8-20-2003

Semiparametric Estimation of Unimodal Distributions

Jason K. Looper
University of South Florida

Semiparametric Estimation of Unimodal Distributions

by

Jason K. Looper

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Physics
Department of Physics
College of Arts and Science
University of South Florida

Major Professor: David A. Rabson, Ph.D.
Wei Chen, Ph.D.
Chun-Min Lo, Ph.D.

Date of Approval:
August 20, 2003

Keywords: Probability, Bump-hunting, Single Peak, Constrained Least Squares

DEDICATION

To my father who never stopped encouraging me, to my other father (Stephen D. Smith) who never stopped funding me, and to my granny Looper who never lost her sense of wonder.

ACKNOWLEDGMENTS

Without the instruction and supervision of Dr. David Rabson this thesis could not have been accomplished, and many results could not have been concluded.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SEMIPARAMETRIC ESTIMATION OF UNIMODAL DISTRIBUTIONS

Jason Keith Looper

ABSTRACT

One often wishes to understand the probability distribution of stochastic data from experiment or computer simulations. However, where no model is given, practitioners must resort to parametric or non-parametric methods in order to gain information about the underlying distribution. Others have used initially a nonparametric estimator in order to understand the underlying shape of a set of data, and then later returned with a parametric method to locate the peaks. However they are interested in estimating spectra, which may have multiple peaks, where in this work we are interested in approximating the peak position of a single-peak probability distribution.

One method of analyzing a distribution of data is by fitting a curve to, or smoothing them. Polynomial regression and least-squares fit are examples of smoothing methods. Initial understanding of the underlying distribution can be obscured depending on the degree of smoothing. Problems such as under and oversmoothing must be addressed in order to determine the shape of the underlying distribution. Furthermore, smoothing of skewed data can give a biased estimation of the peak position.

We propose two new approaches for statistical mode estimation based on the assumption that the underlying distribution has only one peak. The first method imposes the global constraint of unimodality locally, by requiring negative curvature over some domain. The second method performs a search that assumes a position of the distribution's peak and requires positive slope to the left, and negative slope to the right. Each approach entails a constrained least-squares fit to the raw cumulative probability distribution.[1]

[1]Refer to chapter one for the raw cumulative probability distribution.

We compare the relative efficiencies [12] of finding the peak location of these two estimators for artificially generated data from known families of distributions Weibull, beta, and gamma. Within each family a parameter controls the skewness or kurtosis, quantifying the shapes of the distributions for comparison. We also compare our methods with other estimators such as the kernel-density estimator, adaptive histogram, and polynomial regression. By comparing the effectiveness of the estimators, we can determine which estimator best locates the peak position.

We find that our estimators do not perform better than other known estimators. We also find that our estimators are biased. Overall, an adaptation of kernel estimation proved to be the most efficient.

The results for the work done in this thesis will be submitted, in a different form, for publication by D.A. Rabson and J.K. Looper.

CHAPTER 1

INTRODUCTION AND MOTIVATION

1.1 Introduction

We are interested in the peak position of a single-peak, continuous, univariate probability distribution. A probability distribution, $p(x)$, on some domain $[a, b]$, defines the probability, $p(x)dx$, that a random variable will be measured in the range $[x, x + dx]$. There are two constraints:

$$p(x) \geq 0 \tag{1}$$

$$\int_a^b p(x)dx = 1 \tag{2}$$

Fig. 1 shows an example of a probability distribution. If we have a continuous distribution, $p(x)$, we can describe the continuous cumulative distribution, $C(x)$, by

$$C(x) = \int_{-\infty}^{x} p(x')dx'. \tag{3}$$

An example of a continuous cumulative distribution is shown in Fig. 2 If we have a set of random measurements, x_i, taken from a distribution, $p(x)$, we can use the discrete form of the cumulative distribution

$$C(x) = \sum_{x_i<x} \frac{1}{N} \tag{4}$$

where N is the number of random measurements. An example of a set of random measurements and the discrete cumulative distribution is shown in Fig. 3. Essentially every time a datum is encountered in the random measurements, the height increases by N^{-1}

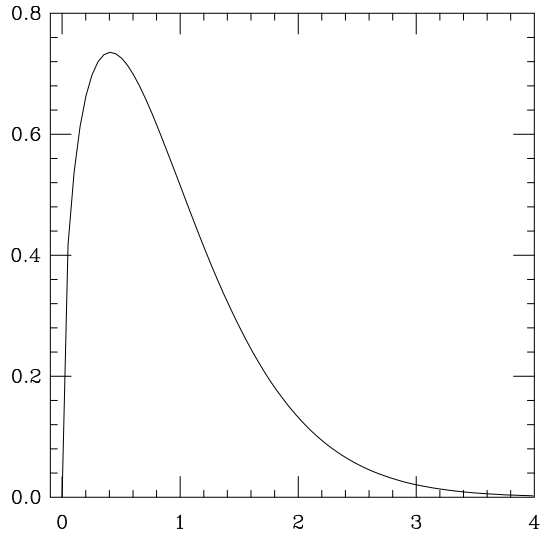

Figure 1. Probability distribution.

in the discrete cumulative distribution. The discrete cumulative distribution is another way of describing how much probability is at, or to the left of, each datum.[1]

It is not difficult to see that when the data are grouped close together, the raw cumulative distribution will have a large slope. Where the slope is largest corresponds to the peak position in the probability distribution. We must first smooth the raw cumulative distribution in order to estimate the peak in the probability distribution.

In analyzing data we are often interested in determining the underlying distribution associated with observed or artificially produced measurements. Given a finite set of measurements, we can employ several different statistical methods in order to estimate the peak of the distribution. These methods can be used individually or together. The relative efficiency of these methods depends on our knowledge of the system being measured, or of the artificially produced data. The two statistical methods that will be discussed are parametric and non-parametric estimators. A combination of these methods to locate a peak position of a unimodal distribution is the main topic of this work.

[1]Random measurements from a distribution will be referred to as raw data sets, and the discrete cumulative distribution will be referred to as the raw cumulative distribution.

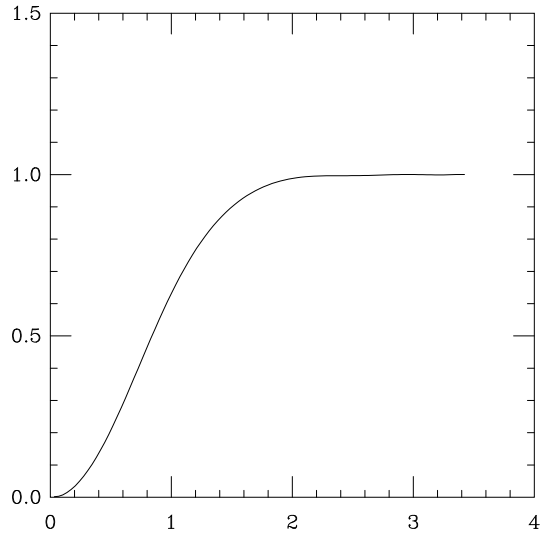

Figure 2. Cumulative distribution.

Often we have partial or no knowledge of the underlying functional form of the distribution associated with a data set. In this situation, non-parametric methods are generally used because they impose no restrictions on the data. Nonparametric estimators can be effective on any family of distributions, even when the distribution is not known. The counterparts are the parametric methods, in which we assume a functional form with a small number of parameters. The estimation then consists of finding the best fit for these parameters. If we know something about the distribution, we can use a hybrid technique, which consists of both parametric and non-parametric methods.

Riedel proposed a piecewise convex method of estimating the shape of an unknown function, which uses a two stage smoothing technique [9]. The first stage uses a non-parametric smoothing method for determining the location of inflection points and then a second-stage method to smooth the function between these points. Riedel's method parallels our proposal in that we also use a first-stage estimation to determine the domain of a peak and then a final-stage estimation to smooth. Riedel is also using an initial smoothing method to gain information about a spectrum of multiple peaks, where we already understand that our distributions have only one peak. Unlike Reidel's method,

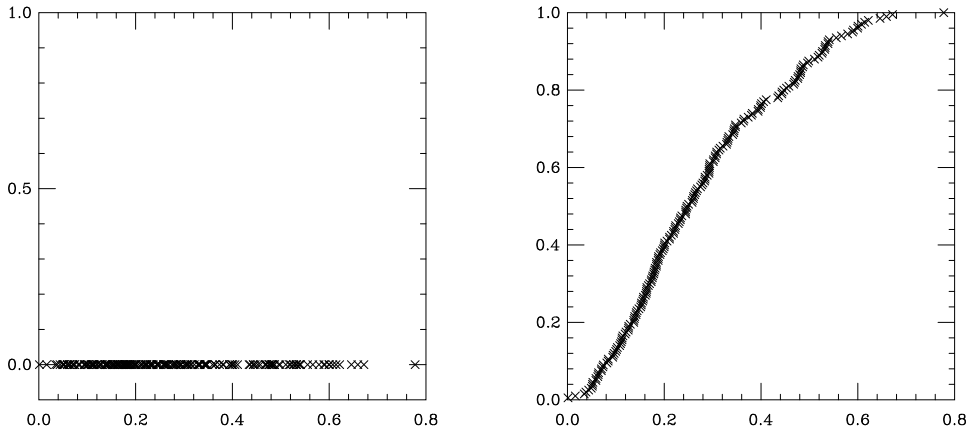

Figure 3. The figure on the left is a set of random measurements taken form a beta distribution having $p = 1.8$ and $q = 5$. The set to the right is its discrete (raw) cumulative distribution.

our first-stage estimator is used for location of a peak and not to determine the number of peaks. Riedel's proposal also differs from ours in the final method of smoothing [10].

We propose two different methods for determining the peak position of an unknown probability distribution. The distributions we are interested in estimating are unimodal and therefore contain no more than two inflection points. The first method, named `Unimodal3`, uses an estimation technique to determine the domain of the peak position. Once this domain is found, a constrained least-squares fit is formed in that domain on the raw cumulative distribution, where the constraint is on the third derivative of the fit [3]. Rasmus, Nicholaos and Sidiropoulos use weighted least squares, however they are also interested in estimating spectra [8]. The second proposed method, named `Unimodal2`, performs an exhaustive or binary search for the peak position, subject to the condition of positive slope to the left, and negative slope to the right of the peak location. Once the fit has been estimated, the residuals from the fitted cumulative distribution can be compared for every proposed peak location.[2]

[2]Residuals are the square difference of the fitted cumulative distribution and the raw cumulative distribution.

1.2 Motivation

One motivation for this project came from modeling finite nanometer-scale quantum wires. It is found that wires with a finite number of hopping sites act unlike systems with an infinite number of sites. In infinite chains, electronic transport is believed to undergo an immediate transition from ballistic conduction to diffusive when infinitesimal second-neighbor coupling is turned on. In small chains, Rabson, Narozhny, and Millis now have numerical evidence instead for a crossover region of the level-spacing probability distribution, $P_\Delta(E)$, where Δ is the mean level spacing [7].

The Hamiltonian with only first-neighbor interactions is integrable. This is due the large number of conserved quantities which determine the behavior of the system. Here energy levels can generically cross.[3] Therefore the level-spacing distributions for the first-neighbor interaction follow Poisson statistics,

$$P_\Delta(E) = \frac{1}{\Delta} e^{(-E/\Delta)}. \tag{5}$$

When we turn on second-neighbor interaction, we break the integrability of the system by reducing the number of conserved quantities. Here the level spacings generically do not cross, so that small and large level spacings become unlikely. The level spacing with the second-neighbor interaction follows Wigner-Dyson statistics,

$$P_\Delta(E) = \frac{\pi E}{2\Delta} e^{\pi E^2/4\Delta^2}. \tag{6}$$

The crossover describes the region of change from Poisson to Wigner-Dyson statistics. Our understanding of this region is limited, but it is likely that the level-spacing distributions possesses only one most probable level. Therefore the unknown distributions within this crossover region each possess only one peak. We hope that the peak of the level-spacing distributions in the crossover region will give us insight into the change of conductances

[3] For example as a function of magnetic flux.

and its finite-size scaling. Since the functional forms of these distributions are not known, we must resort to statistical methods such as parametric and non-parametric estimators.

1.3 Research Plan

In order to test the efficiencies [12] of `Unimodal3` and `Unimodal2` in determining the peak position of an unknown distribution, we generate sets of deviates from unimodal distributions on $[0, \infty)$. Here we will employ a random-number generator to produce deviates from known probability distributions Weibull, gamma, and beta.[4] These artificial data sets will range from a small number of deviates per data set to a large number, all having multiple realizations so that we may analyze the statistics of the estimators. Since scientists are unable to gather an infinite number of measurements in an actual physical system, we are interested in the efficiencies of our estimators on data sets that have a small number of data points. We will compare the relative efficiencies of our estimators to those of other estimators such as kernel smoothing, polynomial smoothing, histograms, and the unbiased estimators of the distributions.

We will vary the parameters of each generated distribution to have a range of skewness and kurtosis.[5] By changing the parameters, we will produce underlying distributions ranging from those with peaks near the y axis to others that are flatter.

[4] Deviate is a simulated measurement or datum.

[5] Refer to Appendix A for examples of distributions having different skewness.

CHAPTER 2

PARAMETRIC AND NON-PARAMETRIC ESTIMATORS

Two types of estimators were pointed out in the introduction, parametric and non-parametric. This chapter focuses on describing traditional parametric and nonparametric estimators that are used in analyzing data. Some of these estimators have been modified to improve their ability to find the peak of a distribution.

2.1 Kernel Density Estimator

The kernel density estimator is a nonparametric approach to estimating a probability distribution. Because there is no imposed parametric model, kernel smoothing allows the data to speak for themselves. The basic principles of the kernel estimator were independently introduced by Fix and Hodges (1951) and Akaike (1954) [13]. If we have a set of data, x_i, from a distribution $P(x)$, then the kernel estimator of that distribution is given by

$$F(x;h) = (nh)^{-1} \sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big) \tag{7}$$

where h is the parameter known as the bandwith and K is the kernel function satisfying $\int K(x)dx = 1$. We can condense this formulation by introducing a rescaling notation

$$K_h(u) = h^{-1}K(u/h), \tag{8}$$

which allows us to write

$$F(x;h) = (n)^{-1} \sum_{i=1}^{n} K_h(x - x_i). \tag{9}$$

Because the kernel function K is chosen to be a symmetric unimodal probability density function, then $F(x;h)$ is also a density function. We have chosen the kernel function, K, to be a parabola.

The central parameter in kernel smoothing is the bandwidth h of the estimator. The bandwidth represents the base size of the parabola, or kernel, that is centered on a datum. The value of the function $F(x;h)$ at each point is simply the average of the kernel functions at each point. The bandwidth determines the degree of smoothing. For example, two smoothing extremes are a consequence of varying the bandwidth size. If the bandwidth is set to a minimal value, undersmoothing of the underlying data is the result. As seen in Fig. 4, obvious problems of undersmoothing include the presence of multiple peaks and spurious structure in the body of the distribution. Taking this problem to the extreme results in delta functions centered at each datum. Contrary to undersmoothing is the problem of oversmoothing. Oversmoothing results when the bandwidth is set to a large value. As seen in Fig. 4, problems pertaining to oversmoothed data sets include the shifting of the estimated mode position and the flattening of features.

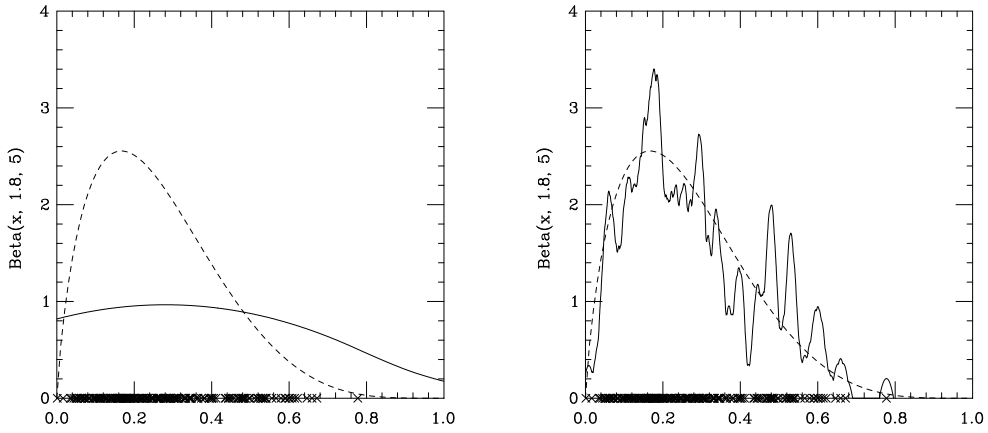

Figure 4. Example of over and undersmoothing by kernel estimation on a 1000-deviate sample drawn from a beta distribution having $p = 1.8$ and $q = 5$. Raw data from the beta distribution are plotted as crosses on the x axis. The dotted line represents the underlying beta distribution. To the left is an example of a distribution that has been oversmoothed. It is noticeable that the smoothed peak position has shifted relative the the true peak position. To the right is an example of undersmoothing.

Heuristically it appears that the most efficient bandwidth for the purpose of locating the position of a single peak is one that lies just over the threshold for undersmoothing. For example, Fig. 4 shows that oversmoothing can result in the shifting of the smoothed peak position. Therefore we wish to find an h that is close to, but does not result in undersmoothing. However undersmoothing can result in multiple peaks, but if these peaks are small compared to a large dominating peak, then these small peaks are not likely to be candidates for the peak of the distribution. Therefore we can ignore small peaks that appear in undersmoothing. Now we can find an h that produces a distribution that is not oversmoothed and not quite undersmoothed. Fig. 5 shows an example.

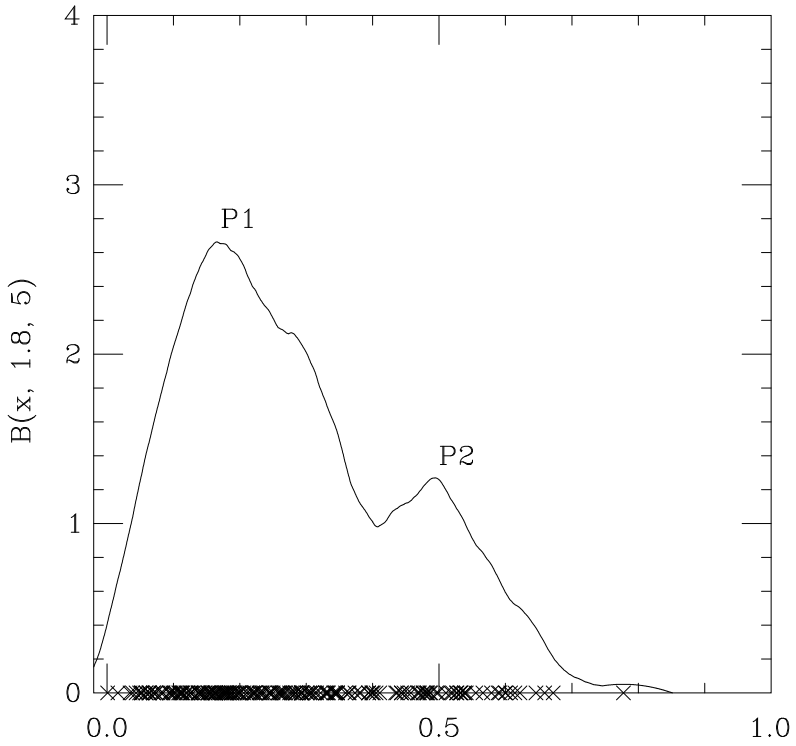

Figure 5. Example of a kernel-smoothed beta distribution $p = 1.8$ and $q = 5$. The raw data are plotted on the x axis. $P1$ is a mode candidate and $P2$ is not. If $P2$ is less than half the size of $P1$, then $P2$ is not a mode candidate. All peaks that are less than half the size of the most dominant peak are ignored. Here the h parameter is set to be as small as possible, consistent with there being only a single peak candidate.

2.2 Optimal Kernel Density Estimator

The Optimal Kernel Smoothing (OKS) method is an alternative to traditional kernel smoothing. Unlike traditional kernel smoothing, where the user must test for the best bandwidth, OKS determines h that best satisfies the criterion for a unimodal distribution. Specifically, two bandwidths must be determined in order to bracket the most efficient bandwidth. The first bracketing bandwidth, h_1, is set so that the resulting smoothed distribution will be oversmoothed. This can be achieved by setting h_1 to the standard deviation of the distribution or some multiple. The second bracketing bandwidth, h_2, is set so that the resulting smoothed distribution is undersmoothed. Here h_2 is a small fraction of the standard deviation. Once these two bandwidths have been found, the optimization routine then searches for a band-width h_o, between h_1 and h_2, that is as small as possible consistent with unimodality. The optimal bandwidth h_o is the one that lies just under the threshold for undersmoothing. Again undersmoothing can result in multiple peaks; therefore OKS also compares the peaks using a height restriction for mode candidates to help determine the optimal h_o. Here the height restriction is $1/2$ the height of the highest peak. Any peak that falls bellow this height is not a mode candidate.

This method determines h_o for every data set. This differs from the kernel estimation in section 2.1 where we hold h constant over all 500 realizations, and run several trials varying h, to determine h_o. We will from now on refer to the kernel estimator in 2.1 as Fixed Kernel Smoothing (FKS).

2.3 Adaptive Histogram

Another nonparametric estimator is the adaptive histogram (AHIST) estimator. This method also has one parameter and does not rely on a parametric model. The central parameter in AHIST estimation is the number of data points in a bin. Unlike traditional histogram bins, each adaptive bin holds the same number of data points. The width of the bin is increased to encompass the given number of points. The area under the bin

equals the number of points and is held constant, therefore the height varies inversely as the width. As shown in Fig. 6, data points that are close produce a taller bin than those that are spread farther apart. One does not wish to set the bin size to be too large because this would decrease the number of bins being used. Similar to the problem of oversmoothing, a small number of bins flattens out any features that might be present in the underlying distribution. Also one does not wish to use a bin size that is too small. This would result in the undersmoothing of the distribution. The optimal bin size is one that puts the smallest number of data in each bin but does not undersmooth the distribution. Fig. 7 shows an example.

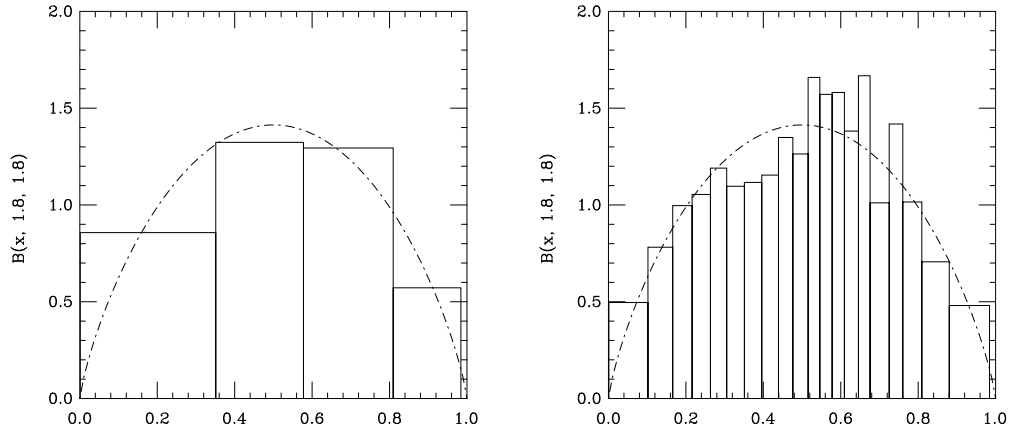

Figure 6. The graph on the left represents a beta distribution that has been oversmoothed by AHIST. The graph on the right represents the same beta distribution undersmoothed. The underlying distribution is plotted as the dot-dashed line. Both graphs are drawn from a sample of 1000-deviates.

2.4 Optimal Chebyshev Polynomial Fit

Optimal Chebyshev Polynomial Fitting (OCPF) is the next non-parametric estimator that we will introduce. In this section we will explain how we use OCPF to smooth the raw cumulative distribution. We are calling this method optimal because we have included a two-test procedure for fitting a curve to the raw cumulative distribution. These tests insure that the fit results in a single-mode probability distribution.

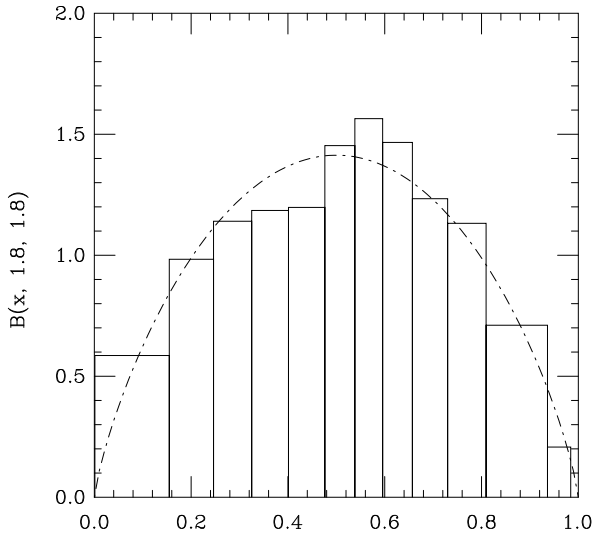

Figure 7. The histogram above is of a beta distribution with an optimal bin parameter of .1, which is 10 percent of the data set. The underlying distribution is plotted as the dot-dashed line.

If we have a raw cumulative distribution, $C(x)$, formed from a set of random deviates from a unknown distribution, OCPF first determines the domain of the data points,[1] then calculates the Chebyshev coefficients using ether a cubic or linear interpolation. Here we have decided to use a linear interpolation, due to smoothing anomalies produced when using a cubic interpolation.[2] The resulting number of coefficients equals $n + 1$, where n is the order of the fit determined by OCPF. OCPF starts with a $n = 4$ order fit and then increases n by 1, depending on the two tests. Once all of the coefficients have been calculated, OCPF can approximate $C(x)$ and then by taking the derivative, the underlying probability distribution. In order to tell whether the final fit to the cumulative distribution has been optimized, OCPF incorporates two tests on its derivative.

As in the previous estimation methods there are two major problems, over and undersmoothing. A large n can result in undersmoothing while a small n can result in oversmoothing. What we would like is the highest-order n that does not result in under-

[1] Appendix C describes a change of variables necessary to use Chebyshev polynomials.

[2] This smoothing anomaly involves a hump at the beginning of the cumulative distribution fit, that is most likely not part of the underlying distribution.

smoothing. Therefore the first test after the approximation of $C(x)$ is one that determines if the 1^{st} derivative of the fit is unimodal. As shown in the kernel estimation, undersmoothing can result in multiple peaks, therefore OCPF incorporates the same peak-candidate tests as OKS. The second test that OCPF performs is a test on the beginning of the 1^{st} derivative of the fit. Because we understand that the distributions we are trying to estimate are unimodal, the 1^{st} derivative of the fit close to the y axis should not have a negative slope.[3] If an n^{th}-order fit produces a curve that has a negative slope at the start of a estimated probability distribution, then OPCF throws away the fit and procceeds to a higher order estimation. What is considered to be an optimal fit is one that produces the highest order estimation, does not have a negative slope at the beginning of its 1^{st} derivative, and does not undersmooth the distribution. Fig. 8 and 9 show examples of a good and a bad fit.

2.5 Unbiased Parametric Estimator

Each of our three test distributions possesses two parameters that determine its shape.[4] The following estimators are parametric unbiased estimators of the distributions. These estimators determine the distributions' parameters in order to locate the peak position for Weibull, gamma, and beta distributions.

Since any distribution or population sample on $[0, \infty)$, can be scaled to have unit mean, we require that two of the test distributions have a mean equal to one. Therefore when we produce a random data set from the Weibull and gamma distributions, we have built in the constraint that the mean have a value of unity. Because we have a finite number of data, the mean of the sample is not exactly 1, but is close. We will use the fact that the mean is equal to one in our Weibull and gamma parametric estimators.

The reason we are calling these parametric estimators unbiased is that the mean error in estimating the mode position has a value very close to zero. Lunneborg gives the bias

[3] This will fail on a Poisson distribution.

[4] Refer to Appendix B for distribution parameters.

of an estimator as

$$B(t|\theta) = \frac{1}{N}\sum_{i=1}^{N}(t_i - \theta) \tag{10}$$

where in our case t is the estimator's predicted peak position and θ is the true peak position [4]. We do not find the peak estimations to be more heavily weighted to either side of the true peak position. However this is the case only for an estimator that acts on its own distribution. For example, if we have an Weibull estimator that can approximate the parameters from a unknown Weibull distribution, then the Weibull estimator will be unbiased in estimating the peak position. However this may not be the case for the Weibull estimator on a beta distribution.

The gamma distribution is defined as

$$G(x, a, b) = \frac{1}{b^a \Gamma(a)} x^{(a-1)} \exp[-(\frac{x}{b})] \tag{11}$$

where a and b are the parameters to be estimated and are greater than zero, and $\Gamma(a)$ is the Gamma function.[5] The mean of a gamma distribution as a function of its parameters is

$$\mu = ab. \tag{12}$$

Since we are dealing with a $\mu = 1$, then $a = 1/b$, and therefore we only have one parameter left to estimate. Johnson and Kotz [1] give an unbiased estimator for the gamma distribution

$$\log(\mu/G_\mu) = \log(a) - \psi(a) \tag{13}$$

where G_μ is the geometric mean defined as

$$G_\mu = \prod_{i=1}^{N} x_i^{1/N} \tag{14}$$

[5]Refer to Appendix B.1 for definition of gamma function

and $\psi(a)$ is the digamma function defined as

$$\psi(a) = \frac{\Gamma'(a)}{\Gamma(a)}. \tag{15}$$

Since the geometric mean, G_μ, is liable to blow up on any significant data set, (13) can be written as

$$\log(\mu) - \frac{1}{N}\sum_{i=1}^{N}\log(x_i) = \log(a) - \psi(a). \tag{16}$$

In order to estimate a, we must first simplify (16):

$$\Phi - \log(a) + \psi(a) = 0 \tag{17}$$

where Φ is a constant number for each data set and is the combination of the mean and the geometric mean. From (17), the problem is reduced to a root-finding routine. In order to find the root of (17), we must first find two a values. The first a_1 must result in a negative value for the left-hand side of (17), and the second a_2 must result in a positive value. Once a_1 and a_2 have been found, the gamma estimator incorporates a root-finding routine that converges on the best a value between a_1 and a_2. After the parameter a has been determined, we set $b = 1/a$. From (11) we estimate the peak position

$$X_M = b(a-1) = 1 - b. \tag{18}$$

Fig. 10 shows a histogram of the absolute errors for the gamma estimator in estimating the peak of an almost-zero-skewness gamma distribution with $a = 100$ and $b = 0.01$. Fig. 11 shows a histogram of the absolute errors for the gamma estimator on estimating the peak of a gamma distribution that has a higher skewness with $a = 2$ and $b = 0.5$.

The Weibull distribution is defined as

$$W(x, a, c) = ca^{-1}(\frac{x}{a})^{(c-1)}\exp[-(\frac{x}{a})^2] \tag{19}$$

where a and c are the parameters to be estimated. The mean of the Weibull distribution is

$$\mu = a\Gamma(1 + c^{-1}) \tag{20}$$

where $\Gamma(c)$ is the Gamma function. Johnson and Kotz [1] give unbiased estimators for the Weibull distribution

$$a = \exp\left[\frac{1}{N}\sum_{i=1}^{N}\log(x_i) + \frac{\gamma}{c}\right] \tag{21}$$

and

$$\frac{1}{c} = \frac{\sqrt{6}}{\pi}\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\log(x_i) - \mu)^2} \tag{22}$$

where γ in (21) is Euler's constant defined by

$$\gamma = \lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{i} - \ln(n) \approx 0.57721566490153286060651209008 \tag{23}$$

and N is the number of data. As can be seen from (21), in order to estimate a we need again the $\log(G_\mu)$, where G_μ is the geometric mean. If we have the special case when $\mu = 1$, then we may use (22) to solve for c, then substitute this into the new equation, $a = c/\Gamma(1/c)$, to find a. With these two parameters we can estimate the peak position as

$$X_M = (\frac{c-1}{c})^{(1/c)}a. \tag{24}$$

The beta distribution is

$$B(x, p, q) = \frac{x^{p-1}(1-x)^{q-1}}{\beta(p, q)} \tag{25}$$

defined on the interval $[0, 1]$, where p and q are the parameters to be estimated. Since it is defined on a fixed finite interval, we cannot rescale it to unit mean as we did with Weibull

and gamma. Johnson and Kotz [1] give unbiased estimators for beta [1]

$$\psi(p) - \psi(p+q) = \frac{1}{N}\sum_j \log(x_j) \tag{26}$$

$$\psi(q) - \psi(p+q) = \frac{1}{N}\sum_j \log(1-x_j). \tag{27}$$

In order to calculate p or q we can either incorporate a two-dimensional root-finding routine or replace one of the parameters. Here we decide to replace the q parameter by using the mean for a beta distribution as a function of its parameters given as

$$\mu = \frac{p}{q+p}. \tag{28}$$

Now $q = p(1-\mu)/\mu$, and (27) can be written as

$$\psi(p) - \psi(\frac{p}{\mu}) = \frac{1}{N}\sum_j \log(x_j). \tag{29}$$

In order to estimate p, we must simplify (29)

$$\psi(p) - \psi(\frac{p}{\mu}) - \Phi = 0 \tag{30}$$

where Φ is the $\log(G_\mu)$. The problem is reduced again to a root-finding routine for p. This routine is the same procedure as the gamma root finding for p. When the p value has been estimated we can use (28) to calculate q and then estimate the peak position as

$$X_M = \frac{p-1}{q+p-2}. \tag{31}$$

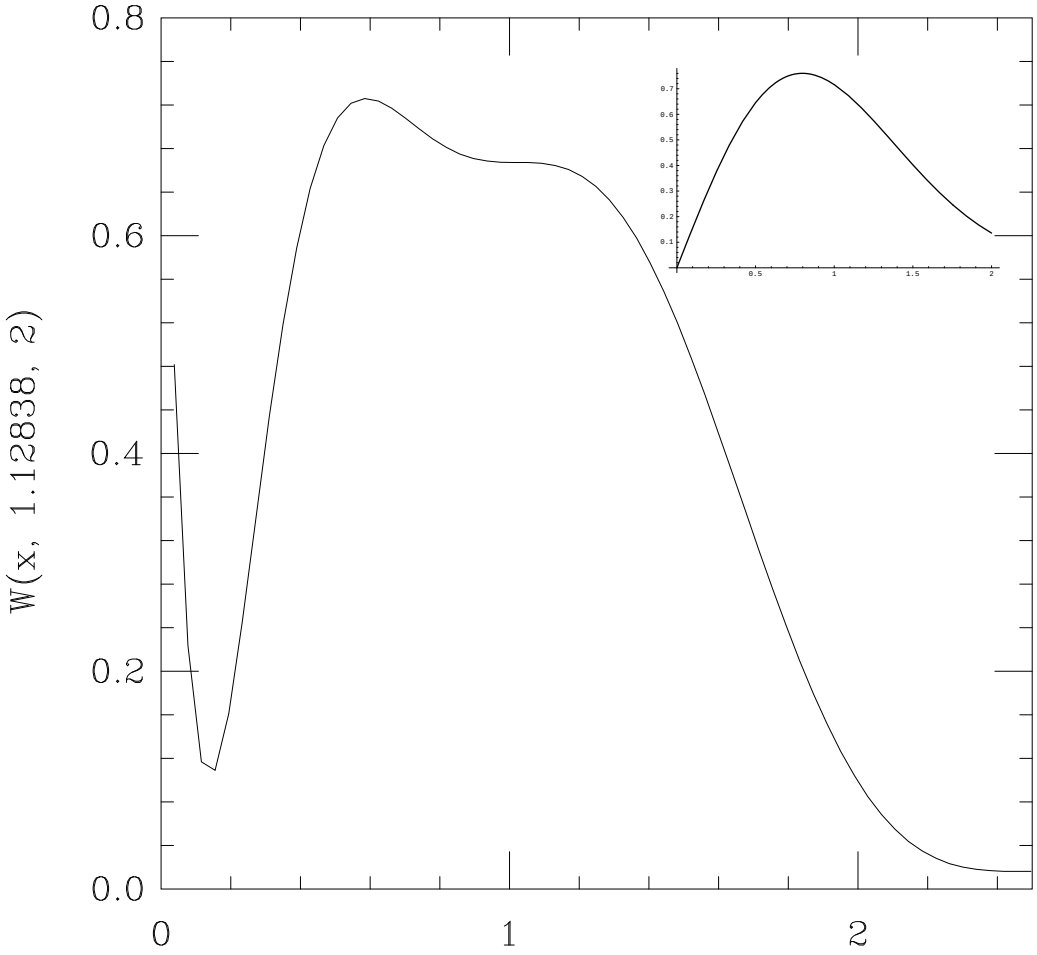

Figure 8. Chebyshev polynomial fit to a Weibull distribution with $a = 1.12838$ and $c = 2$ having 1000 deviates. The order of the fit, $n = 12$, is not optimal because of the negative slope at the beginning of the distribution. The true distribution is shown in the top right.

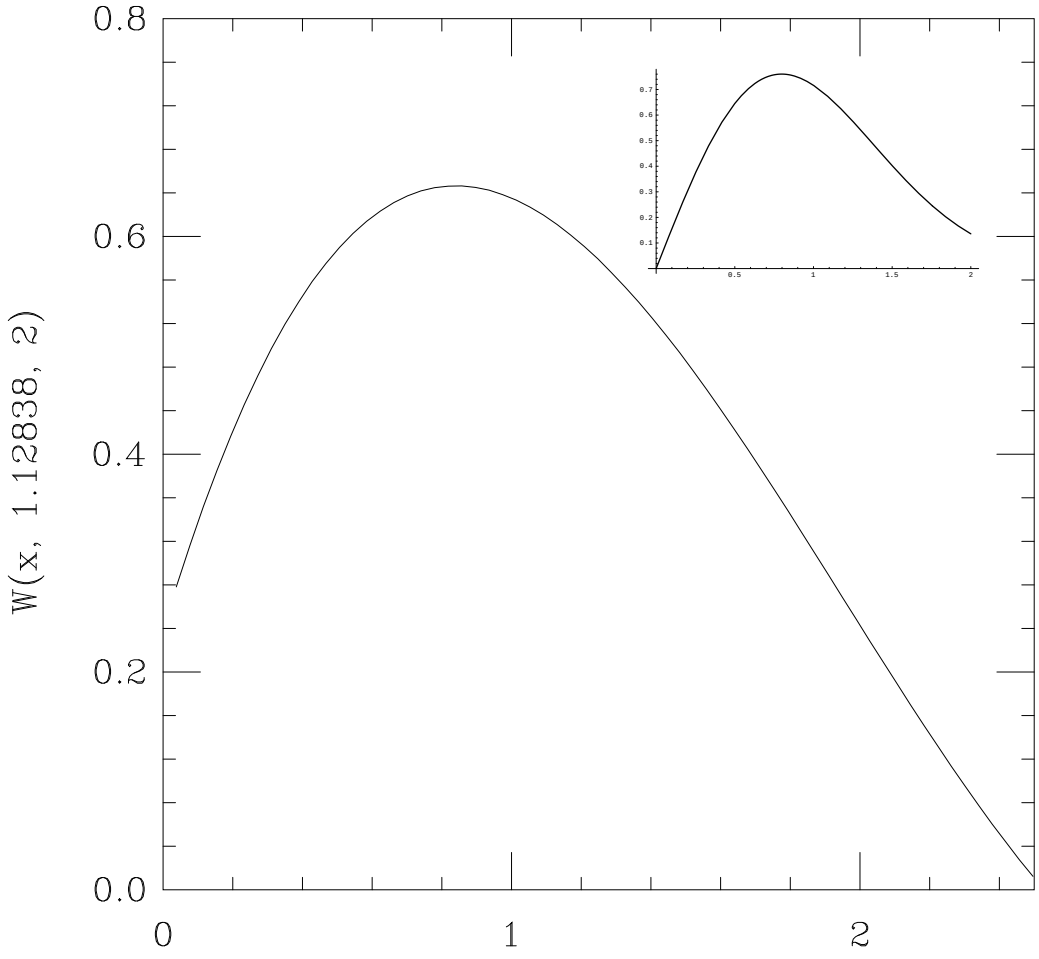

Figure 9. OCPF to a Weibull distribution with $a = 1.12838$ and $c = 2$ having 1000 deviates. Here the order of the fit is $n = 4$, and there are no negative-slope problems at the start of the distribution and no multiple peaks. The true distribution is shown in the top right.

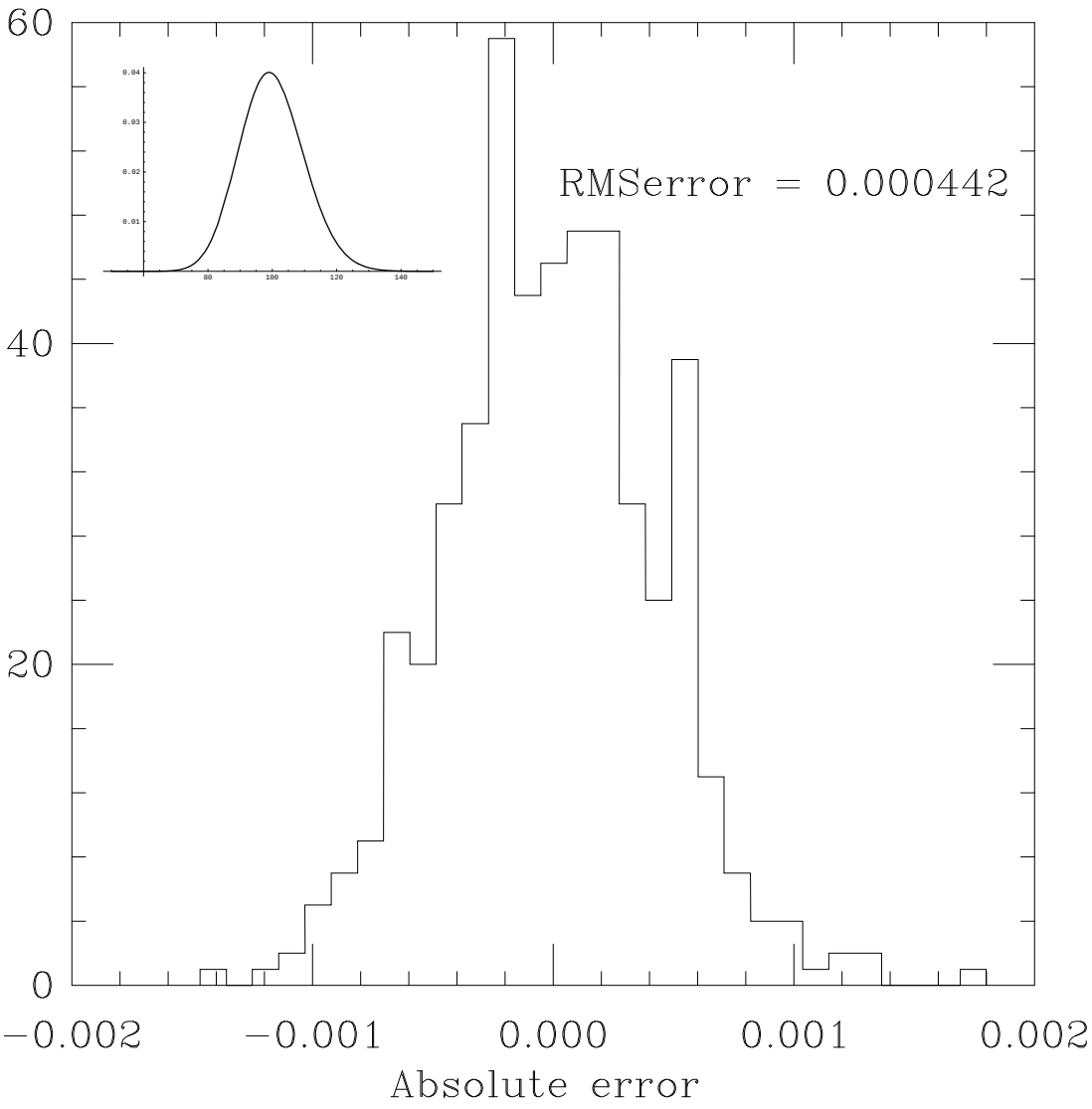

Figure 10. The histogram bin shows the absolute error for the gamma estimator on a sample of 1000 deviates from a gamma distribution with $p = 100$ and $b = 0.01$. The underlying distribution is represented in the top left corner. As can be seen, the absolute errors are centered about the 0 point on the x axis, indicating a lack of bias. The small spread about zero shows furthermore that the estimation is efficient.

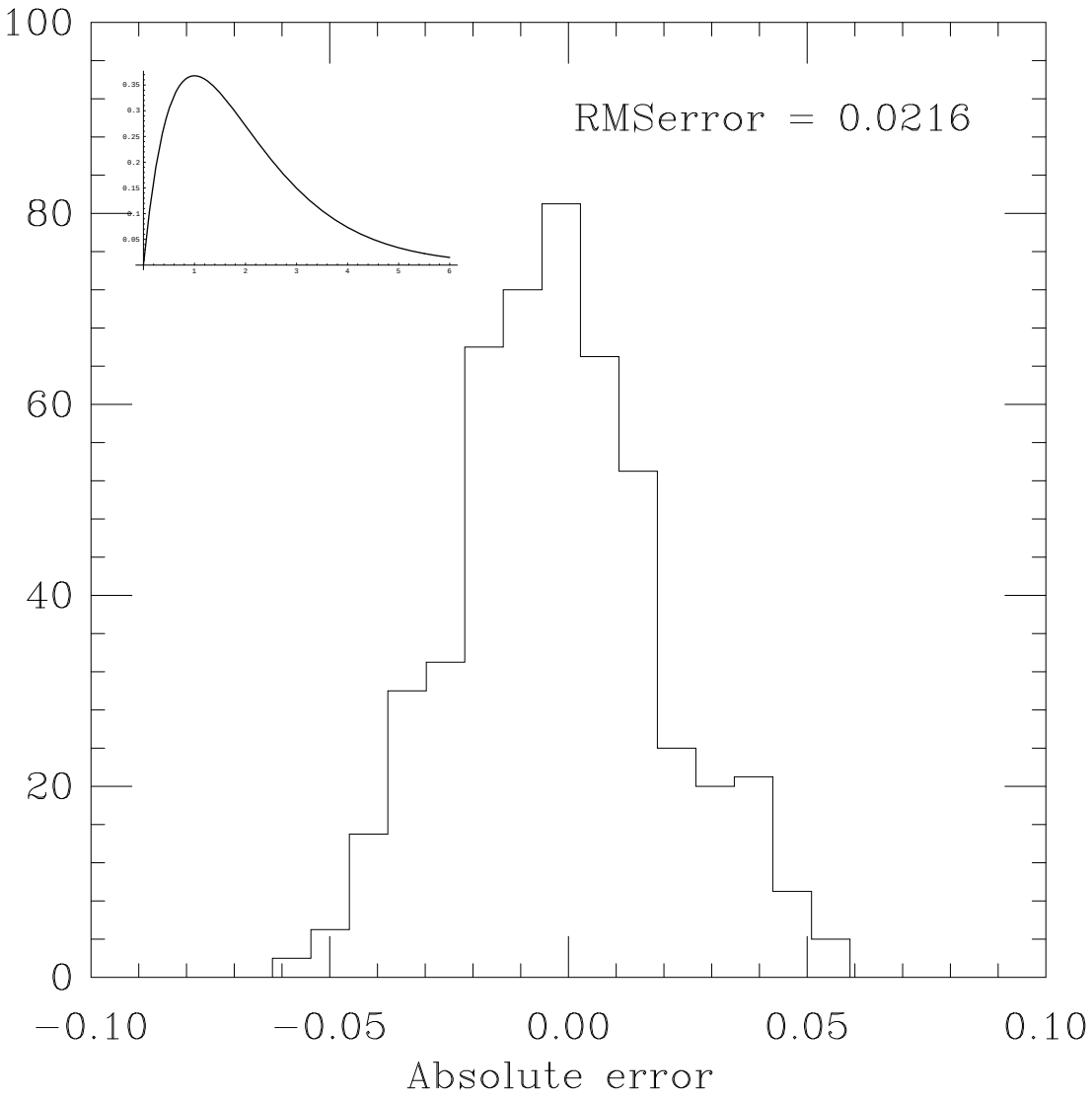

Figure 11. The histogram shows the absolute error for the gamma estimator on a sample of 1000 deviates from a gamma distribution with $p = 2$ and $b = 0.5$. While the estimator again appears (numerically) to be unbiased, it is also less efficient. The underlying distribution is represented in the top left corner. Here the underlying distribution has a higher skewness than in Fig. 10.

CHAPTER 3

SEMI-PARAMETRIC ESTIMATORS

In the previous chapter, we introduced parametric and nonparametric estimators. We will introduce two estimation algorithms that use partial knowledge of the geometry of the underlying distribution. The two estimators, `Unimodal3` and `Unimodal2`, use parametric fitting and exhaustive search routines to locate the domain around the peak of a unimodal probability distribution and then estimate the peak location. We would like to note here that all of the derivatives are estimated using finite differences as explained in appendix D.

3.1 `Unimodal3`

`Unimodal3` is the first estimator we will discuss. As mentioned before, this estimator uses a two-stage smoothing technique in order to locate the peak position of a probability distribution. The fundamental idea in this estimator is imposing unimodality. `Unimodal3` achieves this with a local constraint in its second-stage estimation. The second stage-estimation is made by fitting a curve to a domain in the raw cumulative distribution, using constrained least-squares, the constraint being that the third derivative to the fit have only negative values.[1] The first stage is only a necessary adjunct. Since fitting all of the data will most likely not result in negative values in the third derivative, we would like first to estimate the domain of the data with only negative curvature and thus containing the peak.

After we have collected the data into a raw cumulative distribution, we can apply an OCPF routine to fit the data. By observing where the third derivative to the fit of the raw

[1] Refer to appendix D for finite differences related to the 3^{rd} derivative.

cumulative distribution drops into the negative, we locate the domain in the probability distribution that has only negative curvature. Fig. 12 shows an example of a cumulative distribution and its 1^{st} and 3^{rd} derivatives.

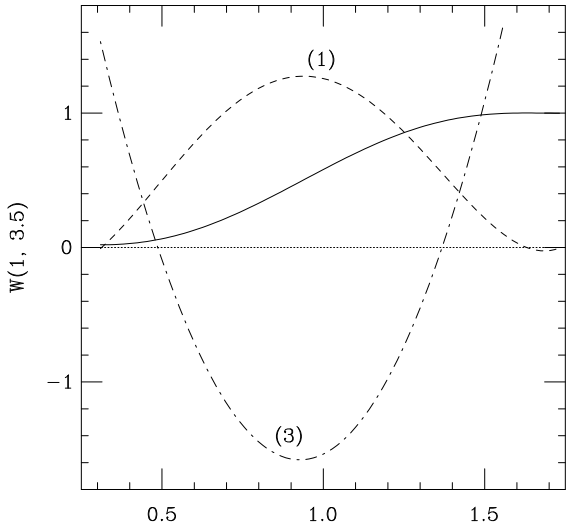

Figure 12. The cumulative distribution is represented by the solid line, the 1^{st} derivative by the dashed line, and the 3^{rd} derivative by the dot-dashed line. We need to find the domain along the x axis in which the 3^{rd} derivative drops into the negative. The plots are an ideal case, in order to illustrate the cumulative distribution and its 1^{st} and 3^{rd} derivatives. The 1^{rd} and 3^{st} derivatives have been plotted in arbitrary units.

We are looking for the turning points which bracket the mode of the probability distribution. At the turning points to the probability distribution, the third derivative of the fit to the cumulative crosses the x axis. Fig. 13 shows the output of the first-stage estimation on the raw cumulative distribution. We will call the output of the first-stage estimator for `Unimodal3` the filtered raw cumulative distribution.

The second stage-estimator of `Unimodal3` makes use of the well known method of constrained least-squares fitting. Here `Unimodal3` performs a constrained least-squares fit to the filtered raw cumulative distribution, where the 3^{rd} derivative is constrianed to possess only negative values. Fig. 14 shows the least-squares fit to the filtered raw cumulative distribution and the derivative of that fit. As can be noticed in Fig. 14, `Unimodal3` is not useful in determining the shape of the probability distribution about

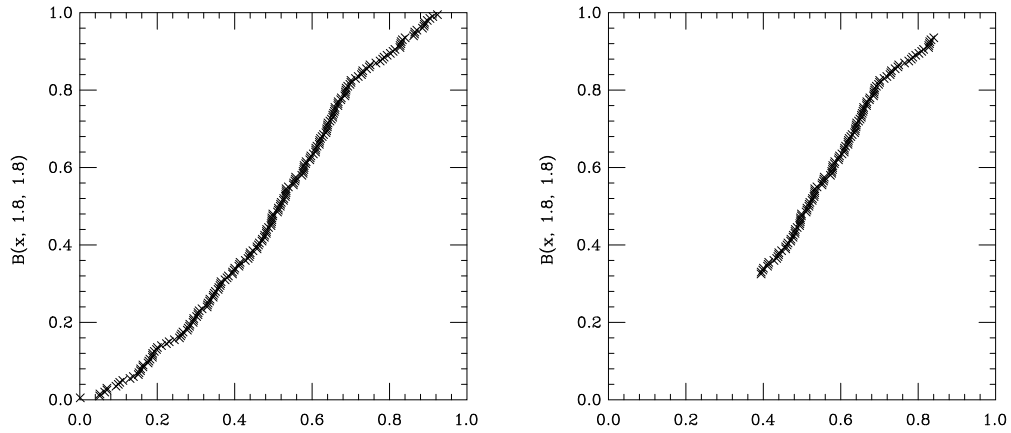

Figure 13. The graph to the left is the raw cumulative distribution taken from a beta distribution, $p = 1.8$ and $q = 1.8$, having 1000 deviates. The graph to the right is the raw cumulative distribution after the first-stage estimator has estimated the range of negative curvature in the probability distribution. The true mode position for this distribution is at $x = .5$. The mode position falls within the filtered domain.

the peak. This is because the derivative of the constrained least-squares fit empirically is piecewise-linear. Increasing the size of the data set being used does little to improve the smoothness of the output of `Unimodal3`.

3.2 `Unimodal2`

Similar to `Unimodal3`, `Unimodal2` can also use a two-stage smoothing technique to determine the position of the peak for an unknown distribution. The second stage of the estimator performs an exhaustive or a binary search for the peak position. The second-stage estimation assumes a peak position, then fits the raw cumulative data subject to the constraint that the second derivative be positive to the left and negative to the right of the assumed peak position. The peak position resulting in the smallest residual is the estimated mode. Fig. 15 shows an example of a cumulative distribution and its 1^{st} and 2^{nd} derivatives. If we choose to use a first-stage estimation, it is to decrease the runtime of the second stage.

There are three different methods we can use to search for the peak position, the choice entailing a trade-off between accuracy and runtime. We can perform an exhaustive search

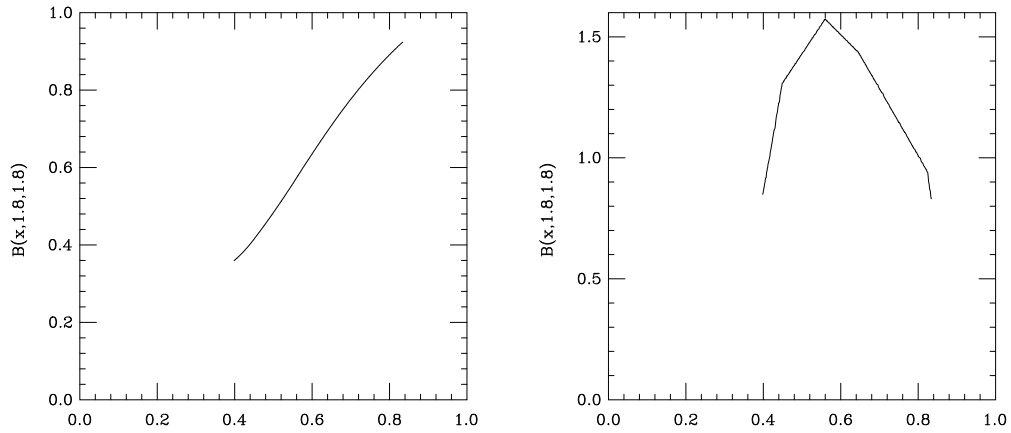

Figure 14. The graph to the left is the constrained least-squares fit to the filtered raw cumulative distribution of Fig.13. The graph to the right is the derivative of the least-squares fit. The mode position for this distribution is at $x = .5$, and `Unimodal3` estimated the peak to be at $x = 0.56034$.

for the peak position, which tests every datum as a peak estimation, or we can perform a binary search, which starts at both ends of the domain and converges on the peak estimation. The binary is more time efficient than the exhaustive search; however it could converge on a local (rather than global) minimum in the objective function. If we choose not to decimate the data (first-stage estimation), exhaustive search takes prohibitively long. The binary search on undecimated data for a range of number of deviates from 400–1000 averages 14 seconds per realization. For 500 realizations, it would take 14 hours to run a full set of samples.[2] The exhaustive search on decimated data averaging 300 deviates averages 4 seconds per realization. For 500 realizations this would only take 4 hours to run on a full set of samples. Figs. 17 and 18 show that for 200-deviate samples, the RMS errors are indistinguishable (within error bars) between the binary search on undecimated data and the exhaustive search on decimated data. Therefore using binary search on decimated data proves to be the most efficient method for runtime with no trade-off in accuracy. Fig. 19 shows that we can improve the runtime efficiency without a trade-off in accuracy by using a binary search on the decimated data. Using a binary search on decimated data proves to be more time efficient than exhaustive on decimated.

[2]This runtime is on a Pentium-IV processor 2.4GHz.

In theory binary can never be more efficient than exhaustive. However Fig. 19 shows slightly smaller errors for the binary search. This may be due to the bias of `Unimodal2`. None the less,to reduce the chance of estimations falling into local minima, we choose to run an exhaustive search on decimated data.

The first-stage estimation for `Unimodal2` decimates the number of data points, using a distance scale to compare each datum to its neighbors. For example, when the measurements are grouped closely together, it is likely that this range of grouping can hold a possible peak. Because we are dealing with unimodal distributions, there should be only one major domain of grouped measurements. Here we set a length scale that is a fraction of the average difference between measurements. `Unimodal2` compares every point with its neighbors using this length scale as a limit. If a datum falls within the limit of its neighbors, then that datum is saved. However if that datum is not within the limit of its neighbors, then that datum is removed. This process removes data that fall out on the tails of a distribution, where a peak is less likely. Fig. 16 shows examples before and after the decimation.

As can be seen in Fig. 16, some of the data in the middle of the distribution have been filtered out. At first, this appears to be a problem. We decided to compare `Unimodal2`'s first-stage estimator, which removes data from the edges and the middle, against `Unimodal3`'s first-stage estimator, which removes data from only the edges, and see how the exhaustive search for `Unimodal2` compares. We produced two sets of ninety realizations having numbers of deviates ranging from 400 to 1000 and ran the first-stage estimations for `Unimodal2` and `Unimodal3` on each set. The size of the output files from `Unimodal3`'s first stage ranged from 50 to 800 deviates, while the average size of `Unimodal2`'s first-stage output was 250 deviates.

Therefore it generally takes longer to run `Unimodal2`'s exhaustive search on the output of `Unimodal3`'s first stage than on `Unimodal2`'s first stage. We find that the time it takes to run `Unimodal2`'s exhaustive search on a data set with X number of deviates scales as $X^{3.3}$. So for a data set with 1000 deviates, the run takes 20 times as long as a data set

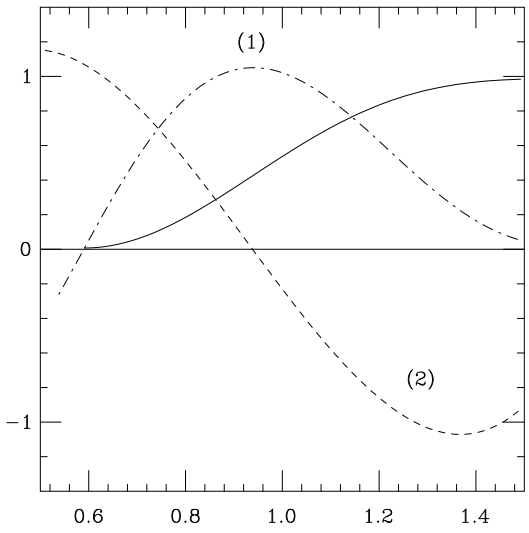

Figure 15. The smoothed cumulative distribution is represented by the solid line, the 1^{st} derivative estimate by the dot-dashed line, and the 2^{nd} derivative by the dashed line. The plots are an ideal case in order to illustrate the point where the 2^{nd} derivative crosses the x axis. This point represents the peak position in the 1^{st} derivative. The 1^{st} and 2^{nd} derivatives are plotted in arbitrary units.

with 400 deviates. However, the overall efficiency in calculating the peak position does not differ by much when we use `Unimodal2`'s or `Unimodal3`'s first-stage estimation. Therefore to run in a reasonable time, we use Unimodal2's first-stage estimator.

After we have assumed a mode position and estimated the cumulative probability distribution subject to the second-derivative constraint, we compare this estimation to the raw cumulative distribution and find the sum of square differences. When we have a peak candidate that produces the smallest sum of square differences, we have found the best estimate for the peak position.

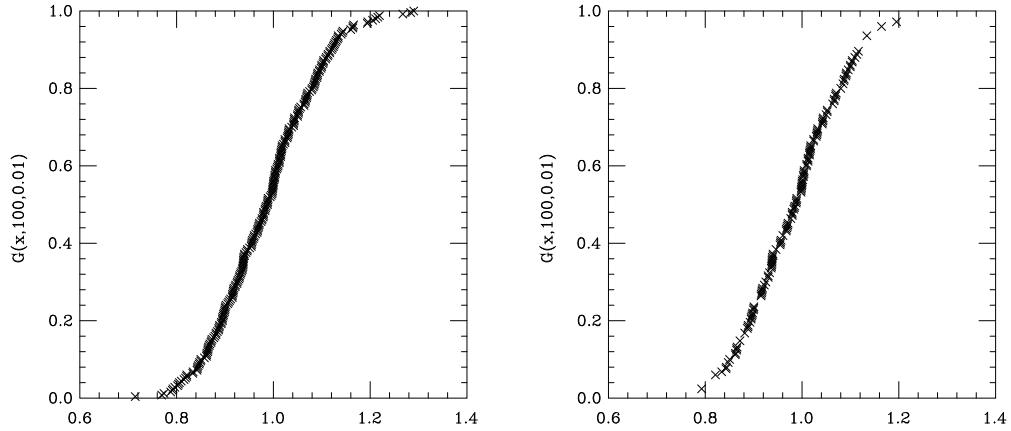

Figure 16. The graph to the left shows the raw cumulative distribution for a gamma distribution with $p = 100$ and $b = 0.01$ taken from a sample of 1000 deviates. The graph to the right is after Unimodal2's first-stage estimation having 311 deviates. As can be seen, the tails of the distribution have been thinned due to the unlikelihood of their containing a peak candidate.

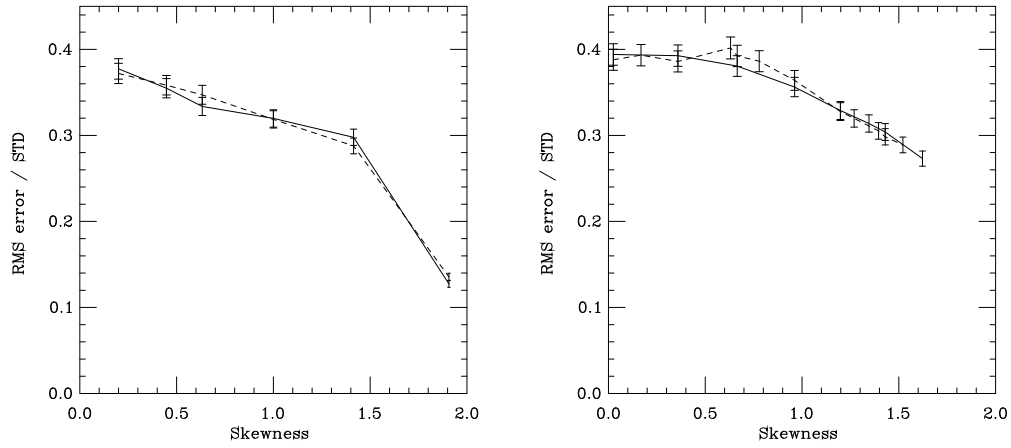

Figure 17. The dashed lines are `Unimodal2`'s exhaustive search on decimated data, and the solid lines are the binary search on undecimated data. The graph to the left is of 200-deviate samples drawn from a family of gamma distributions. There are 500 realizations per point. The graph to the right is of 200-deviate samples drawn from a family of Weibull distributions. There are 500 realizations per point.

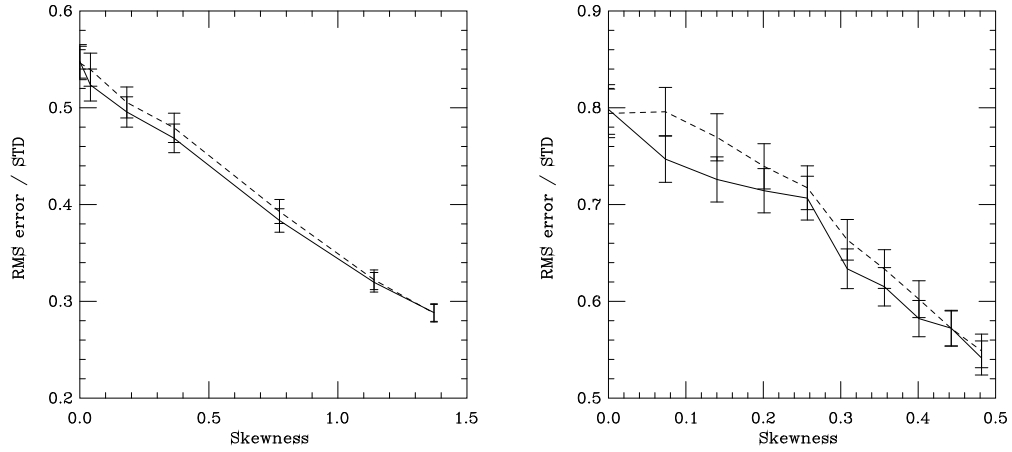

Figure 18. The dashed lines are `Unimodal2`'s exhaustive search on decimated data, and the solid lines are the binary search on undecimated data. The graph to the left is of 200-deviate samples drawn from a family of sharp-peaked beta distributions. There are 500 realizations per point. The graph to the right is of 200-deviate samples drawn from a family of flat-peaked beta distributions. There are 500 realizations per point.

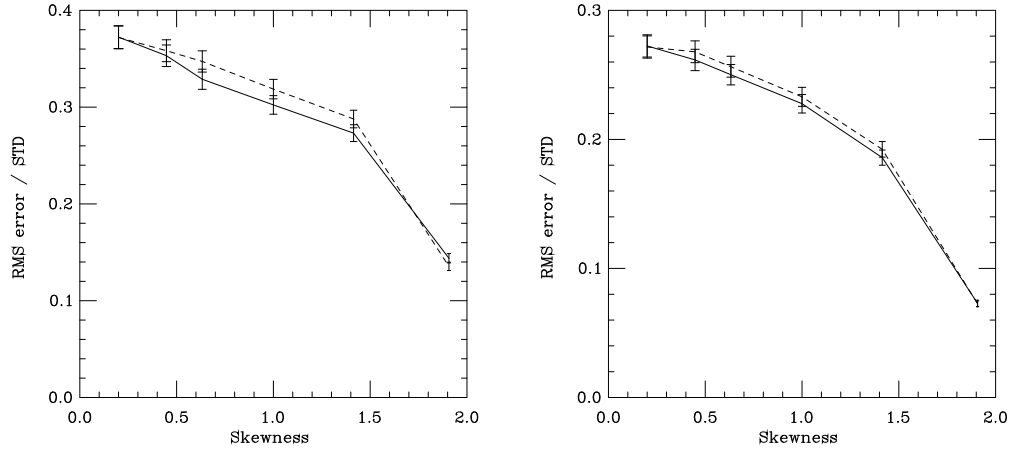

Figure 19. The dashed lines are `Unimodal2`'s exhaustive search on decimated data, and the solid lines are the binary search on decimated data. The graph to the left is of 200-deviate samples drawn from a family of gamma distributions. There are 500 realizations per point. The graph to the right is of 1000-deviate samples drawn from a family of gamma distributions. There are 500 realizations per point.

CHAPTER 4

MODE-ESTIMATION EFFICIENCY

In this chapter we will discuss how we determine the relative efficiency of our estimators in locating the true peak position of an underlying probability distribution.

4.1 Setting a Scale

In this section we will explain why we decide to use the standard deviation as a comparison scale. It will be shown that a set of realizations that have broad flat peaks can produce higher errors than realizations with sharper, pointier peaks. Therefore, we can use the standard deviation of the population as a scale for comparing the estimators' efficiencies. For example, if we have two differently shaped distributions, as in fig.20, without using the standard deviations of the distributions as a scale, all estimators will perform better on the sharp peak than on the broad flat peak. However, if we divide the error by the standard deviation of the population, we can better compare the differences in the estimations for each peak. Therefore we set the comparison function of how well the estimator estimates the true peak position for several realizations to be

$$\theta(\alpha) = \frac{RMS_\alpha}{STD_\alpha}, \tag{32}$$

where α characterizes the distribution, RMS_α is the root mean square error, and STD_α is the standard deviation of α. We would expect the RMS_α and STD_α to decrease with a sharper skinner peak and increase with a flat broader one.

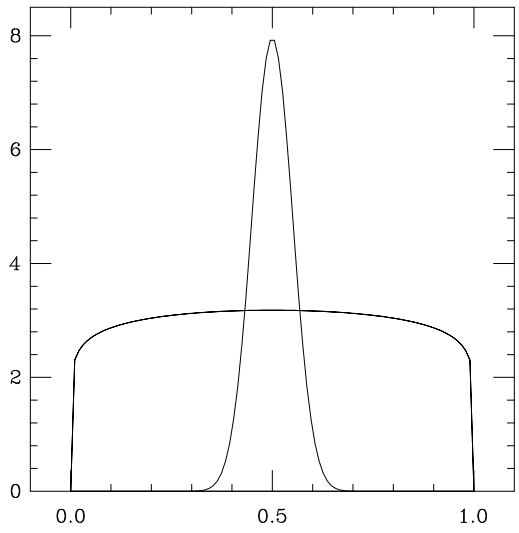

Figure 20. Without using the standard deviation of the population as a comparison scale, all estimators would perform well on this sharp peak and poorly on the flat peak.

4.2 Mirror Distributions

In this section we will explain why we choose to divide the RMS by STD. Even though we are interested in how well our hybrid estimators estimate the true peak position of a highly skewed distribution, we must consider data sets that have negative and positive skewnesses on the same comparison scale.[1] That is, we cannot just compare the estimators' efficiency in locating peak positions that are only close to the y axis. For example, a distribution with a skewness value of S can have a mirror counterpart with skewness $-S$. Fig. 21 shows a beta distribution with skewness values that are absolutely equal but have different signs. These distributions also have identical kurtosis and variance values. The only properties that will differ for these distributions are the mean and the signs of the skewness and higher odd-order cumulants.[2] Our estimators should produce similar errors for a distribution that has $-S$ or S. Therefore we cannot consider the percentage error in estimating a peak position as a way of comparing the efficiencies of our estimators. For example if we have two distributions as in Fig. 21, and we have the same absolute error estimation on both,

[1]Positive skewness is described by the peak position being shifted asymmetrically toward the y axis.
[2]Refer to appendix A for definition of cumulants.

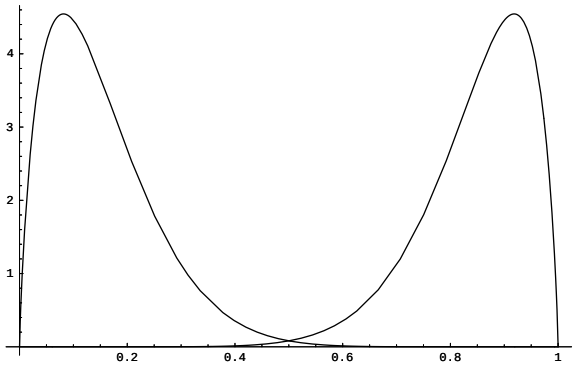

Figure 21. This figure shows a mirror image of a beta distribution centered about $x = .5$. The distribution to the left has parameters $p = 1.8$ and $q = 10$. The distribution to the right has $p = 10$ and $q = 1.8$.

we cannot divide the errors by the peak positions and still have the same fractional error. Because of this, we must find another way of determining the estimators' efficiencies. Since the two distributions in the example have the same standard deviation, we can divide the absolute error by the standard deviation and produce the same fractional result. Therefore our estimators will produce the same errors for mirror distributions.

4.3 STD of the STD as an Error assessment

Because we are going to be plotting the RMS errors over the STD, we need to know how significant our comparisons are. If we assume that the distribution of mode estimations over many realizations is Gaussian about the true mode position, we must assess the standard deviation of the standard deviation of the mode estimate. Keeping gives the

variance of the variance estimate as

$$\sigma_{k_2}^2 = \frac{2}{N+1} k_2^2 \tag{33}$$

where k_2 is the variance of an arbitrary distribution [2]. For our purpose we are going to call k_2 the mean square error given as

$$k_2 = \frac{1}{N} \sum_i (x_i - \theta)^2, \tag{34}$$

where x_i is the mode estimation for the i^{th} realization and θ is the true mode position. If we wish to find the standard deviation σ_{k_2} we have

$$\sigma_{k_2} = k_2 \sqrt{\frac{2}{N+1}} \tag{35}$$

which gives us the standard deviation as a function k_2. However we are interested in $\sigma_{\sqrt{k_2}}$, which is the standard deviation of the standard deviation of the mode estimates. Therefore we can use the relation given by Taylor for uncertainty in power,

$$\frac{\delta q}{|q|} = |n| \frac{\delta x}{|x|}, \tag{36}$$

where x is measured with uncertainty δx and is used to calculate the power $q = x^n$ [11]. The fractional uncertainty in q is $|n|$ times that in x. If we make the relations to (36), δq is $\sigma_{\sqrt{k_2}}$, q is $\sqrt{k_2}$, n is $\frac{1}{2}$, δx is σ_{k_2} and x is k_2. Solving (36) for δq and substituting the previous relations gives

$$\sigma_{\sqrt{k_2}} = \sqrt{\frac{k_2}{2(N+1)}}. \tag{37}$$

Now we have an expression for the standard deviation of the standard deviation of the mode estimate where $\sqrt{k_2}$ is the RMS error of the mode estimate for N realizations.

4.4 Choice of α

The final decision to be made on how to compare the relative efficiencies of our estimators is what should the set α be in (32)? What property of the distribution should we compare against? In the quantum-wire level-spacing distributions, we are interested in probability distributions that lie between the Poisson and Wigner-Dyson distributions. Some may have high skewness. Therefore one of the α sets we will use is the skewness of the underlying distribution. Since we know nothing outside of unimodality in these distributions, we can also use the kurtosis of the underlying distributions as a second α set.

CHAPTER 5

RESULTS FOR DISTRIBUTIONS HAVING UNIT MEAN

In this chapter we will show the results of peak location for each estimator on three different family of distributions. Our data sets are produced from the Weibull, beta, and gamma distributions. We varied each of the distributions' parameters in order to achieve a range of skewness and kurtosis values, thereby changing the shapes of the distributions. In order to observe the statistics of these estimators, we ran a series of tests on data sets each containing 500 realizations. Each realization within a set contained the same number of deviates, and for different sets, the number of deviates ranged from 100 to 1000. We will show how each estimator compared in relative efficiency in locating the peak positions of data sets with 1000 and 200 deviates by plotting the mean of the fractional error, (32), over all of the realizations, versus the skewness of the distributions.

Because any data set on the semi-infinite interval $[0, \infty)$ can be scaled so that the mean is equal to one, we have prepared data sets from Weibull and gamma distributions where the mean is equal to one. We compare the relative efficiencies of eight different estimators on locating the peak position of these distributions. These eight estimators are kernel estimation (FKS), optimal kernel smoothing (OKS), adaptive histogram (AHIST), optimal Chebyshev polynomial (OCPF), Weibull, gamma, `Unimodal2`, and the `Unimodal3` estimator.

We compare the fractional error of (32) against the skewness of the distributions for a set of gamma and Weibull distributions listed in Tables 1 and 2. A flatter peak has a negative kurtosis and a sharper peak a positive kurtosis. As to be expected, if we are given more random deviates in a data set, the mode estimations have a lower fractional error. The more information we have, the better the estimation. Figs. 22, 23, 24, and

	a	c	$Skewness$	$Kurtosis$
Weibull	1.11142	3.5	0.0249969	-0.286522
	1.11985	3	0.168098	-0.270617
	1.12706	2.5	0.358619	-0.143214
	1.12838	2	0.631109	0.245085
	1.12777	1.95	0.665277	0.311061
	1.1245	1.8	0.778745	0.556757
	1.11536	1.6	0.961966	1.04394
	1.09719	1.4	1.19847	1.83907
	1.09053	1.35	1.26919	2.11427
	1.08275	1.3	1.34593	2.4322
	1.07747	1.27	1.39518	2.6475
	1.07367	1.25	1.42955	2.80217
	1.06309	1.2	1.5211	3.2356
	1.05075	1.15	1.62209	3.74807
	1.03636	1.1	1.73397	4.35981

Table 1. Fifteen different Weibull distribution, with unit mean. $W(x,a,c) = ca^{-1}(\frac{x}{a})^{(c-1)}\exp[-(\frac{x}{a})^2]$.

	p	b	$Skewness$	$Kurtosis$
Gamma	100	0.01	0.2	0.06
	20	0.05	0.447214	0.3
	10	0.1	0.632456	0.6
	4	0.25	1	1.5
	2	0.5	1.41421	3
	1.1	0.909091	1.90693	5.45455

Table 2. Six different gamma distributions with unit mean. $G(x,a,b) = \frac{1}{b^a\Gamma(a)}x^{(a-1)}\exp[-(\frac{x}{b})]$.

25 show the averaged results over 500 realizations of all of the estimators for gamma and Weibull data sets containing samples of 1000 and 200 deviates.

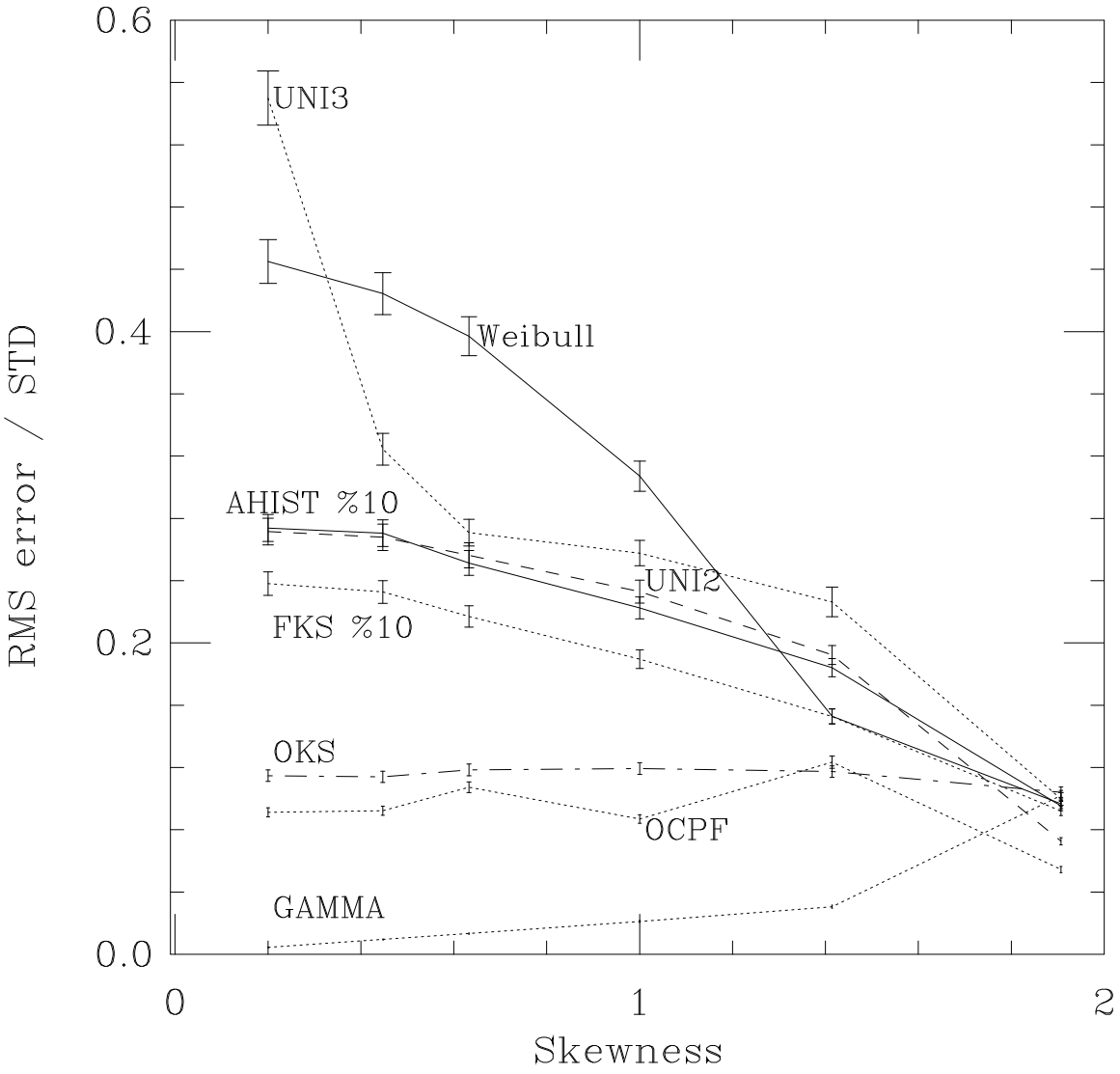

Figure 22. Results from a family of gamma distributions for realizations drawing 1000 deviates each. See also Table 2. For UNI2 we have used the exhaustive search method for decimated data.

5.1 Weibull and Gamma Estimators

In both the 1000- and 200-deviate samples, the gamma and Weibull estimators produced the smallest fractional errors over all ranges of skewness on their respective distributions. This is to be expected since these are parametric unbiased estimators of their distributions. However, for a small range of skewness values in the Weibull distributions, the gamma estimator produced lower fractional errors than Weibull. This was not expected. If we observe the Weibull distributions at these skewness values and then observe the gamma distributions with the parameters estimated from the gamma estimator on these Weibull distributions, we can see that the gamma and Weibull distributions are similar. This would explain why the gamma estimator does so well for the Weibull distributions in this region of skewness. Fig. 26 shows an example.

5.2 OKS and OCPF

OKS and OCPF had the next-lowest fractional errors over all skewness values. However in the 1000-deviate case, OKS's fractional error differed over all the skewness range by only four times the average[1] $\sigma_{\sqrt{k_2}}$, while OCPF's fractional error differed by eighteen times its average $\sigma_{\sqrt{k_2}}$. As will be shown in the next chapter, OCPF suffers when estimating the peaks of distributions that have negative kurtosis.[2] However, for the positive kurtosis values of the gamma and Weibull distributions, listed in Table 1, OCPF is flat within $.02\sigma_{\sqrt{k_2}}$.

In the 200-deviate case, OKS suffers at high skewness values. The explanation is simple. When OKS tries to estimate a data set that does not have enough information, the optimal routine must keep increasing the size of the kernel parameter to overcome undersmoothing. This is often the case with a small number of deviates. If the kernel parameter is set too high, oversmoothing is the result, and therefore there is a shift of the smoothed peak position relative to the true peak position. This shift does not affect

[1] $\sigma_{\sqrt{k_2}}$ is represented by the error bars in figs. 22, 23, 24, 25; refer to (37) for definition.
[2] Negative kurtosis is represented by a flatter peak.

distributions that are near zero skewness as much as those that are highly skewed. Fig. 27 shows an example.

5.3 FKS and AHIST

Kernel estimation (FKS)[3] and AHIST estimation follow similar fractional error patterns. Both of these are nonparametric estimators, and from several trials we can choose the best parameter value for each. With FKS, 10 percent of the standard deviation of the sample appears to be a good over-all parameter. This percentage of the standard deviation of the sample yielded the lowest fractional error over all ranges of skewness. With the AHIST method, 10 percent of the population of deviates was a good over-all parameter value. This is a little more intuitive considering that the parameter value represents the percentage of data in each bin. Too great or too small of a percentage can result in under or oversmoothing. Fig. 28 shows these estimates.

As to be expected, when there is a smaller number of deviates in each realization, the fractional error increases for each estimator. We see the same error pattern in the 1000-deviate case as we do in the 200-deviate case, except the error has increased when there are only 200 deviates in each realization.

5.4 Unimodal3 and Unimodal2

Our two estimators, `Unimodal2` and `Unimodal3`, showed similar fractional-error problems in both the Weibull and gamma distributions.

We encountered several problems with the `Unimodal3` estimator. The first problem is that Unimodal3 shows biased results for the mode estimates. Fig. 29 shows an example of fourteen `Unimodal3` outputs on a gamma distribution with $p = 100$ and $b = 0.01$. As can be seen from this figure, `Unimodal3` shows a positive bias, meaning that the majority of the peak estimates are to the right of the true mode position. We can observe this positive bias over all 500 realizations through a histogram plot of the absolute error in Fig. 30

[3]Fixed-width kernel smoothing (FKS), from chapter 2.2 .

The second problem with `Unimodal3` is shown in the estimations on a gamma distribution with low skewness and large number of deviates. Fig. 31 shows us a RMS error plot versus number of deviates. We would expect the RMS error to be high for a small number of deviates in each realization and decrease as the number of deviates increase. However we do not observe this effect. Here we notice a dip in the RMS error at 200 deviates and then a rise in the RMS error as the number of deviates increase. This is not what we would expect, since more information should lead to a better approximation. If we produce another 500 realizations with the same distribution parameters shown also in Fig. 31, we see the same systematic results. This problem could be the result of the positive bias of `Unimodal3` on the gamma distribution.

We do not notice the same systematic effect, in Fig. 31, when `Unimodal3` is estimating a Weibull distribution. However we do notice a bias in estimating the mode for distributions with small skewness. Fig. 32 shows a histogram plot of the absolute error over 500 realizations for a Weibull distribution having $a = 1.11142$ and $c = 3.5$ and how this bias varies with skewness. For greater skewness, the bias is much lower than for the near-zero-skewness distributions. We can see from Fig. 32 that the magnitude of this bias is less for the Weibull distribution. However, the fractional error for skewness values ranging between $0.5 \leq \gamma_1 \leq 2$ are similar on both distributions in the 1000- and 200-deviate cases. Again this is most likely due to the similarities in the shapes of the distributions within this skewness range.

In Fig. 33, `Unimodal2` shows small changes in bias over a range of skewness for both gamma and Weibull families. However these changes are small. `Unimodal2` also shows little to no bias for distributions having near-zero skewness. This is different from `Unimodal3` in Figs. 32 and 30. Figs. 34 and 35 show histograms of the bias of `Unimodal2` on extremes in skewness for the gamma and Weibull families.

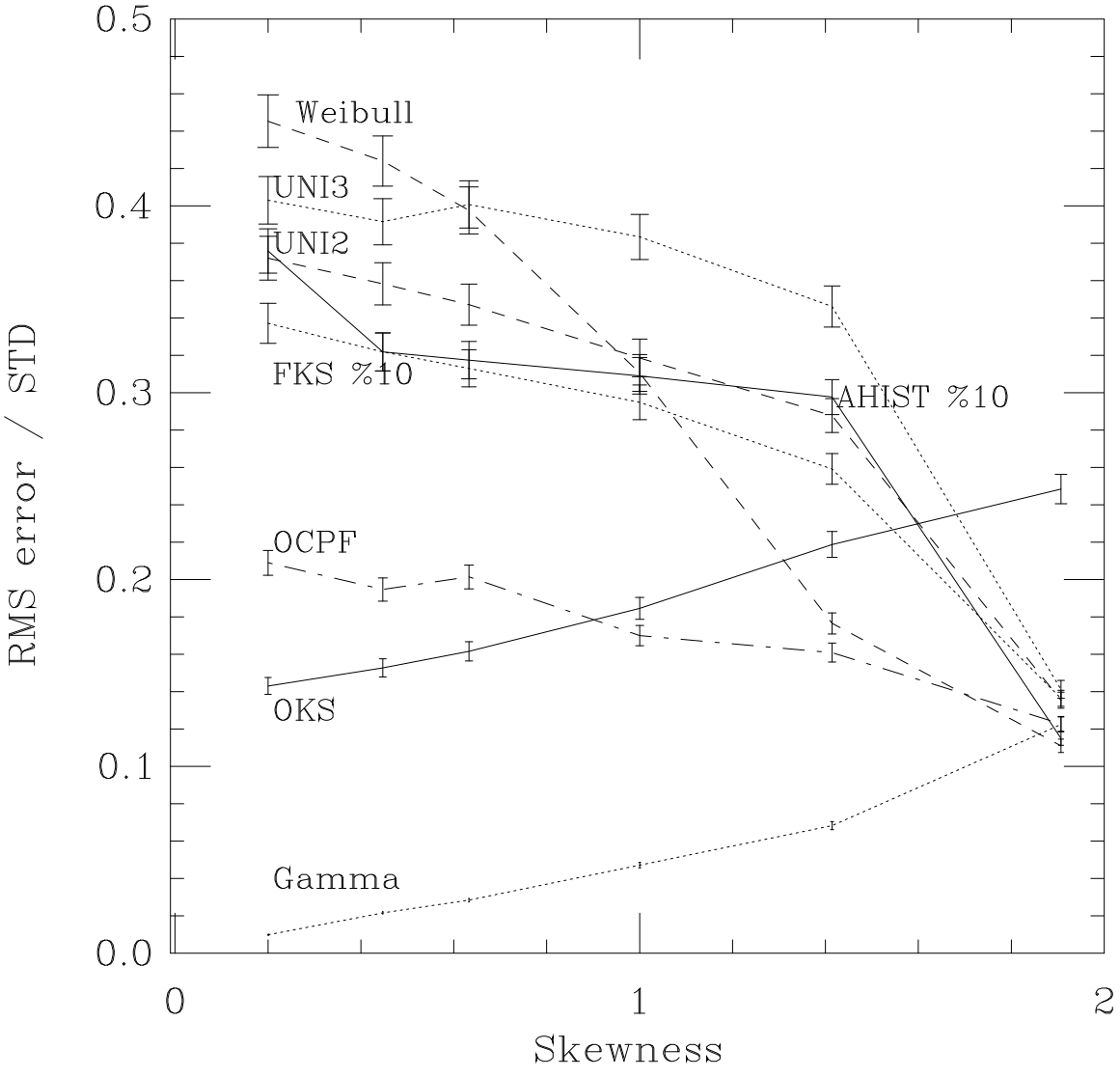

Figure 23. Results from a family of gamma distributions for realizations drawing 200 deviates. See also Table 2. For UNI2 we have used the exhaustive search method for decimated data.

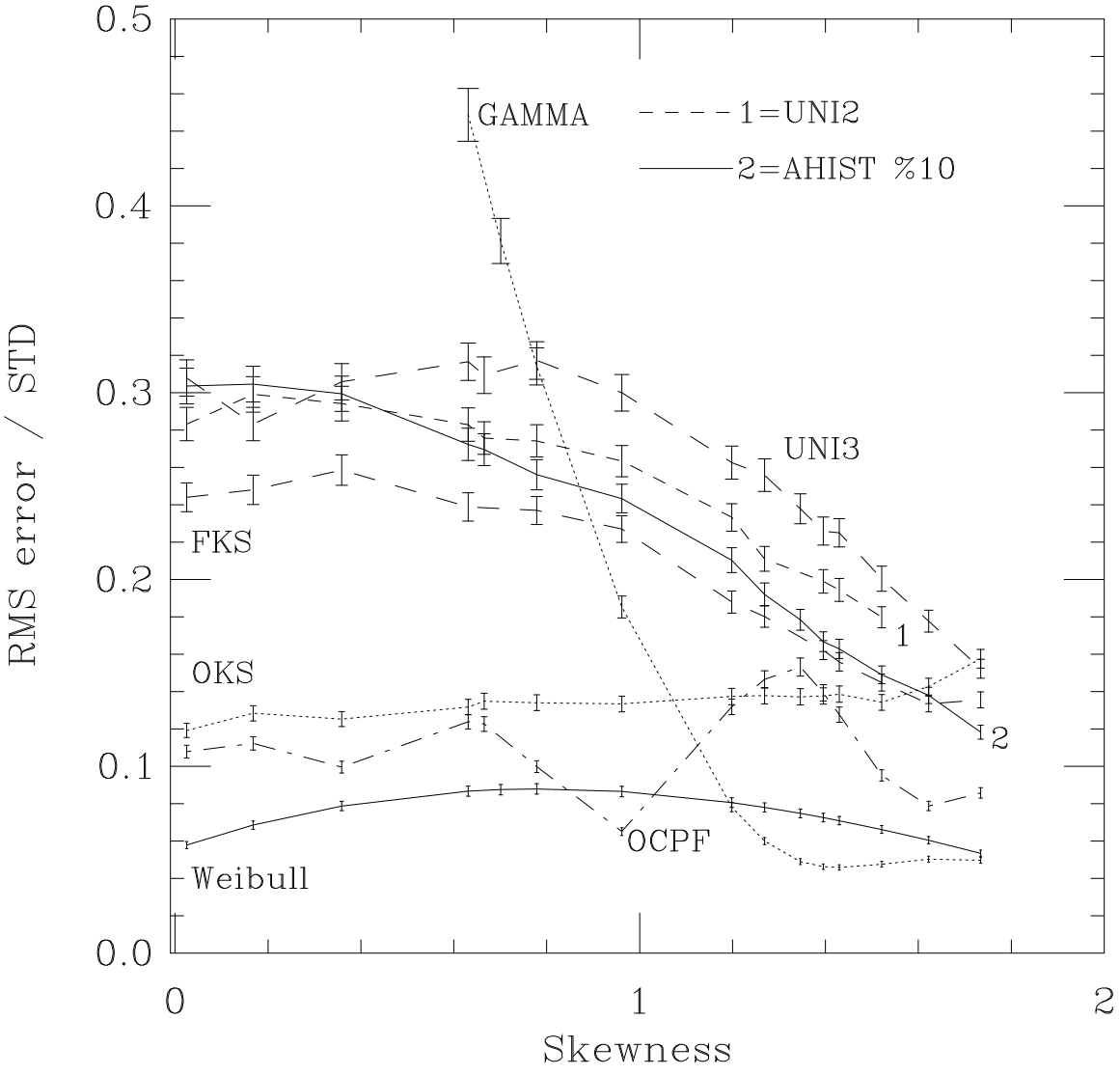

Figure 24. Results from a family of Weibull distributions for realizations drawing 1000 deviates each. See also Table 1. For UNI2 we have used the exhaustive search method for decimated data.

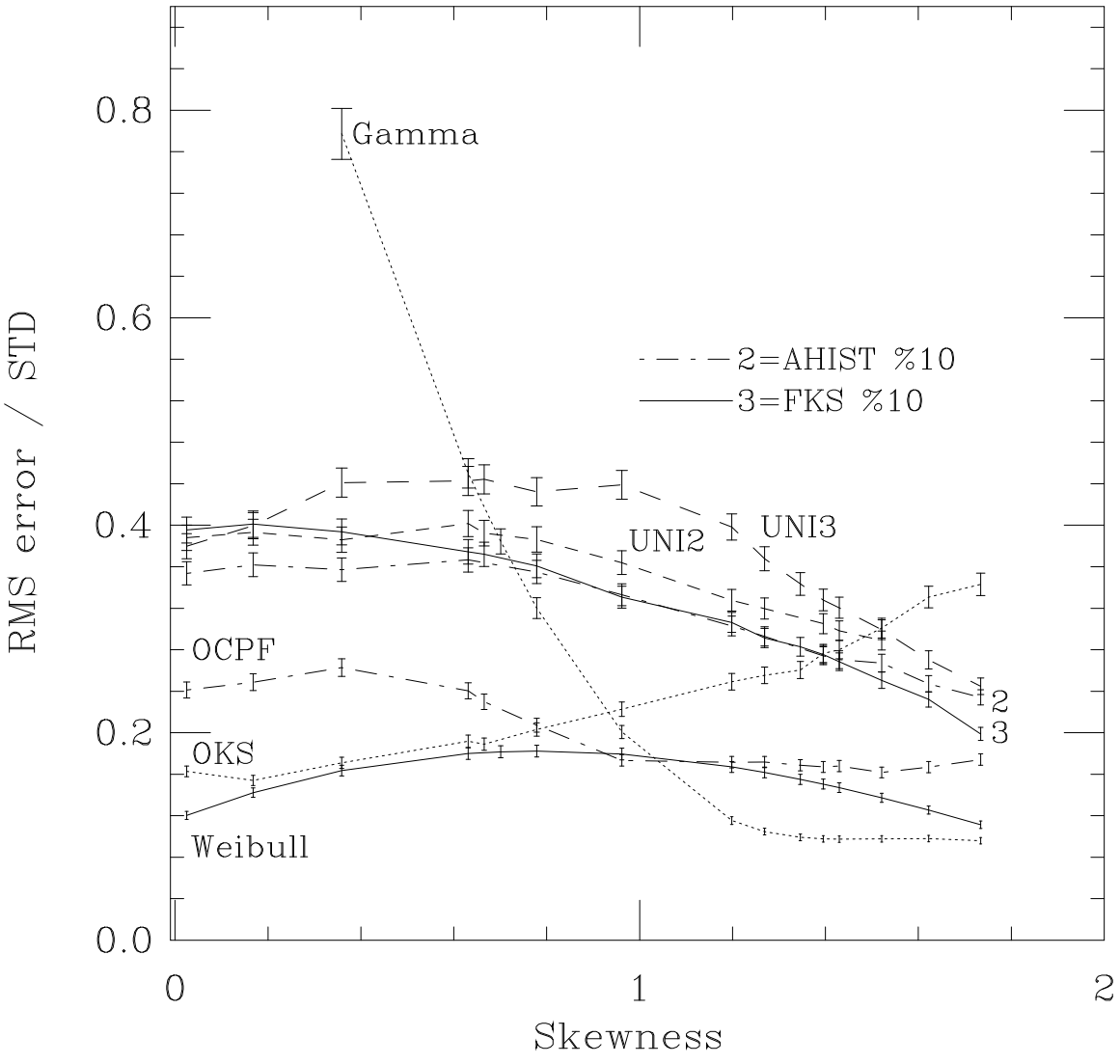

Figure 25. Results from a family of Weibull distributions for realizations drawing 200 deviates each. See also Table 1. For UNI2 we have used the exhaustive search method for decimated data.

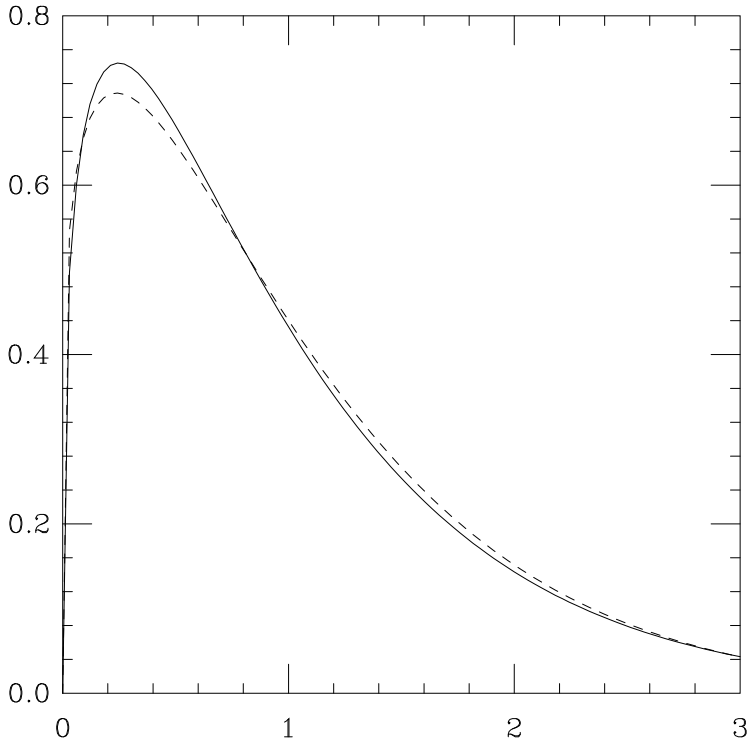

Figure 26. The dotted line represents a Weibull distribution with $a = 1.06309$ and $c = 1.2$. The solid line is a gamma distribution with $p = 1.33$ and $b = .75$. Both are taken from 1000-deviates sample. These parameters for the gamma distribution were calculated by the gamma estimator on the Weibull distribution shown by the solid line. As can be seen these distributions are close to the same shape.

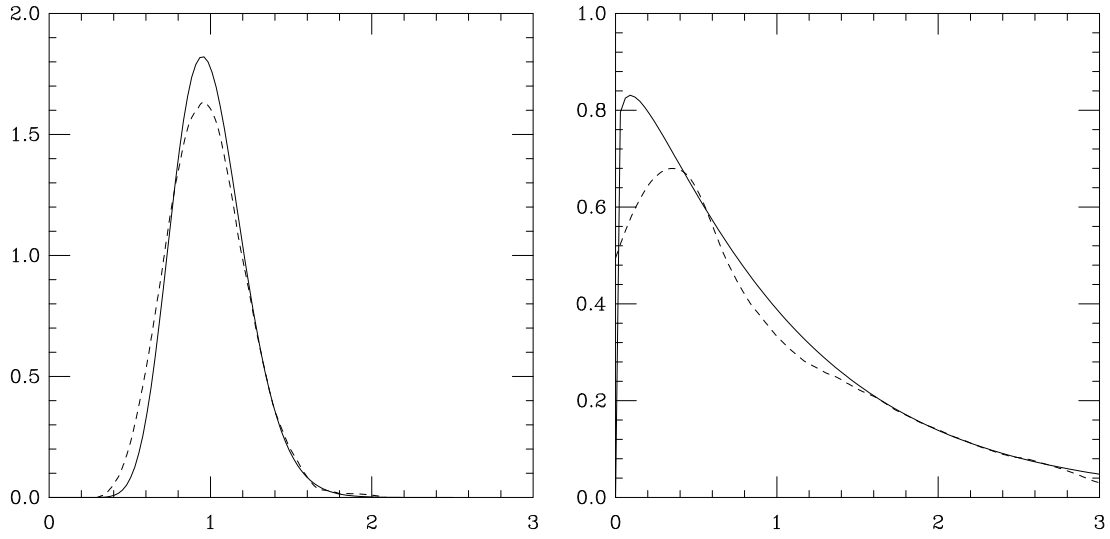

Figure 27. In the graph to the left, the dotted line represents optimal kernel estimation on a gamma distribution with $p = 20$ and $b = 0.05$, and the solid line represents the true gamma distribution with the same parameters. This distribution has a low skewness, and therefore oversmoothing flattens the peak position but does not shift it much. The graph to the right is another gamma distribution having, $p = 1.1$, $b = 0.909091$. This distribution has a high skewness, and therefore oversmoothing flattens the peak and shifts its position. Both are taken from 200-deviate samples.

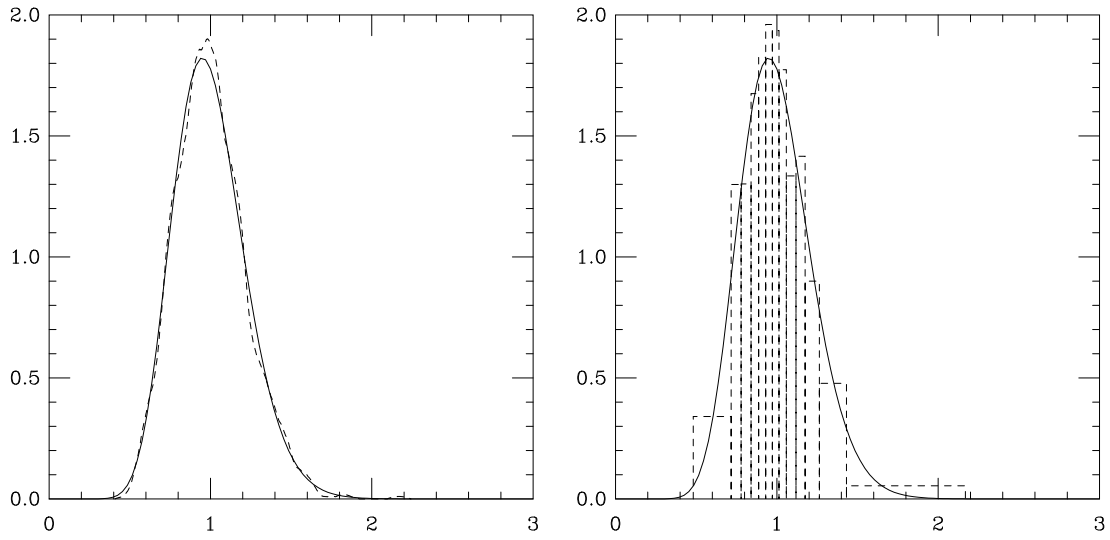

Figure 28. In the graph to the left, the solid line represents a gamma distribution with $p = 20$ and $c = 0.05$. The dotted line is the kernel approximation with the kernel parameter equal to 10 percent of the standard deviation of the population. The graph to the right is the adaptive histogram estimator with the same gamma distribution. Both are taken from 1000-deviate samples.

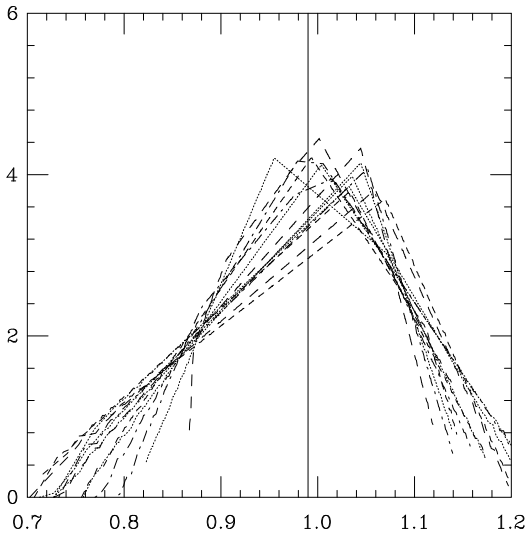

Figure 29. The solid vertical line represents the true mode position for a gamma distribution with $p = 100$ and $b = 0.01$. The dotted lines are the output of the `Unimodal3` distribution on fourteen realization. As can be seen, the majority of the peak estimations fall to the right of the true mode position, indicating a positive bias. All are taken from 1000-deviate samples.

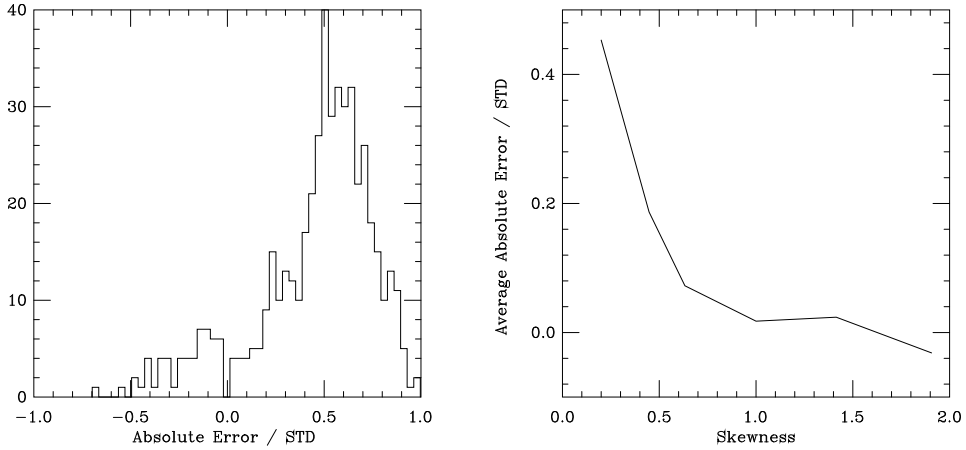

Figure 30. The histogram above is of the absolute error over all 500 realizations of the `Unimodal3` estimator on a gamma distribution taken from 1000-deviates samples with $p = 100$ and $b = 0.01$ and skewness of $\gamma_1 = 0.2$. We can see in the graph to the right how the bias changes with skewness.

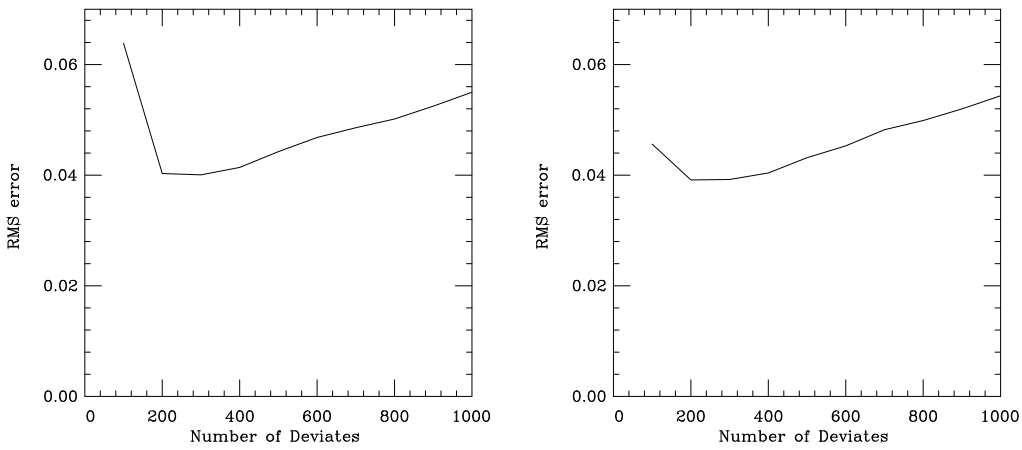

Figure 31. The graph to the left shows the `Unimodal3` results for the average of 500 realizations on a gamma distribution $p = 100$ and $b = 0.01$. We see the same systematic results for the average of another 500 realizations in the graph to the right. All realizations taken from 1000-deviate samples.

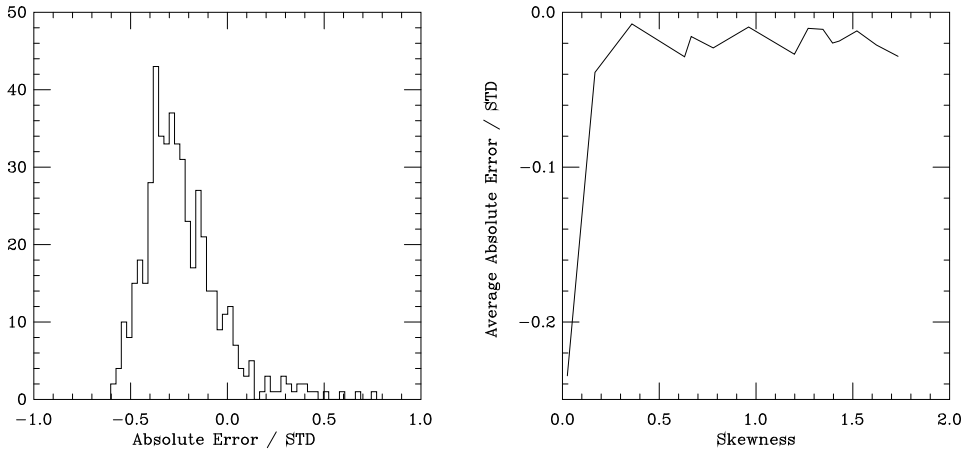

Figure 32. In the Wiebull distribution we can see how `Unimodal3` has a negative bias for the mode estimations on distributions that have near zero skewness. The graph to the left is a histogram of the biases for all 500 realizations taken from a sample of 1000 deviates for $a = 1.11142$ and $c = 3.5$ having a skewness of $\gamma_1 = 0.025$. The graph to the right demonstrates how this bias changes with skewness.

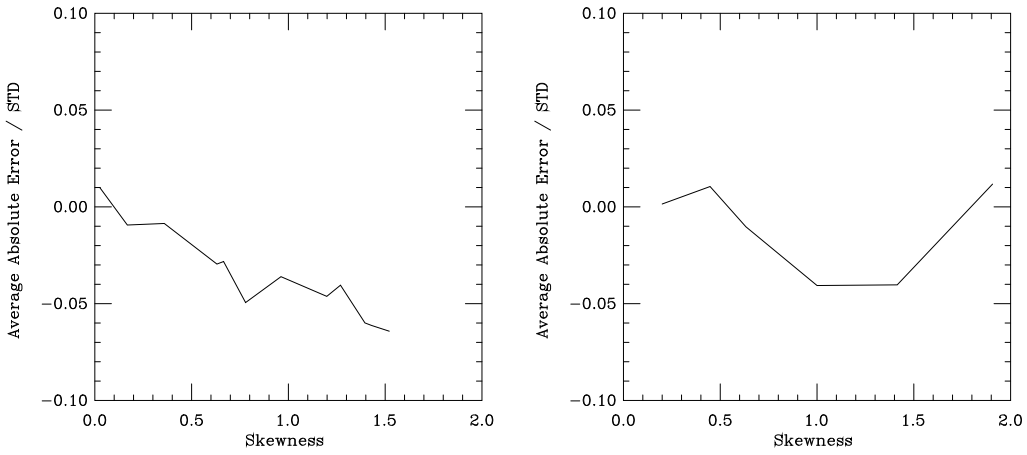

Figure 33. The graph to the left shows a plot of the bias of `Unimodal2` in estimating the peak position versus the skewness for the Weibull distribution taken from 1000-deviate samples. The graph to the right shows the same for a gamma distribution. As can see seen from the graphs, both distributions show a bias over a range of skewness values; however the bias is small. For `Unimodal2` we have used the exhaustive search method on decimated data.

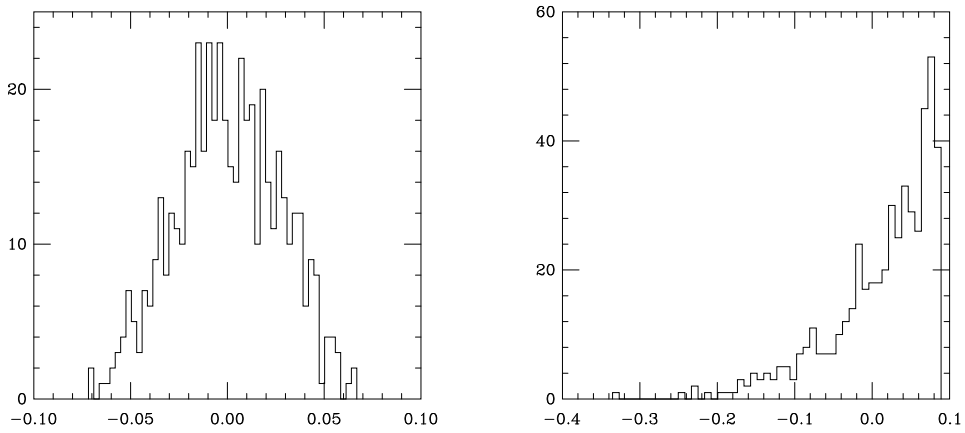

Figure 34. The histogram to the left shows the absolute errors from `Unimodal2` of all 500 realizations for a gamma distribution taken from 1000-deviate samples having a zero skewness value. To the right is the same for a gamma distribution having a large skewness of $\gamma_1 = 1.90693$. For `Unimodal2` we have used the exhaustive search method on decimated data.

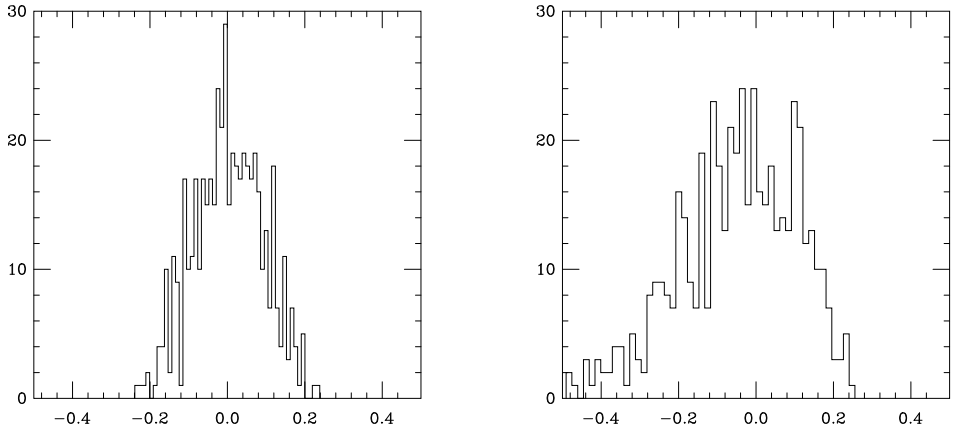

Figure 35. The histogram to the left shows the absolute errors from `Unimodal2` of all 500 realizations for a Weibull distribution taken from 1000-deviate samples having a zero skewness value. To the right is the same for a Weibull distribution having a high skewness value of $\gamma_1 = 1.73397$. For `Unimodal2` we have used the exhaustive search method on decimated data.

CHAPTER 6

RESULTS FOR BETA DISTRIBUTIONS HAVING NEGATIVE AND POSITIVE KURTOSIS

In this chapter we will concentrate on the estimators' relative efficiency on two families of beta distributions. These distributions are different from the Weibull and gamma distributions in that the means of these distributions are not equal to one, and the domain is constrained to $0 \leq x \leq 1$. We will look at a family of beta distributions where the peaks are flat, Table 3, and a family of distributions where the peaks are sharp, Table 4.

6.1 Beta Estimator

The beta estimator produced the lowest fractional error over most ranges of skewness. However for the family of flat distributions, in Fig. 38, the beta estimator performs poorly compared to OKS, and in Fig. 39, the beta estimator performs poorly compared to AHIST, UNI2, and FKS in estimating the peaks of near-zero skewness. We found that (30) has multiple roots, only one of which results in a peak in the interval $0 \leq x \leq 1$. We found several realizations resulted in peak estimates that where outside this interval therefore resulting in large fractional errors. Because of these errors we modified (30) instead to converge on the mode position and not the p value. We used the fact that the domain of the beta distribution is limited to $0 \leq x \leq 1$ in order to set the upper and lower mode positions limits, m_1 and m_2. These limits are similar to the p_1 and p_2 limits in chapter 2.5. We can determine these m's by calculating the mean, μ, of each realization. If μ for a realization is less than $1/2$, then $m_1 = 0$ and $m_2 = \mu - \epsilon$, where ϵ is some tolerance.[1] If μ for a realization is greater than $1/2$, then $m_1 = \mu + \epsilon$ and $m_2 = 1$. Once we have

[1] We used $\epsilon = 10^{-8}$.

determined these limits we can use the relation

$$p = \mu \frac{2m - 1}{m - \mu} \tag{38}$$

substituted into (30) to give

$$\Phi - \log(\mu \frac{2m - 1}{m - \mu}) + \psi(\mu \frac{2m - 1}{m - \mu}) = 0 \tag{39}$$

again where Φ is a constant number for each data set and is the combination of the mean and the geometric mean. By using the same root-finding routine we can converge on the estimated mode for each realization. However, on some realizations we ran into problems when finding the upper and lower limits of m. If our limits did not produce opposing signs in (39) and $\mu < 1/2$, then we set the estimated mode for that realization equal to zero. If our limits did not produce opposing signs in (39) and $\mu > 1/2$, then we set the estimated mode for that realization equal to one. By observing the spikes on the edges of the histogram of absolute errors in Fig. 36, we can see several realizations of the beta distribution having $p = 1.1$ and $q = 1.1$ produced this problem.

6.2 OKS and OCPF

OKS and OCPF resulted in the next smallest fractional errors over all skewnesses. However, in the case of large negative kurtosis, OCPF produces a larger fractional error than `Unimodal3`, FKS, and AHIST. Fig. 42 shows a flat-peaked distribution, having steep sides, with a skewness of zero. OCPF must use a high-order polynomial in order to fit the steep sides of this distribution. When the order is increased, the peak of the distribution is undersmoothed. Therefore OCPF cannot fit the sharp sides without undersmoothing the distribution. OCPF does not suffer from this problem when the skewness increases because the flatness of the peak decreases. Fig. 38 and Fig. 40 show examples.

OKS results in the lowest fractional error for all of the estimation methods at the largest negative kurtosis. In Fig. 43, we can see that OKS does not suffer from the

	p	q	$Skewness$	$Kurtosis$
Beta	1.1	1.1	0	-1.15385
	1.1	1.2	0.0735413	-1.12549
	1.1	1.3	0.140178	-1.08709
	1.1	1.4	0.201008	-1.04132
	1.1	1.5	0.256887	-0.990119
	1.1	1.6	0.308494	-0.934923
	1.1	1.7	0.356378	-0.876821
	1.1	1.8	0.400988	-0.816641
	1.1	1.9	0.442697	-0.755024
	1.1	2	0.481818	-0.69247

Table 3. Ten different beta distribution families. $B(x,p,q) = \dfrac{x^{p-1}(1-x)^{q-1}}{\beta(p,q)}$

	p	q	$Skewness$	$Kurtosis$
Beta	1.8	1.8	0	-0.909091
	1.8	1.9	0.0531526	-0.893363
	1.8	2.3	0.181951	-0.802405
	1.8	3	0.36578	-0.594268
	1.8	6	0.773718	0.259259
	1.8	15	1.14018	1.54851
	1.8	50	1.37243	2.6643

Table 4. Seven different beta distribution families. $B(x,p,q) = \dfrac{x^{p-1}(1-x)^{q-1}}{\beta(p,q)}$

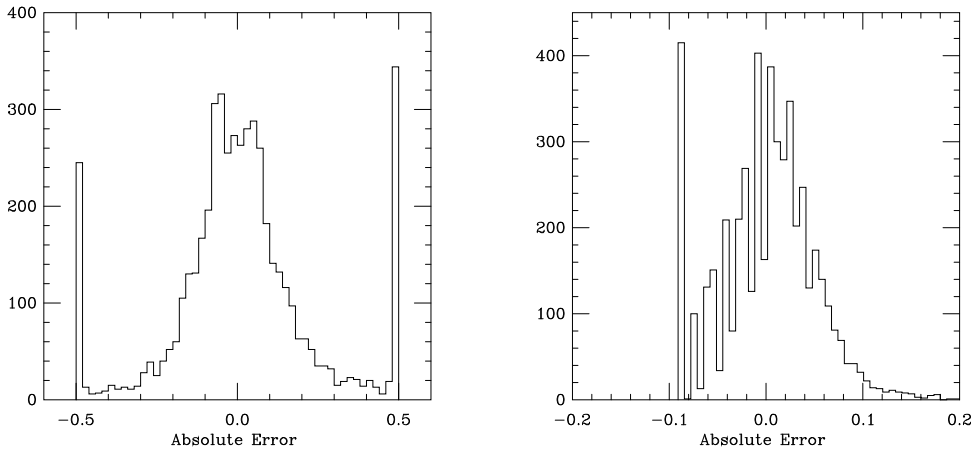

Figure 36. The histogram on the left is of absolute errors in beta mode estimations for a beta distribution with $p = 1.1$ and $q = 1.1$ having 1000 deviates, 500 realizations and sknewess $\gamma_1 = 0$. The histogram on the right is of the beta mode estimations for a beta distribution with $p = 1.1$ and $q = 2$ having 1000 deviates, 500 realizations and skewness $\gamma_1 = .481818$. The spikes on either side of the histograms are due to mode estimation errors found when estimating flat-peaked families.

undersmoothing problems of OCPF. However, from Fig. 43, OKS does a poor job in estimating the steep sides of the flat-peaked beta distribution. This is not a problem that affects the peak estimation.

If we reduce the number of deviates to 200 in each realization, OKS's and OCPS's fractional errors increase. However OKS's fractional error in the flat-peaked families, Fig. 38, increases much more within the range of skewness than in the sharper-peaked families, Fig. 39. This is because OKS suffers from oversmoothing when trying to estimate a realization having a small number of deviates.

6.3 FKS and AHIST

FKS and AHIST produced nearly the same fractional errors for the two families of beta distributions. As to be expected, the fractional errors increased as the number deviates decreased.

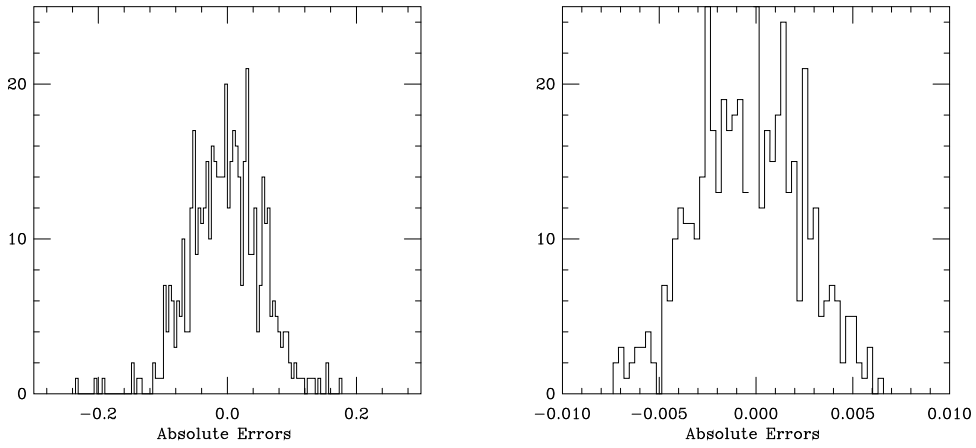

Figure 37. The histogram on the left is of absolute errors in beta mode estimations for a beta distribution with $p = 1.8$ and $q = 1.8$ having 1000 deviates, 500 realizations and sknewess $\gamma_1 = 0$. The histogram on the right is of the beta mode estimations for a beta distribution with $p = 1.8$ and $q = 50$ having 1000 deviates, 500 realizations and skewness $\gamma_1 = 1.37243$.

6.4 `Unimodal3` and `Unimodal2`

`Unimodal2`[2] and `Unimodal3` produced the largest fractional errors on both families of beta distributions over all ranges of skewness. For the family of distributions in Table 3, `Unimodal3` suffers from problems in the first-stage estimations. `Unimodal3` produces the largest fractional error for all ranges of skewness because the first-stage estimation performs poorly on the negative-kurtosis distributions. Since `Unimodal3` uses OCPF as its first-stage estimator, problems in locating the peak domain can parallel problems in OCPF estimations. Fig. 45 shows an OCPF estimation of the first derivative of the raw cumulative distribution. The output for the first-stage estimation does not include the true mode position. Therefore the second-stage will fault on estimating the mode. When the skewness increases, `Unimodal3`'s first-stage estimator no longer has the problem of excluding the mode, and therefore the fractional error decreases.

For the sharp-peaked families, `Unimodal2` shows no bias in estimating the mode. For this family of distributions, Fig. 48 shows the bias for all skewness, and Fig. 46 shows

[2]For `Unimodal2` we have used the exhaustive search method on decimated data.

the histograms of these biases for the extremes in skewness. However, for the flat-peaked families, `Unimodal2` does show a bias as the skewness increases. Fig. 48 shows the bias for all skewnesses, and Fig. 47 shows the histograms of these biases for the extremes in skewness.

For the sharp-peaked families, `Unimodal3` shows no bias in estimating the mode. Fig. 51 shows this bias for all skewnesses, and Fig. 50 show the histograms of these biases for the extremes in skewness. However, for the flat-peaked families, `Unimodal3` does show a bias as the skewness increases. Fig. 51 shows this bias for all skewnesses, and Fig. 49 shows the histograms of these biases for the extremes in skewness.

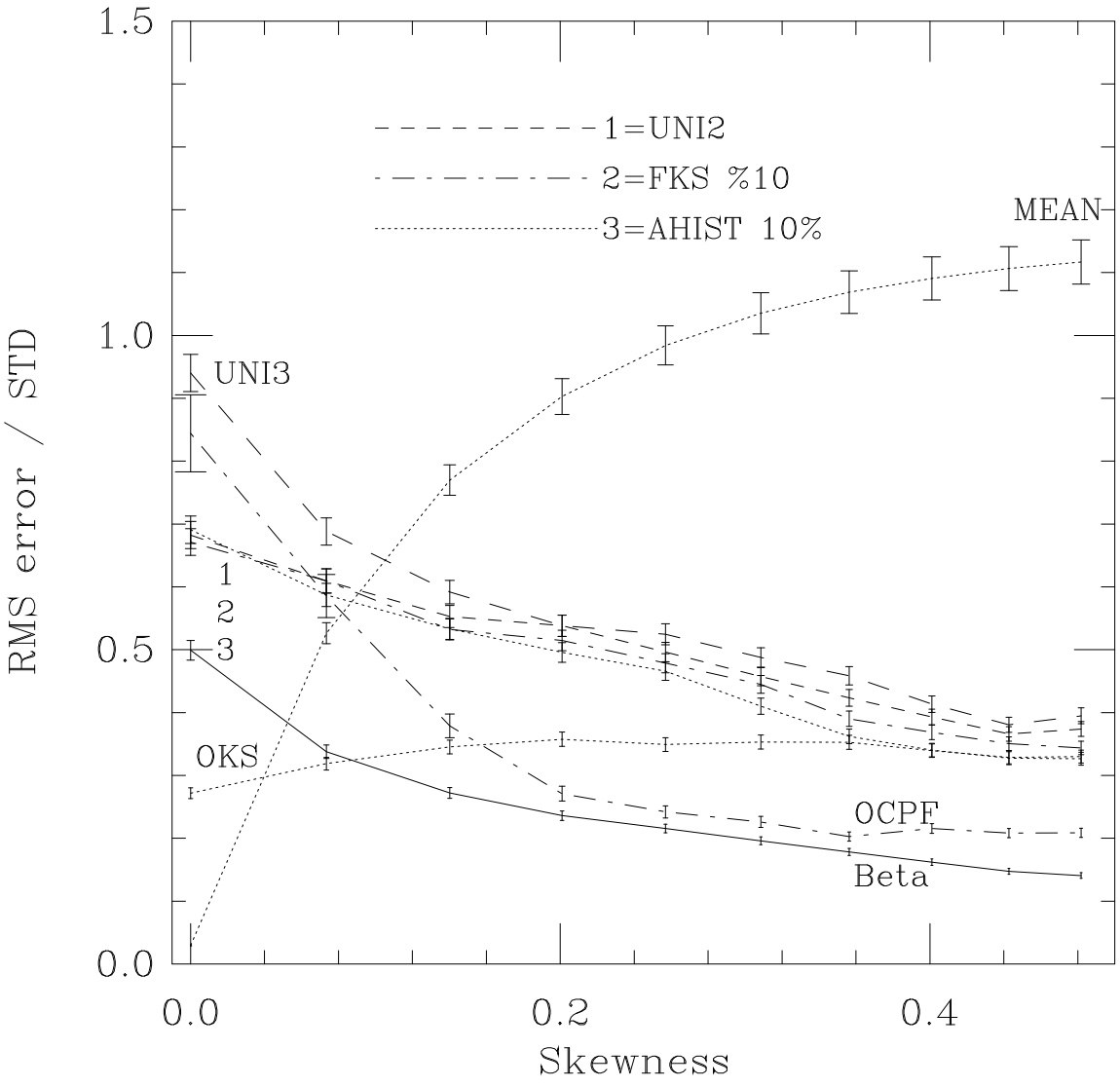

Figure 38. Results for a family of beta distributions $p = 1.1$ drawing 1000 deviates. Each point has 500 realizations. See also Table 3. For `Unimodal2` we have used the exhaustive search method on decimated data.

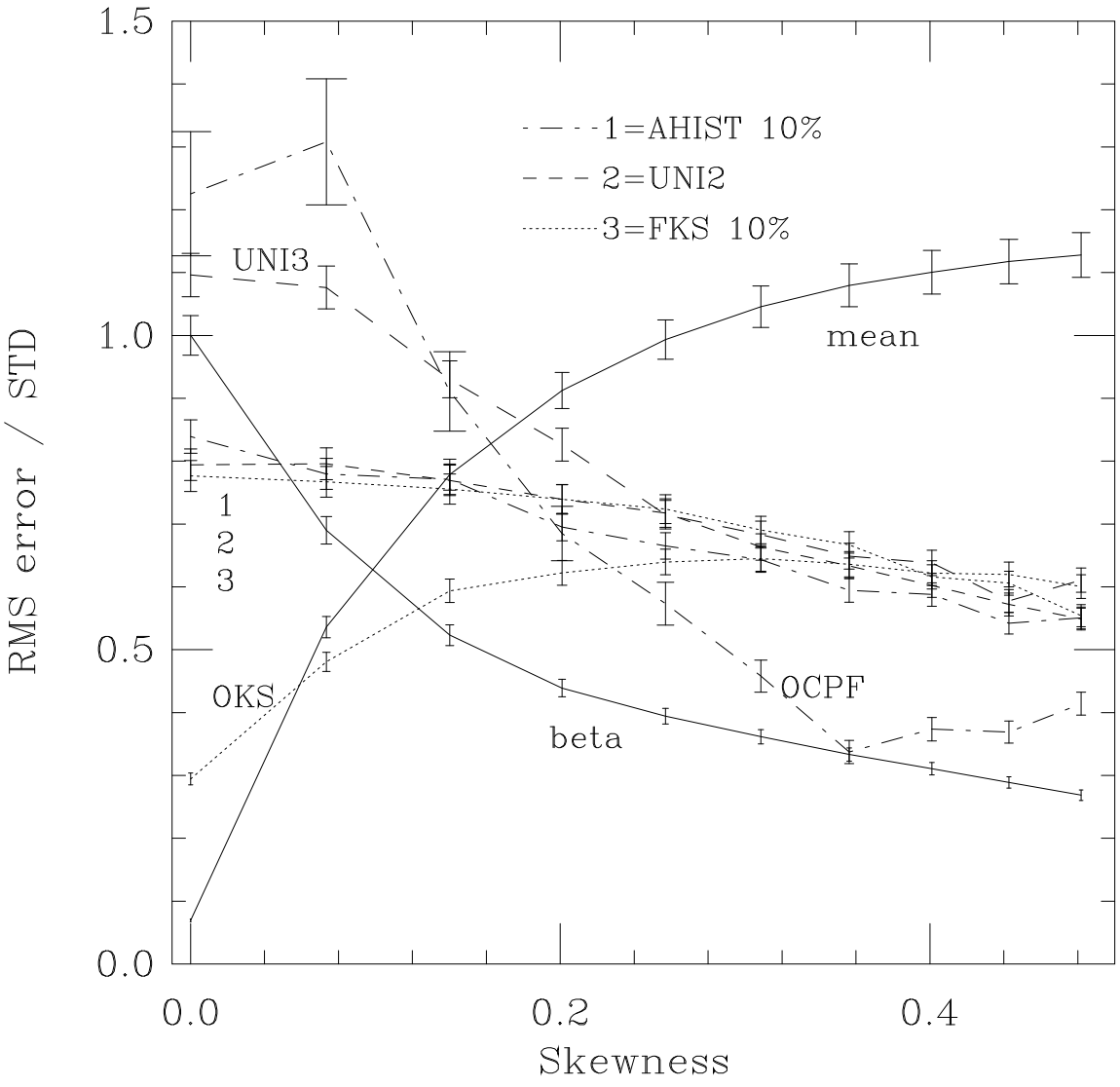

Figure 39. Results for a family of beta distributions with $p = 1.1$ drawing 200 deviates. Each point has 500 realizations. See also Table 3. For `Unimodal2` we have used the exhaustive search method on decimated data.

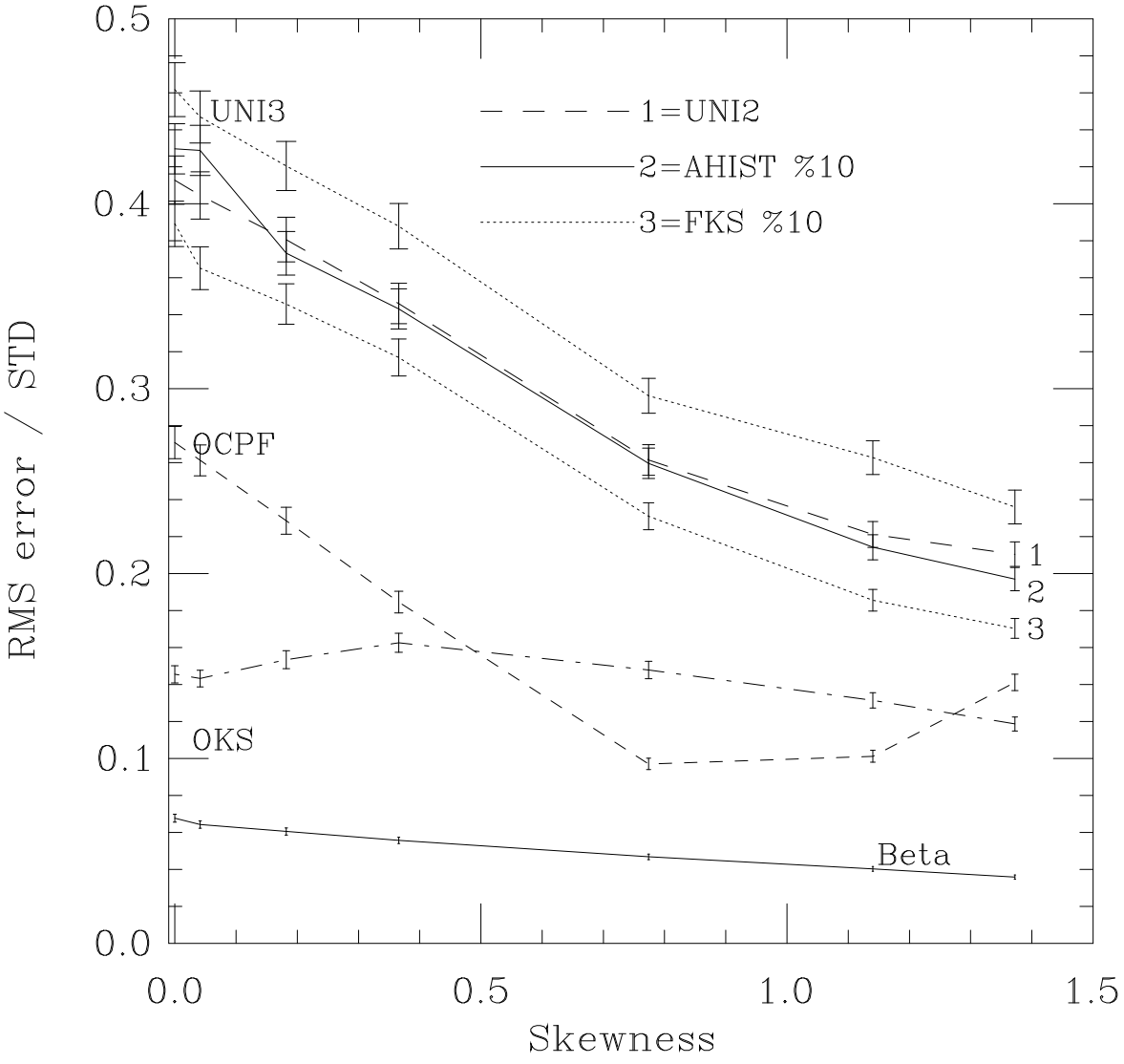

Figure 40. Results for a family of beta distributions with $p = 1.8$ drawing 1000 deviates. Each point has 500 realizations. See also Table 4. For `Unimodal2` we have used the exhaustive search method on decimated data.

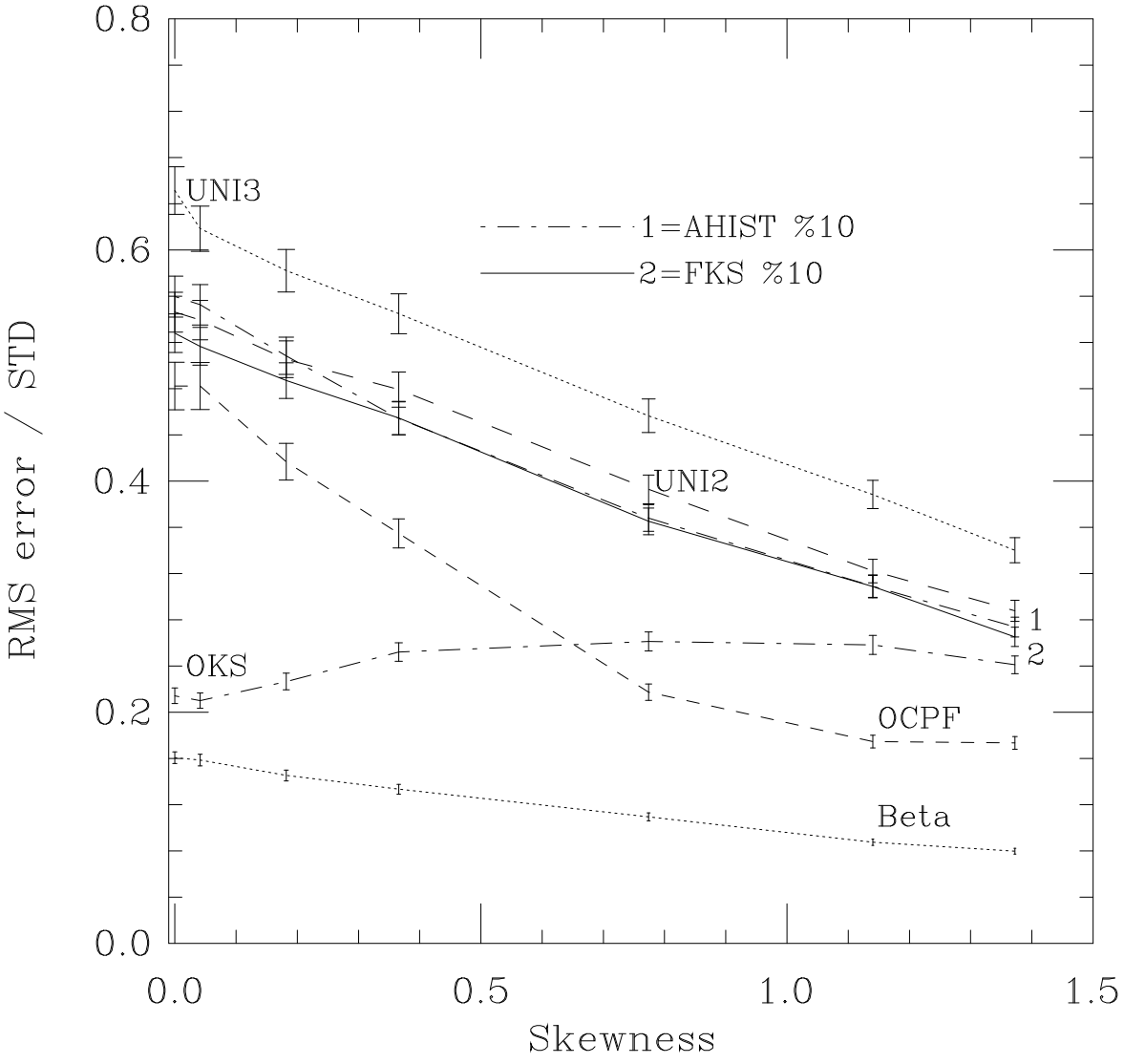

Figure 41. Results for a family of beta distributions with $p = 1.8$ drawing 200 deviates. Each point has 500 realizations. See also Table 4. For `Unimodal2` we have used the exhaustive search method on decimated data.

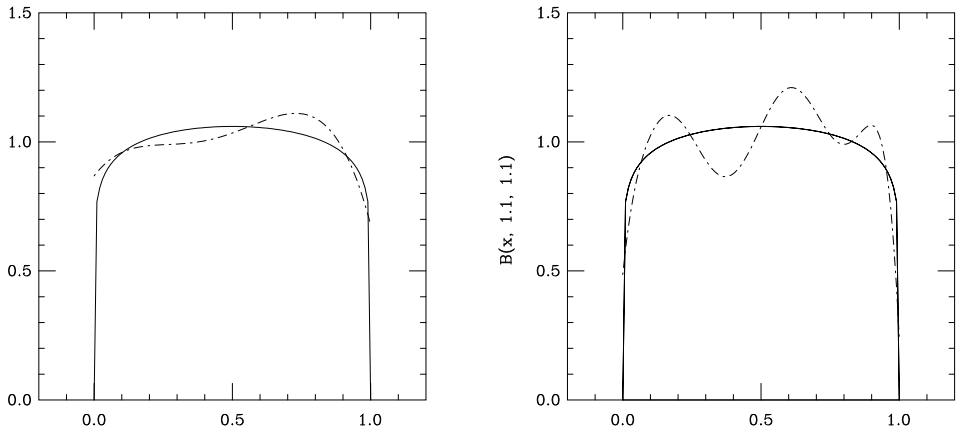

Figure 42. In the graph to the left is an OCPF estimation on a beta distribution having $p = 1.1$ and $q = 1.1$ taken from a 1000-deviate sample. To the right is an example of how the polynomial fit has to undersmooth a flat peak distribution in order to fit the sharp sides.

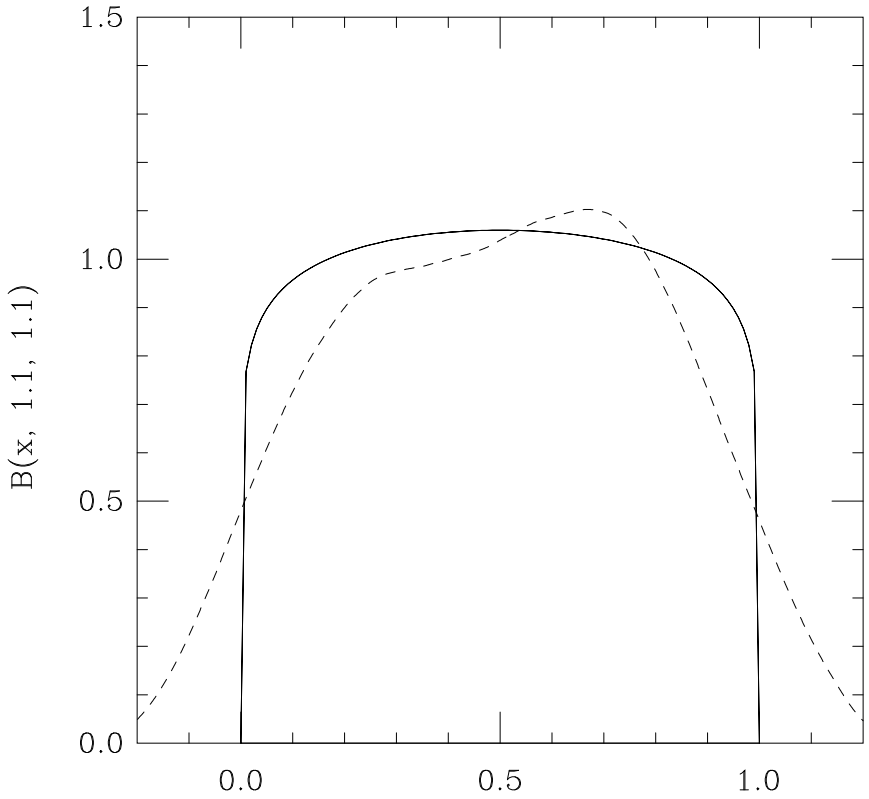

Figure 43. The solid line is the actual beta distribution with $p = 1.1$ and $q = 1.1$ having 1000 deviates. The dotted line is the OKS estimation.

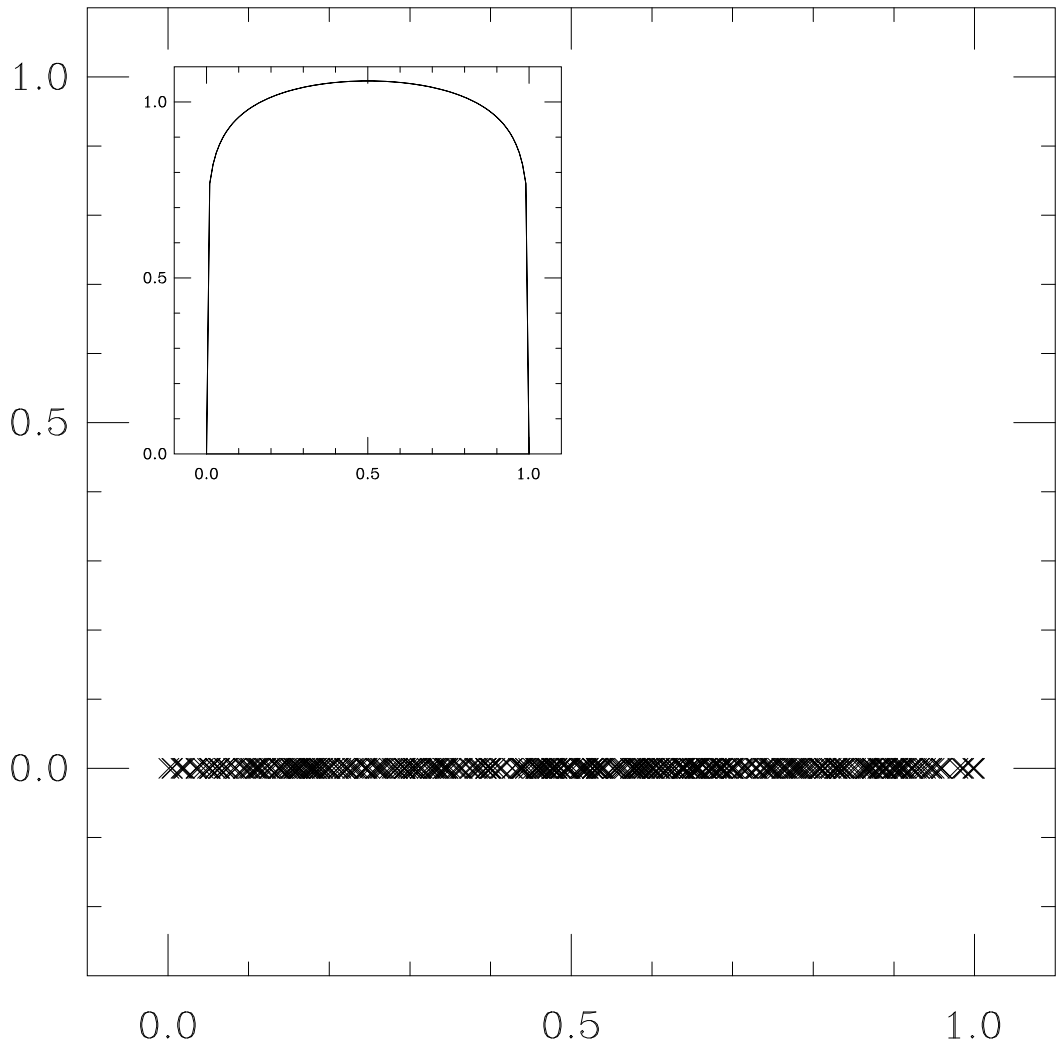

Figure 44. Plotted as x's on the x axis is the 500-deviate sample from a beta distribution having $p = 1.1$ and $q = 1.1$. The actual distribution is shown in the top left hand corner.

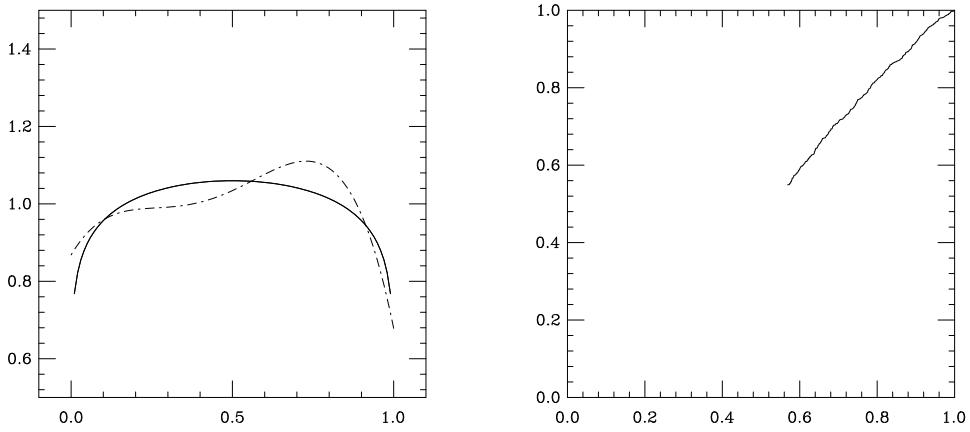

Figure 45. The graph on the right shows an OCPF to a 1000-deviate sample drawn from a beta distribution with $p = 1.1$ and $q = 1.1$. The solid line is the underlying distribution. The graph on the left is the output from `Unimodal3`'s first-stage estimation. The true mode position at .5 is not in the data domain.

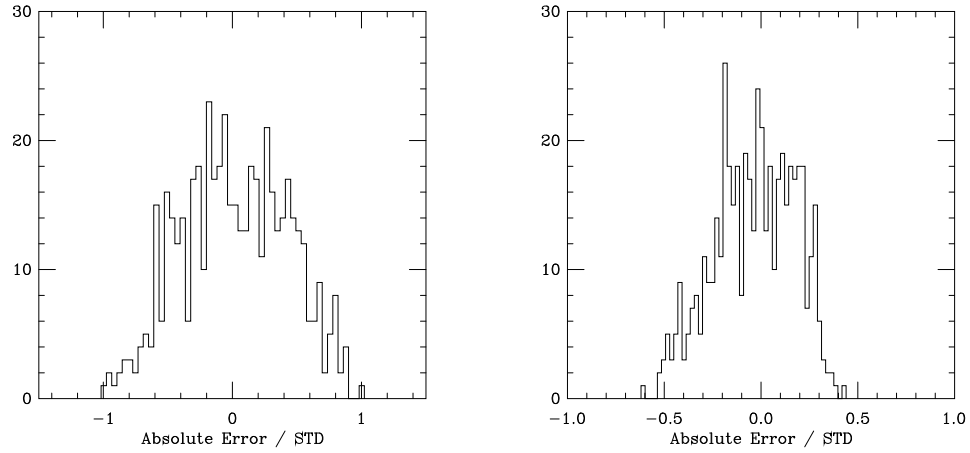

Figure 46. The histogram to the left shows the absolute errors of mode estimation divided by the standard deviation for `Unimodal2` on a beta distribution with $p = 1.8$ and $q = 1.8$ and a skewness $\gamma_1 = 0$. Each of the histograms has 500 realizations of 1000-deviate samples. The histogram to the right shows the absolute errors of mode estimation for `Unimodal8` on a beta distribution with $p = 1.8$ and $q = 50$ and a skewness $\gamma_1 = 1.37243$. Again there are 500 realizations of 1000-deviate samples. For `Unimodal2` we have used the exhaustive search method on decimated data.

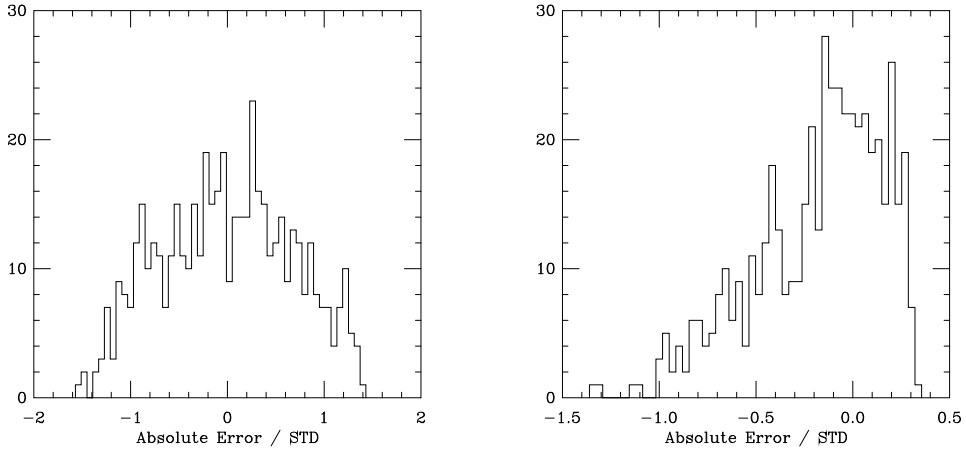

Figure 47. The histogram to the left shows the absolute errors of mode estimation divided by the standard deviation for `Unimodal2` on a beta distribution with $p = 1.1$ and $q = 1.1$ and a skewness $\gamma_1 = 0$. The histogram to the right shows the absolute errors of mode estimation for `Unimodal2` on a beta distribution with $p = 1.1$ and $q = 2$ and a skewness $\gamma_1 = 0.481818$. Both are from 500 realizations of 1000-deviate samples. For `Unimodal2` we have used the exhaustive search method on decimated data.

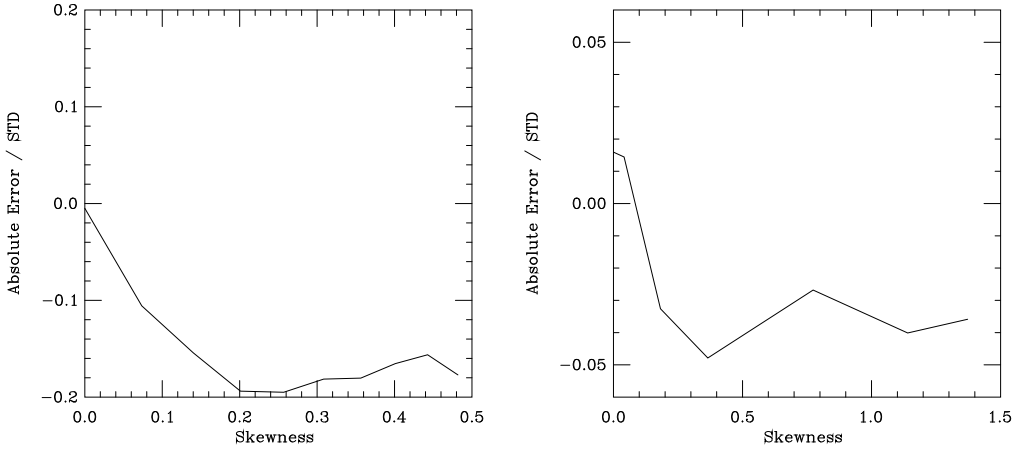

Figure 48. The graph to the left shows the absolute errors divided by standard deviation for `Unimodal2` over all skewnesses, for the family of beta distributions with $p = 1.1$ and q varying from 1.1 to 2. The graph to the right is the absolute errors for `Unimodal2` over all skewnesses for the family of beta distributions with $p = 1.8$ and q varying from 1.8 to 50. Each are from 500 realizations of 1000-deviates samples. For `Unimodal2` we have used the exhaustive search method on decimated data.

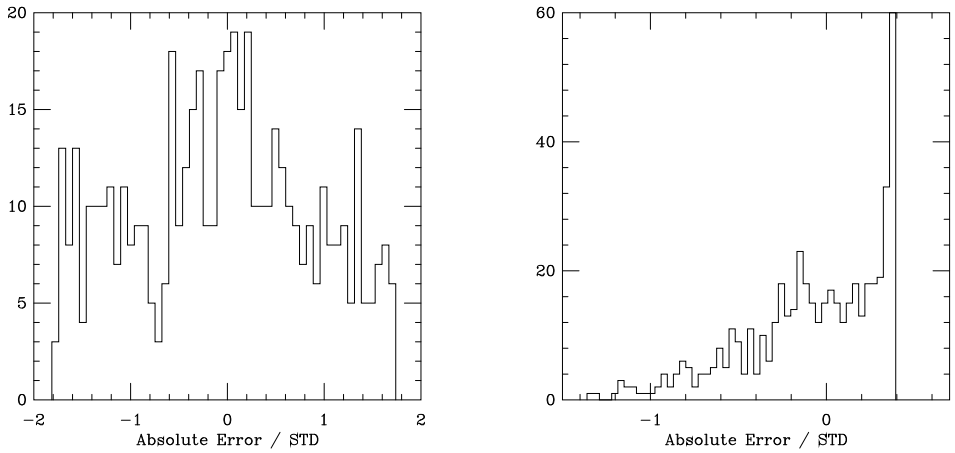

Figure 49. The histogram to the left shows the absolute errors of mode estimation divided by the standard deviation for `Unimodal3` on a beta distribution with $p = 1.1$ and $q = 1.1$ and a skewness $\gamma_1 = 0$. The histogram to the right is the absolute errors of mode estimation divided by the standard deviation for `Unimodal3` on a beta distribution with $p = 1.1$ and $q = 2$ and a skewness $\gamma_1 = 0.481818$. Each are taken from 500 realizations of 1000-deviate samples.

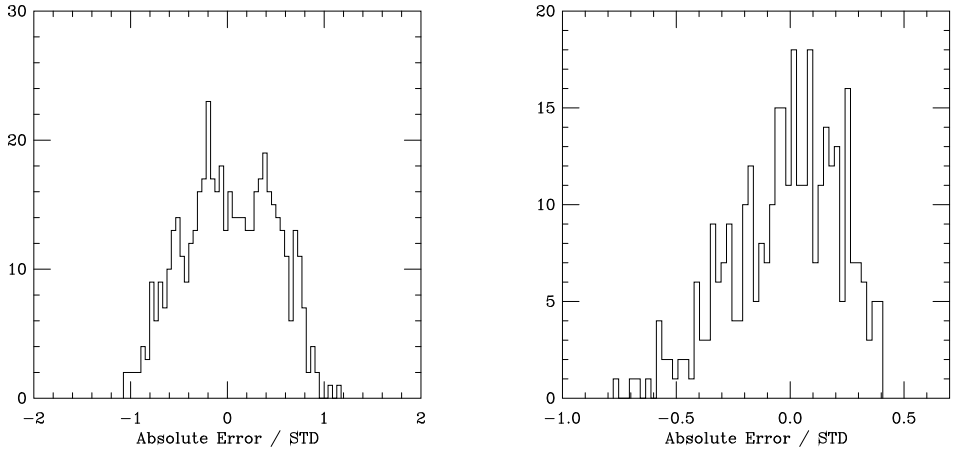

Figure 50. The histogram to the left shows the absolute errors of mode estimation divided by the standard deviation for `Unimodal3` on a beta distribution with $p = 1.8$ and $q = 1.8$ and a skewness $\gamma_1 = 0$. The histogram to the right is the absolute errors of mode estimation divided by the standard deviation for `Unimodal3` on a beta distribution with $p = 1.8$ and $q = 50$ and a skewness $\gamma_1 = 1.37243$. Each are taken from 500 realizations of 1000-deviate samples.

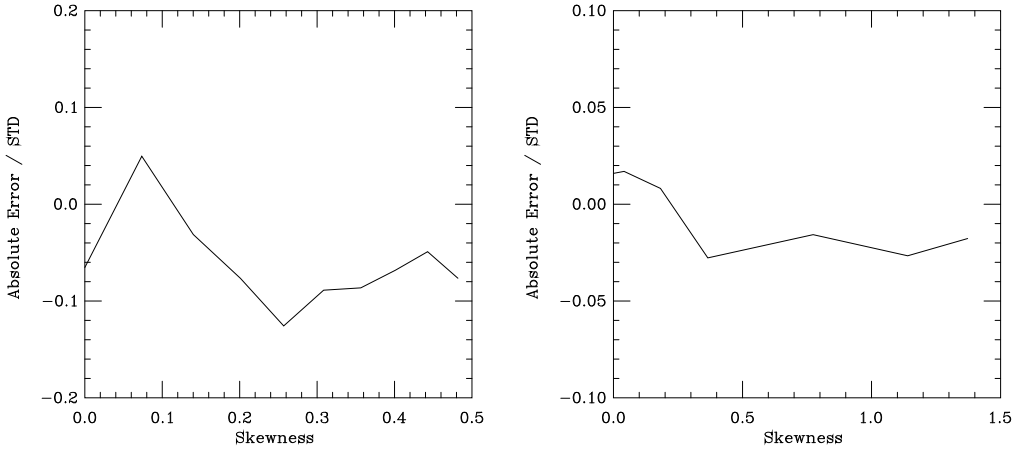

Figure 51. The graph to the left shows the absolute errors for `Unimodal3` over all skewnesses for the family of beta distributions with $p = 1.1$ and q varying from 1.1 to 2. The graph to the right is the absolute errors for `Unimodal3` over all skewnesses for the family of beta distributions with $p = 1.8$ and q varying from 1.8 to 50. Each is taken from 500 realizations of 1000-deviate samples.

CHAPTER 7

CONCLUSION AND FUTURE DIRECTIONS

`Unimodal3` and `Unimodal2` show a bias in their estimates on distributions having flat peaks and zero skewness, which contributed to their large fractional errors. These errors were larger than estimates on more highly skewed data. This was contrary to our initial assumption that a higher skewness would lead to larger fractional errors. Most estimators, excluding OKS, gave large fractional errors for zero skewness and smaller fractional errors as the skewness increased. OKS and OCPF proved to be the most efficient estimators.

Possible future work could include a change in the first-stage estimator for `Unimodal3`. It was shown that OCPF fails in extracting the true peak domain for families of distributions having a flat peak and zero skewness. This contributed to the magnitude of `Unimodal3`'s second-stage estimation errors. However, OKS did not suffer in locating the peak of flat distributions with zero skewness, and therefore using this method in `Unimodal3`'s first-stage estimator may help in correctly extracting the true peak domain for flat-peaked distributions. Problems to overcome, if OKS is used for the `Unimodal3`'s first-stage estimation, would include extracting the domain of negative curvature from OKS.

REFERENCES

[1] N. Johnson and S. Kotz, *Continuous univariate distributions*, Wiley, 1970.

[2] E.S. Keeping, *Introduction to statistical inference*, Van Nostrand, 1962.

[3] C. Lawson and R. Hanson, *Solving least squares problems*, Prentice-Hall, 1974.

[4] C. Lunneborg, *Data analysis by resampling: Concepts and applications*, Duxbury, 2000.

[5] P. Pitman, *Probability*, Springer, 1992.

[6] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical recipes in c*, Cambridge, 1992.

[7] David A. Rabson, B.N. Narozhny, and A.J. Millis, *Crossover from poisson to wigner-dyson level statistics in spin chains with integrability breaking*, Submitted for publication.

[8] Bro Rasmus, Nicholaos D. Sidiropoulos, and Age K. Smilde, *Maximum likelihood fitting using ordinary least squares algorithms*, J. Chemometrics **16** (2002), 387–400.

[9] Kurt S. Riedel, *Piecewise convex function estimation and model selection*, Approximation theory VIII, Vol. 1 (College Station, TX, 1995), Ser. Approx. Decompos., vol. 6, World Sci. Publishing, River Edge, NJ, 1995, pp. 467–475.

[10] Alexander Sidorenko and Kurt S. Riedel, *Adaptive kernel estimation of the spectral density with boundary kernel analysis*, Approximation theory VIII, Vol. 1 (College Station, TX, 1995), Ser. Approx. Decompos., vol. 6, World Sci. Publishing, River Edge, NJ, 1995, pp. 519–527.

[11] J. Taylor, *An introduction to error analysis*, University Science Books, 1982.

[12] D. Wackerly, W. Mendenhall III, and R. Scheaffer, *Mathematical statistics with applications*, Duxbury, 1996.

[13] M. Wand and M. Jones, *Kernel smoothing*, Chapman & Hall, 1995.

APPENDICES

Appendix A
Moments, Properties and Cumulants

Many distributions have the same mean and standard deviation. Because of this fact we must consider other numerical descriptions that characterize data set and its distribution. The raw moments μ_n' of a data set define a probability distribution. The n^{th} moment of a random variable Y taken about the origin is defined by

$$E(Y^n) = \int y^n p(Y = y)dy = \mu_n^{'} \tag{40}$$

where $E(Y)$ is the expectation value of $Y = y$. The moment-generating function $m(t)$, packs all the moments of a random variable Y into a nice simple expression and is defined by

$$m(t) = E(e^{tY}). \tag{41}$$

This function exists if there is a positive constant b such that $m(t)$ is finite for $|t| \leq b$. The importance of $m(t)$ is that if it exists, we can calculate any of the moments for Y. For example, if $m(t)$ exists, then for any k_{th} derivative with respect to t

$$\frac{d^k m(t)}{dt^k} = m^{(k)}(0) = \mu_k' \tag{42}$$

when k is positive and t is set to zero.

However when the range in the distribution is large, the higher-order moments increase in size rapidly. Therefore instead of the moments being expressed about zero, they may also be expressed about the mean of the distribution

$$E(Y^n) = \int (y - \mu)^n P(Y = y)dy = \mu_n. \tag{43}$$

If one can find the moment-generating function of a random variable Y, then one may use the raw moments, or the moments about the mean, to determine the properties of the probability distribution such as the mean, variance, skewness, and kurtosis. These

properties expressed as functions of the moments about the mean are as follows:

$$\begin{aligned} \mu &= \mu_1 \\ \sigma^2 &= \mu_2 \\ \gamma_1 &= \frac{\mu_3}{\mu_2^{3/2}} \\ \gamma_2 &= \frac{\mu_4 - 3\mu_2^2}{\mu_2^2}. \end{aligned} \tag{44}$$

We describe a probability distribution by its mean, variance, skewness, and kurtosis. The mean of a sample, x_i, is given by

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{45}$$

where N is the total number of sample measurements. This is simply the average over all of the sample measurements. The variance of a sample x_i is given by

$$\sigma^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(x_i - \mu)^2. \tag{46}$$

The variance of a sample of measurements is the sum of the square differences between a measurements and their mean, divided by $N - 1$. One may also use skewness and kurtosis to describe a probability distribution. The skewness is a measure of asymmetry in a sample distribution. Fig. 52 show a family of Weibull distributions with different skewnesses. The kurtosis is a measure of flatness or sharpness of the peak of a probability distribution.

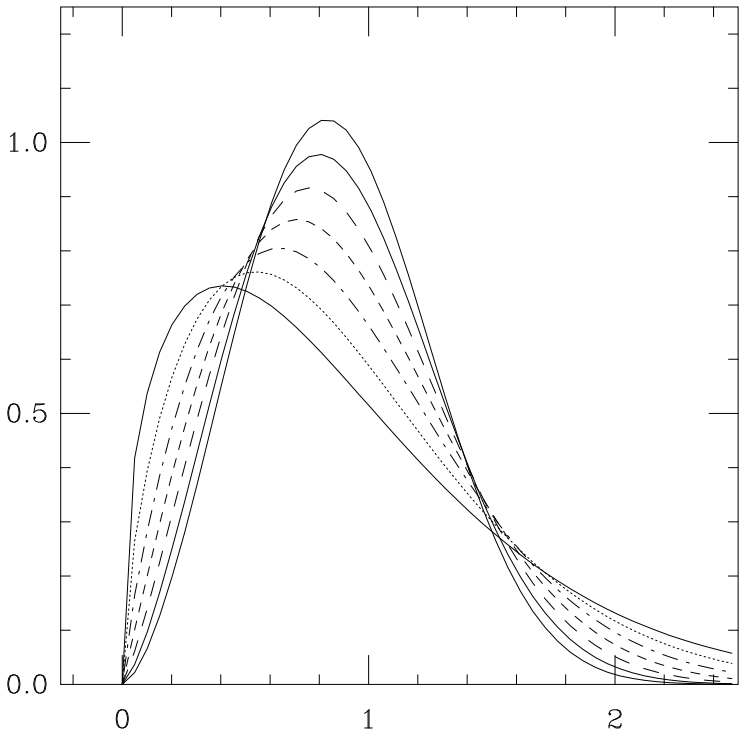

Figure 52. Weibull distributions with $a = 1$ and c ranging from 1.4 to 2.6 . As the c parameter increases the peak position moves away from the y axis. Therefore when $c = 1.4$, the distribution has a larger skewness, and the peak is closer to the y axis.

Another numerical description of the uniqueness of a distribution, is by the use of the cumulants κ_n. The cumulants are derived from the raw moments, μ'_n, and are defined by

$$\ln[\phi(t)] = \sum_{n=0}^{\infty} \kappa_n \frac{(it)^n}{n!} \tag{47}$$

where $\phi(t)$ is the characteristic function defined as the Fourier transform of the probability density function $p(x)$

$$\phi(t) = F_x[p(x)](t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx. \tag{48}$$

Appendix A (Continued)

Here the cumulants are the real coefficients of a Maclaurin series of $\ln[\phi(t)]$. The first four cumulants expressed as functions of raw moments μ_n, are as follows:

$$\begin{aligned}
\kappa_1 &= \mu_1 \\
\kappa_2 &= \mu_2 - \mu_1^2 \\
\kappa_3 &= 2\mu_1^3 - 3\mu_1\mu_2 + \mu_3 \\
\kappa_4 &= -6\mu_1^4 + 12\mu_1^2\mu_2 - 3\mu_2^2 - 4\mu_1\mu_3 + \mu_4.
\end{aligned} \tag{49}$$

The unique properties of a distribution such as mean, variance, skewness, and kurtosis can also be expressed as a combination of cumulants

$$\begin{aligned}
\mu &= \kappa_1 \\
\sigma^2 &= \kappa_2 \\
\gamma_1 &= \frac{\kappa_3}{\kappa_2^{(3/2)}} \\
\gamma_2 &= \frac{\kappa_4}{\kappa_2^2}.
\end{aligned} \tag{50}$$

Appendix B
Test Probability Distributions

The Weibull distribution is named after Waloddi Weibull, who offered it as an analytical tool for modeling the breaking strengths of materials. Other uses include reliability and lifetime modeling [5]. The Weibull distribution is defined as

$$W(x,a,c) = ca^{-1}(\frac{x}{a})^{(c-1)}\exp[-(\frac{x}{a})^2], \tag{51}$$

where $a, c > 0$ and $0 \leq x < \infty$. The a parameter alters the variance and the scale of the distribution, while the c parameter alters the skewness as well as the kurtosis. The moments of the Weibull distribution about the mean are

$$\begin{aligned}
\mu_1 &= a\Gamma(1+c^{-1}) \\
\mu_2 &= a^2\Gamma(1+2c^{-1}) \\
\mu_3 &= a^3\Gamma(1+3c^{-1}) \\
\mu_4 &= a^4\Gamma(1+4c^{-1}),
\end{aligned} \tag{52}$$

and the mean, variance, skewness and kurtosis are

$$\mu = a\Gamma(1+c^{-1}) \tag{53}$$

$$\sigma^2 = a^2[\Gamma(1+2c^{-1}) - \Gamma(1+c^{-1})^2] \tag{54}$$

$$\begin{aligned}
\gamma_1 = \frac{2\Gamma(1+c^{-1}) - 3\Gamma(1+c^{-1})\Gamma(1+2c^{-1})}{[\Gamma(1+2c^{-1}) - \Gamma(1+c^{-1})^2]^{(3/2)}} \\
+ \frac{\Gamma(1+3c^{-1})}{[\Gamma(1+2c^{-1}) - \Gamma(1+c^{-1})]^{(3/2)}}.
\end{aligned} \tag{55}$$

$$\begin{aligned}
\gamma_2 = -6\Gamma(1+c^{-1})^4 + 12\Gamma(1+c^{-1})^2\Gamma(1+2c^{-1}) - 3\Gamma(1+2c^{-1})^2 - \\
\frac{4\Gamma(1+c^{-1})\Gamma(1+3c^{-1}) + \Gamma(1+4c^{-1})}{[\Gamma(1+2c^{-1}) - \Gamma(1+c^{-1})^2]^2}.
\end{aligned} \tag{56}$$

Here $\Gamma(p)$ is the Gamma function defined as

$$\Gamma(p) = \int_0^\infty t^{p-1} \exp[-t] dt. \tag{57}$$

If we integrate $\Gamma(p)$ from $0 \leq t < \infty$, we can see that $\Gamma(1) = 1$, and for any $p > 1$, $\Gamma(n) = (n-1)!$, as long n is an integer.

The Gamma distribution is based on two parameters and is used to model skewed frequency distributions. Such examples of skewed frequency populations include the lengths of time between malfunctions for aircraft engines, and arrivals at a supermarket checkout counter [12]. The Gamma distribution is defined as

$$G(x, a, b) = \frac{1}{b^a \Gamma(a)} x^{(a-1)} \exp[-(\frac{x}{b})]. \tag{58}$$

The skewness and kurtosis of the distribution are governed by the value of a, while the mean and variance depend on both a and b. The moments of the distribution about the mean are

$$\begin{aligned} \mu_1 &= \frac{b\Gamma(1+a)}{\Gamma(a)} \\ \mu_2 &= \frac{b^2\Gamma(2+a)}{\Gamma(a)} \\ \mu_3 &= \frac{b^3\Gamma(3+a)}{\Gamma(a)} \\ \mu_4 &= \frac{b^4\Gamma(4+a)}{\Gamma(a)}, \end{aligned} \tag{59}$$

and the mean, variance, skewness, and kurtosis are

$$\mu = ab \tag{60}$$

$$\sigma^2 = ab^2 \tag{61}$$

$$\gamma_1 = \frac{2}{\sqrt{(a)}} \tag{62}$$

$$\gamma_2 = \frac{6}{a}. \tag{63}$$

Unlike the gamma and Weibull distributions, the beta distribution is defined over a closed interval $0 \leq x \leq 1$. The beta distribution function is often used to model proportions. Such examples include the proportion of chemical impurities in a sample, and the proportion of time a machine is under repair [12]. The beta distribution is defined as

$$B(x, p, q) = \frac{x^{p-1}(1-x)^{q-1}}{\beta(p, q)}, \tag{64}$$

where $\beta(p, q)$ is the Beta function,

$$\beta(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt, \tag{65}$$

and $p, q > 0$. Compared to the gamma and Weibull distributions, the beta distribution can assume a wide range of shapes.

The raw moments about the mean for the beta distribution are

$$\begin{aligned} \mu_1 &= \frac{\Gamma(1+p)\Gamma(p)}{\beta(p,q)\Gamma(1+p+q)} \\ \mu_2 &= \frac{\Gamma(2+p)\Gamma(p)}{\beta(p,q)\Gamma(2+p+q)} \\ \mu_3 &= \frac{\Gamma(3+p)\Gamma(p)}{\beta(p,q)\Gamma(3+p+q)} \\ \mu_4 &= \frac{\Gamma(4+p)\Gamma(p)}{\beta(p,q)\Gamma(4+p+q)}, \end{aligned} \tag{66}$$

and the mean, variance, skewness, and kurtosis are

$$\mu = \frac{p}{(p+q)} \tag{67}$$

Appendix B (Continued)

$$\sigma^2 = \frac{pq}{(p+q)^2(p+q+1)} \tag{68}$$

$$\gamma_1 = \frac{2(q-p)}{(p+q+2)}\sqrt{\frac{p+q+1}{pq}} \tag{69}$$

$$\gamma_2 = \frac{6(a^3 + a^2(1-2c) + c^2(1+c) - 2ac(2+c))}{ac(2+a+c)(3+a+c)}. \tag{70}$$

Appendix C
Chebyshev Polyomial

The Chebyshev polynomial is defined as

$$T_n(x) = \cos(n \arccos x), \tag{71}$$

where n is the degree of the polynomial. $T_n(x)$ can have explicit expressions; Press, Teukolsky, Vetterling and Flannery show the first four for $n = 0$ to $n = 3$ [6]:

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_n + 1(x) &= 2xT_n(x) - T_n - 1(x)n \geq 1. \end{aligned} \tag{72}$$

These polynomials are orthogonal on the the interval [-1,1] over a weight $(1 - x^2)^{-1/2}$,

$$\int_{-1}^{1} \frac{T_i(x)T_j(x)}{\sqrt{1-x^2}} dx \quad = \begin{cases} 0 & : \quad i \neq \text{J} \\ \frac{\pi}{2} & : \quad i = j \neq 0 \\ \pi & : \quad i = j = 0 \end{cases} \tag{73}$$

and have n zeros on the interval [-1,1] located at

$$x = \cos\left(\frac{\pi(k - 1/2)}{n}\right) \quad k = 1, 2, ..., n. \tag{74}$$

The Chebyshev polynomials not only satisfy a continuous relation, Equ. 73, but also a discrete orthogonality relation. If $x_k(k = 1, ..., m)$ are the m zeros of T_m given by equations

5, and if $i, j < m$, then

$$\sum_{k=1}^{m} T_i(x_k)T_j(x_k) = \ = \begin{cases} 0 & : \quad i \neq \text{J} \\ m/2 & : \quad i = j \neq 0 \\ m & : \quad i = j = 0 \end{cases} \tag{75}$$

Press, Teukolsky, Vetterling, and Flannery show us that for any arbitrary function on the interval [-1,1], and if N coefficients, c_j, where $j = 0, ..., N-1$, are defined by

$$c_j = \frac{2}{N} \sum_{k=1}^{N} f(x_k)T_j(x_k) \tag{76}$$

then the approximation formula is

$$f(x) \approx [\sum_{k=0}^{N-1} c_k T_k(x)] - \frac{1}{2}c_0 \tag{77}$$

and is *exact* for x equal to all of the N zeros of $T_N(x)$. However not all of our distributions are defined on the interval [-1,1]. Therefore we must use an effective change of variable given by

$$x' \equiv \frac{x - \frac{1}{2}(b+a)}{\frac{1}{2}(b-a)} \tag{78}$$

where a and b are the arbitrary limits instead of [-1,1].

Appendix D
Finite Differences

In `Unimodal2` we imposed constraints on the 2^{nd} derivative, and in `Unimodal3` we imposed constraints on the 3^{rd} derivative. We use the finite differences of a data set, x_i, where $i = 0...N-1$, in order to estimate the derivatives. We express the 1^{st}, 2^{nd}, and 3^{rd} finite differences of a data set x_i as

$$a_i = \frac{f_{i+1} - f_i}{\delta_i} \tag{79}$$

$$b_i = \frac{a_{i+1} - a_i}{\delta_i} \tag{80}$$

$$c_i = \frac{b_{i+1} - b_i}{\delta_i} \tag{81}$$

where $\delta_i = x_{i+1} - x_i$, and f_i is the discrete cumulative distribution

$$f_i = \frac{1}{N-1} i \ . \tag{82}$$

We wish to express the function f_i in terms of the finite differences. We can invert (79), (80), and (81) to get

$$f_i = f_{i-1} + a_{i-1}\delta_{i-1} \tag{83}$$

$$a_i = a_{i-1} + b_{i-1}\delta_{i-1} \tag{84}$$

$$b_i = b_{i-1} + c_{i-1}\delta_{i-1}. \tag{85}$$

If we are using `Unimodal2`, we will express f_i in terms of a_i and b_i (second and third differences) and get

$$f_i = f_{i-1} + \Delta_i \tag{86}$$

where $\Delta_i = a_{i-1}\delta_{i-1}$. By recursively plugging in the expression for a_{i-1}, taken from (84), into (83), we get

$$\Delta_i = \Big[a_0 + \sum_{k=0}^{i-2} b_k\delta_k\Big]\delta_{i-1}. \tag{87}$$

The first three expressions for f_i are

$$\begin{aligned} f_0 &= f_0 \\ f_1 &= f_0 + a_0\delta_0 \\ f_2 &= f_0 + a_0\delta_0 + a_0\delta_1 + b_0\delta_0\delta_1 \\ &\cdot \\ &\cdot\cdot \end{aligned} \tag{88}$$

We can write (88) in matrix form. The final problem is reduced to a design matrix for least-squares minimization[1] expressed as

$$D\mathbf{w} = \mathbf{f} \tag{89}$$

where D a matrix defined by (88), and the vector $\mathbf{w}$ is the vector with components

$$\begin{aligned} w_0 &= f_0 \\ w_1 &= a_0 \\ w_j &= b_{j-2} \quad where \quad 2 \le j \le p_0 + 2 \\ w_j &= -b_{j-2} \quad where \quad j > p_0 + 2 \end{aligned} \tag{90}$$

and p_0 is the index of the assumed mode position x_{p_0}.

[1]Refer to Numerical Recipes in C for definition of design matrix.

Appendix D (Continued)

The least-squares minimization of

$$R^2 = \sum_i ([D\mathbf{w}]_i - f_i)^2 \tag{91}$$

is identical to the problem (89) [6]. If we replace $\mathbf{w}$ with $\widetilde{\mathbf{w}}$, we can write a new identity

$$D\widetilde{\mathbf{w}} = \widetilde{\mathbf{f}}. \tag{92}$$

If we enforce no constraints on $\widetilde{\mathbf{w}}$, then $\widetilde{\mathbf{w}} = \mathbf{w}$, $\widetilde{\mathbf{f}} = \mathbf{f}$, and $\widetilde{\mathbf{f}}$ would reconstruct the original discrete cumulative distribution f. Requiring $\widetilde{\mathbf{w}}$ to be non-negative leads to a constrained least-squares problem, and $\widetilde{\mathbf{f}}$ is the closest cumulative probability distribution (in the least-squares sense) to the original data.

If we are using `Unimodal3`, then we have to extend (87) to include the third finite differences, c_i:

$$\Delta_i = \Big[a_0 + \sum_{k=0}^{i-2} \Big(b_0 + \sum_{l=0}^{k-1} c_l \delta_l\Big) \delta_k] \delta_{i-1} \tag{93}$$

where the problem is again reduce to a design matrix for least-squares minimization, and $\mathbf{w}$ components are now

$$\begin{aligned} w_0 &= f_0 \\ w_1 &= a_0 \\ w_2 &= b_0 \\ w_3 &= -c_0 \\ w_4 &= -c_1. \\ &\cdot \\ &\cdot\cdot \end{aligned} \tag{94}$$

Appendix D (Continued)

Again we can replace $\mathbf{w}$ with $\widetilde{\mathbf{w}}$ and have a new identity as in (92), where $\widetilde{\mathbf{w}}$ is constrained to be non-negative.