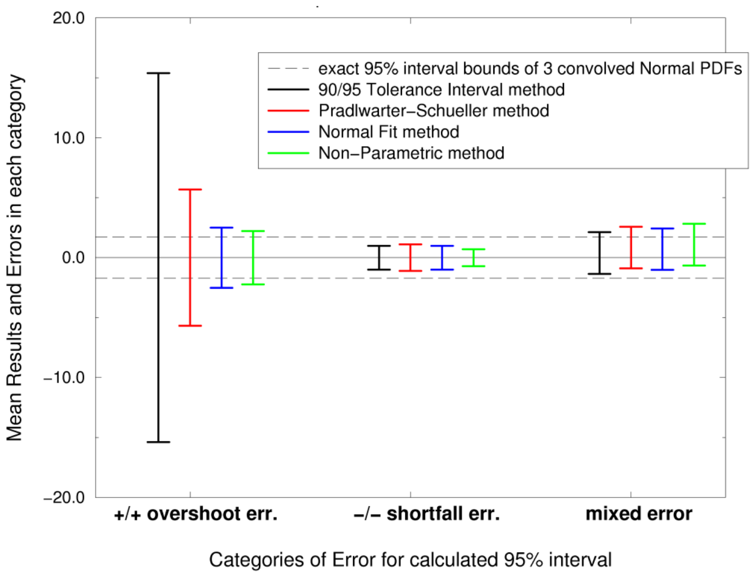
- **The higher the solid lines in the graphs, the more net desirable results (%pos/pos results minus %neg/neg results) occurred.**
- **This desirability measure falls slowly for the better performers (Tol. Intervals and Pr-Sch) with increasing # of samples as the % of less desirable mixed results rises slowly (dashed lines). Proportions of +/+, −/−, and mixed results for Normal-Fit and Non-Para. mthds. improve w/ #samples but still net-undesirable even at 32 samples.**
- **These trends persist for performance in convolution, with slight improvement generally in desirability of percentages of results in the +/+, −/−, and mixed bins.**

Mean Results and Error Magnitudes in each Category for each Method at n=2 samples

Performance in Fitting Normal PDF



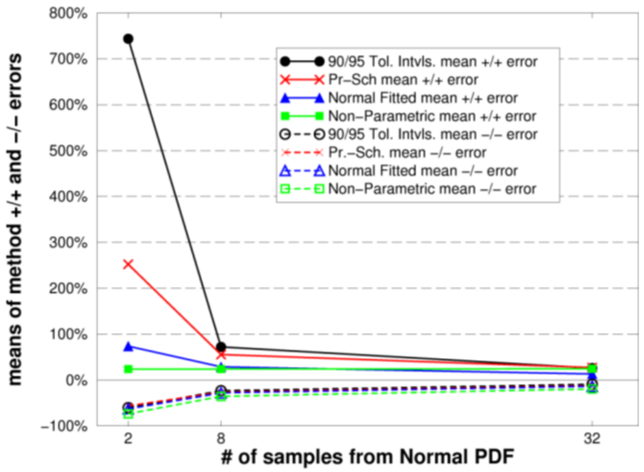
Performance in Convolution

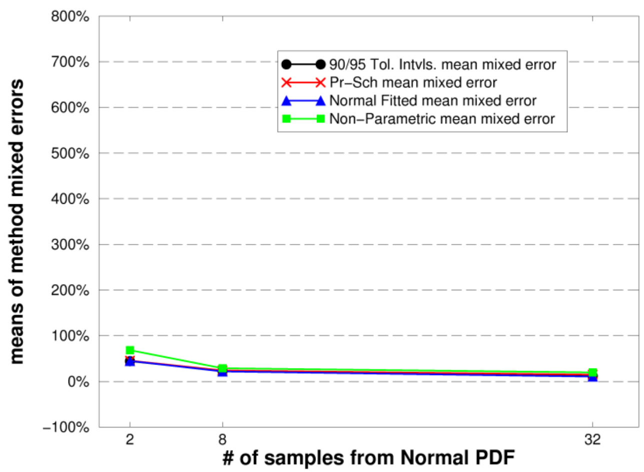


- **Tol. Interval method has much larger average overshoot errors (in +/+ category) than all other methods, with Pr-Sch next largest, then Normal-Fit and Non-Para. Mthds.**
- **Tol. Interval method is least desirable in this performance objective, then Pr-Sch, then the Normal-Fit method, with the Non-Parametric method being most desirable.**
- **Method differences much smaller for −/− shortfall errors and mixed +/− and −/+ errors.**
- **Conclusions same for performance in convolution, where +/+ overshoot errors are generally slightly larger %wise, but the −/− and mixed errors are generally smaller %wise (see next slide).**

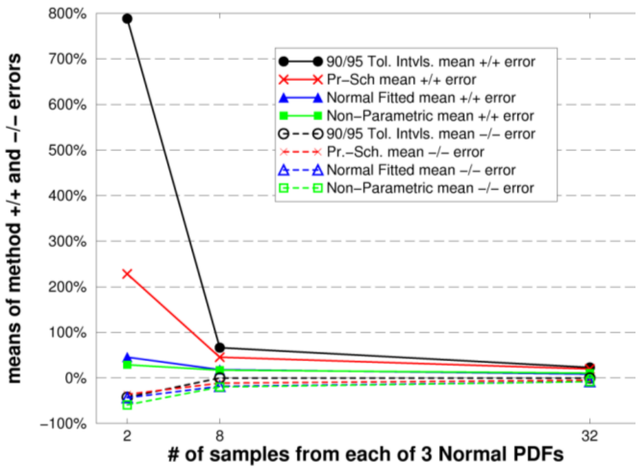
Mean %Error Magnitudes for each method vs. # of samples

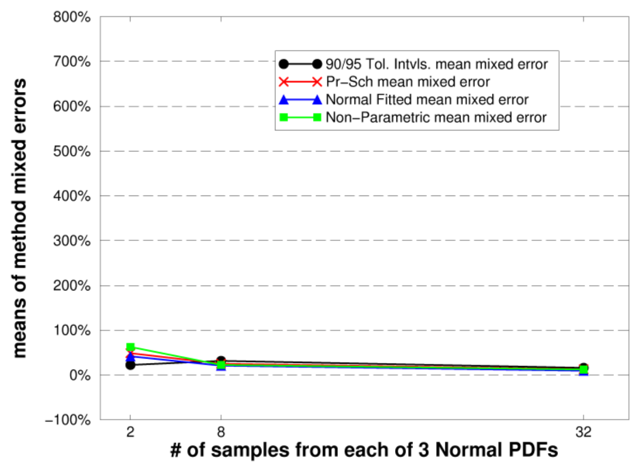
Performance in Fitting Normal PDF



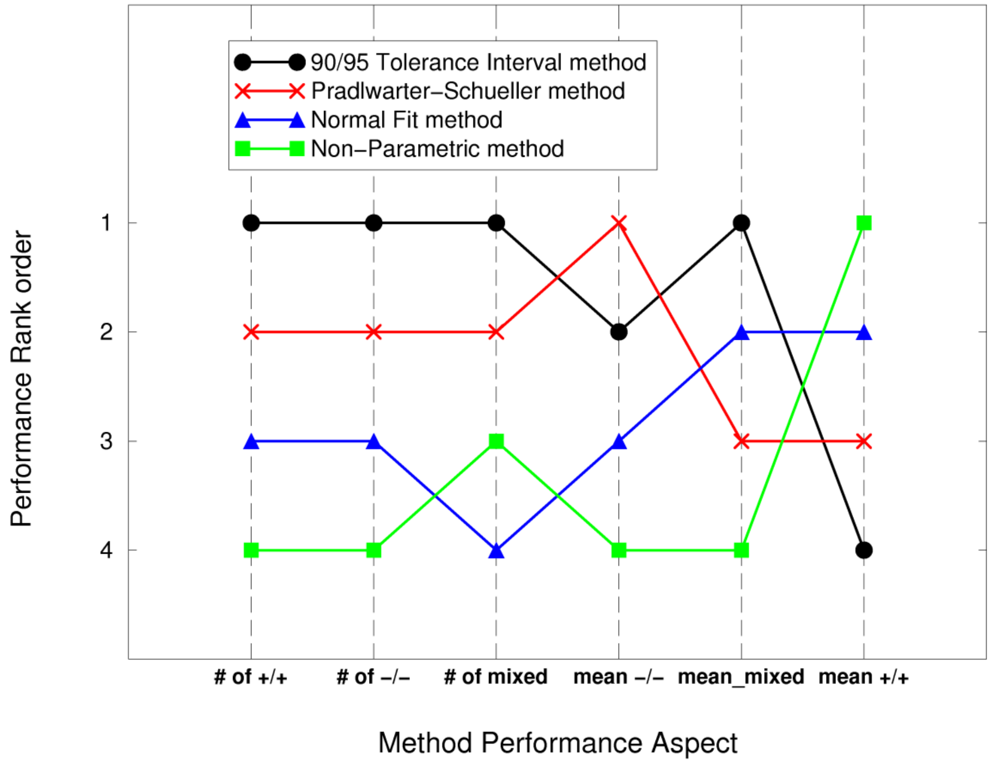


Performance in Convolution





Zeroth-Order Ranking Scheme for Method Performance



- **No method consistently best for all six performance attributes.**
- **Total score across all six attributes yields the following ranking:** **1**-Tol. Intervals, **2**-Pr-Sch, **3**-Normal Approximation, **4**-Non-Parametric mthd.
- **Method ranking on total score is robust whether maximum instead of mean errors are considered, or Normal PDF or Convol., or n=2,8,32.**

Tentative Conclusions based on this Initial Partial Study (Normal PDFs)

Sandia National Laboratories

- **Tolerance Interval method performed best on balance according to the simple ranking scheme employed. The method is also very simple to use.**
- **However, the Zeroth-Order ranking scheme used is very simplistic:**
  - weights all six attributes equally
  - only considers performance order in each attribute, not *how much better* one method performs vs. another.
- **A different ranking could occur with a more sophisticated performance scoring system and when the other aspects of the fuller study are performed.**
  - In particular, Pradlwarter-Schueller KDE general performed well and at low # of samples has less egregious +/+ overshoot errors than Tol. Intvls., but also has significantly less desirable +/+ results and significantly more undesirable −/− & mixed results than Tol. Intvls. Implementation of the Pr-Sch method is also considerably more involved.
- **The disadvantage of large Tol. Intvl. +/+ overshoot errors relative to the other methods diminishes by 8 samples, but the advantage in terms % of net desirable results remains through 32 samples.**
- **The common practice of fitting a Normal distribution to the data underperforms even if the underlying PDF being sampled is Normal!**
  - This common practice is inadequate for dealing with sparse data and should be abandoned in favor of the Tolerance Interval or Pradlwarter-Schueller KDE approaches.
- **The presence of multiple sources of uncertainty in aggregation (convolution) does not substantially lessen the effect of method approximation errors in representation of the individual contributing PDFs.**