SANDIA REPORT
SAND2013-4561
Unlimited Release
Printed September 2013

A Comparison of Methods for Representing Sparsely Sampled Random Quantities

Vicente Romero, Laura Swiler, Angel Urbina, Josh Mullins

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2013-4561
Unlimited Release
Printed September 2013

A Comparison of Methods for Representing Sparsely Sampled Random Quantities

Vicente Romero, Laura Swiler, Angel Urbina
Verification & Validation, Uncertainty Quantification and Credibility Processes Dept.
Sandia National Laboratories[1]
P.O. Box 5800
Albuquerque, New Mexico 87185-0825

Josh Mullins
Vanderbilt University PhD Candidate and Sandia Student Intern

Abstract

This report discusses the treatment of uncertainties stemming from relatively few samples of random quantities. The importance of this topic extends beyond experimental data uncertainty to situations involving uncertainty in model calibration, validation, and prediction. With very sparse data samples it is not practical to have a goal of accurately estimating the underlying probability density function (PDF). Rather, a pragmatic goal is that the uncertainty representation should be conservative so as to bound a specified percentile range of the actual PDF, say the range between 0.025 and .975 percentiles, with reasonable reliability. A second, opposing objective is that the representation not be overly conservative; that it minimally over-estimate the desired percentile range of the actual PDF. The presence of the two opposing objectives makes the sparse-data uncertainty representation problem interesting and difficult. In this report, five uncertainty representation techniques are characterized for their performance on twenty-one test problems (over thousands of trials for each problem) according to these two opposing objectives and other performance measures. Two of the methods, statistical Tolerance Intervals and a kernel density approach specifically developed for handling sparse data, exhibit significantly better overall performance than the others.

[1] Sandia is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

ACKNOWLEDGMENTS

The authors thank the Advanced Simulation and Computing (ASC) and Campaign 7 programs of the National Nuclear Security Administration (NNSA) at Sandia National Laboratories for funding this research.

CONTENTS

FIGURES

TABLES

NOMENCLATURE

ASC	Advanced Simulation and Computing
CDF	cumulative density function
DOE	Department of Energy
JN	Johnson method
KDE	Kernel density estimation
MISE	mean integrated squared error
MLCV	maximum-likelihood cross validation
NF	Normal Fitting
NLLS	non-linear least squares
NNSA	National Nuclear Security Administration
NP	non-parametric method
PDF	probability density function
PS	Pradlwarter-Schuëller
TI	Tolerance intervals

1 INTRODUCTION

This report discusses and tests various statistical concepts and techniques for expressing uncertainty due to random variability (aleatory uncertainty) when limited samples exist of a random-variable quantity. Limited sampling introduces an epistemic contribution of uncertainty to the problem of random-variable characterization. The importance of this topic extends beyond experimental uncertainty characterization to situations where the derived experimental information is used for model validation or calibration purposes.

If the random quantity is fully sampled such that its characteristic probability density function (PDF) is fully known, then well-known approaches can be used to express and work with the probabilistic uncertainty. These include Monte Carlo and Quasi Monte Carlo approaches (often involving constructed response surfaces for non-linear response functions, e.g. references [1]-[4]); linearized-response function approaches popular in experimental uncertainty quantification (e.g. [5]-[10]); and more recent approaches like polynomial chaos and stochastic expansion methods for uncertainty propagation and aggregation (e.g. [11]).

However, when relatively few samples are available a substantial epistemic contribution of uncertainty exists in addition to the aleatory uncertainty due to the quantity's random variation. This epistemic uncertainty undermines accurate estimation of the underlying variability distribution or PDF. Hence, it is not practical to pursue a goal of accurate PDF estimation from sparse data. Rather, a pragmatic goal is that the uncertainty representation should be ***conservative*** so as to bound a desired percentage of the actual PDF, say 95% included probability, with reasonable reliability. A second, opposing objective is that the representation not be overly conservative; that it minimally overestimate the desired percentile range of the actual PDF. The presence of the two opposing objectives makes the sparse-data uncertainty representation problem an interesting and difficult one.

A classical approach to working with sparse random data is to use statistical tolerance intervals ([12], [13]) to provide a reliably conservative interval representation for the combined epistemic and aleatory uncertainty associated with the limited data. The Tolerance Interval approach is tested in this report, along with the very common practice of simply fitting the sample data with a normal distribution. A recently developed kernel-density estimation technique specifically designed to cope with limited data ([14]) is also tested, along with a "non-parametric" cubic-spline PDF method based on a likelihood function fit to the data ([15]). Finally, taking a cue from [16] and [17] an approach is tried that uses a Johnson-family four-parameter representation of PDFs ([18]). An approach introduced in [19] does not appear to address the type of sparse-data induced uncertainty that we are concerned with here.

Section 2 of this report describes the uncertainty representation problem and the test matrix for assessment of the sparse-data approaches. The approaches are described in detail in Section 3. Section 4 presents results and performance comparisons. Previous results from [20] and [21] are summarized and new results are presented and synthesized in this report. Section 5 summarizes conclusions of the study.

2 TEST PROBLEMS AND PROCEDURES FOR SPARSE-DATA UNCERTAINTY REPRESENTATION METHODS

The set of problems studied is represented in Figure 1. Diverse PDFs of normal, uniform, and right-triangular shapes are randomly sampled and the samples are used by the methods under investigation to identify a range that ostensibly covers the 0.025 to 0.975 percentile ranges of the PDFs. The percentile range 0.025 to 0.975 was selected because it corresponds to approximately ±2σ (±2 standard deviations) for a Normal distribution and ±2σ is used for many engineering purposes (e.g. measurement uncertainty and model validation assessment) to gauge extent of uncertainty ([5]-[10]). The PDFs are sized and located relative to the r_{e_o} reference lines in the figure such that the range between the said percentiles is 2 for all PDFs and the range is centered on the reference line. Accordingly, the 0.025 and 0.975 percentiles of the normal, uniform, and triangular PDFs reside at values of -1 and 1.

The methods are tested for their performance in estimating, from random samples of the PDFs, ranges that contain the true 0.025 to 0.975 percentile range (-1,1). The estimated ranges are compared to the true ranges and method performance is assessed and characterized according to the metrics and procedures presented in Section 4 of this report.

Given our resource constraints the study involves sample sizes of n = 2, 8, and 32 samples from each PDF. Since the increase in sample size is a constant factor (of 4) each time, a rate of decrease of method estimation errors with increasing sample size can be calculated, but this is not done in this document.

The study also assesses method performance for fitting multiple sources of sparsely sampled probabilistic uncertainty (from the sets of PDFs in each row in Figure 1) that aggregate to a total resultant uncertainty for a system. It was surmised that the aggregation of multiple sources of uncertainty may smooth out and reduce, or alternatively could amplify, the performance differences between the methods, so the test plan investigates this aspect. Since often only a handful of empirical samples of each source distribution exist, will the sparse-sample methods avoid a severe underestimation of the combined variability in the resultant exact PDFs loosely depicted at right in the figure? More specifically in this study, do the methods avoid underestimation of the true 0.025 to 0.975 percentile range of the combined variability at the system level?

Broad experience finds that most system response quantities of project-level engineering interest are dominated by a few random variables even if many stochastic variables exist in the system. Therefore the study was designed to consider three dominant random variables having equal impact on the variability of system response (or nearly equal impact in the case of the bottom row). A system response that is linear in the ranges of the uncertain variables is adopted to avoid any confounding of sparse-sample methodology results with errors from any inexactly captured non-linearities in uncertainty propagation procedures that would otherwise have to be employed. The three random variable sources are also specified to be independent of each other. The problem then defaults to a linear convolution problem. The four different convolution problems shown in Figure 1 are studied, each with a different combination of three source uncertainties of normal, uniform, and/or right-triangular PDFs. “Exact” resultant PDF shapes are shown in

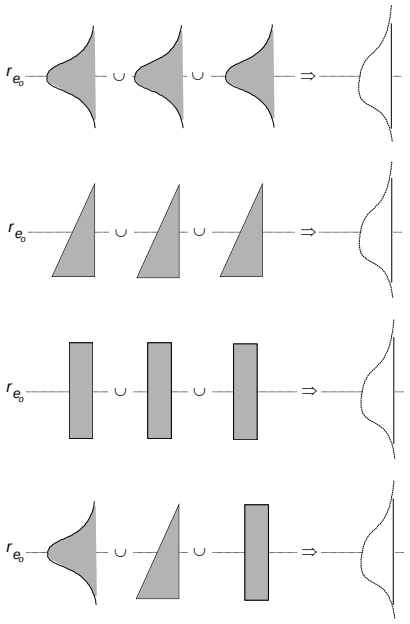


Figure 1. Convolution cases in test matrix for PDF representation under limited sample data. In total, 21 cases are investigated: normal, uniform, and right-triangle PDFs fitted alone, plus the four convolution combinations shown above in rows 1-4, where all seven cases are investigated at three sampling levels n=2, 8, 32.

Figure 2. These are calculated from 10 million Monte Carlo samples as follows. A random sample from each of the three source PDFs per row in Figure 1 is taken and the values of the three samples are added together. This sum constitutes one sample of the resultant PDF. Similarly 9,999,999 other samples of the resultant PDF are obtained. A Normal PDF approximation to each resultant PDF is also shown in Figure 2. The Normal approximations are such that their 0.025 and 0.975 percentiles pass through the exact resultant PDFs' 0.025 and 0.975 percentiles listed in Table 1. It is enlightening to see how close to Normal the resultant PDFs are from three source PDFs of normal, uniform, and/or right-triangular shapes.

We note that the locations of the 0.025 and 0.975 percentiles of the resultant PDF from convolving the three Normal PDFs in row 1 of Figure 1 can be determined analytically (exactly). The resultant is a Normal PDF with zero mean and a variance σ_{res}^2 that is the sum of the variances of the three Normal PDFs convolved. Thus the resultant standard deviation is $\sigma_{res} = [(1/1.96)^2 + (1/1.96)^2 + (1/1.96)^2]^{1/2} = 0.8837$. The 0.025 to 0.975 percentiles of the resultant Normal PDF lie at $0 \pm 1.96\sigma_{res} = (-1.732, 1.732)$. This agrees with the results in Table 1 from 10 million Monte Carlo samples, providing a check on our implementation of the Monte Carlo convolution procedure.

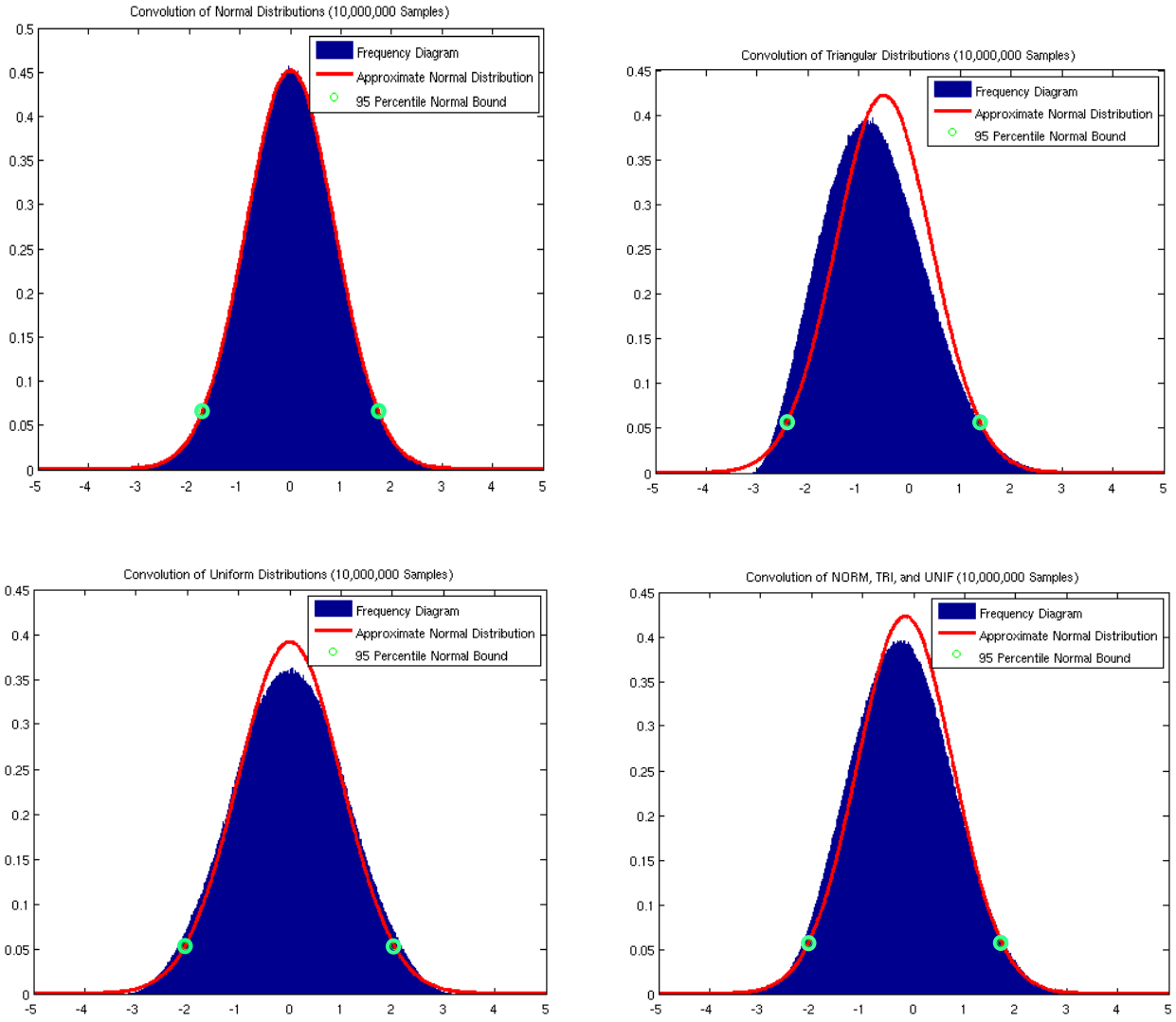


Figure 2. **"Exact" resultant PDFs from linear convolution (by Monte Carlo procedure) of three source uncertainties of normal, uniform, and/or right-triangular shapes shown in rows of Figure 1. Plot title above each resultant PDF identifies the row in Figure 1 it corresponds to. A Normal PDF approximation to each resultant PDF is shown which has 0.025 and 0.975 percentiles that pass through the 0.025 and 0.975 percentiles of the exact resultant PDFs (see Table 1).**

Table 1. Locations of 0.025 to 0.975 percentiles of Exact Resultant PDFs from Monte Carlo Convolution of PDFs in Rows of Figure 1.

PDFs convolved:	Location of 0.025 percentile	Location of 0.975 percentile
Row 1: three Normal PDFs	-1.732	1.732
Row 2: three Right-Triangle PDFs	-2.399	1.384
Row 3: three Uniform PDFs	-2.039	2.039
Row 4: one Normal, one Right Triangle, and one Uniform	-2.052	1.724

The following procedure was undertaken for the convolution problems in the rows of Figure 1 when very limited (sparse) data is available.

1. For a given row in Figure 1 randomly sample each of the three PDFs with a specific number of samples prescribed in the test plan (n = 2, 8, or 32).
2. shows an example for n = 8 samples and the problem depicted in the top row of Figure 1.
3. Use one of the sparse-data treatment methods to estimate from the three sets of n samples (obtained in step 1) three ranges that ostensibly bound the 0.025 to 0.975 percentile range of the exact PDFs from which the samples are drawn.
4. Use the three estimated PDF percentile ranges to estimate the aggregate 0.025 to 0.975 percentile range of the resultant PDF from convolving the three PDFs in the given row in Figure 1. The aggregation procedure for each method is explained in Section 3.
5. Using metrics and procedures from Section 4.1, assess and characterize the estimated aggregate percentile range against the true aggregate 0.025 to 0.975 percentile range of the exact resultant PDF.
6. Perform steps 1 – 4 1000 times (1000 "*trials*") to characterize the random variability of method performance under different sets of random samples of the three involved PDFs.
7. Upon completion of step 5, 3000 estimated bounds will exist for the 0.025 to 0.975 percentile range of the source triangular, uniform, and Normal PDFs. Assess the 3000 estimated bounds against the true range (-1,1) using the metrics and procedures presented in Section 4.1.
8. Perform steps 1 - 6 for each of the sparse-data treatment methods.
9. Perform steps 1 – 7 for n = 2, 8, and 32 samples per source PDF.
10. Perform steps 1 – 8 for each of the four rows in Figure 1.

Normal PDF
being sampled

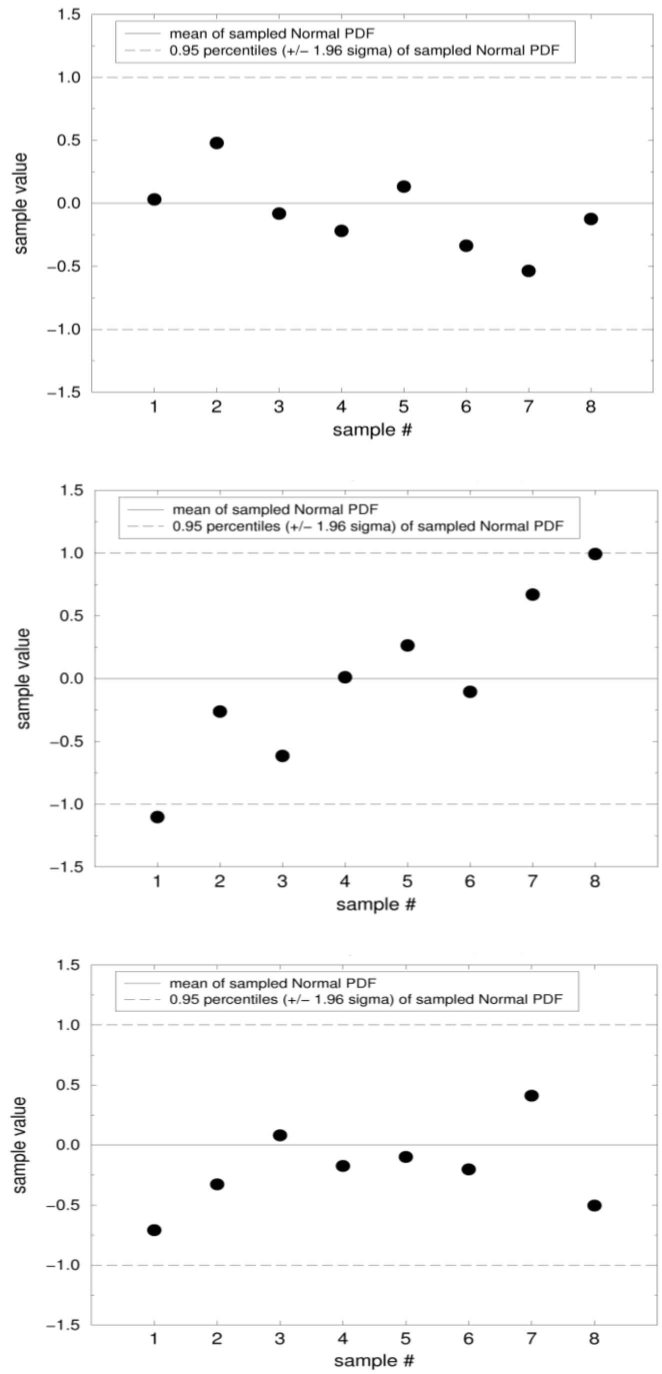


Figure 3. Three sets of 8 random samples from a Normal PDF.

3 UNCERTAINTY ESTIMATION METHOD DESCRIPTIONS

3.1 Fitting a Normal PDF to the Sample Data

A very common approach in engineering practice is to simply fit the sample data with a normal PDF. This practice will be analyzed in Section 4 for effectiveness with respect to the performance objectives previously discussed. The 0.025 and 0.975 percentiles of the fitted sample data triangular, uniform, and Normal PDFs are assessed against the true 0.025 and 0.975 percentile ranges (-1,1) of the source PDFs.

To provide an estimated range for the 0.025 to 0.975 percentiles of the resultant PDF from convolving the three source PDFs depicted in a row of Figure 1, the following procedure is used. For the three Normal PDF fits produced per "trial" (step 5 in Section 2), their means relative to the reference line r_{e_o} in Figure 1 are vectorially added to obtain the mean of the resultant PDF that comes from convolving the three approximating Normal PDFs. This makes use of a standard rule for convolution of PDFs. Likewise the variance of the resultant convolved PDF is obtained by summing the variances of the contributing Normal PDFs. Thus the mean and variance of the resultant PDF are easily obtained. It is also a fact that convolving Normal PDFs produces a resultant that is Normal PDF. Because it is Normal, the obtained mean and variance of the resultant PDF fully define its density function. The 0.025 to 0.975 percentile range of the resultant PDF constructed in each trial is then compared against the exact 0.025 to 0.975 percentile range of the true resultant PDF (from convolving the exact source PDFs in the applicable row of Figure 1).

3.2 Tolerance Interval Method

Another simple approach is to use statistical tolerance intervals ([12], [13]) to provide a reliably conservative interval representation for the combined epistemic and aleatory uncertainty associated with sparse data. The length of tolerance intervals accounts for both the epistemic and aleatory elements of uncertainty accompanying limited data samples. Tolerance intervals (TIs) are characterized by two user-prescribed settings: one for “coverage” of a prescribed subset of the variability, and one for statistical “confidence” in covering or bounding at least that subset of the variability. For instance a 0.95_coverage/0.90_confidence TI prescribes lower and upper values of a range of response that is said to have at least 90% odds that it covers or spans the 0.025 and 0.975 percentiles of the “true” PDF from which the random samples were drawn (for a large array of PDF types).

A 0.95/0.9 tolerance interval is constructed by multiplying the calculated standard deviation σ of the data samples by the appropriate factor f in Table 2 to create an interval of total length $2f\sigma$, where the interval is centered about the calculated mean of the data samples. The produced tolerance intervals are compared in Section 4 to the true 0.025 and 0.975 percentile range (-1,1) of the sampled PDF. Although constructed for Normal PDFs, 0.95/0.90 TIs also span, with greater than 90% odds, the 0.025 to 0.975 percentile ranges of many other PDF types when sparsely sampled. This is shown in [21] for uniform and right-triangular PDFs and for PDFs resulting from convolutions in rows 2 – 4 in Figure 1.

Table 2. 0.95_coverage/0.9_assurance Tolerance Interval Factors versus Number of Samples of PDF (from [13]).

	n = 2 samples	n = 8 samples	n = 32 samples
factor, f	18.8	3.264	2.395

Figure 4 and Table 3 present a fuller look at how tolerance interval size varies with the number of data samples. The 0.95/0.9 tolerance intervals are very large at n=2 samples but quickly decrease in size when more samples are available. A knee in the rate of uncertainty decrease (per added sample) occurs somewhere between 4 to 6 samples, with the rate of decrease being fairly small after n=8 samples. The tolerance interval has an asymptotic standard-deviation multiplier value of 1.96 for an infinite number of samples. This corresponds to the exact central 0.95 percentile range of the sampled normal PDF.

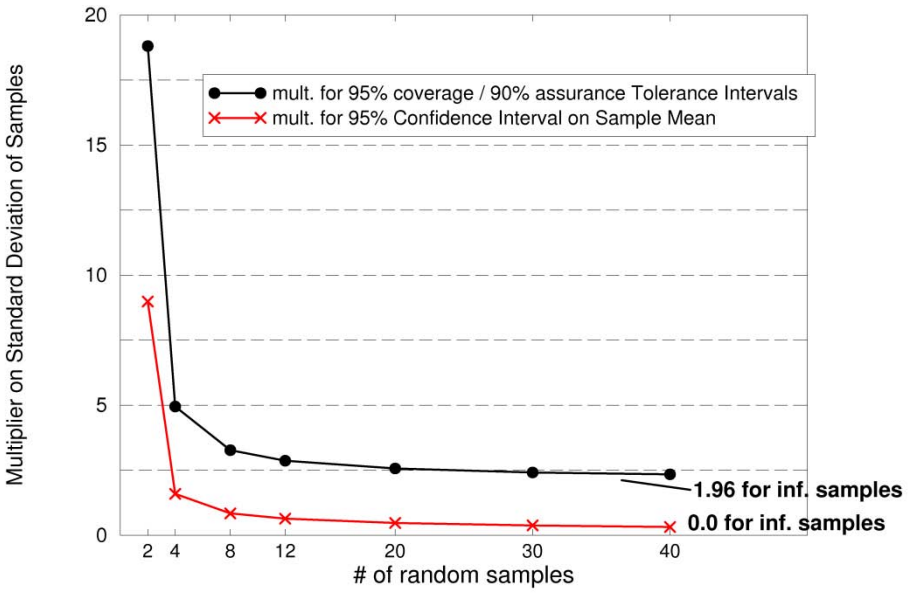


Figure 4. Multiplier on calculated standard deviation used to form uncertainty interval ranges for tolerance intervals. (Figure reproduced from [22], ignore confidence interval curve.)

Table 3. 0.95_coverage /0.90_assurance Tolerance Interval Factors (standard deviation multipliers) vs. # of Samples of PDF

# samples	$f_{0.95/0.9}$
2	18.80
3	6.92
4	4.94
5	4.15
6	3.72
8	3.26
12	2.86
20	2.56
30	2.41
40	2.33
∞	1.96

The following procedure is used to provide estimated bounds on the exact 0.025 to 0.975 percentiles of the exact resultant PDFs in Figure 1. For the three 0.95/0.9 tolerance intervals produced per trial (step 5 in Section 2), their means are added to obtain the value of a resultant mean. This gives the center of a net tolerance interval that is compared in Section 4 to the exact 0.025 to 0.975 percentile range of the true resultant PDFs. Using the classical linear-independent propagation rule of Kline & McClintock [5], the length of the net tolerance interval is the root sum of squares of the lengths of the three TIs calculated for that particular trial. The endpoints of this net TI identify the 0.025 and 0.975 percentiles of a Normal PDF that would result from convolving three Normal PDFs whose 0.025 and 0.975 percentiles are defined by the endpoints of the three 0.95/0.9 TIs in that trial.

3.3 PDF Estimation by the Pradlwarter-Schuëller Kernel Density Method

Kernel density estimation (KDE, seminal papers by Rosenblatt [23], and Parzen [24]) is a technique used to estimate the density of a random variable X given n independent samples X_1, ..., X_n of it. Let $K(\cdot)$ be a kernel function which is a non-negative, real-valued, symmetric function that integrates to one $\int K(x)dx = 1$. Then the KDE is

$$f^{\{KDE\}}(x) = \hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \quad (1)$$

Here, h is known as the *bandwidth parameter* which controls the influence of each sample in providing a density estimate at a near-by point. Small h corresponds to a small region of influence; a large h to a large one. Common kernel functions include the Gaussian, uniform, triangular, and Epanechnikov kernels. In this report, we use a Gaussian kernel.

Selection of the bandwidth parameter is the subject of a vast literature. Popular methods include cross validation (see [25], [26]) and asymptotic analysis (see [26], [27]). A common approach to measuring the error in the estimation process is the mean integrated squared error (MISE):

$$MISE = E\int (\hat{f}_h(x) - f(x))^2 dx \quad . \tag{2}$$

However, it is not feasible to minimize the MISE with respect to the bandwidth h unless the true density function $f(x)$ is known—which is not usually the case. Thus, in practice, people use cross-validation measures. In this approach, one maximizes a criterion such as the maximum-likelihood cross validation (MLCV) measure. At each observation X_i, the log-likelihood of the density at this point X_i is estimated based on all remaining observations except X_i. These log-likelihoods are then averaged over all observations:

$$MLCV = \frac{1}{n}\sum_{i=1}^{n} \log(\hat{f}_{-i}(X_i)) \, . \tag{3}$$

Bandwidth estimation needs to be automated in some fashion. However, the use of automated methods such as optimization methods which maximize a log-likelihood function (for example) may be subject to local optima depending on the starting bandwidth. Figure 5 shows two cases where a kernel density estimate was constructed over the same data set of four points, denoted with red X marks. The kernel density estimate shown with the green line is a KDE constructed with a small bandwidth (bandwidth value of 0.2), resulting in multimodal behavior of the output where the modes are around the actual data points. The density estimate shown with the blue line is constructed using a large bandwidth in the kernel (0.99), resulting in a smoothing of the density estimate.

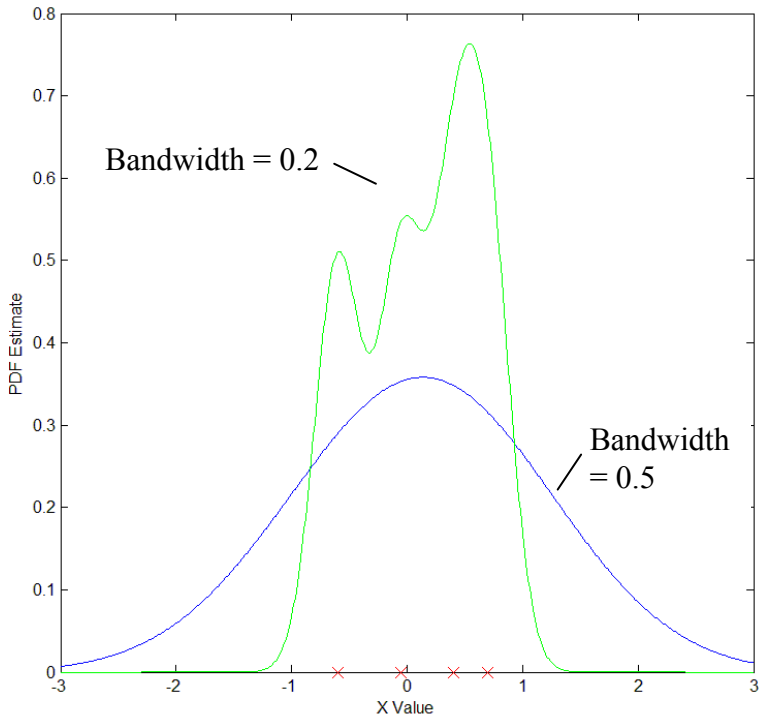


Figure 5. Kernel Density Estimate based on four points at the red X marks. The green line is a KDE constructed with a small bandwidth. The blue line is a KDE constructed with a large bandwidth.

In this work we did not want to use the cross-validation methods because of the small data sets (e.g. "leave one out" when you only have two data points leaves only one data point with which to construct the density estimate). We wanted to have an approach that was robust to small numbers of data points. Also, we wanted to ensure that density estimates based on the sparsest data sets were the most conservative. That is, we wanted to ensure that the tails of distributions derived from fewer data points would bracket the tails of the distributions derived from more data points. As Pradlwarter & Schuëller ([14]) state "In general, this condition is not fulfilled." To handle low numbers of data points robustly, we used a bandwidth estimation approach that they proposed and implemented in [14]. This approach was designed to minimize the probability that future measurements will lead to data points outside the domain of existing data points. Specifically, Pradlwarter & Schuëller define the bounds *a* and *b*:

$$\begin{aligned} a &= x_{\min} - \frac{1}{2n-2}(x_{\max} - x_{\min}) \\ b &= x_{\max} + \frac{1}{2n-2}(x_{\max} - x_{\min}) \end{aligned} \tag{4}$$

where x_{min} and x_{max} are the minimum and maximum values of the existing *n* data points. Given these bounds and the assumption that the existing data points are equivalent to *n* independent and identically distributed realizations of sample points from an unknown distribution $f(x)$, the probability that a point will fall outside the bounds *a* and *b* is:

$$p = 1 - \int_a^b f(x)dx \quad . \tag{5}$$

The probability that *n* independent data points will fall inside the domain is $\left(\int_a^b f(x)dx\right)^n$. To achieve a confidence level (1-α) that the probability of a new point falling outside the bounds will not exceed *p*, Pradlwarter and Schuëller suggest interpreting $\alpha = (1-p)^n$ as the confidence level or level of significance, and thus $P(\alpha,n)=1-\alpha^{1/n}$. Given this framework, the goal is to find a bandwidth *h* that will satisfy the following condition:

$$\int_{-\infty}^{a} f(x;h)dx + \int_{b}^{+\infty} f(x;h)dx = P(\alpha,n) \quad . \tag{6}$$

They use a Gaussian kernel, so that the KDE is specified as:

$$f^{\{KDE\}}(x) = \hat{f}(x;h) = \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^{n} \exp\left(\frac{x - X_i}{2h^2}\right) \quad . \tag{7}$$

We found this formulation to be easy to optimize. We specified α=0.1, and then calculated $P(\alpha,n)$ for a given number of data points *n*. Then, we found *h* which minimized the following expression:

$$abs\left[\left(\int_{-\infty}^{a} f(x;h)dx + \int_{b}^{+\infty} f(x;h)\right) - P(\alpha,n)\right] \tag{8}$$

This is an easy optimization calculation, and appears robust to a limited number of data points. The KDE densities constructed in this fashion have the property that the density estimates are likely narrower for increasing number of points: the density estimates based on a few data points likely bracket the density estimates based on more points. Also, the density estimates are wide and smooth when constructed with only a few points.

We constructed the KDE 0.025 to 0.975 percentile range by calculating the KDE values at input values between -5 and 5, at increments of 0.01. We took the KDE PDF and calculated the CDF (cumulative density function) from it. Then we constructed 10,000 random samples and interpolated where they would fall on the CDF curve (that is, we did an inverse mapping from the CDF back to the X values for the 10K samples). These 10K samples of a given KDE PDF formed the basis for locating its 0.025 to 0.975 percentiles. The percentile range was then compared against the exact 0.025 to 0.975 percentile range (-1,1) of the true PDF to yield the error characterization described in Section 4.

To provide an estimated bounding range for the exact 0.025 to 0.975 percentiles of the resultant PDF from convolving the three source PDFs depicted in a row of Figure 1, the 10K samples from each of the three KDE PDFs per "trial" (step 5 in Section 2) were added in the following manner. The 1st sample from each of the three KDE PDFs was taken and the values of the three samples were added together. This sum constituted one sample of the resultant PDF corresponding to a convolution of the three KDE PDFs. Similarly, 9999 other such samples of the resultant PDF were obtained. The obtained 10K samples portray the resultant PDF for a given trial, 1000 of which trials were performed in the study. For each trial the 0.025 to 0.975 percentile range of the aggregate KDF PDE was compared against the exact 0.025 to 0.975 percentile range of the true resultant PDF in Figure 2 to yield the error characterization described in Section 4.

3.4 PDF Estimation by the Non-Parametric Method

The following methodology was developed by Sankararaman and Mahadevan [15] for the use of non-parametric distributions to fit point data. The following summary of the technique is paraphrased from [15]. For a more complete explanation of this approach see that reference.

Discretize the domain of X into a finite number of points, say $\theta_i, i = 1{:}\,Q$. The domain is chosen based on the available data; the lowest value and the highest value are chosen as the lower bound and the upper bound of the domain, respectively. Assume that the PDF values at each of these Q points are denoted by $f_X(x = \theta_i) = p_i$ for $i = 1{:}\,Q$. Using an interpolation technique, the entire PDF $f_X(x)$ can be approximated for all $\theta \in X$, i.e. over the entire domain of X. Then the probability of observing the given point data is given by the likelihood, $L(\boldsymbol{p})$. This likelihood is a function of the following: (a) The discretization points selected, i.e. $\theta_i, i = 1{:}\,Q$. (b) The corresponding PDF values p_i; and (c) The type of interpolation technique used. For this work, the discretization is fixed, i.e. uniformly spaced p_i values $(i = 1{:}\,Q)$ over the domain of X are chosen in advance and the likelihood is maximized over the various values of p_i. The value of Q

(number of discretization points) is chosen based on computational power—the larger the Q, the finer (in terms of flexibility) is the resulting interpolation of the PDF. In this report, for the purpose of illustration, Q has been chosen to be equal to 11. Also, θ_1 is equal to the minimum of the available data (Xmin) and θ_Q is equal to the maximum of the available data (Xmax), and the intermediate $\theta's$ are uniformly interspersed between Xmin and Xmax.

The optimization problem is formulated as:

$$\begin{gathered} \text{Given } \theta_i \in X \forall i, i = 1{:}Q \\ Max\ L(\boldsymbol{p}) \end{gathered} \tag{9}$$

$where\ \boldsymbol{p} = \{p_1, p_2, p_3 \dots p_{Q-1}, p_Q\}\ \&\ f_X(x = \theta_i) = p_i$
subject to:

$$\begin{gathered} (1)\ p_i \geq 0 \forall i \\ (2)\ f_X(x) \geq 0 \forall x \\ (3) \int f_X(x)dx = 1 \end{gathered}$$

Note: $p_i\ at\ \theta_i\ (i = 1{:}Q)$ is used to interpolate and calculate $f_X(x)$.

This optimization problem maximizes the likelihood function subject to three constraints. The first constraint states that the vector **p** (that contains probability values) needs to be positive. The second and third constraints state that $f_X(x)$ must be positive and the area under this curve should be equal to unity, so as to satisfy the properties of a PDF. The PDF may be constructed using various interpolation methods. Several interpolation techniques, such as linear interpolation, spline-based interpolation and Gaussian process interpolation could be used. In this report a cubic spline is used for interpolation. The term ‘‘spline’’ refers to a wide variety of functions used for purposes of data interpolation and/or smoothing. Spline functions for interpolation are calculated as the minimizer of some suitable measure (e.g., the integral of the squared curvature) subject to some constraints (the interpolation constraints). For a more detailed explanation of splines, refer to Ahlberg et al. [28]. For the example presented here, eleven discretization points are chosen and the PDF is constructed. To find the .025 and .975 percentiles of the constructed PDFs and their convolved resultant PDFs, a similar procedure to the one described in the last paragraph of Section 3.3 is performed.

3.5 Johnson PDF Method

The Johnson family of distributions was first introduced in [18] and has been further explored by many researchers. The Johnson family is comprised of four types of distributions: lognormal, normal, bounded Johnson, and unbounded Johnson. In practice, the special cases of exact lognormal and normal distributions are rarely encountered when fitting sample data. Accordingly these two distributions were ignored in this study and each data set was fit to either a bounded or unbounded Johnson distribution. The bounded and unbounded Johnson distributions can assume many different PDF shapes, depending on the values of the four shape-defining parameters in the Johnson family of PDFs. It has been shown in [16] and elsewhere that sparse data often leads to a bounded distribution, and that result is also observed in the present study.

Four methods of estimating the Johnson parameter values are described in [29]: they are: (1) moment matching; (2) percentile matching; (3) least squares estimation; and (4) minimum L_p norm estimation. For this study the moment-matching approach was employed. The mean and the second, third, and fourth central moments of the data are calculated and equated to the respective quantities from a Johnson PDF parameterized by its four parameters. The system of four nonlinear simultaneous equations is solved to determine the four parameter values that cause the Johnson PDF's mean and moments to match those of the data set. In general the solution cannot be found analytically. A non-linear least squares (NLLS) optimization solver was used in this study. An NLLS solver is particularly well suited to this type of problem because it directly tracks the value of the entire residual vector at each iteration and calculates gradient information for each equation independently. Other solvers require a scalar objective function which would generally be some norm of the residual vector. Since it utilizes more of the residual information, the NLLS algorithm can be vastly more efficient than some of its counterparts. When it is well-behaved it converges rapidly to a solution. However, in some cases numerical noise arising from numerical integration to calculate moments of the Johnson PDFs can cause convergence issues and lead to very poor parameter estimates. In such cases it is useful to employ a gradient-free method to select a good starting point for the NLLS solver. A divided rectangles (DIRECT [30]) global optimization algorithm was utilized in this study for such cases. A steep increase in computational expense comes with such solvers, so DIRECT should not be used indiscriminately in all cases unless computational expense is of little or no concern. The solver convergence issues were rare enough that an efficient and robust algorithm was implemented which started with NLLS and switched to DIRECT on a case-by-case basis if/when convergence difficulty arose.

The study involved data sets of 2, 8, and 32 data points as previously mentioned. Numerical issues with the Johnson approach generally decrease as the number of data points increases. However, numerical issues do not completely vanish even with the relatively large sets of 32 samples. If the data set contains as few as 2 data points there is an additional complication stemming from uniqueness considerations. It can easily be shown that the third moment (skewness) of the data for any set of two samples is zero, and the fourth moment (kurtosis) of any such set is one. These are attributable to the fact that the two samples will be on opposite sides of the mean, and equidistant from it. Then a bounded Johnson distribution will always be selected from the family. The γ parameter, which governs skewness, was automatically set to zero for all cases of two data points, and the mean was forced to fall at the midpoint of the two distribution bounds (governed by ξ and $\xi+\lambda$ respectively). Under these assumptions it is possible to achieve a converged solution for the four distribution parameters. No such assumptions are necessary when 8 or 32 data points are present. To find the .025 and .975 percentiles of the constructed Johnson PDFs and their convolved resultant PDFs, a similar procedure to the one described in the last paragraph of Section 3.3is performed.

4 UNCERTAINTY ESTIMATION RESULTS, ERROR MEASURES, AND PERFORMANCE COMPARISONS

4.1 Measures of Uncertainty Estimation Error

Figure 6 defines discrepancy or error quantities to be calculated by comparing the exact 0.025 to 0.975 percentile range of the true PDF against the percentile range calculated from the estimation methods being tested. For the *i*th trial the calculated discrepancies at the upper and lower ends of the exact percentile range are given by:

$$\Delta_{\text{U-i}} = r_{\text{U-i}} - r_{\text{U-exact}}$$
$$\Delta_{\text{L-i}} = r_{\text{L-exact}} - r_{\text{L-i}} \; . \tag{10}$$

These equations return a positive value of discrepancy $\Delta_{L\text{-}i}$ when the estimated range extends beyond or "bounds" the true percentile at the lower end. The discrepancy $\Delta_{U\text{-}i}$ at the upper end is similarly positive when the estimated range extends beyond the true percentile at the upper end. In the opposite case where the true range extends beyond the estimated range, the discrepancy Δ is a negative value.

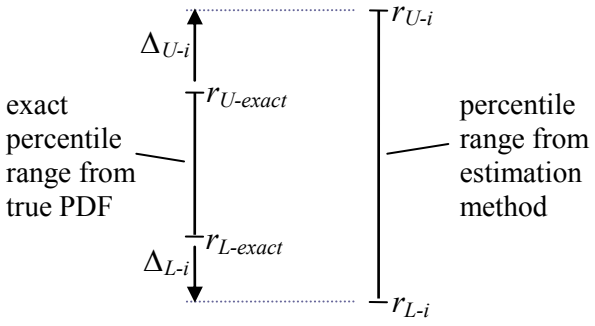


Figure 6. Definition of discrepancy or error between true percentile range and range yielded by estimation method.

A primary goal for the methods here is that the estimated percentile range be conservative so as to envelope or bound the actual percentile range of the sampled PDF. Therefore, positive discrepancy values Δ from over-prediction of the actual percentile range are looked upon as more desirable errors than are negative values of Δ. Hence, we are looking for methodologies that will produce positive values of $\Delta_{U\text{-}i}$ and $\Delta_{L\text{-}i}$ with high reliability (i.e., in a high percentage of the trials). Nonetheless, a second opposing objective is that the positive values of Δ be small; that the method minimally over-estimate the true percentile range of the sampled PDF. This presence of the two opposing objectives makes the sparse-data uncertainty representation problem interesting and difficult. A clear "winner" would be the estimation method that brackets the exact result the most times in the trials and simultaneously has the least overshoot error (smallest average and maximum positive Δ values). Unfortunately, it will be seen that no method enjoys the best of both worlds, but some score significantly better than others over the set of attributes or measures of method performance characterized in the following.

4.2 Example of Results for Normal PDFs

The Normal PDFs in Figures 1 and 3 have a mean μ = zero and a standard deviation σ = 1/1.96 such that the $\mu \pm 1.96\sigma$ extents of the Normal PDFs, which mark their 0.025 to 0.975 percentiles, have values of -1 and 1. Figures 7-9 plot the exact 0.025 to 0.975 percentiles at -1 and 1, along with the first 20 of each method's estimated bounds for the said percentiles. For a fair comparison, each method is applied to the same sets of random samples. For example, the first three sets of samples for the n=8 trials are shown in Figure 3. These three sets of samples are used by each method to yield the first three intervals (in order left to right) in each method's plot in Figure 8.

The first 20 results from each method in Figures 7-9 give a visual indication of the relative performance of the methods. The TI method yields the largest proportion of conservative intervals at each # of trials, n=2, 8, 32. This will be quantified later. The Pradlwarter-Schuëller (PS) method yields the next highest proportions of conservative intervals for n=2, 8, 32. The Normal Fitting (NF) method, the non-parametric (NP) method, and the Johnson (JN) method all yield significantly smaller numbers of conservative intervals, even at the relatively large number of 32 samples in Figure 9.

For n=2 samples the TI method yields the largest intervals on average, with many excessively large ones. The PS method yields more moderately sized conservative intervals, although significantly more non-conservative intervals than the TI method (a detailed quantitative analysis is presented in the next section). For n=2 the NF, NP, and JN methods all yield a high proportion of non-conservative intervals that significantly underestimate the true 0.025 and 0.975 percentiles. Figures 8 and 9 show that the differences in interval sizes diminish markedly as the number of samples increases first to n=8 and then to n=32 samples. For n=32 samples the conservative intervals are actually slightly larger on average for PS than for TI. For n=32 samples the NF, NP, and JN methods yield intervals that are not problematic in terms of significantly underestimating the true 0.025 and 0.975 percentiles.

Figures 10-13 show error histograms for TI, PS, NF, and NP estimates of the 0.025 to 0.975 percentile range of the Normal source PDFs in the top row of Figure 1 and of the convolution resultant PDF shown at top-left in Figure 2. (The histograms are from [20] where the JN method was not studied.) The charts in each figure have the same scale so that they can be compared side-by-side to see at a glance the appreciable qualitative differences between the methods' results. Some qualitative observations are made below. Quantitative information from the plotted histograms is extracted and processed in the next section to provide a quantitative basis for assessing the methods' performance. Each chart in Figures 10-13 has estimation errors separated into three histograms as defined next. The integrated area under each histogram is proportional to the fraction of trial results binned into it, as follows.

- The green histograms contain the +/+ or "pos/pos" trial sums S_i (= $\Delta_{U\text{-}i} + \Delta_{L\text{-}i}$) where both $\Delta_{U\text{-}i}$ and $\Delta_{L\text{-}i}$ are positive in a trial. These results represent a trial in which the estimated interval bounds the true interval.
- The red histograms contain the -/- or "neg/neg" trial sums S_i (= $\Delta_{U\text{-}i} + \Delta_{L\text{-}i}$) where both $\Delta_{U\text{-}i}$ and $\Delta_{L\text{-}i}$ are negative in a trial. These results represent a trial in which the

estimated interval falls short of the true interval range at both upper and lower ends of the range.

- The blue histograms contain the "mixed" trial sums S_i (= $\Delta_{U\text{-}i} + \Delta_{L\text{-}i}$) where one of $\Delta_{U\text{-}i}$ and $\Delta_{L\text{-}i}$ is positive and the other is negative. These results represent a trial in which the true interval is bounded by the estimated interval at only one end.

The following observations apply for the representations of the three source (Normal) PDFs and for the convolved resultant Normal PDF. The trends of method performance at n=2 samples (Figures 10 and 11) also generally apply at n=8 and n=32 samples (Figures 12 and 13 respectively) as will be quantified by more detailed processing in the next subsection. (The detailed processing in the next subsection will include Johnson method results even though JN histograms are not presented here.) The performance distinctions between the various methods decrease as more data samples are taken.

By observing the relative areas of the histograms it is apparent that the PS and TI methods contain much greater proportions of positive results than do the Normal-Fitting and Non-Parametric methods. The detailed results processing in the next subsection will show that the TI method also contains significantly more positive results than does the PS method. In a correlated result, the TI method contains significantly less negative and mixed results than all other methods.

The extents of the green +/+ histograms along the positive abscissa axis indicate the relative conservatism of the various methods' +/+ results. The greater the extent into positive territory, the more conservative on average the +/+ interval estimates are (this will be quantified in the next subsection). Likewise, the more negative the extents of the red -/- histograms, the worse the average shortfall of the unconservative intervals.

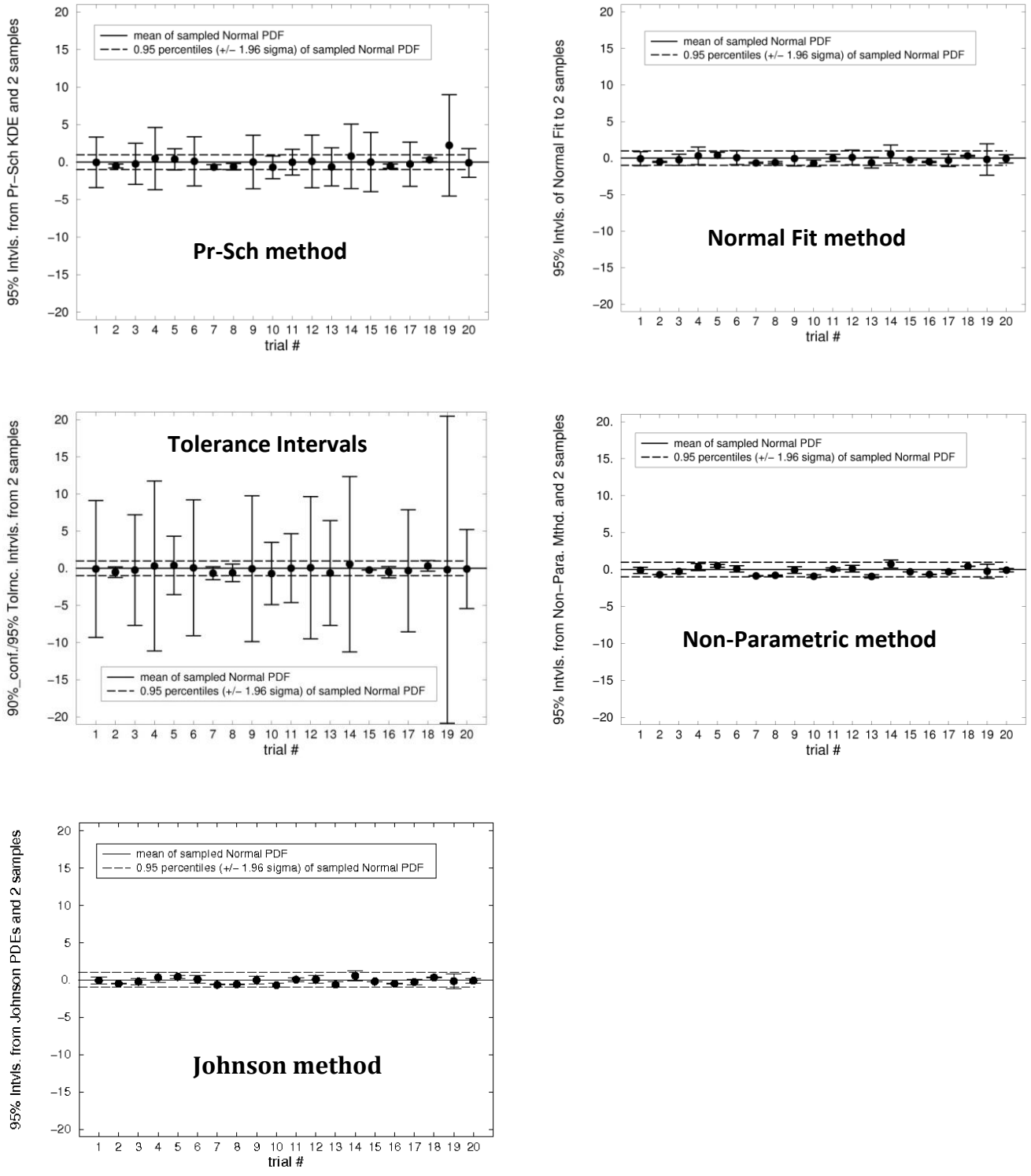


Figure 7. Each method's first 20 estimates of 0.025 to 0.975 percentile range of exact Normal PDF —results for *n*=2 samples of exact Normal PDF.

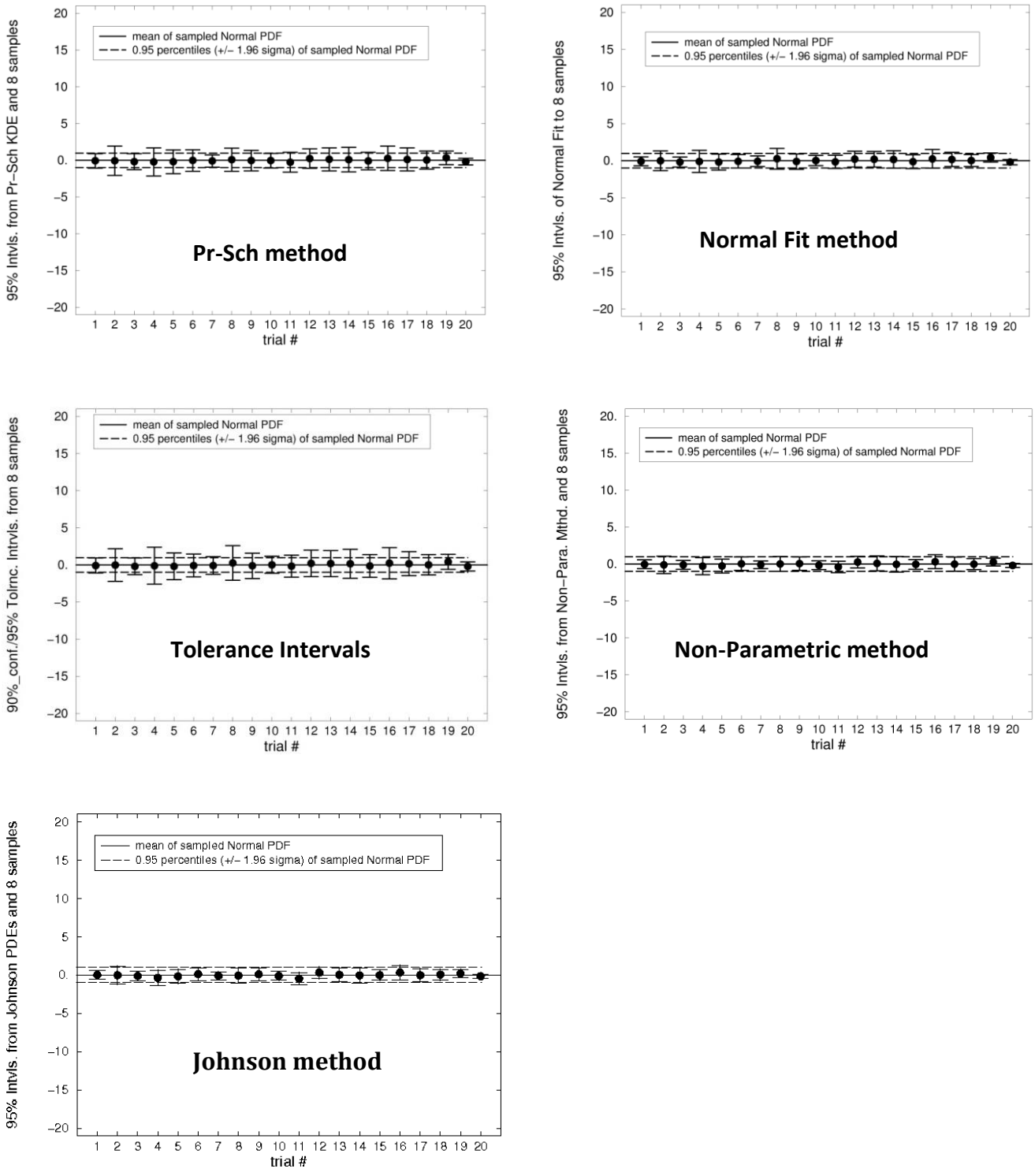


Figure 8. Each method's first 20 estimates of 0.025 to 0.975 percentile range of exact Normal PDF —results for *n*=8 samples of exact Normal PDF.

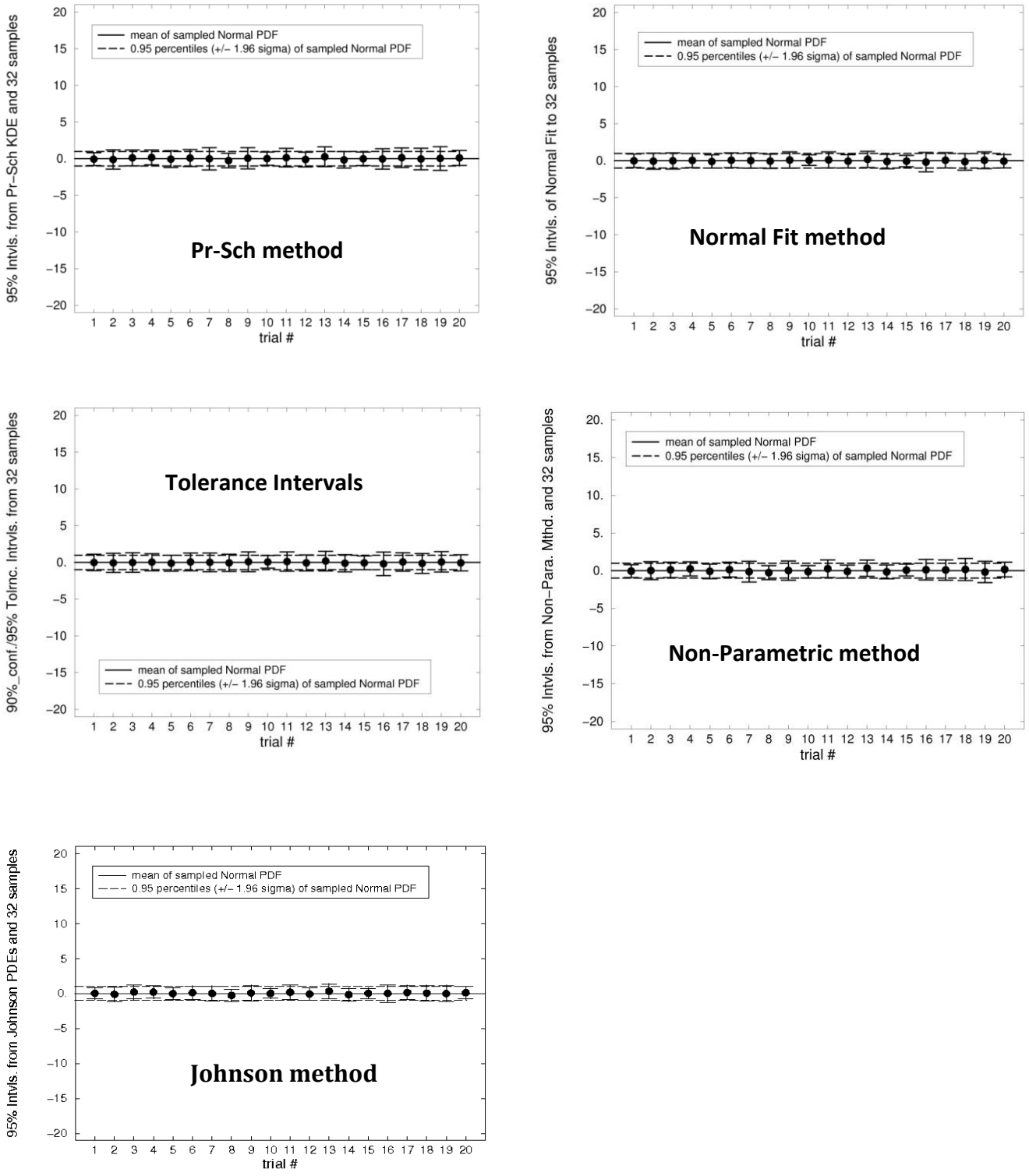


Figure 9. Each method's first 20 estimates of 0.025 to 0.975 percentile range of exact Normal PDF —results for *n*=32 samples of exact Normal PDF.

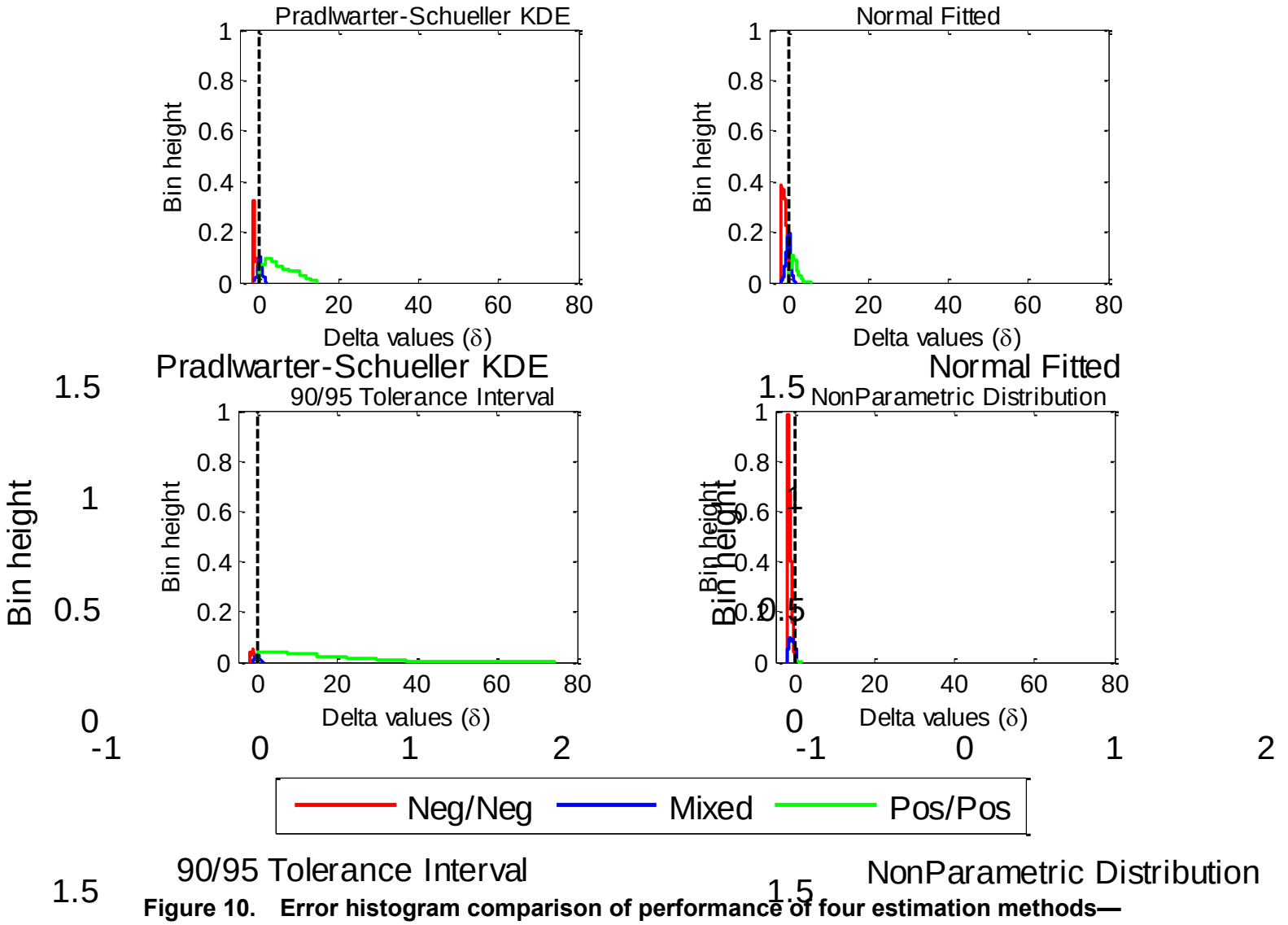


Figure 10. Error histogram comparison of performance of four estimation methods—results for 3000 trials of *n*=2 samples of a normal PDF.

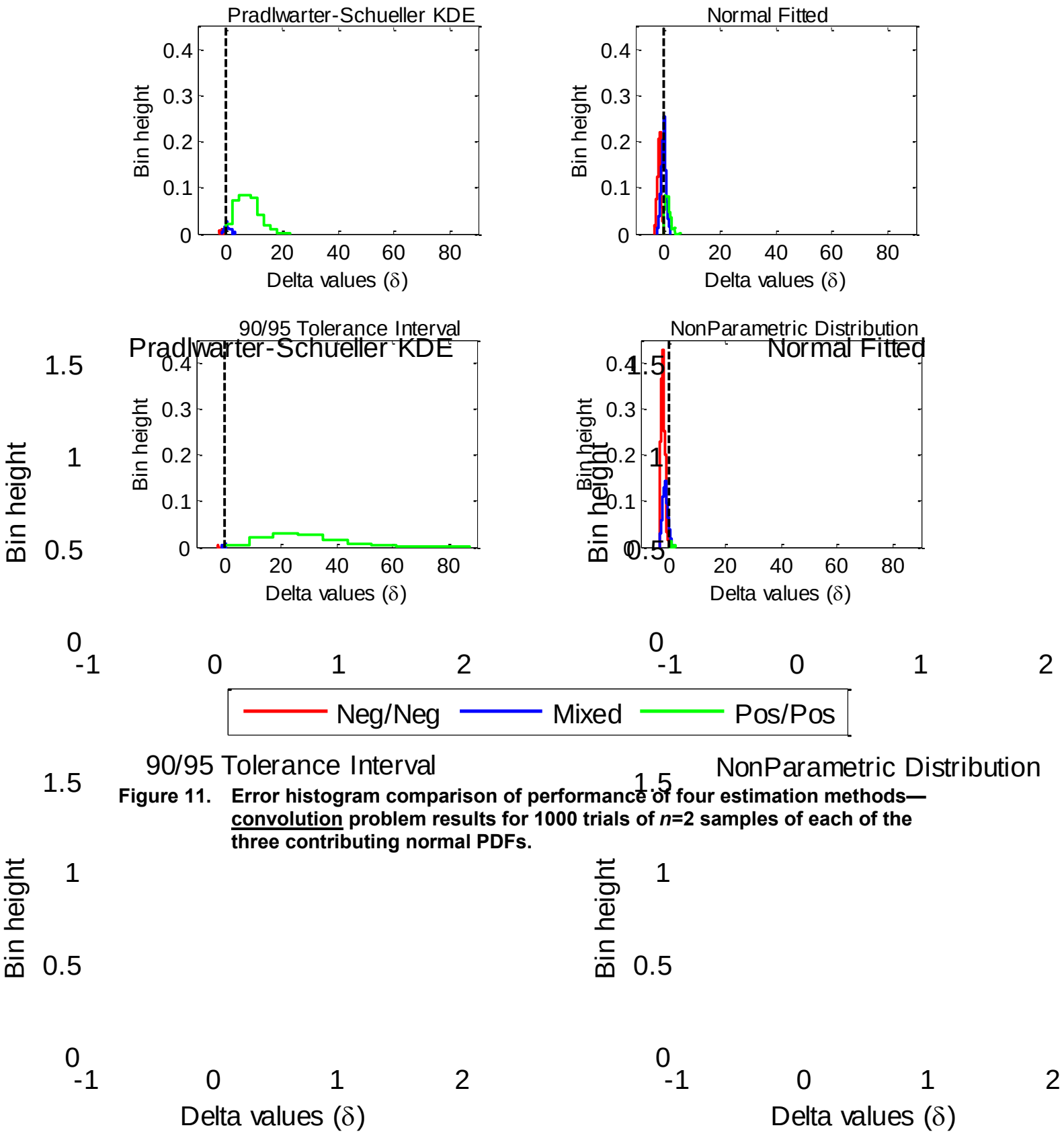


Figure 11. Error histogram comparison of performance of four estimation methods—convolution problem results for 1000 trials of *n*=2 samples of each of the three contributing normal PDFs.

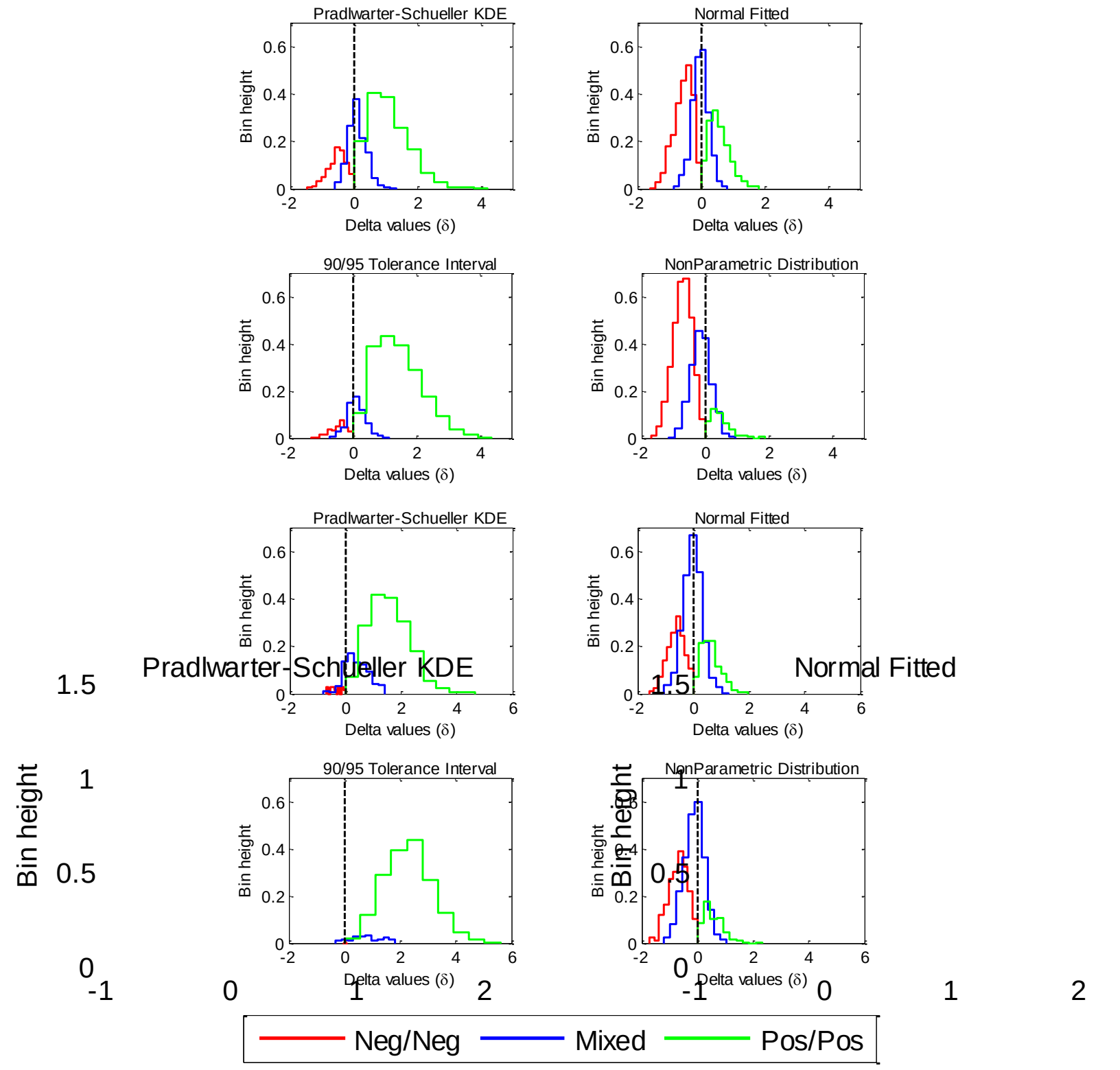


Figure 12. Error histogram comparison of performance of four estimation methods for *n*=8 samples. Top four plots are for 3000 trials of fitting a normal PDF, bottom four plots are for 1000 trials of the 3PDF convolution problem.

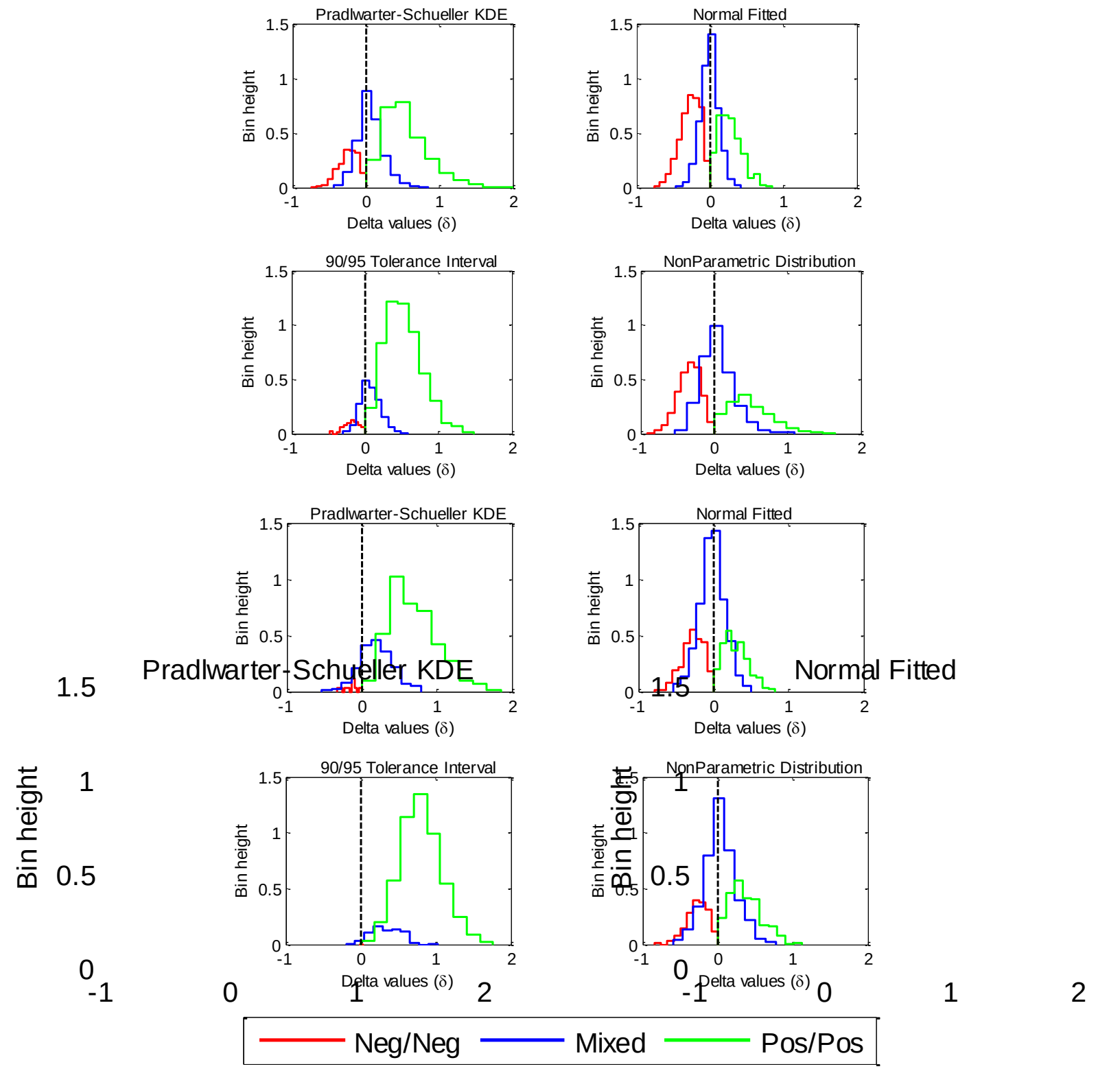


Figure 13. Error histogram comparison of performance of four estimation methods for *n*=32 samples. Top four plots are for 3000 trials of fitting a normal PDF, bottom four plots are for 1000 trials of the 3PDF convolution problem.

4.3 Detailed Processing and Comparison of Results

4.3.1 Example of Normal PDF Results for Row 1 in Figure 1, All Methods

Relative Frequencies of +/+, -/-, and mixed errors

Figures 14 and 15 plot the proportion of results in +/+, -/-, and mixed categories as a function of the number of data samples used in fitting Normal PDFs. (Numerical values of the plotted results for Figures 14 – 17 are listed in Appendix B.) The solid lines in Figures 14 and 15 show the net of taking the number of +/+ results (desirable occurrences) and subtracting the number of -/- results (undesirable occurrences) to get a net number of desirable results (as a percent of the total number of trials). The higher a solid line lies on the plots, the better the method's performance with respect to this measure. The Tolerance Interval method always ranks best by a considerable margin according to this measure, followed by the Pr-Sch method, then generally the Normal Fitting method, then the Non-Parametric method, and the Johnson method last. The NF, NP, and JN methods all have more undesirable -/- results than desirable +/+ results, giving them net negative scores for this performance measure, even at the relatively large number of 32 samples.

The dashed lines show the percentage of trials in which mixed results occurred. Mixed results are considered to be undesirable because one end of the estimated interval does not bracket the true percentile range. However, mixed results are not as undesirable as -/- results. Because mixed results are undesirable, the lower the dashed line lies on the plots (i.e., the closer to zero), the better the method's performance with respect to this performance attribute. The TI method always performs best by a substantial margin according to this measure, followed by the PS method, then relatively far behind are the NF, NP, and JN methods.

If two or more methods' solid-line results occupy the same ordinate value at a given abscissa value of say n=2 samples, each method's dashed line can still have a different ordinate value there. For example, let one method yield (60% +/+, 30% -/-, 10% mixed) results and another method yield (55% +/+, 25% -/-, 20% mixed) results. Then both methods have the same solid-line plotted value of 30%, reflecting an equal score of "goodness" with regard to the number of very desirable +/+ occurrences minus the number of very undesirable -/- occurrences. This tie score can be broken by considering the number of undesirable results comprising the mixed population. The method with 10% mixed results would be considered better than the method with 20% mixed results, hence the method that yielded (60% +/+, 30% -/-, 10% mixed) results is considered the overall better performer.

Accordingly, the virtual tie between the solid-line results of the Non-Parametric and Normal Fit methods at n=32 in Figure 14 is broken by considering their dashed-curve values at that abscissa value. The Normal Fit method has about 10% less mixed results than does the Non-Parametric method, making the Normal Fit method's performance better overall at n=32.

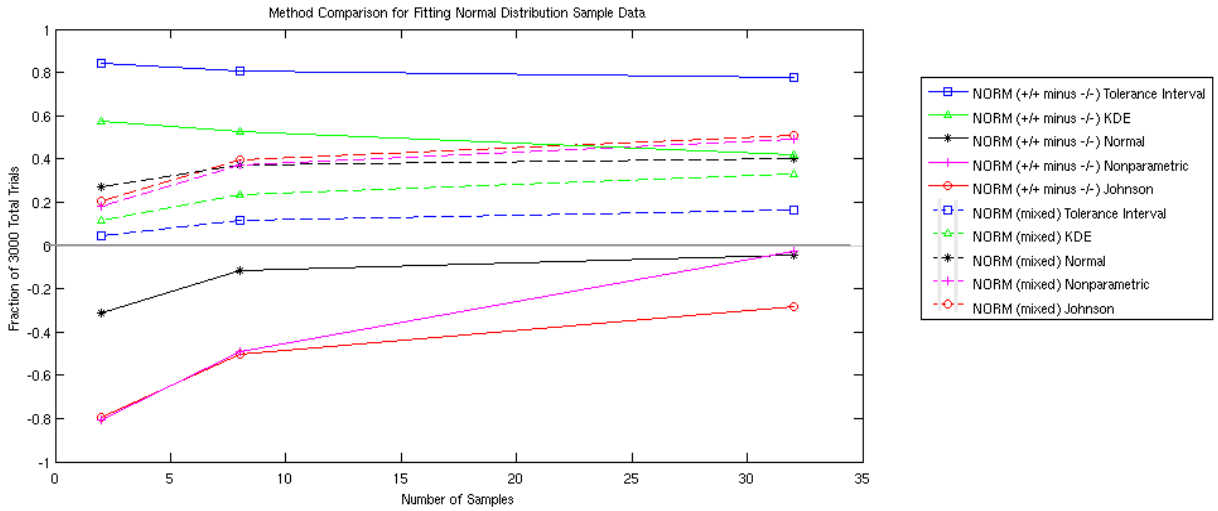


Figure 14. Comparison of proportions of results in +/+, -/-, and mixed categories (for the five estimation methods) as a function of number of data samples from a Normal PDF.

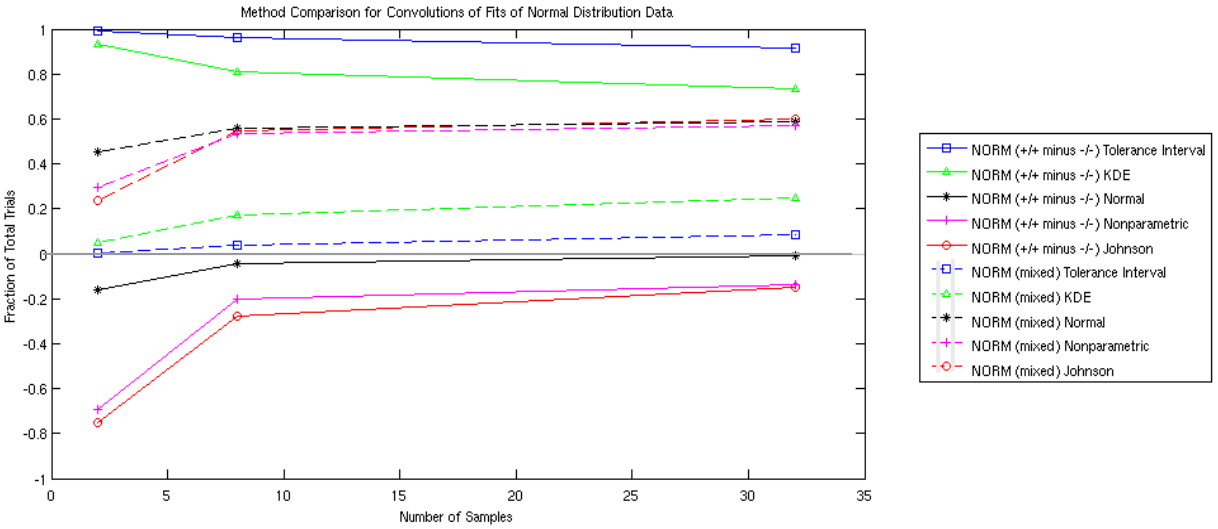


Figure 15. Comparison of proportions of results in +/+, -/-, and mixed categories (for the five estimation methods) as a function of number of data samples— performance in convolution of fits to three Normal PDFs.

If one gives the same undesirability weight to the mixed results as is given to the -/- results, then the graph and the ranking of method performance could be made substantially simpler by just plotting the net value [(% of +/+ results) minus (% of -/- results) minus (% of mixed results)]. The larger the value, the better the method has performed. But in Figures 14 and 15 we elect to

treat a mixed result as less bad or undesirable than a -/- result, so the two types of results are split out in the figures. Ultimately though, for quantitative scoring in Figure 18 we use a very simple equally-weighted performance ranking scheme described at the end of this subsection.

Method performance differences in % of +/+, -/-, and mixed errors when convolving 3 fitted Normal PDFs vs. fitting a single Normal PDF

By comparing Figures 14 vs. 15 a tendency can be seen that the aggregation (convolution) of multiple sources of variability somewhat washes out or mitigates the relative number of undesirable -/- and mixed errors of the individual contributing sources. However, this tendency is not universal, especially for the Normal-fitting method. A close examination of the numerical results (presented in extensive tables in Appendix B) that underlie the plots in Figures 14 and 15 reveals the following.

- % of -/- undesirable results declines (gets better) for all methods and all # of samples (2,8,32) when convolving 3 fitted Normal PDFs vs. fitting a single Normal PDF
- % of mixed +/- undesirable results declines (gets better) for TI and PS methods for all # of samples (2,8,32) when convolving 3 fitted Normal PDFs vs. fitting a single Normal PDF, but increases (gets worse) for all other methods (NF, JN, NP) when convolving 3 fitted Normal PDFs vs. fitting a single Normal PDF
- % of +/+ desirable results increases (gets better) for all methods and # of samples when convolving 3 fitted Normal PDFs vs. fitting a single Normal PDF, except for NF method @ 2,8,32 samples and PS @ 32 samples.

Method performance differences in % of +/+, -/-, and mixed errors when # of samples changes from n=2, 8, 32 (for Normal PDFs)

Considering method performance trends as the # of samples increases from n=2 to 8 to 32, Figures 14 and 15 show the same trends whether fitting a single Normal PDF or convolving 3 fitted Normal PDFs:

- % of -/- undesirable results significantly declines (gets better) as # of samples increases for all methods except PS (PS has relatively flat behavior in % -/- as # of samples changes)
- % of +/- undesirable results increases significantly as samples are added for all methods (but with anything near 32 samples or more, the next section shows that the magnitudes of the mixed errors are small)
- % of +/+ desirable results improves significantly for the three underperforming methods NF, NP, and JN when # of samples increases, while the +/+ percentages of the better performing TI and PS methods degrades significantly as samples are added. But even at n=32 their percentages of desirable +/+ results are still more than double those of the NF NP, and JN methods.

Magnitudes of +/+, -/-, and mixed Overshoot and Shortfall Errors

Now the desirability of the methods is considered with respect to magnitudes of the overshoot and shortfall errors. For this performance category, mixed errors are treated by averaging the absolute values of $\Delta_{U\text{-}i}$ and $\Delta_{L\text{-}i}$.

Figures 16 and 17 plot the methods' average predicted 0.025 to 0.975 percentile ranges for results that lie in the +/+, -/-, and mixed categories. The corresponding mean overshoot and shortfall errors are evident. First consider the results for n=2 samples. The magnitudes of the mean -/- and mixed errors are not dramatically different among all the methods. But for the +/+ overshoot errors, the Tolerance Interval method has large average overshoot in its large +/+ population of results. The mean overshoot error is much larger than with the Pr-Sch method, which also has a relatively large +/+ population of desirable results. The +/+ overshoot errors of the NF, NP, and JN methods are considerably less than for the other two methods. Unfortunately the proportions of desirable +/+ results are relatively small for the NF, NP, and JN methods.

The relative trends observed for n=2 samples are generally maintained at n=8 and n=32, within a context of decreasing error magnitude distinctions between the various methods as the number of data samples increases. At n=32 the distinctions between method performance are small; the mean -/- undershoot and mixed undershoot-overshoot errors are relatively minor for all methods, and the +/+ overshoot errors are relatively insignificant for the NF, NP, and JN methods and are only slightly bigger (but still small) for the TI and PS methods.

Any mitigation or washing-out of PDF representation error magnitudes in the aggregation (convolution) of multiple PDFs can be assessed by comparing Figures 16 vs. 17. *Normalized* error magnitudes (as a % of the true percentile ranges) for the convolved results are typically slightly smaller than the individual PDF representation errors. This is usually the case for all methods, numbers of samples, and categories of +/+, -/-, and mixed errors. However, some exceptions arise (see lettered observations below).

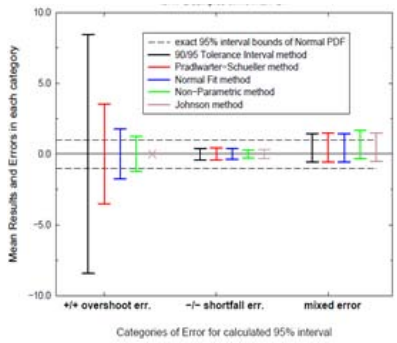
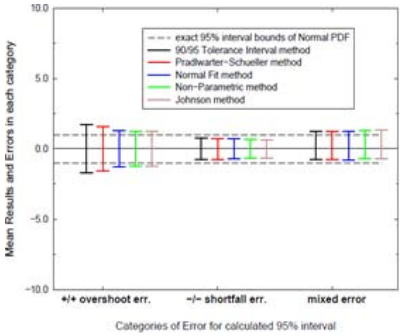
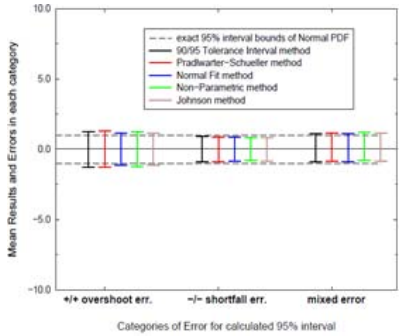
n = 2 samples	8 samples	32 samples
		

Figure 16. Mean predicted percentile ranges and mean +/+ overshoot, -/- shortfall, and mixed errors for the various methods—performance in representing a normal PDF. (On the left plot for *n*=2 samples there were no +/+ results in the 3000 trials of the Johnson method.)

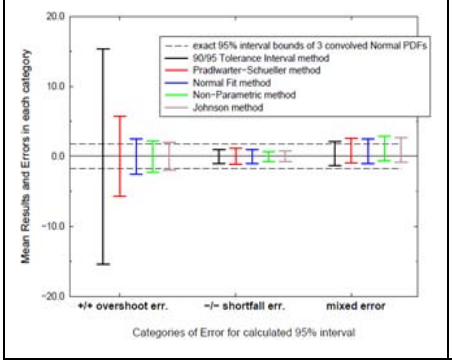
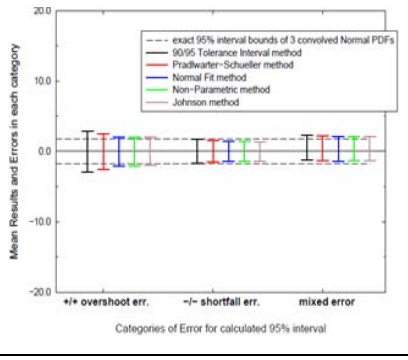
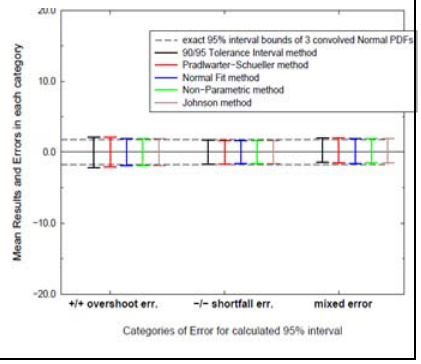
n = 2 samples	8 samples	32 samples
		

Figure 17. Mean predicted percentile ranges and mean +/+ overshoot, -/- shortfall, and mixed errors for the various methods—performance in convolution of fits to three Normal PDFs.

From an examination of the numerical results in tables in Appendix B the following more precise statements can be made.

Method performance differences in error magnitudes when convolving 3 fitted Normal PDFs vs. fitting a single Normal PDF

- Normalized magnitudes of -/- errors decrease significantly for all methods and # of samples 2,8,32 when convolving 3 fitted Normal PDFs vs. fitting a single Normal PDF.
- Normalized magnitudes of +/- mixed errors decrease when convolving 3 fitted Normal PDFs vs. fitting a single Normal PDF when NF and NP methods are used @ all sample sizes, but increases for TI, PS, and JN at two of the three sample sizes 2,8,32.
- Normalized magnitudes of +/+ errors decrease when convolving 3 fitted Normal PDFs vs. fitting a single Normal PDF for all methods and # of samples, except for TI and NP methods at 2 samples.

Method performance differences in error magnitudes when # of samples changes from n=2, 8, 32 (for Normal PDFs)

Figures 16 and 17 reveal the following similar trends whether fitting a single Normal PDF or convolving 3 fitted Normal PDFs.

- Normalized magnitudes of -/- errors significantly decrease as the # of samples increases, for all methods.
- Normalized magnitudes of +/- mixed errors quickly decrease as the # of samples increases, for all methods.
- Normalized magnitudes of +/+ errors quickly decrease as the # of samples increases for all methods except NP. For NP the overshoot errors are relatively small and unchanged whether n=2, 8, or 32.

Method Performance Scoring Approach

Figure 18 summarizes the relative performance of the methods for n=2-sample fits of a Normal PDF according to the performance aspects discussed above and listed along the bottom of the figure. A performance rank of 1 in the figure is best and a rank of 5 is worst. The methods' performance ranks are also presented in parentheses in the 1st column of numerical values in Table B.1 of Appendix B. Similar performance ranks are given in the other numerical columns of Tables B.1 – B.3 for the 20 other test problems.

A simple overall score for each method can be obtained by adding the method's rank values across the six attributes, and then dividing by six. For example, in the first column of values in Table B.1 the ranks for TI in the six performance aspects average to (1+5+1+2+1+1)/6 = 1.83. The average performance scores for all methods are, from Figure 18 or equivalently from the ranks in the 1st column of values in Table B.1: 1.83 for the TI method, 2.33 for PS, 3.17 for NF, 3.67 for JN, and 4.0 for NP (listed in order of best (lowest) to worst (highest) average scores).

The simple performance ranking scheme gives all performance aspects equal weight, which over-simplifies things. Beyond the simple "zeroth-order" consideration here of whether one method is better than another in a performance attribute, a more refined judgment of performance would consider *how much better* one method does than another, say in terms of the number of realizations in the desirable +/+ category versus the very undesirable -/- category, and in terms of the magnitudes of overshoot and shortfall errors. Moreover, the various attributes can be differentially weighted according to importance to various types of analysis results and purposes. Finally, complexity of method implementation should also be factored in. Then a variety of decision theory methods could be used to help make a more rigorous decision according to user preferences. This might result in a different ranking of methods than arrived at here. Nonetheless, the simple zeroth-order performance ranking scheme and resulting numerical scores provided in the next subsection along with the observations there and those previously given, provide initial insight into the relative performance of the methods.

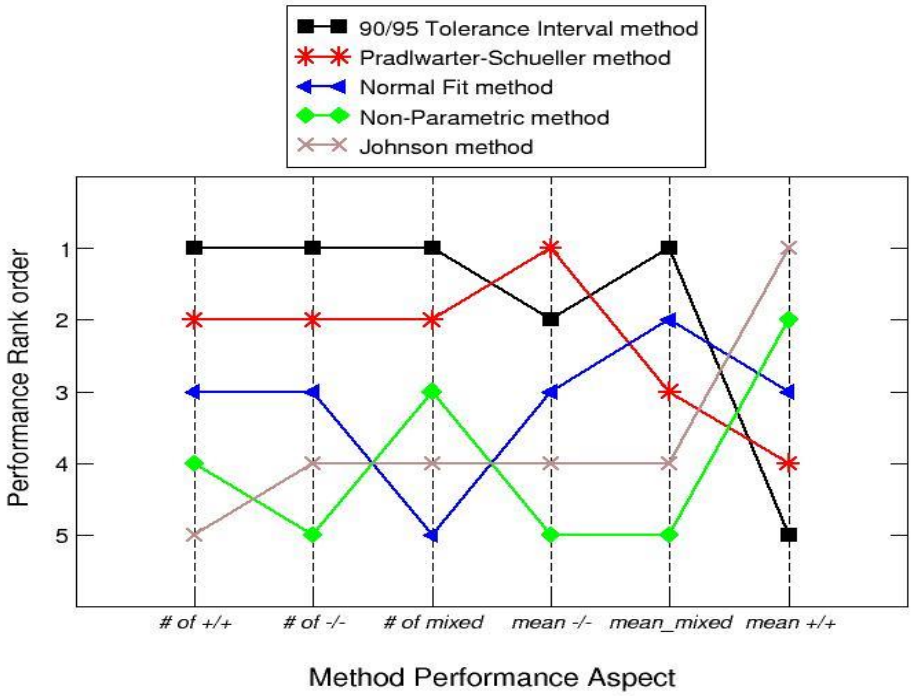


Figure 18. Performance rankings of the listed methods according to the performance aspects cited at bottom of plot (for performance on fitting a Normal PDF with *n*=2 samples).

4.3.2 Overall Performance Summary for all PDF shapes and Convolution Cases

We now additionally consider the fits to uniform and right-triangle PDFs and convolution cases rows 2 – 4 in Figure 1. Table 4 summarizes method rankings for the various test cases. For example, average ranking scores cited in the last sentence in the penultimate paragraph of Section 4.3.1 for the test case corresponding to Figure 18 are listed in the 1st column of numerical values @ *n*=2 samples in Table 4.

Plots similar to Figure 18 can be envisioned for each of the summary results sets in Table 4. Appendix B lists the full data set for creating such plots but Table 4 only lists each method's average ranking score. Table 4 also indicates rankings (in parentheses) of each methods' average performance rank over the 6 performance aspects, for each of the 21 cases studied.

Table 4. Numerical Performance Scoring and Ranking Summary for Sparse-Data Fitting Methods

Method and # of samples	Fitted PDF Shape			3 Fitted PDFs in Convolution				row-avg. of (ranks)
	Normal	Right Triangle	Uniform	3 Normal PDFs (row1 Fig.1)	3 Rt. Tri. PDFs (row2 Fig.1)	3 Uniform PDFs (row3 Fig.1)	Norm., Tri, Unif. PDFs (row4 Fig.1)	
***n*=2 samples**								
TI	1.83 (1)	2.0 (2)	2.0 (2)	1.83 (1)	1.67 (1)	1.67 (1)	1.67 (1)	(1.29)
PS	2.33 (2)	1.83 (1)	1.83 (1)	2.33 (2)	2.0 (2)	2.0 (2)	2.17 (2.5)	(1.79)
NF	3.17 (3)	2.17 (3)	2.17 (3)	3.17 (3)	2.33 (3)	2.33 (3)	2.17 (2.5)	(2.93)
JN	3.67 (4)			3.58 (4)				
NP	4.0 (5)			4.08 (5)				
***n*=8 samples**								
TI	1.83 (1)	1.67 (1)	2.0 (2)	2.33 (1)	1.67 (1)	1.67 (1)	1.67 (1)	(1.14)
PS	2.5 (2)	2.0 (2)	1.83 (1)	2.67 (2)	2.0 (2)	2.0 (2)	2.0 (2.5)	(1.86)
NF	2.83 (3)	2.33 (3)	2.17 (3)	3.0 (3)	2.33 (3)	2.33 (3)	2.33 (2.5)	(3.0)
JN	4.33 (4)			3.67 (5)				
NP	3.5 (5)			3.33 (4)				
***n*=32 samples**								
TI	1.67 (1)	1.67 (1)	1.67 (1)	2.33 (1)	1.67 (1.5)	1.67 (1.5)	1.33 (1)	(1.07)
PS	2. 67 (2.5)	1.83 (2)	2.33 (3)	2.67 (2)	1.67 (1.5)	1.67 (1.5)	2.17 (2)	(2.14)
NF	2. 67 (2.5)	2.5 (3)	2.0 (2)	2.83 (3)	2.67 (3)	2.67 (3)	2.5 (3)	(2.79)
JN	4.17 (4)			4.0 (4)				
NP	3.83 (5)			3.17 (5)				

The six cases discussed in Section 4.3.1 are summarized in the 1st and 4th columns of numerical values in Table 4. The method performance rankings in these columns are consistent with the

results and observations in Section 4.3.1. These columns show that whether n= 2,8,32 samples and whether fitting a Normal PDF or convolving three fitted Normal PDFs, TI always ranks 1st in the six cases, PS ranks 2nd and NF ranks 3rd except for a tie between the two in the 1st column @ n=32 in Table 4, and JN and NP always rank 4th or 5th.

JN also performs poorly in comparison to the other methods when fitting uniform and right-triangle PDFs and applied to convolution cases rows 2 – 4 in Figure 1. On these types of PDFs and convolution cases NP is conjectured to likewise perform relatively poorly, based on the following observations.

Appendix A and ref. [21] show that JN performs poorly compared to PS on Normal, uniform, and right-triangle PDFs and all convolution cases (rows 1 – 4) in Figure 1. PS yields significantly more desirable results and less undesirable results than JN for all PDF shapes and numbers of samples. PS similarly dominates JN in the convolution cases (usually).[2]

We conjecture that NP would likewise closely echo JN's relative under-performance in uniform and right-triangle PDFs and convolution rows 2 – 4. This is based on the similar performances (in Section 4.3.1) of NP and JN methods on Normal PDFs and on row 1 convolutions.

Therefore we concentrate on TI, PS, and NF methods in what follows, and in the method performance trend analysis in Appendix B. Given the findings in Section 4.3.1 that NF performance substantially lagged that of TI and PS, we might have also dropped NF from further attention, but we decided to retain it because it is easy to use, is fairly commonly used in engineering practice, and it performed considerably better than NP and JN in Section 4.3.1.

In Table 4 the performance ranking data for uniform and right-triangle PDFs and convolutions in rows 2 – 4 in Figure 1 are only reported for TI, PS, and NF methods. For these cases TI, PS, and NF usually rank 1st , 2nd, and 3rd, respectively, but in four of the 15 cases PS ties or ranks better than TI and in three other cases NF ties or ranks better than PS. In Table 4's last column the overall ranks averaged over the columns are presented for the top three methods TI, PS, and NF. In this last column, TI, PS, and NF rank overall 1st, 2nd, and 3rd at all sample sizes, n=2,8,32.

We now focus on the fundamental objective of attaining a high proportion of the desirable (conservative) +/+ results. Table 5 presents for TI, PS, NF methods the proportions of +/+ results averaged in various ways from the +/+ proportions listed in tables B.1-B.3 in Appendix B. The results show that the 0.9_confidence/0.95_coverage TIs hover on average around their advertised 90% reliability/confidence in bracketing the actual 0.025 to 0.975 percentile range. PS averages about 20% less reliable than TI whether single PDFs or convolutions of them are involved. With a relatively low overall average reliability of 28%, NF averages less than 1/2 the reliability of TI when the single PDFs are involved and less than ¼ the reliability of TI for the convolution cases.

[2] Incidentally, Appendix A and ref. [21] also show that the uniform and right-triangle PDFs give greater fractions of desirable results (#'s of +/+ minus # of -/- results) than when fitting Normal PDFs. This performance trend does not extend to the undesirable 'mixed' results, which show varying method performance with respect to PDF shape. Nonetheless, PS and JN perform best overall for uniform PDF shapes, then for right triangle PDFs, then for Normal PDFs. However this does not appear to be generally the case for the other fitting methods (see Appendix B), so we make no general conclusions about method performance vs. PDF shape.

Table 5. Averages of % +/+ results = Method Reliability in Bracketing True 0.025 to 0.975 percentiles

Method	Avg.* % +/+ over PDF cases Normal, Right Tri., Unifrm.	Avg.** % +/+ over convolution cases rows 1 – 4 in Figure 1	Weighted avg. over all 21 cases
TI	87.1%	96.0%	92%
PS	67.2%	72.1%	70%
NF	36.3%	21.7%	28%

* averaged over fits of single Normal, right-tri., and unif. PDF shapes for n= 2,8,32 samples (9 cases)
** averaged over the four convolution cases rows 1-4 in Figure 1 for n= 2,8,32 samples (12 cases)
21 cases all together in * and **

Table 5 also reveals that for TI and PS the +/+ proportions for fitted PDFs used in convolution are somewhat better on average (more reliably conservative) over n= 2,8,32 and over the various PDF types in convolution than when singular PDFs are fitted. This is good for TI and PS because it is expected that the convolution case of several sources of fitted input PDFs is more common in practice than having just one important input PDF to be fitted. But this is bad for the NF method, which Table 5 shows performs relatively poorly on average for single PDFs and even worse on average for multiple convolved PDFs. This dynamic, combined with its low overall average of just 28% reliability in capturing the true 0.025 to 0.975 percentile ranges in the 21 test problems, leads us to dismiss NF as an acceptable method for general engineering use if one has the objectives outlined at the start of this report.

Indeed, even if the underlying random process being sampled is Normal, for n=2 samples the Normal Fitting method produced 52% very undesirable -/- errors, with a mean magnitude of 63% shortfall from the true 95-percentile range of the Normal PDF. For n=8 samples, these numbers fall to 37.2% -/- errors with a mean shortfall of 28%—still very substantial even for the relatively large number (in engineering practice) of 8 data samples. This significant representation error is only slightly mitigated in uncertainty aggregation (convolution) with multiple similarly sized and represented sources of Normal uncertainty (compare Figures 16 and 17).

We next examine the +/+ overshoot error magnitudes of TI and PS methods for the test 21 cases. We do not further consider the NF method because of its low reliability relative to TI and PS methods. We do not further analyze the magnitudes of -/- or mixed -/+ errors. For these we presume that for TI and PS and all 21 test cases the magnitude trends vs. # of samples (2,8,32), and the magnitude trends between fitting single PDFs vs. use of three such fitted PDFs in convolution, are similar to those in Figures 16 and 17 involving Normal PDFs and convolutions of Normals. In any case, because TI and PS have relatively high reliabilities or odds of getting +/+ errors, the issue of magnitudes of -/- and mixed -/+ errors is of diminished concern for TI and PS methods.

Table 6 presents average overshoot error magnitudes normalized in terms of % excess of the true 0.025 to 0.975 percentile ranges. For example, for n=2 samples the average +/+ overshoot error magnitude for fitting a Normal PDF by the TI method is 14.87 (from the 1st column of numbers in Table B.1 in Appendix B). This excess of 14.87 is divided into average upper and lower overshoot errors $\Delta_{U\text{-}i}$ and $\Delta_{L\text{-}i}$ (see Figure 6) of equal magnitude = 14.87/2 = 7.44. These upper and lower average overshoot errors are added to the true 0.025 - 0.975 percentile range of ±1 to get the upper and lower limits of ±8.44 plotted as the leftmost interval in Figure 16. The combined average overshoot error magnitude of 14.87 is divided by the true 0.025 to 0.975 percentile range of two (a.k.a. ±1) to yield 14.87/2 = 7.44 = 744% in excess of the true 0.025 to 0.975 percentile range of 2. The 744% normalized overshoot error is listed in the 1st column of numbers in Table 6 for n=2 samples, TI method, single Normal PDF fitted.

Table 6. Average +/+ Overshoot Error Magnitudes for TI, PS, and NF Methods (presented as normalized % excess of the true 0.025 to 0.975 percentile ranges)

Method and # of samples	Fitted PDF Shape			3 Fitted PDFs in Convolution			
	Normal	Right Triangle	Uniform	3 Normal PDFs (row1 Fig.1)	3 Rt. Tri. PDFs (row2 Fig.1)	3 Uniform PDFs (row3 Fig.1)	Norm., Tri, Unif. PDFs (row4 Fig.1)
***n*=2 samples**							
TI	744%	850%	922%	790%	798%	804%	798%
PS	253%	297%	310%	230%	233%	227%	238%
***n*=8 samples**							
TI	73%	94%	100%	66%	71%	69%	69%
PS	56%	56%	55%	46%	38%	33%	42%
***n*=32 samples**							
TI	26%	41%	46%	23%	26%	24%	24%
PS	27%	12%	8%	20%	7%	6%	12%

Figure 19 plots the TI and PS +/+ normalized % overshoot error magnitudes from Table 6 against the TI and PS % reliabilities (proportions of +/+ errors) from Tables B.1-B.3 of Appendix B. For each fitting method, two performance curves are plotted: one for performance in fitting single PDFs and the other for performance in fitting multiple PDFs used in the

convolutions in Figure 1. Each curve is defined by (goes through) the mean data points of data "clusters" that correspond to n= 2, 8, and 32 samples as indicated in Figure 19. Each cluster has vertical and horizontal ranges of data points defined in Table 7.

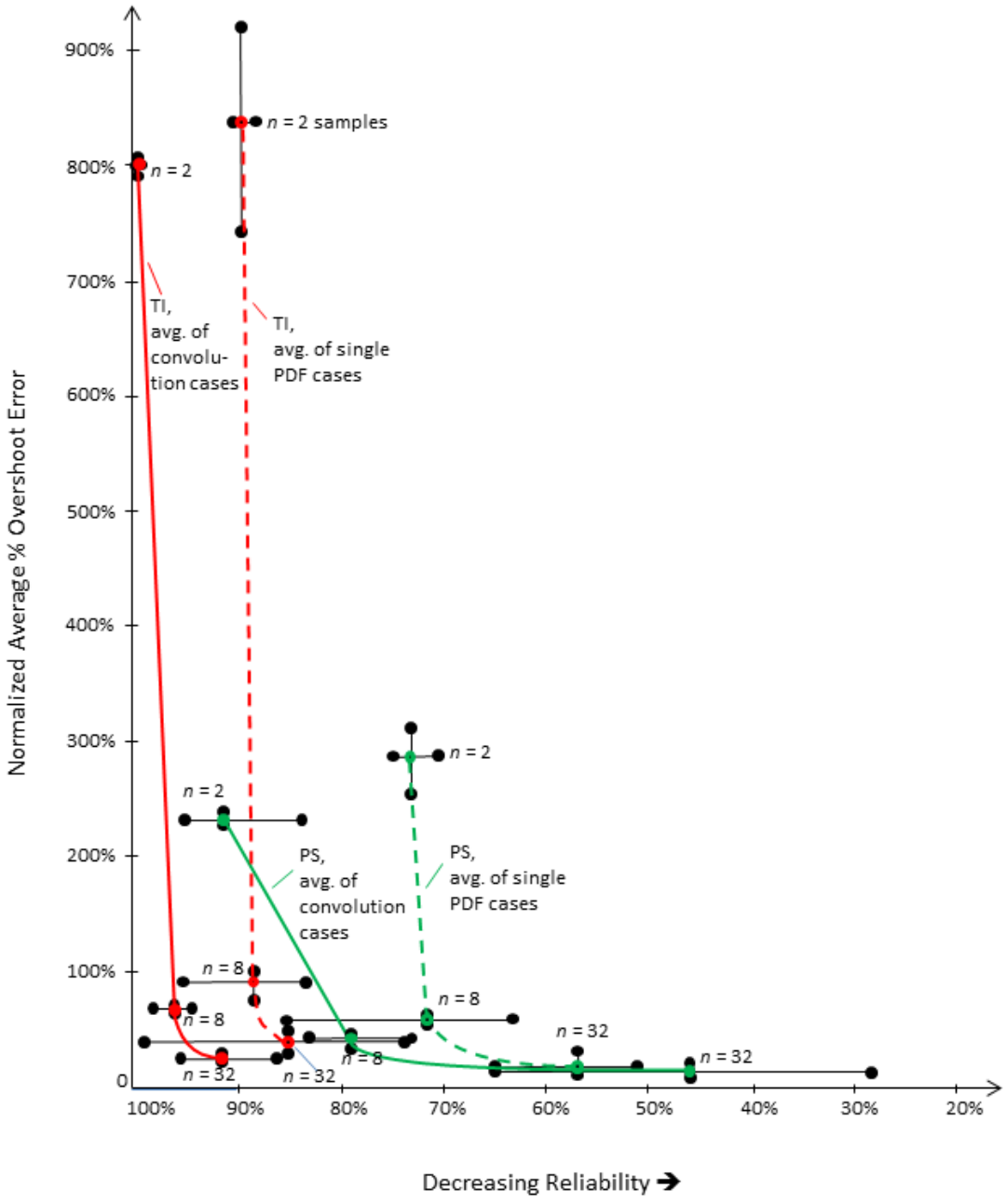


Figure 19. Tradeoff curves of TI and PS +/+ Error Magnitudes vs. Reliabilities in obtaining conservative +/+ errors.

Table 7. Data defining Data-Point "Clusters" in Figure 19

normalized % overshoot error magnitudes (+/+) defining min., average, and max. points in <u>Vertical</u> ranges of data clusters (derived from data in Table 6)

	0.9/0.95 Tolerance Interval Method						Pradlwarter-Schueller Method					
# of Samples	Single PDFs			Convolution Cases			Single PDFs			Convolution Cases		
	Low	Avg. of 3 shapes	High	Low	Avg. of 4 conv. cases	High	Low	Avg. of 3 shapes	High	Low	Avg. of 4 conv. cases	High
2	744%	839%	922%	790%	798%	804%	253%	287%	310%	227%	232%	238%
8	73%	89%	100%	66%	69%	71%	55%	56%	56%	33%	40%	46%
32	26%	38%	46%	23%	24%	26%	8%	16%	27%	6%	11%	20%

% reliability (proportions of +/+ overshoot errors) defining min., average, and max. points in <u>Horizontal</u> ranges of data clusters (derived from data in tables B.1-B.3 of Appendix B)

	0.9/0.95 Tolerance Interval Method						Pradlwarter-Schueller Method					
# of Samples	Single PDFs			Convolution Cases			Single PDFs			Convolution Cases		
	Low	Avg. of 3 shapes	High	Low	Avg. of 4 conv. cases	High	Low	Avg. of 3 shapes	High	Low	Avg. of 4 conv. cases	High
2	87.6%	89%	90%	99.3%	99.4%	99.5%	70.4%	73%	74.8%	83.7%	92%	95.3%
8	83.7%	88%	95.3%	94.8%	96.4%	98.1%	64.6%	72%	85.3%	72.6%	79%	83.2%
32	74.3%	85%	98.9%	86.1%	92%	95.8%	51.1%	57%	64.7%	27.3%	46%	74.3%

The following observations come from examination of Figure 19 and the supporting data tables.

<u>TI and PS performance when convolving 3 fitted PDFs vs. fitting a single PDF</u>

- For both TI and PS the normalized magnitudes of +/+ overshoot errors are typically larger when fitting a single PDF than when using three such fitted PDFs in convolution. (Compare column 2 vs. column 5, column 3 vs. column 6, column 4 vs. column 7 in Table 6.) The overshoot magnitudes in columns 2, 3, 4 are an average of 34% greater than their convolution counterparts in columns 5, 6, 7. Hence in Figure 19 each method's convolution related data clusters for n= 2, 8, and 32 samples are typically significantly lower than the corresponding data clusters from fitting single PDFs. The convolution

related mean curves (solid) in the plot therefore lie lower than the corresponding mean curves (dashed) for fitted single PDFs.

- TI and PS reliabilities are also usually significantly better when fitting three PDFs used in convolution than when fitting a single PDF. Hence each method's convolution related data clusters and mean curves (solid) are usually significantly leftward of the corresponding data clusters and mean curves (dashed) from fitting single PDFs.

- In both performance aspects above, TI and PS perform appreciably better when fitting a single PDF than when fitting three PDFs used in convolution. Both of these performance aspects contribute to separation (vertically due to magnitudes of +/+ overshoot errors and horizontally due to +/+ % reliabilities) between convolution related data clusters and mean performance curves their corresponding data clusters and mean performance curves for fitted single PDFs.

TI vs. PS performance vs. # of samples

- In all of the 21 individual cases but one, PS +/+ errors are smaller (usually considerably smaller) than TI errors, see tables B.1-B.3 in Appendix B. Graphically, for convolution and single PDF cases represented by the four curves/data-sets in Figure 19, the +/+ overshoot errors are largest at n=2 samples and decrease dramatically as n increases to 8 and then 32. At n=2 samples the TI average +/+ overshoot is ≥ 3X that of PS (see Table 7). When n increases from 2 to 8, TI average errors decrease by ~10-11X while those for PS decrease by a much lesser 5-6X but are still significantly smaller than TI average errors at n=8. In going from 8 to 32 samples, PS average errors decrease by a smaller factor of ~3-4X and TI average errors decrease by an even smaller factor of 2-3X. Average TI +/+ errors are 2-3X those of PS @ n=32. Nonetheless, in absolute terms the size of TI +/+ errors for n=32 range from 23 - 46% overshoot over all cases (Table 7), which is not excessively large.

- Concerning method reliabilities, i.e. proportion of +/+ conservative results: for all four curves/data-sets in Figure 19, reliability is largest at the sparsest # of samples (n=2) and typically drops off significantly as n increases to 8 and then 32. At n=2, TI reliability averages 89% over the three single PDF cases and 99.4% over the four convolution cases (Table 7). These compare respectively to 73% and 92% for PS, a combined average of about 12 percentage points lower/worse reliability than TI @ n=2. When n increases from 2 to 8, PS reliability decreases more on average than TS reliability does. At n=8, TI reliability averages 88% (single PDFs) and 96.4% (convolutions). These compare respectively to 73% and 92% for PS, a combined average of about 17 percentage points lower/worse reliability than TI @ n=8. When n increases from 8 to 32, PS reliability again decreases more on average than TS reliability does. At n=32, TI reliability averages 85% (single PDFs) and 92% (convolutions). These compare respectively to 57% and 46% for PS, a combined average of about 37 percentage points lower/worse reliability than TI @ n=32.

Discussion

At *n*=32 samples the absolute size of TI +/+ errors are 23 - 46% overshoot and those of PS are generally < 1/2 as large, 6 – 27% (see Table 7). But consideration of method reliabilities @ *n*=32 in the last bullet above gives a strong advantage to TI in this category. In many usage settings TI's greater average reliability of 85% (single PDFs) and 92% (convolutions) would be viewed as a better tradeoff to the smaller overshoot errors of PS but its comparatively low reliability of 57% for single PDFs and 46% for convolutions.

At *n*=8 samples the absolute size of TI +/+ errors are 66 - 100% overshoot and those of PS are ~1/2 as large, 33 – 56% (Table 7). But again, TI's much better reliability @ *n*=8 gives a strong advantage to TI in this category: 88% average reliability for single PDFs and 96.4% average reliability for convolutions compared to PS's lower average reliabilities of 72% (single PDFs) and 79% (convolutions). Which method to use, TI or PS, would depend more sensitively on the particular circumstances and purposes of the application at hand.

At *n*=2 samples the absolute size of TI +/+ errors are 744 - 922% overshoot and those of PS are ~1/3 as large, 227 – 310% (Table 7). TI again has a much better reliability of 89% average for single PDFs and 99.4% average reliability for convolutions, compared to PS's average reliabilities of 73% (single PDFs) and 92% (convolutions). Here a fairly obvious choice exists for the convolution cases. For these PS achieves >90% reliability on average (the desired level prescribed for the TI method) but does not have egregiously large +/+ overshoot errors like TI does @ *n*=2 samples. But for the single PDF cases a greater tradeoff exists with respect to PS reliability (only 73% average). Nonetheless the egregious ~3X greater +/+ overshoot errors of TI @ *n*=2 makes for a very severe tradeoff in this category, which makes it difficult to prefer TI over PS when very few samples are involved.

5 CONCLUSIONS

From the foregoing analysis the Tolerance Interval and Pradlwarter-Schuëller methods significantly outperform the other three methods tested (Normal-Fitting, Johnson, and Non-Parametric methods) on the set of 21 test problems. The differences in the methods' performance is much greater for low (e.g. 2) and moderate (e.g. 8) numbers of samples than for numbers of samples approaching 30 or more, where performance differences are not large in practical terms.

The TI and PS methods typically yield somewhat larger errors when the problem involves only a single PDF to be estimated from sparse data vs. when the problem involves multiple sources of uncertainty whose PDFs are to be estimated from sparse data (where each uncertainty source has roughly equal impact on response uncertainty). The +/+ conservative error reliabilities were usually appreciably greater, and the normalized +/+ overshoot error magnitudes were typically significantly less, when fitting three equally dominant PDFs in the convolution cases than when fitting a single PDF.

True to the advertised 0.95_coverage/0.9_confidence tolerance intervals used, the TI method exhibited an average of 92% reliability in bracketing the actual 0.025 to 0.975 percentile ranges in the 21 test problems, with a low of 74% reliability and a high of 99%. PS yielded substantially lower average reliability of 70% with a low of 27% reliability and a high of 95%. The classical TI method is also much simpler to implement than the PS method.

The PS method generally yields much smaller +/+ overshoot errors. Hence a significant performance tradeoff exists between the two methods, as depicted in Figure 19. At larger numbers of samples (i.e. n=32), the much better reliability and non-excessive +/+ overshoot errors of TIs may favor them over PS in many use settings. At moderate numbers of samples (i.e. n=8) the absolute size of TI +/+ errors averaged about 2X those of PS. But TI's much better reliability of 88% average for single PDFs and 96.4% average for convolution cases is significantly better than PS's respective averages of 72% and 79%. Which method to use with moderate numbers of samples depends fairly sensitively on the particular circumstances and purposes of the application at hand. At n=2 samples the absolute size of TI +/+ errors averaged about 3X those of PS. This tends to push the advantage to PS for n=2, especially for the convolution cases where PS achieves >90% reliability on average (which is the prescribed target for the TI method). But PS's average reliability for the single PDFs was only 73% while TI's average reliability was 89%, so a more difficult method selection decision exists in this circumstance.

Clearly the PS method deserves more study and characterization. In further research it would be informative to compare +/+ overshoot error magnitudes for a TI confidence level set to the same reliability level exhibited by the PS method on average, 70%. Alternatively or in addition it would be good to see whether PS can be modified to achieve higher reliability, say 90% on average, and then see how its +/+ overshoot error magnitudes for n=2,8,32 samples compare to the average +/+ overshoot magnitudes of .90confidence/.95coverage TIs.

Another important finding from this work is that the common practice of fitting sparse data with a Normal PDF is not very reliable, even if the underlying random process being sampled is Normal. Hence, it is important to bring into common engineering practice better methods for

representing stochastic uncertainty when only relatively few data samples exist. The study here provides clear motivation for the use of appropriate methods for representing the combined aleatory and epistemic uncertainty associated with limited data samples of a stochastic quantity or system. Moreover, it provides an initial glimpse into the treatment options available, their implementation, and their relative performance tendencies for sparse samples of random-variable data. A recent extension of the simple Tolerance Interval approach to address sparse realizations of random *functions* has proven easy and useful to handle multiple discrete stress-strain curves of material variability [31].

6 REFERENCES

[1] Krishnamurthy, T., and Romero, V.J., "Construction of Response Surface with Higher Order Continuity and its Application to Reliability Engineering," paper AIAA-2002-1466, 43rd Structures, Structural Dynamics, and Materials Conference, Denver, CO, April 22-25, 2002.

[2] Romero V.J., Burkardt, J.S., Gunzburger M.D., Peterson J.S., and Krishnamurthy T., "Initial application and evaluation of a promising new sampling method for response surface generation: Centroidal Voronoi Tessellation," paper AIAA-2003-2008, 44th Structures, Structural Dynamics, and Materials Conference, Apr. 7-10, 2003, Norfolk, VA.

[3] Romero, V.J., Painton-Swiler, L., and Giunta, A.A., "Construction of response surfaces based on progressive-lattice-sampling experimental designs with application to uncertainty propagation,'' *Structural Safety* 26 (2004) 201-219.

[4] Romero V.J., Slepoy R., Swiler L.P., Giunta A.A., Krishnamurthy T., "Error estimation approaches for progressive response surfaces -more results," paper #114 in Proceedings of the 2006 International Modal Analysis Conference (IMAC XXIV), Jan. 30 - Feb.2, 2006, St. Louis, MO.

[5] Kline, S.J., and McClintock, F.A., "Describing Uncertainties in Single-Sample Experiments," *Mechanical Engineering*, Jan. 1953, pp. 3 – 8.

[6] International Organization for Standardization, *Guide to the Expression of Uncertainty in Measurement*, 1995 corrected and reprinted edition.

[7] U.S. Dept. of Commerce National Institute of Standards and Technology (NIST) Technical Note 1297, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, 1994 Edition, B.N. Taylor and C.E. Kuyatt.

[8] American Society of Mechanical Engineers, ASME PTC 19.1-2005, *Test Uncertainty*, 2006.

[9] Coleman, H.W., and Steele, Jr., W.G., Experimentation and Uncertainty Analysis for Engineers ─ 2nd Edition, John Wiley & Sons, New York, NY, 1999.

[10] American Society of Mechanical Engineers, V&V 20 – 2009 *Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer*.

[11] Eldred, M.S., and J. Burkardt, "Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification," 47th AIAA Aerospace Sciences Meeting and Exhibit, paper AIAA-2009-0976, January 5-8, 2009, Orlando, FL.

[12] Hahn, G.J., and W.Q. Meeker, *Statistical Intervals—A Guide for Practitioners*, Wiley & Sons, 1991.

[13] Montgomery, D.C., and G.C. Runger, *Applied Statistics and Probability for Engineers*, Wiley & Sons, 1994.

[14] 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.

[15] Sankararaman, S., and S. Mahadevan, "Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data," accepted for *Reliability Engineering and System Safety*, doi:10.1016/j.ress.2011.02.003.

[16] Marhadi, K., S. Venkataraman, S. Pai, "Quantifying Uncertainty in Statistical Distribution of Small Sample Data Using Bayesian Inference of Unbounded Johnson Distribution," paper AIAA-2008-1810, 10th AIAA Non-Deterministic Approaches Conference, April 7–10, 2008, Schaumburg, Illinois.

[17] Zaman, K., M. McDonald, S. Rangavajhala, S. Mahadevan, "Representation and Propagation of both Probabilistic and Interval Uncertainty," paper AIAA-2010-2853, 12th AIAA Non-Deterministic Approaches Conference, April 12–15, 2010, Orlando, Florida

[18] Johnson, N.L., "Systems of frequency curves generated by methods of translation", *Biometrika*, 1949, 36:149-176.

[19] An, J., E. Acar, R.T. Haftka, N.H. Kim, P.G. Ifju, T.F. Johnson, "Being Conservative with a Limited Number of Test Results," *AIAA J. of Aircraft*, Vol. 45, No. 6, Nov.-Dec. 2008.

[20] Romero, V., L. Swiler, A. Urbina, "An Initial Comparison of Methods for Representing and Aggregating Experimental Uncertainties involving Sparse Data," 13th AIAA Non-Deterministic Approaches Conference, April 4-7, 2011, Denver, CO.

[21] Romero, V., J. Mullins, L. Swiler, A. Urbina, "A Comparison of Methods for Representing and Aggregating Experimental Uncertainties involving Sparse Data—More Results," *SAE Int. J. of Materials and Manufacturing,* 6(3):2013, doi:10.4271/2013-01-0946, also Society of Automotive Engineers 2013 World Congress paper 13IDM-0043, April 16-18, 2013, Detroit, MI.

[22] Romero, V., B. Rutherford, J. Newcomer, "Some Statistical Procedures to Refine Estimates of Uncertainty when Sparse Data are Available for Model Validation and Calibration," paper AIAA-2011-1709, 13th AIAA Non-Deterministic Approaches Conference, April 4-7, 2011, Denver, CO.

[23] Rosenblatt, M., "Remarks on Some Nonparametric Estimates of a Density Function," *The Annals of Mathematical Statistics*, Vol. 27, 1956, pp. 832-835.

[24] Parzen, E., "On Estimation of a Probability Density Function and More," *The Annals of Mathematical Statistics*, Vol. 33, No. 3, September 1962, pp. 1065-1076.

[25] Silverman, B.W, *Density Estimation,* Chapman and Hall, 1986.

[26] Jones, M.C., Marron, J.S., Sheather, S.J., "A Brief Survey of Bandwidth Selection for Density Estimation," *Journal of the American Statistical Association*, Vol. 91, No. 433 (March 1996), pp. 401-407.

[27] Wand, M.P. and Jones, M.C., *Kernel Smoothing*, Chapman and Hall, 1994.

[28] Alhberg, J.H., E.N. Nilson, J.L. Walsh, *The Theory of Splines and Their Applications*, Academic Press Inc., New York, NY, 1967.

[29] DeBrota, D.J., Dittus, R.S., Roberts, D. S., Wilson, J. R., Swain, J. J., and S. Venkataraman, "Input modeling with the Johnson System of distributions," *Proceedings of the Winter Simulations Conference*, 1988, eds. M. Abrams, P. Haigh, and J. Comfort.

[30] Jones, D.R., Perttunen, C.D., and Stuckman, B.E., "Lipschitzian optimization without the Lipschitz constant," *Journal of Optimization Theory and Applications*, 1993, **79**(1), 157-181.

[31] Romero, V., J.F. Dempsey, G. Wellman, B. Antoun, "A Method for Projecting Uncertainty from Sparse Samples of Discrete Random Functions ─ Example of Multiple Stress-Strain Curves," 14th AIAA Non-Deterministic Approaches Conference, April 23-26, 2012, Honolulu, HI.

Appendix A Head-to-Head Performance Comparison of Pradlwarter-Schuëller and Johnson Methods

Here we consider results from Pr-Sch and JN methods for all rows 1 – 4 in Figure 1. Figures A.1 and A.2 show that for both the Pr-Sch and JN methods the representations of 0.025 and 0.975 percentiles for Uniform and Right-Triangle PDFs have greater fractions of desirable results (#'s of +/+ minus # of -/- results) than when fitting Normal PDFs. The methods perform best for Uniform PDF shapes, then for Right Triangle PDF shapes, then for Normal PDF shapes. This ordering of performance does not extend to the undesirable 'mixed' results, which show a mix of method performance with respect to PDF shape.

Comparing Figures A.1 and A.2 reveals that Pr-Sch yields significantly more desirable results and less undesirable results than the Johnson method, for all PDF shapes and numbers of samples. Pr-Sch similarly dominates JN (generally) when considering convolution results in Figures A.3 and A.4.

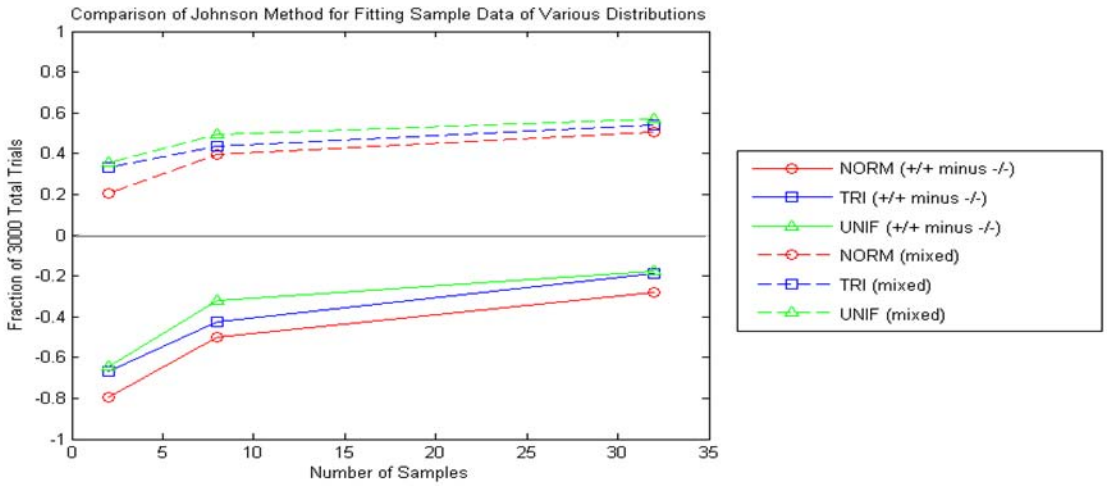


Figure A.1. Comparison of proportions of results in +/+, -/-, and mixed categories for Johnson method as a function of number of data samples from Normal, Right Triangle, and Uniform PDFs.

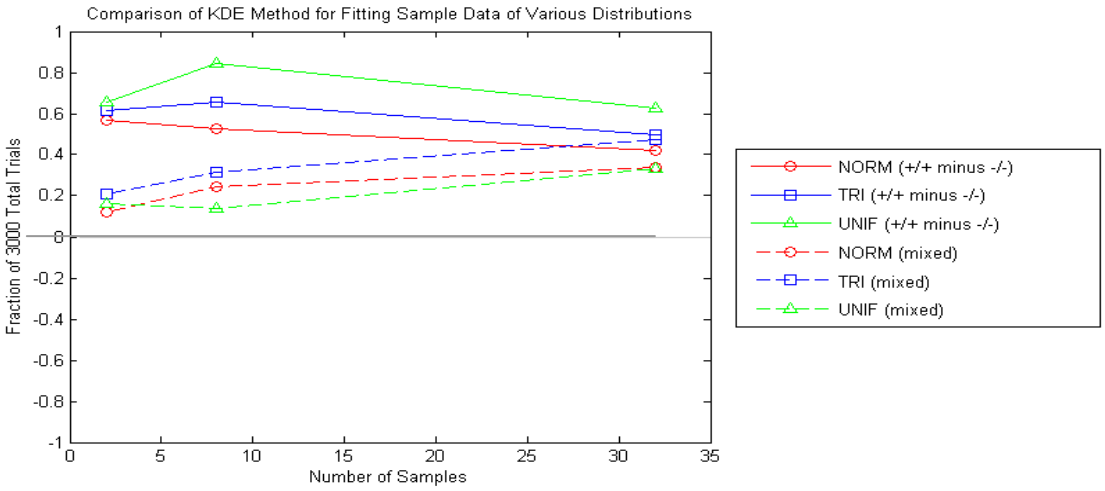


Figure A.2. Comparison of proportions of results in +/+, -/-, and mixed categories for Pradlwarter-Schueller method as a function of number of data samples from Normal, Right Triangle, and Uniform PDFs.

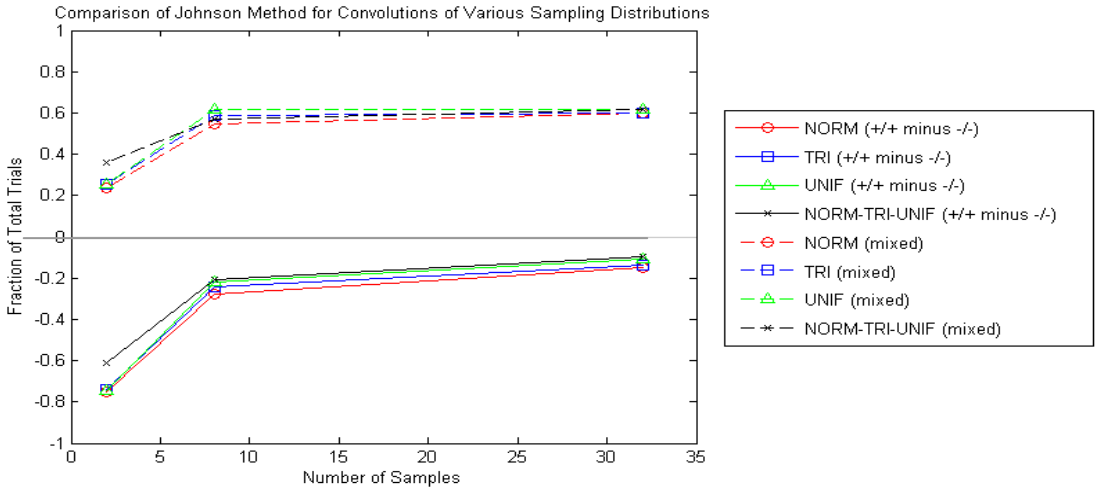


Figure A.3. Comparison of proportions of results in +/+, -/-, and mixed categories for Johnson method as a function of number of data samples—performance in convolution of fits to Normal, Right Triangle, and Uniform PDFs (rows 1 – 4 in Figure 1).

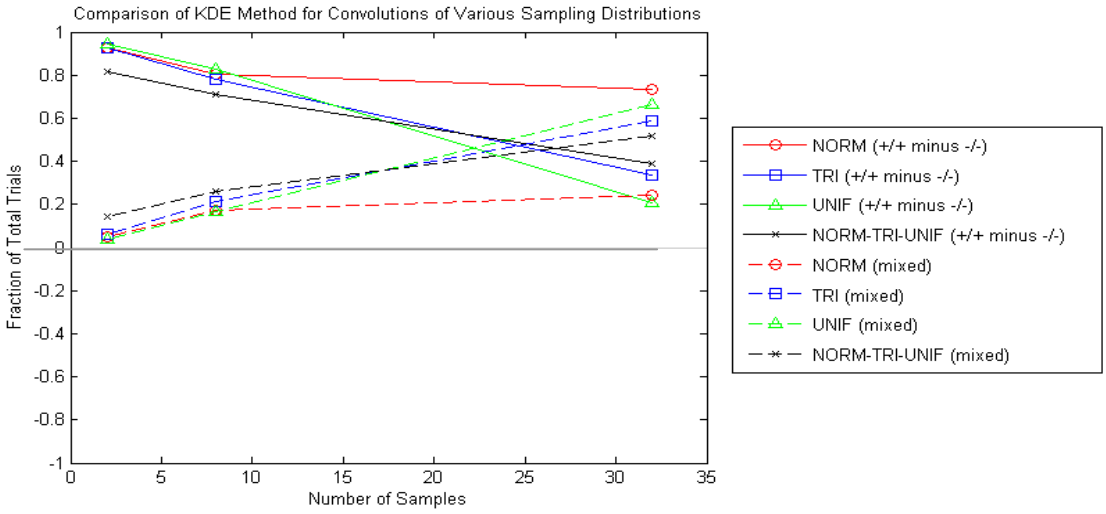


Figure A.4. Comparison of proportions of results in +/+, -/-, and mixed categories for Pr-Sch method as a function of number of data samples—performance in convolution of fits to Normal, Right Triangle, and Uniform PDFs (rows 1 – 4 in Figure 1).

Appendix B Method Performance Trends and Numerical Values of Magnitudes and Proportions of +/+, -/-, and Mixed Errors

Tables B.1 – B.3 list the values for the results plotted in Figures 14 – 17 in Section 4.3.1. Those figures concern Normal PDFs and convolutions of them (row 1 in Figure 1) for all methods studied: TI, PS, NF, NP, JN. Tables B.1 – B.3 also list the results of TI, PS, and NF fits to uniform and right-triangle PDFs and convolution rows 2 – 4 in Figure 1. The results for TI, PS, and NF methods are used for further method performance analysis in Section 4.3.1. NP and JN methods are not analyzed extensively in Section 4.3.1 because of their relative under-performance established in Section 4.3.1 and in Appendix A.

The following performance observations are based on the numerical results in Tables B.1 – B.3 and are analogous to observations A – F in Section 4.3.1 but consider all cases (Normal, uniform, and right-triangle PDFs and convolution cases rows 1 – 4 in Figure 1) and only TI, PS, and NF methods.

Method performance differences in % of +/+, -/-, and mixed errors when fitting a single PDF vs. using 3 such PDFs in convolution

AA. % of -/- undesirable results – When fitting a single uniform or right-triangle PDF or convolving multiple such fitted PDFs in rows 2 and 3 of Figure 1, TI and PS % -/- results are better in the convolution cases for 2 and 8 samples (same as Observation A above regarding fitted Normal PDFs vs. convolution of three fitted Normals). But the agreement with Observation A does not hold for n=32: for Normal PDFs, TI and PS perform slightly better in convolution, but for uniform and right-triangle PDFs, TI and PS perform slightly worse in convolution (though this is of little concern because the size of any -/- errors in either case are relatively minor for n=32). The NF method performed worse in convolutions of uniform and right-triangle PDSs for any # of samples 2,8,32, but performed better in convolution of Normal PDFs for any # of samples (see Observation A).

BB. % of +/- undesirable results – For n=2 samples TI and PS perform better in convolution of three fitted uniform or right-triangle PDFs than when fitting a single uniform or right-triangle PDF. This is the same as Observation B concerning TI and PS and Normal PDFs and all sample sizes 2,8,32. But this trend is reversed for PS @ n=8,32 and for TI @ n=32 by performing worse in convolutions of three fitted uniform or right-triangle PDFs than when fitting single uniform or right-triangle PDFs. For NF the trend cited in Observation B for Normal PDFs holds closely as well for uniform and right-triangle PDFs: for % of +/- results NF performs somewhat worse in convolution than in fitting the singular PDFs.

CC. % of +/+ desirable results – same as in Observation C, +/+ proportion increases (gets better) for all methods and # of samples when convolving three fitted Normal, right-triangle, or uniform PDFs vs. fitting singular PDFs, except for NF @ 2,8,32 samples and PS @ 32 samples.

Method performance differences in % of +/+, -/-, and mixed errors when # of samples changes from n=2, 8, 32

DD. % of -/- undesirable results – consistent with Observation D, -/- proportion significantly declines (gets better) for all except PS methods as # of samples increases (PS method % of -/- errors rises slightly in going from 8 to 32 samples)

EE. % of +/- undesirable results – consistent with Observation E, +/- proportion increases significantly as samples are added for all methods except when applied to a single uniform PDF, for which mixed increases and decreases occur for TI, PS, and NF methods

FF. % of +/+ desirable results – For NF the +/+ proportion improves significantly with increasing samples for Normal PDFs and convolutions of Normals (Observation F) and also uniform PDFs and convolutions of right-triangles, but not for single right-triangle PDFs or convolutions of uniform or mixed PDF (rows 3 and 4). The +/+ percentages of TI and PS methods degrade with increasing samples for Normal PDFs and convolutions of Normals (Observation F) and likewise for all other singular PDF shapes and convolutions except for single uniform PDFs. For the latter TI +/+ proportion improves with increasing samples and PR shows a mixed trend with increasing samples.

Table B.1. Numerical Performance for Sparse-Data Fitting Methods, *n*=2 samples

Metric/Method	Fitted PDF Shape			3 Fitted PDFs in Convolution			
	Normal	Right Triangle	Uniform	3 Normal PDFs (row1 Fig.1)	3 Rt. Tri. PDFs (row2 Fig.1)	3 Uniform PDFs (row3 Fig.1)	Norm., Tri, Unif. PDFs (row4 Fig.1)
proportion of +/+ overshoot errors							
TI	90% (1)	87.6% (2)	89.3% (1)	99.4% (1)	99.4% (1)	99.3% (1)	99.5% (1)
PS	72.7% (2)	70.4% (2)	74.8% (2)	94.3% (2)	93.2% (2)	95.3% (2)	83.7% (2)
NF	30.7% (3)	26.1% (3)	30.6% (3)	19.5% (3)	22.8% (3)	21.6% (3)	21.8% (3)
JN	0% (5)			0.50% (4,5)			
NP	0.6% (4)			0.50% (4,5)			
avg. magnitude of +/+ overshoot errors							
TI	14.87 (5)	17.0 (3)	18.44 (3)	27.29 (5)	30.2 (3)	32.8 (3)	30.14 (3)
PS	5.05 (4)	5.93 (2)	6.2 (2)	7.9 (4)	8.81 (2)	9.24 (2)	9.0 (2)
NF	1.42 (3)	1.55 (1)	1.62 (1)	1.57 (3)	1.56 (1)	1.55 (1)	1.57 (1)
JN	0. (1)			0.55 (1)			
NP	0.483 (2)			0.99 (2)			
proportion of -/- undershoot errors							
TI	5.8% (1)	4.3% (1)	3.8% (1)	0.2% (1)	0.1% (1)	0% (1)	0.1% (1)
PS	15.4% (2)	9.0% (2)	8.8% (2)	1% (2)	0.6% (2)	0.7% (2)	2.1% (2)
NF	52% (3)	36.0% (3)	33.9% (3)	35.3% (3)	30.0% (3)	28.7% (3)	30.5% (3)
JN	29.3% (4)			75.8% (5)			
NP	91.4% (5)			69.8% (4)			
avg. magnitude of -/- undershoot errors							
TI	-1.2 (2)	-1.37 (3)	-1.41 (3)	-1.46 (2)	-0.84 (1)	0. (1)	-1.39 (1)
PS	-1.13 (1)	-1.25 (1)	-1.25 (1)	-1.24 (1)	-1.35 (2)	-1.31 (2)	-1.70 (3)
NF	-1.25 (3)	-1.66 (2)	-1.32 (2)	-1.5 (3)	-1.60 (3)	-1.7 (3)	-1.67 (2)
JN	-1.36 (4)			-1.98 (4)			
NP	-1.47 (5)			-2.04 (5)			
proportion of mixed -/+, +/- errors							
TI	4.2% (1)	8.1% (1)	6.9% (1)	0.4% (1)	0.5% (1)	0.7% (1)	0.4% (1)
PS	11.9% (2)	20.6% (2)	16.3% (2)	4.7% (2)	6.2% (2)	4.0% (2)	14.2% (2)

Metric/Method	Fitted PDF Shape			3 Fitted PDFs in Convolution			
	Normal	Right Triangle	Uniform	3 Normal PDFs (row1 Fig.1)	3 Rt. Tri. PDFs (row2 Fig.1)	3 Uniform PDFs (row3 Fig.1)	Norm., Tri, Unif. PDFs (row4 Fig.1)
NF	27.3% (5)	37.9% (3)	35.5% (3)	45.2% (5)	47.2% (3)	35.5% (3)	47.7% (3)
JN	20.7% (4)			23.7% (3)			
NP	18% (3)			29.7% (4)			
avg. magnitude of mixed -/+, +/- errors							
TI	.89 (1)	1.37 (3)	1.32 (3)	0.77 (1)	2.92 (3)	2.94 (3)	2.88 (3)
PS	.91 (3)	1.32 (2)	1.28 (2)	1.69 (3)	2.17 (2)	2.32 (2)	2.94 (2)
NF	.9 (2)	1.08 (1)	1.05 (1)	1.44 (2)	1.64 (1)	1.05 (1)	1.56 (1)
JN	.94 (4)			1.81 (4)			
NP	1.37 (5)			2.17 (5)			

Table B.2. Numerical Performance for Sparse-Data Fitting Methods, *n*=8 samples

Metric/Method	Fitted PDF Shape			3 Fitted PDFs in Convolution			
	Normal	Right Triangle	Uniform	3 Normal PDFs (row1 Fig.1)	3 Rt. Tri. PDFs (row2 Fig.1)	3 Uniform PDFs (row3 Fig.1)	Norm., Tri, Unif. PDFs (row4 Fig.1)
proportion of +/+ overshoot errors							
TI	84.4% (1)	3.7% (2)	95.3% (1)	96.1% (1)	94.8% (1)	98.1% (1)	96.7% (1)
PS	64.6% (2)	67.0% (2)	85.3% (2)	81.9% (2)	78.2% (2)	83.2% (2)	72.6% (2)
NF	25.6% (3)	34.6% (3)	54.5% (3)	19.9% (3)	24.4% (3)	20.6% (3)	22.5% (3)
JN	5.1% (5)			8.9% (5)			
NP	8.4% (4)			13% (4)			
avg. magnitude of +/+ overshoot errors							
TI	1.45 (5)	1.8 (3)	1.99 (3)	2.3 (5)	2.69 (3)	2.8 (3)	2.6 (3)
PS	1.11 (4)	1.12 (2)	1.09 (2)	1.58 (4)	1.43 (2)	1.35 (2)	1.57 (2)
NF	0.58 (3)	0.68 (1)	0.65 (1)	0.63 (3)	0.60 (1)	0.53 (1)	0.59 (1)
JN	0.44 (1)			0.52 (1)			
NP	0.49 (2)			0.59 (2)			
proportion of -/- undershoot errors							
TI	4% (1)	0.17% (1)	0.07% (1)	0% (1)	0% (1)	0% (1)	0% (1)
PS	11.7% (2)	1.3% (2)	1.2% (2)	1.1% (2)	0.2% (2)	0.2% (2)	1.4% (2)
NF	37.2% (3)	6.5% (3)	7.0% (3)	24.4% (3)	11.8% (3)	13.4% (3)	15.7% (3)
JN	55.2% (5)			36.5% (5)			
NP	54.5% (4)			33.2% (4)			
avg. magnitude of -/- undershoot errors							
TI	-0.45 (1)	-0.40 (1)	-0.6 (3)	0 (1)	0 (1)	0 (1)	0 (1)
PS	-0.53 (2)	-0.55 (2)	-0.41 (1)	-0.39 (2)	-0.30 (2)	-0.23 (2)	-0.41 (2)
NF	-0.56 (3)	-0.56 (3)	-0.42 (2)	-0.63 (3)	-0.56 (3)	-0.53 (3)	-0.55 (3)
JN	-0.93 (5)			-0.72 (5)			
NP	-0.71 (4)			-0.69 (4)			
proportion of mixed -/+, +/- errors							
TI	11.6% (1)	16.1% (1)	4.7% (1)	3.9% (1)	5.2% (1)	1.9% (1)	3.3% (1)
PS	23.7% (2)	31.7% (2)	13.5% (2)	17% (2)	21.5% (2)	16.6% (2)	26% (2)

Metric/Method	Fitted PDF Shape			3 Fitted PDFs in Convolution			
	Normal	Right Triangle	Uniform	3 Normal PDFs (row1 Fig.1)	3 Rt. Tri. PDFs (row2 Fig.1)	3 Uniform PDFs (row3 Fig.1)	Norm., Tri, Unif. PDFs (row4 Fig.1)
NF	37.2% (4)	58.9% (3)	38.6% (3)	55.7% (5)	63.8% (3)	66% (3)	61.8% (3)
JN	39.6% (5)			54.6% (4)			
NP	37.1% (3)			53.8% (3)			
avg. magnitude of mixed -/+, +/- errors							
TI	0.46 (2)	0.92 (3)	0.78 (3)	1.09 (5)	1.51 (3)	1.63 (3)	1.52 (3)
PS	0.48 (3)	0.72 (2)	0.60 (2)	0.89 (4)	1.06 (2)	1.05 (2)	1.18 (2)
NF	0.44 (1)	0.68 (1)	0.52 (1)	0.72 (1)	0.81 (1)	0.77 (1)	0.74 (1)
JN	0.64 (5)			0.77 (2)			
NP	0.58 (4)			0.78 (3)			

Table B.3. Numerical Performance for Sparse-Data Fitting Methods, *n*=32 samples

Metric/Method	Fitted PDF Shape			3 Fitted PDFs in Convolution			
	Normal	Right Triangle	Uniform	3 Normal PDFs (row1 Fig.1)	3 Rt. Tri. PDFs (row2 Fig.1)	3 Uniform PDFs (row3 Fig.1)	Norm., Tri, Unif. PDFs (row4 Fig.1)
proportion of +/+ overshoot errors							
TI	80.7% (1)	74.3% (2)	98.9% (1)	91.3% (1)	86.1% (1)	95.8% (1)	95.3% (1)
PS	54.4% (2)	51.1% (2)	64.7% (3)	74.3% (2)	37.4% (2)	27.3% (2)	43.3% (2)
NF	27.8% (3)	26.2% (3)	80.2% (2)	21.2% (4)	24.9% (3)	21.1% (3)	20.5% (3)
JN	10.5% (5)			12.5% (5)			
NP	24.1% (4)			28.4% (3)			
avg. magnitude of +/+ overshoot errors							
TI	0.52 (4)	0.82 (3)	0.91 (3)	0.80 (5)	0.99 (3)	0.97 (3)	0.91 (3)
PS	0.54 (5)	0.24 (1)	0.16 (1)	0.69 (4)	0.28 (1)	0.23 (1)	0.47 (2)
NF	0.27 (2)	0.47 (2)	0.43 (2)	0.29 (1)	0.32 (2)	0.26 (2)	0.28 (1)
JN	0.256 (1)			0.307 (2)			
NP	0.49 (3)			0.37 (3)			
proportion of -/- undershoot errors							
TI	3.1% (1)	0% (1)	0% (1)	0% (1)	0% (1)	0% (1)	0% (1)
PS	12.4% (2)	1.6% (3)	2.2% (3)	0.9% (2)	3.9% (2)	6.4% (2)	4.7% (2)
NF	32% (4)	0.3% (2)	0.73% (2)	20.3% (4)	4.7% (3)	11.0% (3)	12.8% (3)
JN	38.8% (5)			27.6% (5)			
NP	26.9% (3)			14.7% (3)			
avg. magnitude of -/- undershoot errors							
TI	-0.18 (1)	0. (1)	0 (1)	0. (1)	0. (1)	0 (1)	0 (1)
PS	-0.25 (2)	-0.21 (2)	-0.125 (3)	-0.17 (2)	-0.19 (2)	-0.17 (2)	-0.23 (2)
NF	-0.27 (3)	-0.23 (3)	-0.120 (2)	-0.280 (3)	-0.24 (3)	-0.23 (3)	-0.25 (3)
JN	-0.33 (5)			-0.31 (5)			
NP	-0.33 (4)			-0.282 (4)			

Metric/Method	Fitted PDF Shape			3 Fitted PDFs in Convolution			
	Normal	Right Triangle	Uniform	3 Normal PDFs (row1 Fig.1)	3 Rt. Tri. PDFs (row2 Fig.1)	3 Uniform PDFs (row3 Fig.1)	Norm., Tri, Unif. PDFs (row4 Fig.1)
proportion of mixed -/+, +/- errors							
TI	16.2% (1)	25.7% (1)	1.1% (1)	8.7% (1)	13.9% (1)	4.2% (1)	4.7% (1)
PS	33.2% (2)	47.3% (2)	33.1% (3)	24.8% (2)	58.7% (2)	66.3% (2)	52% (2)
NF	40.2% (3)	73.5% (3)	19.1% (2)	58.5% (4)	70.4% (3)	67.9% (3)	66.7% (3)
JN	50.7% (5)			59.9% (5)			
NP	49% (4)			56.9% (3)			
avg. magnitude of mixed -/+, +/- errors							
TI	0.23 (2)	0.65 (3)	0.53 (3)	0.56 (5)	0.79 (3)	0.97 (3)	0.72 (3)
PS	0.3 (3)	0.26 (2)	0.16 (1)	0.45 (4)	0.34 (1)	0.33 (1)	0.48 (2)
NF	0.21 (1)	0.52 (1)	0.32 (2)	0.34 (1)	0.51 (2)	0.39 (2)	0.38 (1)
JN	0.33 (4)			0.378 (2)			
NP	0.39 (5)			0.42 (3)			

External Distribution

1	Joshua G. Mullins, co-author, Vanderbilt University PhD Candidate			(electronic copy)

Internal Sandia Distribution

1	MS0352	1344	C.E. Hembree	(electronic copy)
1	MS0812	1544	V.J. Romero	(electronic copy)
1	MS0825	1514	M. Pilch	(electronic copy)
1	MS0828	1541	R. Bond	(electronic copy)
1	MS0828	1544	W.R. Witkowski	(electronic copy)
1	MS0828	1544	A.R. Black	(electronic copy)
1	MS0828	1544	B. Carnes	(electronic copy)
1	MS0828	1544	K.J. Dowding	(electronic copy)
1	MS0828	1544	A.C. Hetzler	(electronic copy)
1	MS0828	1544	R.G. Hills	(electronic copy)
1	MS0828	1544	K. Hu	(electronic copy)
1	MS0828	1544	G.E. Orient	(electronic copy)
1	MS0828	1544	J.R. Red-Horse	(electronic copy)
1	MS0828	1544	A. Urbina	(electronic copy)
1	MS0829	431	B.M. Rutherford	(electronic copy)
1	MS0829	431	J.T. Newcomer	(electronic copy)
1	MS0845	1540	D.E. Womble	(electronic copy)
1	MS0845	1545	T.L. Galpin	(electronic copy)
1	MS0845	1542	J. Jung	(electronic copy)
1	MS0897	1543	T.D. Blacker	(electronic copy)
1	MS0897	1544	K.D. Copps	(electronic copy)
1	MS1138	6923	B.S. Paskaleva	(electronic copy)
1	MS1138	6921	M.L. Wilson	(electronic copy)
1	MS1177	1355	J.P. Castro	(electronic copy)
1	MS1318	1440	T.G. Trucano	(electronic copy)
1	MS1318	1441	J.R. Stewart	(electronic copy)
1	MS1318	1441	B.M. Adams	(electronic copy)
1	MS1318	1441	M.S. Eldred	(electronic copy)
1	MS1318	1441	L.P. Swiler	(electronic copy)
1	MS1320	1441	V.G. Weirs	(electronic copy)
1	MS1323	1443	W.J. Rider	(electronic copy)
1	MS9042	8250	M.E. Gonzalez	(electronic copy)
1	MS9159	8954	G.A. Gray	(electronic copy)
1	MS9159	8954	P.D. Hough	(electronic copy)
1	MS0899	9536	Technical Library	(electronic copy)


Sandia National Laboratories