ANALYSIS TECHNIQUES FOR DIFFUSION TENSOR IMAGING DATA

Meagan E. Clement

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctorate of Public Health in the School of Public Health Department of Biostatistics.

Chapel Hill
2008

Approved by:
David Couper
Keith Muller
J. Steve Marron
Hongtu Zhu
Joseph Piven
ABSTRACT

MEAGAN CLEMENT: Analysis Techniques for Diffusion Tensor Imaging Data
(Under the direction of David Couper)

A recent protocol innovation with magnetic resonance imaging (MRI) has resulted in diffusion tensor imaging (DTI). The approach holds tremendous promise for improving our understanding of neural pathways, especially in the brain. MRIs work by recording displacements at a molecular level. The DTI protocol highlights the distribution of water molecules (in three dimensions). In a medium with free water motion, the diffusion of water molecules is expected to be isotropic, the same in all directions. With water embedded in nonhomogeneous tissue, motion is expected to be anisotropic, not the same in all directions, and might show preferred directions of mobility. DTI fully characterizes diffusion anisotropy locally in space, thus providing rich detail about tissue microstructure. However, little has been done to define metrics or describe credible statistical methods for analyzing DTI data.

This dissertation will show that the Geisser-Greenhouse sphericity estimator can be approximated by a squared beta distribution. Noise will also be added to show these fits also work for simulated diffusion tensors. Diagnostics are extremely important prior to analyzing these data. There are various regions, especially in the brain, where the distribution of the fractional anisotropy values could be bimodal. This is most likely due to partial voluming affects in imaging, where a voxel (volume of space) may incorporate more than the region of interest. However, the bimodal distribution can also be the result of picking up both white and grey matter in the region. If checks are not done prior to the analysis, all the results may be incorrect, since the main assumption (approximate $F$) would not be valid. By using
diagnostic approaches like QQ-envelop and SiZer, one can examine whether the approximations are reasonable. If appropriate, the methodology previously discussed can be used. However, if these approximations do not apply, new methods will be necessary to analyze the data. Different methods for analyzing the data will be considered, these methods will include: finding an approximate bimodal distribution and the DiProPerm (Direction PROjection PERMutation) test.
ACKNOWLEDGEMENTS

This dissertation has involved many people, besides myself, and I am indebted to them for their help, guidance, and patience. I would like to thank Dr. Heather C. Hazlett and Rachel Gimpel Smith in the Department of Psychiatry for all of their help in acquiring the necessary data, as well as researching individual subjects. I am especially grateful to my parents, friends, and family for all of their encouragement and support.
# TABLE OF CONTENTS

**LIST OF FIGURES** ........................................................................................................................... vii

**LIST OF ABBREVIATIONS** ............................................................................................................... ix

**Chapter**

I. INTRODUCTION ................................................................................................................................. 1

Motivation ........................................................................................................................................... 1

DTI Summary Measures ..................................................................................................................... 3

Wishart Sphericity Tests .................................................................................................................... 4

II. NOTATION AND KNOWN RESULTS ............................................................................................. 6

Matrices .............................................................................................................................................. 6

DTI Definitions ................................................................................................................................... 7

Distribution Theory ........................................................................................................................... 9

Diffusion Measures in Terms of Wishart Matrices ....................................................................... 11

First Moment Properties of Eigenvalues (Component Variances) .............................................. 12

Second Moment Properties of Eigenvalues .................................................................................. 13

Fractional Anisotropy ....................................................................................................................... 13

Diagnostic Techniques .................................................................................................................... 13

Analysis Techniques ......................................................................................................................... 21

III. APPROXIMATIONS OF THE GEISSER-GREENHOUSE SPHERICITY ESTIMATOR DISTRIBUTION: PAPER 1 .............................................................. 24

IV. DIAGNOSTIC TECHNIQUES FOR DIFFUSION TENSOR IMAGING DATA: PAPER 2 ......................................................................................................................... 33
LIST OF FIGURES

Figure 1. Diffusion in two types of samples................................................................. 2
Figure 2. DTI voxel representations............................................................................... 4
Figure 3. Illustration of two class discrimination with separating hyperplane............. 21
Figure 4. Approximate and simulated densities of $\hat{c}$ when $\epsilon \approx 0.353$ and $p = 3$......... 27
Figure 5. Approximate and simulated densities of $\hat{c}$ when $\epsilon \approx 0.496$ and $p = 3$......... 27
Figure 6. Approximate and simulated densities of $\hat{c}$ when $\epsilon \approx 0.889$ and $p = 3$......... 28
Figure 7. Approximate and simulated densities of $\hat{c}$ when $\epsilon \approx 0.89$ and $p = 4$.......... 29
Figure 8. Approximate and simulated densities of $\hat{c}$ when $\epsilon \approx 0.496$ and SNR = 5 .......... 30
Figure 9. Approximate and simulated densities of $\hat{c}$ when $\epsilon \approx 0.889$ and SNR = 5 .......... 31
Figure 10. Six histograms of the fits of the transformed right cerebellum data.............. 35
Figure 11. QQ-envelop plot of the subject with the worst fit from Figure 10..................... 36
Figure 12. QQ-envelop plot of the subject with the best fit from Figure 10...................... 37
Figure 13. Six histograms of the fits of the transformed corpus callosum data.............. 38
Figure 14. QQ-envelop plot of the subject with the worst fit from Figure 13..................... 39
Figure 15. SiZer plot of the subject with the worst fit from Figure 13............................ 40
Figure 16. QQ-envelop plot of the subject with the best fit from Figure 13...................... 41
Figure 17. SiZer plot of the subject with the best fit from Figure 13............................... 43
Figure 18. Raw FA corpus callosum data for all 32 subjects once DWD was performed..... 47
Figure 19. Transformed FA corpus callosum data for all 32 subjects once DWD was performed.................................................................................................................. 48
Figure 20. Transformed FA corpus callosum data for all 32 subjects once DWD was performed on 31 subjects (all but outlier)................................................................. 50
Figure 21. DiProPerm results for the raw FA data of the corpus callosum on all 32 subjects............................................................................................................................ 52
Figure 22. DiProPerm results for the transformed FA data of the corpus callosum on all 32 subjects........................................................................................................................................... 53

Figure 23. DiProPerm results for the transformed FA data of the corpus callosum on 31 subjects (all but the outlier).................................................................................................................. 54

Figure 24. Approximate and simulated densities of $\hat{\epsilon}$ when $\epsilon \approx 0.496$ and SNR = 10........ 57

Figure 25. Approximate and simulated densities of $\hat{\epsilon}$ when $\epsilon \approx 0.889$ and SNR = 10........ 57

Figure 26. Approximate and simulated densities of $\hat{\epsilon}$ when $\epsilon \approx 0.496$ and SNR = 15....... 58

Figure 27. Approximate and simulated densities of $\hat{\epsilon}$ when $\epsilon \approx 0.889$ and SNR = 15 ....... 58

Figure 28. Approximate and simulated densities of $\hat{\epsilon}$ when $\epsilon \approx 0.496$ and SNR = 20 ...... 59

Figure 29. Approximate and simulated densities of $\hat{\epsilon}$ when $\epsilon \approx 0.889$ and SNR = 20 ...... 59
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>DiProPerm</td>
<td>Direction projection permutation</td>
</tr>
<tr>
<td>DTI</td>
<td>Diffusion tensor imaging</td>
</tr>
<tr>
<td>DW</td>
<td>Diffusion weighted</td>
</tr>
<tr>
<td>DWD</td>
<td>Distance weighted discrimination</td>
</tr>
<tr>
<td>FA</td>
<td>Fractional anisotropy</td>
</tr>
<tr>
<td>HDLSS</td>
<td>High dimension low sample size</td>
</tr>
<tr>
<td>LBI</td>
<td>Locally best invariant</td>
</tr>
<tr>
<td>SiZer</td>
<td>Significance of zero crossing of derivatives</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to noise ratio</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 Motivation

A recent protocol innovation with magnetic resonance imaging (MRI) has resulted in diffusion tensor imaging (DTI). The approach holds tremendous promise for improving our understanding of neural pathways, especially in the brain. MRIs work by recording displacements at a molecular level. The DTI protocol highlights the distribution of water molecules (in three dimensions). In a medium with free water motion, the diffusion of water molecules is expected to be isotropic, the same in all directions (Figure 1). With water embedded in nonhomogeneous tissue, motion is expected to be anisotropic, not the same in all directions, and might show preferred directions of mobility (Figure 1). DTI characterizes diffusion anisotropy locally in space, thus providing rich detail about tissue microstructure. DTI allows tracking fibers in the brain, a result which has many potential applications. Combining fiber tracking with functional MRI seems likely to elucidate many structure-function relationships. Due to the fact that MRI protocols are noninvasive and are deemed to provide essentially no risk to the participant, longitudinal studies of both diseased and normal participants seem especially promising. DTI has already been used to show subtle abnormalities in a variety of diseases; for example, stroke, multiple sclerosis, dyslexia, and schizophrenia (Le Bihan et. al. 2001). The work to be presented was stimulated directly by a longitudinal study in progress at the University of North Carolina Neurodevelopmental Disorders Research Center. The study focused on whether a difference in brain white matter integrity between autistic, developmentally delayed, and normal children could be detected. DTI images and data from 53 independent patients were acquired; however, a method for describing each individual patient and analyzing differences between the groups was needed.
Unfortunately, little has been done to define metrics or describe credible statistical methods for analyzing DTI data. Clement (2005) developed a methodological sequence guided by basic principles to help address what measurement should be used to analyze DTI data. It was shown that one-to-one transformations of the fractional anisotropy (FA) measures lead to accurate representations of their observed distributions in terms of only two estimated parameters each. Using this transformed value will lead to outcomes in statistical models that avoid the “curse of dimensionality” (having far more variables than independent sampling units). Clement also described exact distributional results and a similar analysis for the average diffusion coefficient (ADC).

Once FA is transformed and a distribution is approximated, it is of the utmost importance to have the diagnostic tools available to determine if the approximations are reasonable for the region of interest. Issues of combining different tissue types or different regions of the brain in an analysis could lead to a multimodal distribution of the transformed FA values. If this is true, new analysis approaches must be considered. This paper will
address whether the approximations are adequate, presenting diagnostic tools for showing if
the approximations work on the data, and then discussing new ways of analyzing data if
multimodality occurs.

1.2 DTI Summary Measures

With early diffusion MRI techniques, diffusion was described using a single, scalar
parameter, the diffusion coefficient, D. However, since diffusion can occur in all three
dimensions, it is more fitting to use a diffusion tensor, $D$. Westin et. al. (1999) showed how
to calculate $D$ and then presented a decomposition of $D$ based on its symmetric properties.
Relationships among the eigenvalues of a diffusion tensor classify it according to
geometrically meaningful criteria. Hence Westin et. al. described how to compute the
closeness of a diffusion tensor to the generic cases of a line, plane and sphere. In turn, the
ratio of the smallest and largest eigenvalues gives a measure of anisotropy, which
corresponds to describing the deviation from the spherical case.

Le Bihan et. al. (2001) took a different approach to extracting information from DTI
data. They thought of $D$ as a $3 \times 3$ estimated covariance matrix, $\Sigma$, at the location of
interest. Diffusion data could be analyzed in three ways to provide information on tissue
microstructure and architecture for each voxel. 1) The mean diffusivity characterizes the
overall mean-squared displacement of molecules and the overall presence of obstacles to
diffusion. 2) The degree of anisotropy describes how much the molecular displacements vary
in space and how they are related to the presence of oriented structures. 3) The main
direction of diffusivity reflects the orientation in the space of structures. To obtain an
accurate evaluation of the probability distribution of diffusion in a region, one must avoid
anisotropic diffusion effects and use an orientation invariant measure. Any such invariant
measure can be expressed as a function of the eigenvalues of $\Sigma$. A commonly used invariant
index is fractional anisotropy. Fractional anisotropy is a measurement of the fraction of the
"magnitude" that can be ascribed to anisotropic diffusion.
Figure 2 represents how DTI data can be viewed. The graphic on the left shows individual voxels, $\mathbf{D}$, represented as ellipsoids; ellipsoids that appear more elliptical are anisotropic, while those that appear spherical are isotropic. The data can then be looked at in many different ways. The first is a MD (mean diffusivity) map (Figure 2 right top), which displays the mean diffusivity of each $\mathbf{D}$. The second is an FA (fractional anisotropy) map, which displays the fractional anisotropy measure of each $\mathbf{D}$. The third is a fiber extraction map, where fiber bundles in the brain can be depicted.

### 1.3 Wishart Sphericity Tests

A multivariate Gaussian has a sample covariance following a Wishart distribution (Muller and Stewart, 2006). Later, a one-to-one correspondence with DTI analysis will be shown.

Generally, sphericity corresponds to the statement that all correlations among differences among directions are constant. In the absence of sphericity, Box (1954) proposed
quantifying the deviation from sphericity with a parameter that is defined as the square of the trace of the covariance matrix divided by the trace of the squared covariance matrix. Using this definition, the reciprocal sample value is a simple multiple of the locally best invariant (LBI) test statistic for sphericity (John 1972). The likelihood ratio (LR) test statistic for sphericity is a function of the determinant of the covariance matrix over the squared trace of the covariance matrix. Adjusting the formula that Khatri and Srivastava (1971) derived for the exact non-null distribution for the LR test statistic and extending the exact null density function obtained by John (1972) for the LBI test, Sugiura (1995) derived exact formulas for the non-null density function of the LBI and LR tests for testing sphericity in trivariate normal distributions. Power comparisons were also made and Grieve's (1984) conjecture that the LBI test has more power if the population deviation from sphericity is large was confirmed.
2. NOTATION AND KNOWN RESULTS

2.1 Matrices

A vector (a column) is lower case bold, \( \mathbf{y} \), and a matrix is upper case bold, \( \mathbf{D} \), with transpose \( \mathbf{D}' \). Here, \( \mathbf{I}_n \) is an \( n \times n \) identity matrix, \( \mathbf{1}_n \) is an \( n \times 1 \) vector of 1's, and \( \text{Diag}(\mathbf{x}) \) is a diagonal matrix with \( (j, j) \) element \( x_j \). The rank of a matrix is the maximum number of linearly independent rows or columns. A square and full rank matrix, \( \mathbf{A} \), has a unique and full rank inverse, \( \mathbf{A}^{-1} \). Schott (1997) has details of matrix properties that are not formally given. The trace of \( \mathbf{D} \), \( \text{tr}(\mathbf{D}) \), is equal to the sum of the diagonal elements of \( \mathbf{D} \) and the determinant of \( \mathbf{D} \), \( \det(\mathbf{D}) \), will also be denoted as \( |\mathbf{D}| \). The eigenvalues of \( \mathbf{D} \), namely \( \lambda \), are defined as the roots to the characteristic equation \( |\mathbf{D} - \lambda \mathbf{I}| = 0 \). The trace and determinant of a matrix have simple relationships to the eigenvalues: \( \text{tr}(\mathbf{D}) = \sum_{i=1}^{n} \lambda_i \), and \( |\mathbf{D}| = \prod_{i=1}^{n} \lambda_i \).

If \( \mathbf{D} \) is a symmetric \( n \times n \) matrix, there exists an \( n \times n \) orthonormal matrix, \( \mathbf{V} \), and an \( n \times 1 \) vector, \( \lambda \), such that \( \mathbf{D} = \mathbf{V} \text{Diag}(\lambda) \mathbf{V}' \), which provides the spectral decomposition. Here, \( \mathbf{VV}' = \mathbf{V}' \mathbf{V} = \mathbf{I} \) and \( \text{Diag}(\lambda) \) is the diagonal matrix of eigenvalues. The columns of \( \mathbf{V} = [\mathbf{v}_1 \ldots \mathbf{v}_n] \) are the eigenvectors of \( \mathbf{D} \), corresponding to the eigenvalues. The trace and determinant of \( \mathbf{D} \) are invariant to orthonormal transformations of the form \( \mathbf{ODO}' \), for \( \mathbf{OO}' = \mathbf{O'O} \).

The following notation defines the arithmetic mean, geometric mean, mean squared-value, and variance for any set of \( p \) values, \( \lambda_i \):
\[ \mu_1 = \bar{\lambda} = \sum_{k=1}^{p} \lambda_k / p \]  
\[ \mu_2 = \prod_{k=1}^{p} \lambda_k \]  
\[ \mu_2^2 = \frac{\sum_{k=1}^{p} \lambda_k^2 / p}{\sum_{k=1}^{p} \lambda_k / p} = \frac{\sum_{k=1}^{p} \lambda_k - \bar{\lambda}}{p} = \frac{\sum_{k=1}^{p} \lambda_k^2}{p} - \bar{\lambda}^2 \]  

Although expressed here in terms of population constants, the second central moment will be referred to as the variance.

### 2.2 DTI Definitions

Diffusion tensors are estimated from the raw data contained in diffusion-weighted (DW) images using a relationship between the measured echo attenuation in each voxel and the applied magnetic field gradient sequence. From the following formula, the diffusion tensor is related to the measured echo magnitudes:

\[ \ln(A(\mathbf{b}) / A(\mathbf{b} = 0)) = -\sum_{i=1}^{3} \sum_{j=1}^{3} b_{ij} D_{ij} \]  
\[ = -\text{tr}(\mathbf{b}D) \]

where \( A(\mathbf{b}) \) and \( A(\mathbf{b} = 0) \) are the echo magnitudes of the diffusion weighted and non-diffusion weighted signals, respectively, and \( b_{ij} \) is the component of the \( b \)-matrix, where the \( b \)-matrix summarized the attenuating effect of all gradient waveforms applied in the \( x \), \( y \), and \( z \) directions. Each DW image and its corresponding \( b \)-matrix is used to estimate \( \mathbf{D} \) using multivariate linear regression of (5) (Basser and Jones 2002). Thus, each voxel is dependent on the \( b \)-matrix, which is a function of the diffusion sensitizing gradient strengths and duration. Diffusion tensor measurements require that images be acquired with at least six different \( b \)-value matrices, \( \mathbf{b} \). A seventh measurement is required with no diffusion weighting to provide a reference measure of signal intensity without a diffusion gradient \( (A(\mathbf{b} = 0)) \) (Basser and Pierpaoli 1998).

Here \( n \) is the number of images that are collected with different diffusion weightings and non-collinear gradient directions, \( \mathbf{S}_0 \) is the signal intensity in the absence of a diffusion-
sensitizing field gradient, and $S$ is the signal intensity in the presence of gradient 

$$ g = (g_x, g_y, g_z)' \). The loss of signal intensity due to diffusion is given by the Stejskal-Tanner formula: 

$$ \ln(S) = \ln(S_0) - \gamma^2 \delta^2 (\Delta - \delta/3) g' D g, $$

where $\gamma$ is the gyromagnetic ratio of $^1\text{H}$ (protons), $\delta$ is the duration of the diffusion sensitizing gradient pulses and $\Delta$ is the time between the centers of the two gradient pulses. Using the $n$ images, a system of equations is used to solve for the unknowns, the 6 elements of the symmetric diffusion tensor, $D$, and $S_0$ (Westin et al. 1999).

Zhu et al. (2007) proposed a semi-parametric model to fit the log-transformed signal intensities in diffusion-weighted MRI data that also characterizes the random noise in the magnitude of the observed signal intensity. If there are $n$ DW images for each subject, with each image containing $N$ voxels, each of those voxels consists of $n$ diffusion-weighted measurements. Let $S_i$, $r_i$, and $b_i$ be the $n$ DW measurements at a single voxel in the human brain, where $S_i$ is the signal intensity of the MR image, $r_i$ is a $1 \times 3$ vector that represents the $i$th direction of the diffusion gradient such that $r_i' r_i = 1$, and $b_i$ is the corresponding $b$ factor of each $i^{th}$ DW MRI (Stejskal and Tanner 1965; Anderson 2001; Kingsley 2006). A weighted least squares (WLS) estimate of the diffusion tensors is then provided from a semi-parametric model.

The following heteroscedastic linear model to fit log-transformed signal intensities was considered:

$$ \log S_i = \log S_o - b_i r_i' D r_i + \eta_i = z_i \theta + \exp(-z_i' \theta) \sigma \epsilon_i, $$

where $i \in \{1, \ldots, n\}$, $\theta'$ is a column vector with $\log S_o$ as its first entry and the six unique elements of $D$ as its other components (in the order: $D_{11}$, $D_{12}$, $D_{13}$, $D_{22}$, $D_{23}$, $D_{33}$), $\eta_i = \exp(-z_i' \theta) \sigma \epsilon_i$, and the errors $\epsilon_i$ are independent random variables that have mean zero and finite variances. $z_i$ is a $7 \times 1$ vector with the first row equal to 1 and the subsequent rows to be as follows: $-b_i r_{i,1}^2$, $-2b_i r_{i,1} r_{i,2}$, $-2b_i r_{i,1} r_{i,3}$, $-b_i r_{i,2}^2$, $-2b_i r_{i,2} r_{i,3}$, $-b_i r_{i,3}^2$. The WLS algorithm for the above model is as follows:

1) Set $k = 0$ and calculate the initial $\hat{\theta}_{(k)} = \left( \sum_{i=1}^{n} z_i z_i' \right)^{-1} \sum_{i=1}^{n} z_i \log S_i$. 

8
2) Calculate $\omega_i^{(k)} = \exp\left(2z_i'\hat{\theta}^{(k)}\right)$ for all $i$.

3) Calculate $\hat{\theta}^{(k+1)}$ using the following equation

$$
\hat{\theta}^{(k+1)} = \left(\sum_{i=1}^{n} \omega_i^{(k)} z_i z_i'\right)^{-1} \sum_{i=1}^{n} \omega_i^{(k)} z_i \log S_i. \tag{6}
$$

4) Repeat steps 2 and 3 for $k_0$ iterations to get the estimate $\hat{\theta}^{(k_0)}$.

This estimate will contain the unique element of $\hat{D}$, the estimated diffusion tensor (Zhu et al. 2007).

The DTI variable that is the most useful in analyzing data is fractional anisotropy (Clement 2005). Fractional Anisotropy, $\phi$, is defined in terms of

$$
\phi^2 = \frac{3\sum_{k=1}^{3} (\lambda_k - \bar{\lambda})^2}{2\sum_{k=1}^{3} \lambda_k^2} = \frac{3 \cdot 3 [\mu_2' - (\mu_1')^2]}{2 \cdot 3 \mu_2'} = \frac{3 \left[\mu_2' - (\mu_1')^2\right]}{2 \mu_2'}. \tag{7}
$$

Here $\phi$ is a measure of the dispersion (variance) of the variances of the diffusion tensor (Le Bihan et al. 2001).

### 2.3 Distribution Theory

Johnson, Kotz and Balakrishnan (1995) and Kotz, Balakrishnan and Johnson (2000) provide detailed properties of the following random variables. Writing $B \sim \beta(\nu_{s1}, \nu_{s2})$ indicates that $B$ follows a Beta distribution with $\nu_{s1}$ and $\nu_{s2}$ as the shape parameters. Writing $X \sim \chi^2(\nu, \omega)$ indicates that $X$ follows a chi-square distribution, with $\nu$ degrees of freedom and noncentrality $\omega$. Likewise, writing $R \sim F(\nu_1, \nu_2, \omega)$ indicates that $R$ follows a noncentral $F$ distribution, with numerator degrees of freedom $\nu_1$, denominator degrees of freedom $\nu_2$, and noncentrality $\omega$. Writing $\chi^2(\nu)$ and $F(\nu_1, \nu_2)$ implies $\omega = 0$. Finally, writing $y \sim N_p(\mu, \Sigma)$ indicates $y$ is vector Gaussian with mean $E(y) = \mu$ and covariance $\mathcal{V}(y) = \Sigma$. If an $n \times p$ matrix $Y$ has independent rows and row$_i(Y)' \sim N_p(\mu_i, \Sigma)$, then
\[ \mathbf{Y}' \mathbf{Y} = \mathbf{S} \sim \mathcal{W}_p(\nu, \Sigma, \Delta) \] indicates that \( \mathbf{S} \) has a Wishart distribution with \( \nu \) degrees of freedom, covariance \( \Sigma \), and \( \Delta = \mathbb{E}(\mathbf{Y}')\mathbb{E}(\mathbf{Y}) \). Additionally, \( \mathbf{Y} \sim \mathcal{N}_{\nu, p}(\mathbb{E}(\mathbf{Y}), \mathbf{I}_\nu, \Sigma) \) indicates that \( \mathbf{Y} \) has a matrix Gaussian distribution with mean of \( \mathbb{E}(\mathbf{Y}) \), covariance structure of the columns, within a row, of \( \Sigma \), and the covariance structure of the rows, within a column, of \( \mathbf{I}_\nu \).

A population or sample covariance matrix is always symmetric and can be expressed as an inner product. The covariance matrix will always have a spectral decomposition with only positive or zero eigenvalues. If \( \Upsilon \) is the matrix of eigenvectors of the covariance matrix, then \( \Sigma = \Upsilon \mathbf{D}_g(\lambda) \Upsilon' \) with \( \Upsilon' \Upsilon = \mathbf{I}_b \). Also, if \( \mathbf{Z} \sim \mathcal{N}_{\nu, b}(\mathbf{M}_Z, \mathbf{I}_\nu, \mathbf{I}_b) \) and \( \Sigma = \Phi \Phi' \), where \( \Phi = \Upsilon \mathbf{D}_g(\lambda)^{1/2} \), then \( \mathbf{Y} = \mathbf{Z} \Phi' \sim \mathcal{N}_{\nu, b}(\mathbf{M}_Y, \mathbf{I}_\nu, \Sigma) \) with \( \mathbf{M}_Y = \mathbf{M}_Z \Phi' \) and

\[ \mathbf{Y}' \mathbf{Y} \sim \mathcal{W}_b(\nu, \Sigma, \mathbf{M}_Y' \mathbf{M}_Y) \] (Muller and Stewart 2006).

The locally best invariant (LBI) test for testing sphericity (\( H_0 : \Sigma = \sigma^2 \mathbf{I}_p \)) for unknown \( \sigma^2 \), against all alternatives, is to reject the null hypothesis for large values of \( U = \text{tr}(\mathbf{S}^2)/(\text{tr}(\mathbf{S}))^2 \) (John 1971). For this paper, \( \epsilon \) is defined as:

\[
\epsilon = \frac{\text{tr}^2(\Sigma)}{[p \text{tr}(\Sigma^2)]} = \frac{\left(\sum_{j=1}^{p} \lambda_j\right)^2}{\left(p \sum_{j=1}^{p} \lambda_j^2\right)} = \frac{\left(\sum_{j=1}^{p} \lambda_j/p\right)^2}{\left(p \sum_{j=1}^{p} \lambda_j^2/p\right)} = \frac{\left(\bar{\lambda}\right)^2}{\bar{\lambda}^2} = \frac{(\mu_1')^2}{\mu_2'}.
\]

Thus, the maximum likelihood estimate (MLE), \( \hat{\epsilon} \), of the parameter \( \epsilon \) is a one-to-one function of the LBI test for sphericity, \( \hat{\epsilon} = 1/(pU) \).

When \( p = 3 \), which is the case in diffusion tensors, the exact density function of \( U \) (Sugiura 1995), under the null hypothesis is:
\[ b_U(u) = \begin{cases} c_\nu \{ 5 - 9u + \sqrt{2(3u - 1)^{3/2}} \}^{\nu/2 - 1} & \frac{1}{3} \leq u \leq \frac{1}{2} \\ c_\nu \{ 5 - 9u - \sqrt{2(3u - 1)^{3/2}} \}^{\nu/2 - 1} & \frac{1}{2} < u \leq 1 \end{cases} \] (10)

with \( c_\nu = \sqrt{\pi} \Gamma(3\nu/2) / \{ \Gamma(\nu/2) \Gamma((\nu - 1)/2) \} \) where \( \Gamma(x) = (x - 1)! \) when \( x \) is a positive integer.

Let
\[ g_0(x, u) = \frac{\sqrt{\pi} \Gamma(3\nu/2)}{\Gamma(\nu/2) \Gamma((\nu - 1)/2) \Gamma(\nu - 2)/2} \left( \frac{1 - u}{2} \right)^{\nu/2 - 2} \times \left( \frac{1 - u}{2} - 2x + 3x^2 \right) \] (11)

and
\[ g_1(x, u) = (\gamma_1 \gamma_2)^{-\nu/2} \sum_{k=0}^{\infty} \frac{(3\nu/2)_k}{(k!)^2 2^{3k} [1 + (\gamma_1^{-1} + \gamma_2^{-1} - 2)x]^{k + 3\nu/2}} \times \sum_\kappa \frac{(2k_1 + 1)(2k_2)! C_\kappa(A) C_\kappa(X)}{(2k_1 - 2k_2 + 1)(3/2)_{k_1} k_2!} \] (12)

with \( \gamma_i = \lambda_i / \lambda_3, (a)_k = a(a + 1) \cdots (a + k - 1), (a)_0 = 1, A = Dg(1 - 1/\gamma_1, 1 - 1/\gamma_2), X = Dg(1 - 3x + \sqrt{2u - 3x^2 + 2x - 1}, 1 - 3x - \sqrt{2u - 3x^2 + 2x - 1}) \). Here \( \sum_\kappa \) stands for the sum of all possible partitions \( \kappa = \{ k_1, k_2 \} \) of non-negative integer \( k \) satisfying \( k = k_1 + k_2 \) and \( k_1 \geq k_2 \geq 0 \). Also, \( C_\kappa(\cdot) \) is the zonal polynomial corresponding to the partition \( \kappa \). Then the non-null density of \( U \) is
\[ f_U(u) = \begin{cases} \int_{1/3}^{1/2} g_0(x, u) \times g_1(x, u) \, dx, & \frac{1}{3} \leq u \leq \frac{1}{2} \\ \int_{1/2}^{1} g_0(x, u) \times g_1(x, u) \, dx, & \frac{1}{2} < u \leq 1 \end{cases} \] (13)

Although the density exists, it involves zonal polynomials, which seems likely to cause problems with computation.

2.4 Diffusion Measures in Terms of Wishart Matrices

Given the assumption that the flow of water follows a Gaussian diffusion model arising from Brownian motion theory \( \Sigma \) can be defined as the covariance matrix (tensor) of the
diffusion. In other words, \( \Sigma \) will indicate the population covariance of diffusion. As a consequence of the assumption, \( \nu \hat{\Sigma} \sim \mathcal{W}_p(\nu, \Sigma) \) (follows a Wishart distribution) with \( \nu \) being determined by the number of replicates used to find \( \hat{\Sigma} \) and \( p \) equaling the number of rows in \( \hat{\Sigma} \). The eigenvalues of \( \hat{\Sigma} \), \( \{\lambda_i\} \), are estimates of variances of underlying principal components and hence measures of diffusion in orthogonal dimensions. The most popular measures of diffusion arising from DTI analysis can be expressed solely as functions of the sample eigenvalues which are, in fact, estimated variances.

### 2.5 First Moment Properties of Eigenvalues (Component Variances)

**Trace and ADC.** Interpreting the eigenvalues, \( \{\lambda_i\} \), as measures of variance, then \( \mu_1 \) is the average variance, or the arithmetic mean of the variances. When \( \mathcal{S} = \nu \hat{\Sigma} \) is a Wishart, \( p^{-1} \text{tr}\left(\hat{\Sigma}\right) \) is often called the generalized variance. Johnson, Kotz and Balakrishnan (2000) expressed the trace of a singular covariance matrix with degrees of freedom less than its dimension in terms of a weighted sum of chi-square random variables. Glueck and Muller (1998) derived that the trace of a Wishart equals a weighted sum of noncentral chi-square random variables and constants. The average diffusion coefficient (ADC) is the trace of the tensor, hence it is exactly distributed as a weighted sum of central chi-square random variables (Glueck and Muller 1998). The exact distribution of the ADC can always be computed. An approximate and highly reliable distribution is also available. Kim, Gribbin, Muller, and Taylor (2005) provide a convenient review of exact and approximate calculations of probabilities for such quadratic forms.

**Determinant.** If \( \{\lambda_i\} \) are thought of as measures of principal variation, then

\[
\mu_\gamma = \left(\mu_\gamma^p\right)^{(1/p)}
\]

is the geometric mean of the variances. The sample generalized variance can also be defined as \( |\hat{\Sigma}| \). Although it is common in statistics to discuss \( |\hat{\Sigma}| \) as the generalized variance, it seems more natural to look at the geometric mean, \( \tilde{\mu}_\gamma = \sqrt[p]{|\hat{\Sigma}|} \). Gupta and Nagar (2000, Chapter 3) showed the following. If \( \mathcal{S} \sim \mathcal{W}_p(\nu, \Sigma) \), \( |\mathcal{S}|/|\Sigma| \sim \prod_{i=1}^p u_i \), with
independent \( \{\mu_i\} \) and \( u_i \sim \chi_{\nu-1}^2 \), where \( i \in \{1, \ldots, p\} \). Also,

\[
E\left(|S|^h\right) = 2^h |\Sigma|^h \prod_{i=1}^{p} \left\{\Gamma[(1/2)(\nu - i + 1) + h] / \{\Gamma[(1/2)(\nu - i + 1)]}\right\}.
\]

### 2.6 Second Moment Properties of Eigenvalues

In order to achieve global scale invariance, the measures of dispersion of diffusion (anisotropy) are standardized; thus the central information will remain unchanged if a linear transformation is applied. The main goal is to see if the variances are the same in all three dimensions. By using the parameter estimates \( \{\hat{\lambda}_i\} \), estimates for \( \epsilon \) can be obtained. Box (1954) showed \( \epsilon \) was a function of sphericity; thus, a one-to-one function of \( \epsilon \) provides the locally best invariant test for sphericity by (9). Using this information, FA can be expressed as one-to-one functions of \( \hat{\epsilon} \), the LBI test statistic for sphericity.

#### 2.7 Fractional Anisotropy, \( \phi \)

By Equation 7,

\[
\phi^2 = \frac{3 \cdot 3 [\mu'_2 - (\mu'_1)^2]}{2 \cdot 3 \mu'_2} = \frac{3}{2} (1 - \epsilon).
\]

Hence \( \hat{\phi}^2 \) is scale invariant and

\[
\hat{\epsilon} = 1 - \frac{2}{3} \hat{\phi}^2.
\]

Thus, a linear function of \( \hat{\phi}^2 \) is a one-to-one function of a LBI test for sphericity (Clement 2005). The LBI test for sphericity will be more powerful with values of \( \epsilon \) near one (Sugiura 1995). Experience with DTI brain data shows that the values of \( \hat{\epsilon} \) fit this case. We will show that \( \hat{\epsilon} \) can be approximated by a squared beta distribution. Thus, FA can be approximated by a squared beta distribution.

### 2.8 Diagnostic Techniques
2.8.1 Kernel Density Estimation

The histogram is a widely used tool for displaying the distributional shape of a set of data. Its usefulness lies in the fact that it indicates the shape of the underlying density function. An alternative to estimate the density function is a smooth curve. In order to discuss the construction of estimators of this type, it is important to first consider the construction of a histogram. However, when viewed as an estimate, the histogram can be criticized in the following ways: 1) information is thrown away when the observed values are replaced by a central point in the interval in which they fall; 2) the underlying density function is usually assumed to be smooth, but the estimator is not smooth, due to sharp edges of the boxes from which it is built; and 3) the behavior of the estimator is dependent on the choice of width of the intervals used, and also to some extent, on the starting position of the grid of intervals (Bowman and Azzalini 1997, Chapter 1).

Whittle (1958) and Parzen (1962) developed an approach to the problem which removes the first two of these issues. A smooth kernel function, rather than a box, is used as the basic building block. These smooth functions are centered directly over each observation. The kernel estimator is then of the form:

$$f(y) = \frac{1}{n} \sum_{i=1}^{n} w(y - y_i; h),$$

where $w$ is a probability density called the kernel function, whose variance is controlled by the parameter $h$. Because of its role in determining the way in which the probability associated with each observation is spread over the surrounding sample space, $h$ is called the smoothing parameter or bandwidth. Since properties of $w$ are inherited by $\hat{f}$, choosing $w$ to be smooth will produce a density estimate which is also smooth.

The basic properties of $\hat{f}$ are well documented (Bowman and Azzalini 1997, Chapter 2). The mean of the density estimator can be written as

$$E(f(y)) = \int w(y - z; h) f(z) \, dz.$$
This is a convolution of the true density function with the kernel function $w$. Smoothing has thus produced a biased estimator, whose mean is a smoothed version of the true density. 

Using a Taylor series approximation argument, we can approximate the expected value

$$E\left(\hat{f}(y)\right) \approx f(y) + \left(h^2/2\right)\sigma_w^2 f''(y), \tag{18}$$

where $\sigma_w^2$ denotes the variance of the kernel function, namely $\int z^2 w(z) dz$. Since $f''(y)$ measures the rate of curvature of the density function, this expresses the fact that $\hat{f}$ underestimates $f$ at peaks and overestimates troughs in the true density. The size of the bias is affected by the smoothing parameter $h$.

Through another Taylor series approximation, the variance of the density estimate can be approximated by

$$\text{var}\left(\hat{f}(y)\right) \approx (1/nh)f(y)\alpha(w), \tag{19}$$

where $\alpha(w) = \int w^2(z) dz$. Note that the variance is inversely proportional to sample size. The term $nh$ can be viewed as governing the local sample size, since $h$ controls the number of observations whose kernel weight contributes to the estimate at $y$. It is also noteworthy that the variance is approximately proportional to the height of the true density function.

These approximate expressions for the mean and variance of a density estimate encapsulate the effects of the smoothing parameter. As $h$ decreases, bias diminishes while variance increases. As $h$ increases, the opposite occurs. The combined effect being that in order to produce an estimator which converges to the true density function, it is necessary that both $h$ and $1/nh$ decrease as the sample size increases.

It must also be noted the third criticism of the histogram still applies to the smooth density estimate, namely that its behavior is affected by the choice of the width of the kernel function. When $h$ is small, the estimate displays the variation associated with individual observations rather than the underlying structure of the whole sample. When $h$ is large, this structure is obscured by smoothing the data over too large a region. The asymptotically optimal choice for $h$ and three of the most common practical strategies are described below.
Optimal Smoothing

An overall measure of the effectiveness of $\hat{f}$ is provided by the mean integrated squared error (MISE). The MISE can be defined as

$$\text{MISE}(\hat{f}) = E\left\{ \int [\hat{f}(y) - f(y)]^2 dy \right\}$$

$$= \int [E\{\hat{f}(y)\} - f(y)]^2 dy + \int \text{var}\{\hat{f}(y)\} dy.$$ 

and can be approximated by

$$\text{MISE}(\hat{f}) \approx (1/4)h^4\sigma_w^4 \int f''(y)^2 dy + (1/nh)\alpha(w).$$

From this approximate expression, the value of $h$ which minimizes MISE is

$$h_{opt} = \{\gamma(w)/\beta(f)n\}^{1/5},$$

where $\gamma(w) = \alpha(w)/\sigma_w^4$, and $\beta(f) = \int f''(y)^2 dy$. However, this optimal value for $h$ cannot immediately be used in practice since it involves the unknown density function, but it is informative in showing how smoothing parameters should decrease with sample size and in quantifying the effect of the curvature of $f$ through the factor $\beta(f)$.

Normal Optimal Smoothing

If it is assumed that $f$ is a normal distribution, the following simple formula for $h_{opt}$ arises:

$$h_{opt} = \sigma(4/3n)^{1/5},$$

where $\sigma$ denotes the standard deviation of the distribution. For distributions not far from the normal, this gives a useful choice of smoothing parameter that requires very little calculation. With this, it also has the potential for being cautious and conservative.

Cross-validation

Since 1931, splitting a sample into two parts and using one for linear model selection and one for assessment has been practiced. In the 1960's, Stone first used the leave-$k$ out
approach, labeled as cross-validation. Stone (1974) gives a general description of the ideas involved in cross-validation. It must be noted that there are distinctions between cross-validation and data splitting. As Picard and Cook (1984) cautioned, "Implementation of cross-validation in the derivation of \( \hat{\beta} \) does not, however, alter the fundamentals of predictive assessment. If a proper evaluation of a selected fitted model is to be realized, the optimism principle cannot be ignored. When using data splitting, this implies that all aspects of model selection (even those that are crossvalidatory) should be confined to analysis of the estimation data and that the validation data be reserved solely for assessment." Thus, the use of the phrase "cross-validation" in this sense is synonymous with a "leaving \( k \)-out" approach. It should be noted that is not a way to validate a model, but is a way to select a model.

Rudemo (1982) and Bowman (1984) applied cross-validation ideas to the problem of bandwidth choice through estimation of the integrated squared error (ISE)

\[
\int \left\{ \hat{f}(y) - f(y) \right\}^2 dy = \int \hat{f}(y)^2 dy - 2 \int f(y) \hat{f}(y) dy + \int f(y)^2 dy. \tag{24}
\]

Note that the last term on the right hand side of the equation does not involve \( h \); thus this term can be ignored in the minimization. The second term on the right hand side of the equation can be split into two terms, one involving \( h \) and the other not (Bowman 1984). The terms that involve \( h \) can be estimated by the following formula:

\[
\frac{1}{n} \sum_{i=1}^{n} \int \hat{f}_{-i}(y) dy - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{-i}(y_i), \tag{25}
\]

where \( \hat{f}_{-i}(y_i) \) denotes the estimator constructed from the data without the observation \( y_i \) (Bowman and Azzalini 1997, Chapter 2). The expectation of this expression is the MISE of \( \hat{f} \) based on \( n - 1 \) observations, omitting the \( \int f^2 \) term. The value of \( h \) which minimizes this expression therefore provides an estimate of the optimal smoothing parameter.

**Plug-in Bandwidth**

Iterative procedures have been proposed in which an estimate of \( \hat{f} \) is used in the formula for the optimal smoothing parameter noted in (22). If normal kernels are used, \( \gamma(w) \) and
\( \beta(\hat{f}) \) can be calculated relatively easily and the value of \( h \) which solves this equation can be found by a suitable numerical algorithm.

Sheather and Jones (1991), extending the work of Park and Marron (1990), described a bandwidth selection procedure based on the estimation of \( f'' \) using an additional smoothing parameter related to \( h \). This estimator has very good finite sample, as well as asymptotic, properties. It is more stable than the cross-validation approach described above. The two techniques take separate approaches to the same problem of minimizing ISE. Cross-validation estimates the ISE function and locates the minimum. The plug-in approach minimizes the function theoretically and then estimates this minimizing value directly.

### 2.8.2 Exploratory Tool

When analyzing data using smoothing methods, it is difficult to determine whether peaks and valleys are important underlying structures or artifacts of the sampling process. Since both of these can be made to disappear by increasing the amount of smoothing or can increase the number of features if the smoothing parameter is decreased, it is hard to determine which of these is true. SiZer (based on studying statistical SIgnificance of ZERO crossings of derivatives) is an exploratory data analysis tool that works in conjunction with smoothing methods to analyze which visible features represent important underlying structures.

Developed by Chaudhuri and Marron (1998), SiZer can be used for both density estimation (smoothed histograms) and for nonparametric regression (scatter plot smoothing). SiZer investigates which of the features seen in smooths are statistically significant by studying derivatives of the smooths. There are two components to SiZer: 1) use of a family of smooths for a broad range of \( h \) and 2) a color map of scale space. By using a family approach of bandwidths, the classical need to choose a bandwidth is avoided. Also departing from the classical view, SiZer avoids the bias problem in doing inference by shifting focus away from the true underlying curve to the true curve viewed at different resolutions.
Scale-space ideas are used to provide a new view on kernel smoothing. The family of all kernel smooths indexed by bandwidth, \( h \), is a model used in computer vision. The idea behind this is that large \( h \) models give a macroscopic view where only large-scale features can be resolved; whereas, small \( h \) models give a microscopic view of the small-scale features. It is important to choose a range of \( h \) that highlights a good data based choice. Marron and Chaudhuri (1998) suggest the following defaults: evaluate the smooths at a grid of 401 equally spaced points with the smallest \( h \) equal to twice the grid spacing and the largest \( h \) equal to the range.

The SiZer map is created based on confidence limits for the derivative in scale space, \( \hat{f}'_h(x) \). These confidence limits are of the form

\[
\hat{f}'_h(x) \pm qS\hat{D}(\hat{f}'_h(x)),
\]

where \( q \) is an appropriate quantile and the standard deviation is estimated using the following definition of the variance.

\[
\text{var}(\hat{f}'_h(x)) = n^{-1}s^2(K'_h(x - X_1), \ldots, K'_h(x - X_n)),
\]

where \( s^2 \) is the usual sample variance of \( n \) numbers and \( K'_h \) is the \( h \)-rescaling of the kernel function \( K' \), such that \( K'_h(\cdot) = 1/h(K(\cdot/h)) \). An \((x, h)\) location is called significantly increasing, decreasing, or not significant when 0 is below, above or within these confidence limits, respectively. This information is displayed in the second component, a color map of scale space where each pixel represents a location with respect to both position, \( x \), and bandwidth, \( h \). A blue color (or dark if in black and white) is used on the pixel of the color map if the smooth is significantly increasing. If the smooth is significantly decreasing, the pixel is colored red (or light). If there is not a statistically significant slope, the pixel is shaded purple (or intermediate gray). Similar to the color map for slope, a color map for curvature can be computed. This plot looks at the second derivative instead of the first. If the smooth is significantly convex, an orange color is used on the pixel of the color map. If the smooth is significantly concave, the pixel is colored cyan and if there is not a statistically
significant curvature, the pixel is shaded green. This gives an additional look at the data.
The SiZer map has two important benefits. First, it speeds up the process of deciding "which features are really there" for an experienced analyst while also being able to quantitatively resolve any "gray area" problems. Second, it allows inexperienced analysts to make inferences about which features are really there (Chaudhuri and Marron 1999).

2.8.3 Q-Q plots

Quantile-quantile (Q-Q) plots are used to compare two distributions. These could be both datasets, both theoretical distributions, or most commonly, a combination of the two. The Q-Q plot is a scatterplot of quantiles of one distribution on each axis, which thus gives direct comparison of the distribution. If the points roughly lie on a plot with slope 1, then the distributions are the same. However, how does one know if values away from the line with slope one are due to variation of the sample or if the data really does not come from the given distribution?

Programs supplied by Marron (2007) delved into an approach called QQ-envelop to use a Q-Q plot to test the distributional form against standard distributions. This method creates a typical Q-Q plot, but then simulates pseudo sets of data from the assumed distribution to look at random variability. All pseudo data points, as well as the original observations, are plotted. If the original data points are enveloped by the pseudo data points and the line with a slope of 1, then the distributional fit works; if not, then the distributional assumption was not correct. This approach was used in Hernández-Campos et. al. (2004) to fit distributions to Internet traffic data. Mihee Lee extended the program written by Marron to do this analysis with beta distributions.

2.9 Analysis Techniques

High dimension low sample size (HDLSS) data is becoming increasingly common in many different fields; medical imaging is one of them. In HDLSS situations, a classification method for groups of subjects is needed; in the area of discriminant analysis this is often done
by finding the hyperplane which best separates populations. The distance between the discriminating hyperplane and the data points must be maximized while also separating the two classes. Figure 3 shows a representation of how this can be done. Let \( p_i \) be the given points from the \( c_i \in \{-1, 1\} \) classes, and \( w \) be the normal vector to the separating hyperplane. The residual, or distance, \( r_i \) is from the points to the hyperplane, can be calculated by the following function where \( \beta \) determines the position of the hyperplane.

\[
    r_i = c_i (p_i'w + \beta)
\]  

(28)

A popular method of discriminant analysis is the support vector machine (SVM). SVM attempts to maximize the minimum \( r_i \). In doing this, SVM tends to use only a small subset of the population to define the discriminating hyperplane, more specifically, those near the opposite class. However, SVM suffers from data piling (when most of the samples from the same population group end up very close to each other after being projected onto the normal of the discriminating axis) at the margin since many of the projections can be the same, which can diminish generalizability (Marron, Todd, and Ahn 2007). This lead to the development of distance weighted discrimination (DWD).
2.9.1 Distance Weighted Discrimination

DWD is a multivariate analysis tool that is able to identify systematic biases present in separate data sets and then make a global adjustment to compensate for them (Benito 2004). DWD is also a classification tool described by Marron, Todd, and Ahn (2007). The method starts by dividing the sample union into two classes by a hyperplane and then classifying this combined set as coming from one of the two populations according to whether a point lies on one side or the other of the hyperplane. DWD differs from SVM in how the hyperplane is selected. Unlike SVM, all the sample points are used in the calculation of the discriminating axis. Also, instead of trying to maximize the minimum $r_i$, DWD attempts to minimize the sum of the reciprocals of $r_i$. Thus, DWD achieves a higher robustness when presented with new samples since each point's contribution to the calculation was weighted proportionally to the distance from that point to the opposite population (Marron, Todd, and Ahn 2007). Hall et. al. (2005) developed asymptotic properties in the limit as $d \to \infty$, of SVM and DWD.

2.9.2 DiProPerm Test

If one wants to know if two subpopulations are from the same distribution, the DiProPerm test (Wichers et. al. 2006) can be used. This is very useful in our imaging scenario because one can see if the brain structure in one group is different from another. DiProPerm (Direction Projection Permutation) uses the following ideas to test if distributions are the same:

1) find an appropriate one-dimensional direction vector where the normal hyperplane effectively separates the populations;

2) project data into that one-dimensional subspace;

3) construct a one-dimensional test statistic;

4) for many permutations of class labels, repeat the first three steps;

5) analyze the significance by comparing the "true" statistic among the population of the
permutation statistics.

First, DWD is performed in order to separate the data groups. Other reasonable direction vectors would be: mean difference or support vector machines. DWD was used for this paper due to its robustness over SVM. Once the direction vector is computed, the data is projected onto the normal vector that is orthogonal to the separating hyperplane. These projections are a one-dimensional representation of the population, the Euclidean distance from the normal vector to an individual in the population. However, because the first two steps violate the traditional assumptions of standard null distributions (for example, the Student's t-distribution for the t-statistic), a permutation test is considered so that the results will be valid. Thus, for the third step, one would compute a 2-sample t-statistic on the projected data to see if there was a difference in means. Other reasonable projected one-dimensional statistics would be: chi-square test for different variances, Kolmogorov-Smirnov, or any other good distribution test. Next, the data would be permuted giving each subject a random group assignment. The first three steps are repeated and the "true" statistic (that resulting from the original dataset) is compared among the population of permuted statistics. The p-value of this test will be the quantile of the "true" statistic.
3. APPROXIMATIONS OF THE GEISSER-GREENHOUSE SPHERICITY ESTIMATOR DISTRIBUTION: PAPER 1

Approximately matching the first two moments of $\hat{\varepsilon}$ to a squared beta random variable results in a simple approximate distribution. The fact that $1/p \leq \hat{\varepsilon} \leq 1$ allows concluding

$$0 \leq \left( \hat{\varepsilon} - \frac{1}{p} \right) \left( 1 - \frac{1}{p} \right)^{-1} \leq 1. \quad (29)$$

This leads to defining

$$B^2 = \left( \hat{\varepsilon} - \frac{1}{p} \right) \left( 1 - \frac{1}{p} \right)^{-1}$$

$$= \hat{\varepsilon} \left( \frac{p}{p-1} \right) - \left( \frac{1}{p-1} \right)$$

$$= \hat{\varepsilon} c_1 - c_0. \quad (30)$$

Obviously $E B^2 = c_1 E \varepsilon - c_0$ and $E \varepsilon = (E B^2 + c_0)/c_1$. With $T_1 = tr^2(\Sigma)$ and $T_2 = tr(\Sigma^2)$, it follows that $\hat{\varepsilon} = T_1/(pT_2)$. While it is known that the following assumptions are not true, they were assumed to derive the approximate results. First, it was assumed that $T_1$ and $T_2$ are independent. Second, it is assumed that $E T_2^{-1} = (ET_2)^{-1}$. Muller, Edwards, Simpson and Taylor (2007) reported that $E(T_1) = 2\nu \sum_{k=1}^{p} \lambda_k^2 + \nu^2(\sum_{k=1}^{p} \lambda_k)^2$ and $E(T_2) = \nu(\nu + 2) \sum_{k=1}^{p} \lambda_k^2 + 2\nu \sum_{k=1}^{p} \sum_{k=2}^{p} \lambda_k \lambda_{k'}$. Also, from (30), it is known that

$$B^2 = (T_1/pT_2)[p/(p-1)] - [1/(p-1)]; \text{ thus, } B^2(p-1) + 1 = T_1/T_2.$$

The special case of sphericity leads to $T_1$ being exactly the square of a scaled, central chi-square. In general, $B' \sim \beta(\nu_{*1}/2, \nu_{*2}/2)$ is true if and only if $B' = X_1/(X_1 + X_2)$, with $X_1$ independent of $X_2$ and both distributed chi-square. It seems reasonable to find $B_* \sim \beta(\nu_{*1}/2, \nu_{*2}/2)$ so that, in some sense, $B_*^2 \approx B^2$. Then,
\[ \mathcal{E} \tilde{\varepsilon} = \left( \mathcal{E} B^2 + c_0 \right) / c_1 \]
\[ \approx \left[ \mathcal{E} B^2 \left( \frac{1}{p - 1} \right) \right] \left( \frac{p}{p - 1} \right) \]
\[ \approx \left( \frac{p - 1}{p} \right) \mathcal{E} B^2 + \left( \frac{1}{p} \right). \]

Moments of a Beta are described in Johnson, Kotz, and Balakrishnan (1995, Chapter 25). Thus, for such a \( B_s \),

\[ B_s^2 = \frac{\lambda^2 \nu^2}{\left( \lambda \nu X_1 \right)^2} \]
\[ = \frac{\lambda^2 \nu^2}{\lambda \nu X_1 + \lambda \nu X_2}, \]

with

\[ \mathcal{E} \left( X_1^2 + 2X_1X_2 + X_2^2 \right) = \mathcal{E} X_1^2 + 2\mathcal{E} X_1 \mathcal{E} X_2 + \mathcal{E} X_2^2 \]
\[ = \lambda^2 \nu \nu \left( \nu + 1 \right) + 2\lambda \nu \nu \nu + \lambda^2 \nu \nu \nu 2 \left( \nu + 1 \right). \]

As a Beta random variable, \( \mathcal{E} B_s^2 = \nu \nu \left( \nu + 1 \right) \left( \nu + \nu \nu + \nu + 1 \right) \left( \nu + 1 \right). \) Hence, by taking the expectation of the numerator and denominator separately,

\[ \mathcal{E} \tilde{\varepsilon} \approx \left( \frac{p - 1}{p} \right) \frac{\nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \nu \num

Without the \( (p - 1)/p \) term, this would be what one would expect for the variance of a squared beta distribution. Thus, the approximation depends on \( p \) such that as \( p \) increases, the approximation gets better and better since \( (p - 1)/p \to 1 \) as \( p \to \infty \).

A simulation was conducted to see how close the approximations were. Since this will be applied to DTI data, the case where \( p = 3 \) was used. Simulations were conducted using
one million replications. Using a known epsilon, random Wishart matrices were generated by first creating random normal matrices. A random number generator produced matrices to create a $Z$ matrix. The random number generator used normal values with mean zero and variance of 1. Since epsilon is a function of eigenvalues, $Dg(\lambda)$ was created to achieve the epsilon of choice. $Y$ was set equal to $ZYDg(\lambda)^{1/2}$. Since $\Sigma$ in this case would be $\Phi'\Phi = Dg(\lambda)^{1/2}$ $\Phi'YDg(\lambda)^{1/2}$ and $Y$ is orthonormal, $\Phi'\Phi = Dg(\lambda)$. Hence only $Dg(\lambda)$ and $Z$ are required to produce the Wishart matrices. The following entries comprised the matrix, $Dg(\lambda)$: $(1.00, 0.02, 0.01)$ for $\epsilon = 0.353$, $(0.80, 0.09, 0.10)$ for $\epsilon = 0.496$ and $(1.00, 0.55, 0.55)$, for $\epsilon = 0.889$. By (8), $Y'Y$ was the resulting random Wishart matrix. For each $Y'Y$, $\hat{\epsilon}$ was then computed. These values were then transformed into $B$ values using the square root of 30. The distribution of the transformed values as well as the corresponding beta distribution were plotted. As $p$ increased, the fit became even better, which is logical since both the mean and variance are dependent on a function of $p$. It must be noted that in a DTI setting $p = 3$; however, the use of $p > 3$ is shown to demonstrate a more general use of this finding.

Figures 4, 5, and 6 show the accuracy of the approximations when $p = 3$. The solid red line represents the approximations based on a Beta random variable, while the histogram is from the sample values from the simulation described above. Figure 4 summarizes the simulation when $\epsilon \approx 0.353$. Since so many replicates were used, using a p-value for a goodness of fit test does not make sense since it is dependent on the sample size. If one were used, the fit would need to be perfect in order to get a non-significant result. However, the Kolmogorov-Smirnov test statistic, $D = \max|\hat{F}(x) - F_{obs}(x)|$ was calculated. Because $D$ already denotes the diffusion tensor, the Kolmogorov-Smirnov test will be given as $KS$ for the rest of the paper. Note that this statistic is bounded by 0 and 1 with a perfect fit $= 0$ and the worst fit $= 1$. In this figure, the fit works reasonably well ($KS = 0.050393$).
Figure 4. Approximate (red) and simulated (black) densities of $\hat{c}$ when $\epsilon \approx 0.353$ and $p = 3$ ($KS = 0.050393$).

Figure 5. Approximate (red) and simulated (black) densities of $\hat{c}$ when $\epsilon \approx 0.496$ and $p = 3$ ($KS = 0.04458$).
Figure 5 summarizes the simulation when $\epsilon \approx 0.496$. In Figure 5, the fit works reasonably well, but is not perfect ($KS = 0.04458$). It is good to note that the LBI test is more powerful for large values of $\epsilon$, hence we would expect better fits as $\epsilon \to 1$.

Figure 6 summarizes the simulation when $\epsilon \approx 0.889$. The figure shows that this distribution works extremely well for larger values of epsilon ($KS = 0.010627$). This again was expected, as the LBI test is the better test for large epsilon. Remember that Figures 5 and 6 are for the case where $p = 3$, or a $3\times3$ matrix. The fits will only improve as $p$ increases.

To show this, a simulation was performed when $p = 4$. Figure 7 shows when $\epsilon \approx 0.89$ and $p = 4$. Note that the fit still works extremely well, if not better, since the overlayed beta distribution has a better fit near the apex of the curve ($KS = 0.0056459$).

![Figure 6. Approximate (red) and simulated (black) densities of $\hat{\epsilon}$ when $\epsilon \approx 0.889$ and $b = 3$ ($KS = 0.010627$).](image)
Figure 7. Approximate (red) and simulated (black) densities of $\widehat{\epsilon}$ when $\epsilon \approx 0.89$ and $p = 4$ ($KS = 0.0056459$).

However, when dealing with DTI data, random noise must be accounted for. Thus, an additional simulation was performed that created random $\widehat{D}$ by using the WLS algorithm in (6) with 100,000 replicates. Simulated diffusion-weighted images were generated as follows: $S_0$ was fixed at 1500, values of $\sigma_o$ were varied to provide differing signal-to-noise ratios (SNR = $S_0/\sigma_o$) of 5, 10, 15, 20, 25 and 30. Similar to the simulations in the Zhu et. al. (2007) paper, an imaging acquisition scheme $\{(b_i, r_i) : i = 1, ..., 30\}$ was used that consists of $m = 5$ baseline images with $b = 0$ s/mm$^2$ and $n - m = 25$ directions of diffusion gradient at $b = 1000$ s/mm$^2$ and $r_i$ equivalent to the matrix provided in the Hardin (1994) web site when $m = 30$. For a given diffusion tensor, $D$, $x_i$, and $y_i$ were generated from a Gaussian random number generator with mean zero and standard deviation $\sigma_o$. Similar to the
simulations above, $D$ was defined with the diagonal elements being as follows (units: $10^{-3}$mm/s): (0.80, 0.09, 0.1), for an $\epsilon = 0.496$ and (1.00, 0.55, 0.55), for an $\epsilon = 0.889$.

Finally, the resulting $i^{th}$ acquisition of the resulting diffusion-weighted data was calculated by $S_i = \sqrt{(S_0\exp(-b_i r'_i D r_i) + x_i)^2 + y_i^2}$. The $\tilde{\epsilon}$ were then calculated for each $\tilde{D}$ computed by the WLS algorithm in (6) with $k = 5$. Similar to the first simulation, the $\tilde{\epsilon}$ values were then transformed into $B$ values using the square root of (30) and plotted. The distribution of the transformed values as well as the corresponding beta distribution were plotted.

![Figure 8](image_url)

**Figure 8.** Approximate (red) and simulated (black) densities of $\tilde{\epsilon}$ when $\epsilon \approx 0.496$ and SNR = 5 ($KS = 0.029183$).

Figures 8 and 9 show the accuracy of the approximations when the simulation accounts for random noise. The solid red line represents the approximations based on a Beta random variable, while the histogram is from the sample values from the simulation. Similar to the first simulation, Figure 8 summarizes the simulation when $\epsilon \approx 0.496$. In Figure 8, the fit works reasonably well, but is not great ($KS = 0.029183$). This result is consistent with the
results from the first simulation. If anything, these fits could be slightly better. A main reason why the addition of the noise did not hinder the fits was because the noise that was added was Gaussian, and Beta random variables can be expressed as functions of chi-squared's, which are functions of Gaussians.

Figure 9 summarizes the simulation when \( \epsilon \approx 0.889 \). The figure shows that this distribution works well for larger values of epsilon (\( KS = 0.0175719 \)). The reason for the fit not being as good as Figure 6 is that this simulation was with SNR = 5, which means a larger \( \sigma_o \) used to define \( x_i \) and \( y_i \). As the SNR increases, the fits become better; these images can be found in Appendix A.

Thus, it has been shown that \( \hat{\epsilon} \) can be approximated by a beta-squared random variable and that this approximation holds in the DTI setting when random noise is incorporated into

![Figure 9. Approximate (red) and simulated (black) densities of \( \hat{\epsilon} \) when \( \epsilon \approx 0.889 \) and SNR = 5 (\( KS = 0.0175719 \)).](image-url)
the diffusion matrix. For region of interest analysis of DTI images this is extremely useful, as there is now a statistical distribution that can be associated with the tensors coming from a given region. Chapter 4 will delve into how these approximations work in a real world example of DTI data.
4. DIAGNOSTIC TECHNIQUES FOR DIFFUSION TENSOR IMAGING DATA: PAPER 2

The DTI variable that is the most useful in analyzing data is fractional anisotropy (FA) (Clement 2005). FA values are a measure of the deviance from sphericity. Sphericity, or isotropy, of the principal variables, holds if $\epsilon = 1$ (and hence all population eigenvalues are equal). This condition represents the event of water molecules dispersing, or diffusing, uniformly over a given space.

By (7) and (9),

$$
\phi^2 = \frac{3 \cdot 3 \left[ \mu'_2 - (\mu'_1)^2 \right]}{2 \cdot 3 \mu'_2} = \frac{3}{2} (1 - \epsilon) .
$$

(37)

Hence $\phi^2$ is invariant and

$$
\bar{e} = 1 - \frac{2}{3} \phi^2 .
$$

(38)

Thus, as FA increases, $\bar{e}$ decreases, and diffusion becomes more anisotropic. Additionally, it is shown that a linear function of $\phi^2$ is a one-to-one function of the LBI test for sphericity (Clement 2005). The LBI test for sphericity is more powerful when values of $\epsilon$ are near one (Sugiura 1995). Experience with DTI brain data shows that this is the case. Paper 1 showed that $\bar{e}$ can be approximated by a squared beta distribution, with the approximation best at highest (and lowest) values. Thus, transformed FA values can be approximated by a squared beta distribution.

There is one main concern of the analysis approach used above. This is that distinct types of data could be mixed together, especially white and gray matter. Such mixing could occur due to combining different regions of the brain along the perimeter of the region of
interest, or different segments along a fiber tract. The goal here centers on evaluating how well the beta approximations fit. The recommended approach also was chosen to provide self-diagnostic information quickly about any problems with regions of interest definitions.

In 1987, Box and Draper wrote: "Essentially, all models are wrong, but some are useful." The parametric approach to analyzing diffusion tensor data shows a usefulness for these approximate distributions. It is useful to determine when the approximations are not drastically off base.

The UNC Neurodevelopmental Disorders Research Center provided data for 32 developmentally delayed and typical children. All scans were acquired on a 1.5T GE Sigma Advantage MR scanner. DTI images were acquired using 4 repetitions of 12-direction spin-echo single-shot echo planar imaging (EPI) sequence with a 128x128x130 image matrix at 1.875mm x 1.875mm x 3.8mm resolution with a 0.4mm gap using a b-value of 1000 s/mm^2. Using a custom program designed to automatically remove slices that fall outside predetermined parameters, each DTI slice was screened for motion and other artifacts. After cleaning, both correction of eddy-current based image distortions using mutual information based unwarping and the calculation of the diffusion tensor elements were performed using another custom software package (Cascio et al. 2008). The resulting eigenvalues and eigenvectors of each diffusion tensor were also calculated and FA values were computed. The FA values from different regions of the brain were transformed into B values from (30). The distribution of the transformed values as well as the corresponding approximating beta distribution were plotted. Diagnostic tests were used to see how the fits were performing. The data is ordered from smallest to largest for each subject. Having the ordered data allowed for simple computation of the empirical quantile function to be used, where the $y_i$ values are equal to the raw FA values and the $x_i$ values are equal to $i/(n - 1)$, where $n$ is equal to the number of voxels in a region. The QQ-envelop plots were used on the ordered
data to see how the fit relates to other samples from the same beta distribution; also, since multimodality could be an issue, SiZer was also used.

The right cerebellum is a structure located between the cerebrum and the brainstem which is the unit of motor control. There are 453 voxels that make up this region. Figure 10 depicts a sample of fits of the transformed right cerebellum data, with the $y$-axis being the percent of data points and the $x$-axis, the $B$ value. All of these fits seem to work well; however, we will still look at the diagnostic tests.

![Figure 10. Six histograms of the fits of the transformed right cerebellum data; actual values (black) and approximate beta distribution (red).](image)
resampling the exact beta distribution. If the results fit, then the red line will be encompassed by the blue lines.

Figures 11 and 12 show the worst and best fits from Figure 10, respectively. One will note that the Q-Q plot for the transformed FA data (red line) in the QQ-envelop plots can deviate quite a bit from the theoretical beta distribution (green line) and is not always encompassed in the 1000 resamplings of the beta distributions (blue lines). In Figure 11, it appears that the data in the lower $5^{th}$ percentile deviates from the estimated distribution. In

![Figure 11. QQ-envelop plot of the subject (ID = 515000402) with the worst fit from Figure 10 (plot on third row, second column). Q-Q plot for the data (red lines), Q-Q plot for the theoretical distribution (green), and Q-Q plot for the resampled data are displayed. Fit works well except for the left tail ( < $5^{th}$ percentile).](image-url)
Figure 12, the red line is encompassed by the blue lines at all points except a small region between 0.6 and 0.65 on the y-axis. SiZer was considered, but will not be shown here because there were no occurrences of multimodality.

Thus, the beta distribution approximation works well in the best case, less well for other subjects in the right cerebellum region of interest. However, if we go back to Box and Draper's statement about modeling, this information is definitely still informative. These fits for the right cerebellum appear to work well enough to be used in analyses, as the fit did work for all but a small fraction of the data, mainly the left tail.

![QQ-Envelop Plot](image)

Figure 12. QQ-envelop plot of the subject (ID = 514900302) with the best fit from Figure 10 (plot on second row, second column). Q-Q plot for the data (red lines), Q-Q plot for the theoretical distribution (green), and Q-Q plot for the resampled data are displayed. The fit works well except for a small region between 0.60 and 0.65 on the y-axis.
The corpus callosum is a region of the brain where this is not the case. The corpus callosum is the largest connective pathway in a human brain that allows the two hemispheres of the brain to communicate with each other. Upon first inspection of the data, the fits did not work well at all. In order to understand why these fits were not working, meetings with the investigators of the study and the DTI team were held. It was agreed that a scientific reason could be that partial voluming (including multiple regions in a voxel) was occurring. This was plausible, especially due to the size of the corpus callosum, and could have caused a bimodal effect in the data. The definition of the region of interest was further investigated and updated. The analysis provided below was performed on the new definition of the corpus callosum region. This region is comprised of 857 voxels.

![Figure 13. Six histograms of the fits of the transformed corpus callosum data. Actual values (black) and approximate beta distribution (red).](image-url)
Figure 13 shows the transformed corpus callosum data from six arbitrarily selected children. The fits are obviously not as good as those of the right cerebellum and might even indicate a bimodal distribution. To understand how good the fits were, QQ-envelop was performed. Figures 14 and 16 show the results of the best and worst subjects from Figure 13.

Figure 14 shows that the beta approximation is not working in the first $10^{th}$ percentile as well as between the $25^{th}$ and $50^{th}$ percentiles for the worst fit from the sample of 6 subjects. This is a worse fit than seen in the right cerebellum data. In order to see if this was the case, SiZer was used.

![QQ-plot](image)

Figure 14. QQ-envelop plot of the subject (ID = 515600402) with the worst fit from Figure 13 (plot on second row, second column). Q-Q plot for the data (red lines), Q-Q plot for the theoretical distribution (green), and Q-Q plot for the resampled data are displayed. The fits do not work well for the first $10^{th}$ percentile and between the $25^{th}$ and $50^{th}$ percentiles.
Figure 15 displays the SiZer plot for the same subject displayed in Figure 14. The first plot shows the overlay of the family of bandwidth values used, with the optimal $h$ value denoted as the bolded black line. The green dots represent each transformed value from each voxel in the corpus callosum region. The horizontal dashed lines represent the percentiles that are plotted in Figure 14. We can see from the family overlay plot that a beta distribution is not likely and that an extra mode could be occurring between the $10^{th}$ and $25^{th}$ percentiles.

![Figure 15. SiZer plot of the subject (ID = 515600402) with the worst fit from Figure 13 (plot on second row, second column). The first graphic is the plot of the family of bandwidths, the second is the slope SiZer map, and the third is the curvature SiZer map. From these plots, there does not seem to be any significant results for bimodality.](image)

The second plot in Figure 15 shows the slope SiZer map for the distribution. This plot shows the first derivative of the curve, with a blue color used on the pixel of the color map if
the smooth is significantly increasing. If the smooth is significantly decreasing, the pixel is colored red, and if there is not a statistically significant slope, the pixel is shaded purple. The black horizontal line denotes the same bandwidth that corresponds to the bolded black curve in the family bandwidth plot. Up until the peak between the $10^{th}$ and $25^{th}$ percentiles, the curve is increasing, but doesn't show a decreasing slope until around the $90^{th}$ percentile. Thus, significant bimodality is not shown even though there are separate modes in the first graphic of Figure 15. However, the dark black line in the first graphic in Figure 15, does not depict what one would expect from a probability density function of a beta distribution, as the pdf of a beta distribution would be an increasing function for the majority of the plot.

**Figure 16.** QQ-envelop plot of the subject (ID = 518300401) with the best fit from Figure 13 (plot on third row, first column). Q-Q plot for the data (red lines), Q-Q plot for the theoretical distribution (green), and Q-Q plot for the resampled data are displayed. The fits work well for the whole range of data.
The third graphic of Figure 15 shows the curvature map. This plot shows the second derivative of the curve with a orange color used on the pixel of the color map if the smooth is significantly convex. If the smooth is significantly concave, the pixel is colored cyan and if there is not a statistically significant curvature, the pixel is shaded green. By Figure 15, the only significant concavity is around the 90\textsuperscript{th} percentile. While SiZer did not confirm bimodality, it did confirm that the plot did not look like a pdf of a beta distribution when kernel density estimation was used.

Figure 16 shows the QQ-envelop plot for the best fit from Figure 13 for the transformed corpus callosum data. Unlike the worst fit (Figure 14 and 15), this QQ-envelop plot shows that the corpus callosum beta approximation works very well for the data, as the Q-Q plot from the transformed data (red line) is completely enveloped by the resamplings from the theoretical distribution (blue line). It would appear that for this subject, the inclusion of other regions of the brain was either insignificant or did not occur.

To make sure there were no underlying issues, SiZer was also used. In Figure 17, the SiZer plot for the same subject as in Figure 16 is shown. From the all of the different plots, we see that this definitely looks more like a beta distribution than the plot in Figure 15. The family distribution plot has the bolded black line looking relatively beta-like without any major secondary modes. The slope and curvature SiZer maps have a strongly increasing look for the majority of the plot, as would be expected in a beta distribution. There is also the lack of any unexpected curvature change. Thus, SiZer does not show any reason that the fits should not be used for this subject.
Figure 17. SiZer plot of the subject (ID = 518300401) with the best fit from Figure 13 (plot on third row, first column). The first graphic is the plot of the family of bandwidths, the second is the slope SiZer map, and the third is the curvature SiZer map. From these plots, there does not seem to be any significant results for bimodality.

From this paper, it is shown that the beta transformed data has a reasonable approximation, depending on region of interest. It appears that the fits work best with the larger regions of interest and work less well for the smaller regions of interest. These fits can work well, but more importantly if there are data problems (e.g. partial voluming), the diagnostic techniques will show a problem with the fit. If the fits do not work well, then the data cannot be analyzed in the parametric approach discussed in Cascio et al. (2008). In the next paper, a nonparametric approach to analyzing DTI data will be addressed.
5. NON-PARAMETRIC ANALYSIS TECHNIQUES FOR DIFFUSION TENSOR IMAGING DATA: PAPER 3

From paper 1, it is known that FA values can be transformed by using the square root of \((30)\). This solves the high dimension, low sample size (HDLSS) problem commonly found in imaging studies by accurately summarizing and characterizing a distribution of hundreds or thousands of FA values with two parameters for each individual (Clement 2005). The beta distribution can then be transformed to a non-bounded \(F\)-distribution,

\[
F = \frac{(1 - \beta)/\gamma)(\beta/\alpha)}{2\gamma, 2\alpha}.
\]

This transformation accomplishes several desired goals. First, as with the beta distribution transformation, the data is reduced from thousands of FA values to two parameters for each subject that summarizes the entire distribution. Second, unlike the beta distribution transformation, the \(F\) random variable is scale free. Third, when using the beta transformation, the \(B\) value has the opposite interpretation of the raw FA values. In particular, an FA value of 0 is equivalent to isotropy and an FA value of 1, complete anisotropy. Whereas, a \(B\) value of 0 relates to anisotropy and a \(B\) values of 1, isotropy. However, using the \(F\) transformation allows the data to follow the same directionality of the FA value. Thus, these properties allow the reader to relate the new measure to the more familiar FA measure without confusion.

The \(F\)-distribution has well-documented and simple statistical properties. Thus, a mean and standard deviation of the individual frequency distributions can be calculated for each subject. A single value representing the mean + standard deviation \((\delta = \mu + \sigma)\), which corresponds to roughly the point of inflection in the probability density function, should then be calculated for each subject, thus further reducing the high dimension of the data to one measure per subject. This process makes use of inherent distributional properties, and
coupled with the greater power for detecting differences between groups, is far superior to the current standard process of simply analyzing the mean FA. This one value per observation, \( \delta \), could then be used as the independent variable in initial analyses (Clement 2005). This method was used in Cascio et. al. (2008) to analyze DTI data. However, this method is only valid if the original assumption of the beta fit works. It has been shown in paper 2, that this is not always the case. Thus, another analysis method will be discussed.

One approach could be to find a bimodal distribution that will fit the data. While potentially useful in theory, for ease of use in the medical field, this is not suggested, as it can be computationally hard to find and must be done for each individual subject. Also, if a bimodal distribution was acquired, one must then question what summary measure to use to analyze the subject.

With the beta distribution fit, the use of \( \delta \) has been shown to be an appropriate summary measure as it corresponds to the point of inflection of the distribution; however, the meaning when the distribution is bimodal is a little less clear. Thus, instead of finding the bimodal distribution that fits each subject's data, a non-parametric approach will be used.

Marron's DiProPerm test was discussed in section 2.9.2. As stated there, this test would be very useful if one would like to know if two groups of distributions are the same. Therefore, if one wanted to test if there were differences in a given region of the brain between autistic children and typical children or developmentally delayed children, then the DiProPerm test could be used. The 32 developmentally delayed and typical children's data from the University of North Carolina Neurodevelopmental Disorders Research Center's longitudinal study were used to analyze if there were differences in varying brain regions. There were 10 developmentally delayed children and 22 typical children in this study. The data were analyzed in two different ways for different regions of the brain: using the raw FA data and using the transformed FA values to \( B \) values using (30). The corpus callosum and right cerebellum will be discussed, as these were the two regions addressed in Paper 2.
Prior to performing the DiProPerm test, the DWD vector must be found. The process will be discussed for the transformed FA data; however the same was done for the raw FA values. First, the \( n \) voxels for each region are acquired for each subject. A transformation using \((30)\) was applied to all voxels. An \( n - 1 \) dimensional DWD hyperplane was then found that separated the two populations in the manner discussed in section 2.9.1.

Once the DWD hyperplane is known, the data can be projected onto the normal vector orthogonal to the hyperplane, leading to a one-dimensional representation of each subject. These data were looked at in reference to the DWD direction as well as the first 3 orthogonal principle component (PC) directions. Figure 18 shows this graphic. The red circles denote each subject in the developmentally delayed group, while the blue circles represent each subject in the typical group. The \( x \)-axes of the four columns are as follows: the DWD direction, the first PC orthogonal to the DWD direction, the second PC orthogonal to the DWD direction, and the third PC orthogonal to the DWD direction, respectively. The \( y \)-axes of the four rows are as follows: the DWD direction, the first PC orthogonal to the DWD direction, the second PC orthogonal to the DWD direction, and the third PC orthogonal to the DWD direction, respectively. The only differences are the diagonal graphics which have the \( y \)-axis as the density. Note that the height of these circles in the diagonal graphics is completely arbitrary and just used to separate the data; however, where the data falls in the \( x \)-direction represents the projection of the data onto the given direction (DWD, PC1, PC2, or PC3). Also in these graphics on the diagonal are the KDE smoothed histogram of the combined data (typical and developmentally delayed children).

From the first graphical viewpoint (first row, first column) in Figure 18, there is no clear distinction between the two population groups, so no noticeable demarcation has been made between the developmentally delayed and typical children. Most of the data points are on top of each other relative to the DWD direction. This is also true for all of the diagonal graphics. There is no real appearance of anything occurring except for the fourth column. In the
graphics in the fourth column, the potential of an outlier in the developmentally delayed group can be seen. The one red circle > 0.4 on the x-axis is clearly apart from the other developmentally delayed subjects.

Figure 18. Raw FA corpus callosum data for all 32 subjects once DWD was performed. DD (red) and typical (blue) children's projected values are shown. The x-axes of the 4 columns and y-axes of the 4 rows are as follows: the DWD direction, the 1st PC orthogonal to the DWD direction, the 2nd PC orthogonal to the DWD direction, and the 3rd PC orthogonal to the DWD direction, respectively. The diagonal graphics have the y-axis as the density.
Figure 19. Transformed FA corpus callosum data for all 32 subjects once DWD was performed. DD (red), the outlying DD (green), and typical (blue) children's projected values are shown. The $x$-axes of the 4 columns and $y$-axes of the 4 rows are as follows: the DWD direction, the 1$^{st}$ PC orthogonal to the DWD direction, the 2$^{nd}$ PC orthogonal to the DWD direction, and the 3$^{rd}$ PC orthogonal to the DWD direction, respectively. The diagonal graphics have the $y$-axis as the density.

Figure 19 is similar to Figure 18, except that it shows the transformed FA values using (30). The red circles denote the developmentally delayed group and the blue circles denote
the typical group; however, the green circle denotes the one subject who is in the developmentally delayed group, but appears to be a strong outlier given the first graphic in this figure. It should be noted that this is the same subject that appeared as an outlier when looking at the third PC direction in Figure 18. Thus, the use of the transformed FA values further emphasizes those with outlying values. The transformation also causes this subject to no longer be an outlier in the third PC direction; however, it definitely shows how different the subject is from the other developmentally delayed subjects in comparison to the DWD direction, as the green circle always is clustered with the blue ones.

Information about this particular child was requested. It was shown that the child was definitely developmentally delayed; however the child was also one of the lower functioning developmentally delayed children. Four tests were performed on this child: developmental level, cognitive performance, adaptive functioning and screening for autism. For the developmental level, cognitive performance and screening for autism, this child scored below the average for the developmentally delayed children, but was within one standard deviation. For the adaptive functioning test, the child performed one point higher than the average. This subject also had a small head circumference, which put her in the category of microcephaly. Microcephaly is a medical condition where the circumference of the head is smaller than normal because the brain has not developed properly or has stopped growing. Depending on the severity of the accompanying syndrome, children with microcephaly may have mental retardation, delayed motor functions and speech, difficulties with coordination and balance, and other brain or neurological abnormalities. Some children with microcephaly will have normal intelligence and a head that will grow bigger; however, they will track below the normal growth curves for head circumference (National Institute of Neurological Disorders and Stroke 2007).

From the first graphic in Figure 19 (first row, first column), there also seems to be more of a difference between the typical and developmentally delayed groups; as, with the
exception of the one outlying subject, the red circles tend to be on the right side of the graph whereas the blue circles are on the left. This begs the question of how this subject influenced the DWD direction? To answer this, the DWD vector was computed without the outlying subject.

Figure 20. Transformed FA corpus callosum data for all 32 subjects once DWD was performed on 31 subjects (all but outlier (green)). DD (red) and typical (blue) children's projected values are shown. The $x$-axes of the 4 columns and $y$-axes of the 4 rows are as follows: the DWD direction, the 1st PC orthogonal to the DWD direction, the 2nd PC orthogonal to the DWD direction, and the 3rd PC orthogonal to the DWD direction, respectively. The diagonal graphics have the $y$-axis as the density.
subject; hence, only 31 subjects’ data were used. Then, the plots were created to look at this DWD direction with respect to the first three orthogonal PC directions. All 32 subjects were included in the plot; the outlier was plotted in green. The results are in Figure 20. The separation between the two groups is just as pronounced as in Figure 19; however, where the outlier lies on the plot changes drastically. This subject now also appears to be an outlier of the typical group, as well.

Now that the data have been investigated, the DiProPerm test could be performed. The same DWD vectors were used as those used for the figures above. A t-test was performed on the 1000 resamplings to get the p-value of the test. Since the null-hypothesis is that the two groups come from the same distribution, a p-value less than 0.05 would show that the two groups were different. First, the DiProPerm test was done for the raw FA values. Secondly, the test was performed on the transformed FA data with the DWD vector from all 32 subjects. Finally, the test was performed on the transformed FA data with the DWD vector from the 31 subjects (all but the outlier).

Figure 21 depicts the results of the DiProPerm test for the raw FA values of all 32 subjects. The first graphic shows the projections on the DWD direction vector. This plot is similar to the first graphic in Figure 18; however, smooth histograms are also displayed for each group; typical (blue), developmentally delayed (red) and all subjects (black). The $x$-axis is the projected value on the DWD direction and the $y$-axis is the density for the proportionally weighted histograms of the projected values for each population. The second graphic in Figure 21 shows the results of the 1000 simulations where the data were randomly relabeled. The $x$-axis is the $t$-value and the $y$-axis is the density for the smooth histogram of $t$-values. Each black dot represents one simulated result. The p-value was then computed using the quantile from the original t-test value (displayed by a green vertical line). The DiProPerm test resulted in a $t$-statistic of 3.0555. It is shown from the second graphic that the p-value for this $t$-statistic is 0.775. Thus the null hypothesis that the raw FA values are
different between groups cannot be rejected. This is not surprising since a clear distinction between the groups in Figure 18 was not seen. Thus, it is recommended to look at the transformed FA data instead of the raw FA data.

Figure 21. DiProPerm results for the raw FA data of the corpus callosum on all 32 subjects. Smooth histograms (left graphic) are displayed for each group; typical (blue), developmentally delayed (red), and all subjects (black). Smooth histogram (right graphic) of t-values with each black dot representing one simulated result. The p-value is the quantile of the original t-test value (green vertical line).

Figure 22 shows the results of the DiProPerm test for the transformed FA values. The t-statistic is 4.5769 with a corresponding p-value of 0.044. Thus, from the transformed FA data, we see that there is a statistically significant difference between developmentally delayed and typical children in the corpus callosum region. The distinction between groups seen in Figure 19 was not random. It must be noted that this result was found even with the outlying subject included in the analysis. When the outlying subject was taken out of the
analysis, the results became even more significant. The DWD vector from the 31 subjects was used and the outlier was taken out of the DiProPerm test.

Figure 22. DiProPerm results for the transformed FA data of the corpus callosum on all 32 subjects. Smooth histograms (left graphic) are displayed for each group; typical (blue), developmentally delayed (red), and all subjects (black). Smooth histogram (right graphic) of \( t \)-values with each black dot representing one simulated result. The p-value is the quantile of the original \( t \)-test value (green vertical line).

Figure 23 shows that the t-statistic is 6.3493 with a corresponding p-value <0.001. Thus, there is a highly statistically significant difference between the two groups. Even though this test was performed, it is not advisable, as there was no medically relevant reason to believe that this subject was not in the developmentally delayed group. This was performed just to show how much influence one subject can have over the analysis.

This type of analysis is useful in looking for differences among groups in regions of the brain where the transformed FA data are not unimodal. Differences can be found between
Figure 23. DiProPerm results for the transformed FA data of the corpus callosum on 31 subjects (all but the outlier). Smooth histograms (left graphic) are displayed for each group: typical (blue), developmentally delayed (red), and all subjects (black). Smooth histogram (right graphic) of $t$-values with each black dot representing one simulated result. The p-value is the quantile of the original $t$-test value (green vertical line).

the developmentally delayed and normal children's corpus callosum when using the transformed FA values. Cascio et. al. (2008) used the parametric approach to analyze the same subjects' data along with the data from the autistic subjects; however, we cannot compare the results as Cascio et. al. did not analyze the corpus callosum. When looking at the other regions of the brain the results also cannot be compared, as Cascio et. al. adjusted for covariates like age and gender. Currently, the DiProPerm method cannot adjust for covariates; however, this would be an excellent area of future research. Cascio et. al. (2008) also combined regions of the brain, for example, the right and left cerebellum, and had an indicator variable for side in the mixed model. This is different from the approach defined above where each region should be looked at individually.
6. CONCLUSIONS AND FUTURE RESEARCH

In paper 1, two simulations were performed. The first simulation showed that, in a general setting, the distribution of the Geisser-Greenhouse sphericity statistic can be approximated by a squared beta distribution. This approximation works best when $\epsilon \approx 1$; this is understandable due to its relationship to the locally best invariant test for sphericity. In order to show that this approximation works well in a DTI setting, another simulation was performed that added random noise to acquire a resulting tensor. This simulation also showed that a function of FA can be approximated by a squared beta distribution. Thus, $\hat{\epsilon}$ can be approximated by a squared random variable and in the DTI setting this results in a way to analyze regions of interest using distributions with known properties. However real world DTI data were looked at to make sure the approximations worked in practice.

Paper 2 used data from a University of North Carolina study that was looking at differences in brain structure between autistic, developmentally delayed, and typical children. Data from 32 subjects were used. It was shown that the beta approximations do not always work. This seems to be the case when the region addresses a very small area of the brain; for example, the corpus callosum. This is most likely due to partial voluming, where voxels are including different regions of the brain. Thus, before any analyzing techniques can be performed, one must use diagnostic techniques to see if the fits work well. The paper includes evaluations of using QQ-envelop and SiZer to see how well the beta approximations work. If the fits do not work well, then SiZer should be performed to check multimodality issues. Multimodality will most likely be a result of a partial voluming issue. If the approximations work well, an analysis approach similar to that in Cascio et. al. (2008) can be used. However, if the fits do not work, then another approach must be considered.
Paper 3 discussed two possible approaches to analyze this data: finding the bimodal distribution and a non-parametric approach. Finding the bimodal distribution does not seem fitting for this type of data, as a different bimodal distribution would need to be fit for every subject, which could be a very labor intensive process. Even if the distributions were found, the δ value does not have as much meaning in a bimodal setting. Thus, a non-parametric approach was decided upon to analyze the difference in groups for individual regions of interest. The DiProPerm method was used, as it handles the HDLSS problem by comparing the histograms of each child to see if there were consistencies across groups. For this data, the use of transformed FA values instead of the raw FA values was preferred as it highlighted outliers. Thus, both non-parametric as well as parametric approaches have been found for analyzing regions of interest in DTI data.

Further research to see how these two approaches relate to each other would be interesting. This would not be a trivial task. While a three group DiProPerm test is in production, it would still not allow adjusting for other covariates like age, gender, and study site. Covariates, however, could be done in the parametric mixed model approach. Further investigation as to how to address this would be needed. Theoretically, it would also be interesting to find the bimodal distributions that occur in the data; however, medically this is probably not useful. Two type mixtures are appealing for viewing the distribution of white and gray matter.

Analyzing the corpus callosum data with the addition of the autistic subjects would be needed. This approach could not be done in Cascio et. al. (2008) since the approximations were not working. Now that a nonparametric approach is known, the differences between the autistic, developmentally delayed and typical groups in this region should be tested. It would be useful to first look at the plots by age group and gender to see if these would be confounders.
Appendix A: Simulation Results For Other SNRs

Figure 24. Approximate (red) and simulated (black) densities of $\zeta=0.496$ and SNR = 10 ($KS = 0.055415$).

Figure 25. Approximate (red) and simulated (black) densities of $\zeta=0.889$ and SNR = 10 ($KS = 0.0093909$).

57
Figure 26. Approximate (red) and simulated (black) densities of $\hat{\tau} = 0.496$ and SNR = 15 ($KS = 0.057239$).

Figure 27. Approximate (red) and simulated (black) densities of $\hat{\tau} = 0.889$ and SNR = 15 ($KS = 0.0104291$).
Figure 28. Approximate (red) and simulated (black) densities of $\bar{z} = 0.496$ and SNR = 20 ($KS = 0.042671$).

Figure 29. Approximate (red) and simulated (black) densities of $\bar{z} = 0.889$ and SNR = 20 ($KS = 0.0088142$).

Since there were no major differences between SNR = 15 and SNR = 20, SNR > 20 were not considered.
Appendix B: Program for First Simulation

TITLE1 "&JOB..SAS-simulate epsilon hat density, epsilon=1,b=3, nu=8";
LIBNAME OUT01 ".\DATA";

PROC IML;
SEED = 73477;
B=3;
NU=8;
LAMBDA=J(B,1,1); *vector of eigenvalues all = 1 to change program, just change this;
NREP=1000000;
STD=SQRT(LAMBDA);
D_STD=DIAG(STD);
EPSILON=(LAMBDA[+]##2)/(B#LAMBDA[#]);
COUNT=J(1000,1,0);
SUM=J(4,3,0);

DO REPLICAT=1 TO NREP;
  Z=NORMAL(J(NU,B,SEED));
  Y=Z*D_STD;
  S=Y*Y;
  Q1=TRACE(S)#2;
  SSQ=S*S;
  Q2=TRACE(SSQ);
  EPSHAT=Q1/(Q2*B);
  ICOUNT=FLOOR(EPSHAT#1000);
  COUNT[ICOUNT]=COUNT[ICOUNT]+1;
END;

HOLDNM={COUNT};
CREATE ONE VAR HOLDNM;
APPEND FROM COUNT;
CASE=NU||B||EPSILON;
CASENM={NU B EPSILON};
CREATE TWO VAR CASENM;
APPEND FROM CASE;

DATA OUT01.&JOB;
LENGTH DEFAULT=4;
RETAIN NU B EPSILON;
IF _N_=1 THEN SET TWO;
SET ONE;
ICOUNT=_N_; 
EPSHAT=_N_/1000;
IF EPSHAT>=1/B THEN OUTPUT;
LABEL NU  ="error df=N-r"
          B  ="# cols Sigmastar"
EPSILON="population value epsilon"
COUNT  ="# obs in interval"
ICOUNT ="floor(1000*epshat)"
EPSSHAT="epsilon hat"
FORMAT EPSILON EPSHAT 5.3;

PROC PRINT DATA=OUT01.&JOB UNIFORM;
Appendix C: Program for Second Simulation

TITLE1 "&JOB..SAS-simulate diffusion tensor using Zhu (2007) method, epsilon=1, b=3, SNR=5";
LIBNAME OUT01 ".\DATA";

PROC IML;
SEED1=58943;
SEED2 = 73477;
SEED3 = 11765;
So = 1500;
SNR = 5;
sigma = So/SNR;
B=3;
LAMBDA=J(B,1,1); *vector of eigenvalues all = 1 to change program, change this;
NREP=100000;
DIAG = diag(lambda);
first = J(5,1,0);
second = J(25, 1, 1000);
b = first // second;
r={-0.006653594567 0.898917674061 0.438067055301, *from Hardin (1994);
0.910432431342 0.375861444128 -0.172745369774, 
-0.739286008328 0.66128200031 -0.670136746531, 
-0.119881642095 0.594206835081 -0.79532800910, 
-0.639562720115 0.726562214570 0.251131191606, 
0.639562709633 -0.726562226844 -0.251131182791, 
-0.667875921118 -0.544735685623 0.507153612626, 
0.296407339459 -0.436029585697 -0.849718123563, 
0.016512944080 -0.873455374227 -0.486624117685, 
-0.503800715885 -0.553035944290 -0.663578241805, 
-0.974931033164 0.216191593878 -0.052637204613, 
0.464845057365 0.775384781553 -0.427431296444, 
0.624261678448 -0.609877744888 0.488207428366, 
0.174067093646 -0.971581735749 0.160404419108, 
0.288299868987 0.446205413102 0.847221343039, 
-0.623981779896 0.663039307921 -0.413552432600, 
0.453897247750 0.268395653788 -0.849671149041, 
0.96507197501 -0.241653702570 0.101194366101, 
-0.834455827389 0.067046499165 0.546981022602, 
-0.902483656446 -0.411365491337 -0.127678041907, 
0.819760700810 -0.157436248876 -0.550641644763, 
0.587501166626 0.769194941197 0.251359347646, 
-0.459446524745 -0.888194833339 -0.004339231413, 
-0.274943879059 -0.145673222845 0.950360550272,
...
epsilon = j(NREP, 1, .);

DO REPLICAT=1 TO NREP;
    call randseed (seed1);
    x = J(30, 1, .);
    CALL RANDGEN( x, 'normal', 0, sigma);
    call randseed (seed2);
    y = J(30, 1, .);
    CALL RANDGEN( y, 'normal', 0, sigma);
    s = J(30, 1, .);
    z = J(30, 7, .);

    DO i = 1 to 30;
        bi = b[i,1];
        xi = x[i,1];
        yi= y[i,1];
        ri = r[i,];
        ri1 = r[i,1];
        ri2 = r[i,2];
        ri3 = r[i,3];
        x1 = -bi#ri1*ri1;
        x2 = -bi#2*ri1*ri2;
        x3 = -bi#2*ri1*ri3;
        x4 = -bi#ri2*ri2;
        x5 = -bi#2*ri2*ri3;
        x6 = -bi#ri3*ri3;
        inner = ri*diag*ri`;
        exp = exp(-bi*inner);
        funct = (so*exp + xi)**2 + yi**2;
        s[i,1] = sqrt(funct);
        z[i,] = 1 || x1 || x2 || x3 || x4 || x5 || x6;
    END;

    slog = log(s);
    za = J(7,7, .);
    sum = J(7, 7, 0);
    suma = J(7, 1, 0);
do k = 1 to 30;
za = z[k,]ʼ*z[k,];
sum1 = sum + za;
sum = sum1;
sum2 = suma + (z[k,]ʼ#slog[k,1]);
suma = sum2;
end;

thetals0 = inv(sum1)*sum2;
wio = J(30, 1, .);

do j = 1 to 30;
inside = z[j,]ʼ*thetals0;
wio[j,1] = exp(2#inside);
end;

za2 = J(7,7,);;
sumb = J(7,7,0);
sumc = J(7,1,0);

do l = 1 to 30;
za2 = wio[l,1]ʼ*(z[l,]ʼ*z[l,]);
sum3 = sumb + za2;
sumb = sum3;
sum4 = sumc + (wio[l,1]ʼ#z[l,]ʼ#slog[l,1]);
sumc = sum4;
end;

thetals = inv(sum3)*sum4;
da1 = thetals[2,1];
da12 = thetals[3,1];
da13 = thetals[4,1];
da22 = thetals[5,1];
da23 = thetals[6,1];
da33 = thetals[7,1];
one = d11 || d12 || d13;
two = d12 || d22 || d23;
three = d13 || d23 || d33;
Dhat = one // two // three;
dhat2 = dhat*dhat;
epsilon[replicat,1] = trace(dhat)#2 / (3#trace(dhat2));
END;

HOLDNM={EPSILON};
CREATE ONE VAR HOLDNM;
APPEND FROM EPSILON;

CASE=So||SNR;
CASENM={So SNR};

CREATE TWO VAR CASENM;
APPEND FROM CASE;

DATA OUT01.&JOB;
LENGTH DEFAULT=4;
RETAIN SNR SO;
   IF _N_=1 THEN SET TWO;
   SET ONE;
LABEL So = 'Initial S'
       SNR = 'Signal to noise ratio'
       EPSILON="Dhat value epsilon";
SQRTEPS = SQRT(EPSILON);
FORMAT EPSILON 5.3;

PROC PRINT DATA=OUT01.&JOB (OBS=40) UNIFORM;
REFERENCES


