Statistics and Its Interface Volume 3 (2010) 81–90

Approximating the Geisser-Greenhouse sphericity estimator and its applications to diffusion tensor imaging

Meagan E. Clement-Spychala*, David Couper, Hongtu Zhu and Keith E. Muller

The diffusion tensor imaging (DTI) protocol characterizes diffusion anisotropy locally in space, thus providing rich detail about white matter tissue structure. Although useful metrics for diffusion tensors have been defined, statistical properties of the measures have been little studied. Assuming homogeneity within a region leads to being able to apply Wishart distribution theory. First, it will be shown that common DTI metrics are simple functions of known test statistics. The average diffusion coefficient (ADC) corresponds to the trace of a Wishart, and is also described as the generalized (multivariate) variance, the average variance of the principal components. Therefore ADC has a known exact distribution (a positively weighted quadratic form in Gaussians) as well as a simple and accurate approximation (Satterthwaite) in terms of a scaled chi square. Of particular interest is that fractional anisotropy (FA) values for given regions of interest are functions of the Geisser-Greenhouse (GG) sphericity estimator. The GG sphericity estimator can be approximated well by a linear transformation of a squared beta random variable. Simulated data demonstrates that the fits work well for simulated diffusion tensors. Applying traditional density estimation techniques for a beta to histograms of FA values from a region allow representing the histogram of hundreds or thousands of values in terms of just two estimates for the beta parameters. Thus using the approximate distribution eliminates the "curse of dimensionality" for FA values. A parallel result holds for ADC.

Keywords and phrases: Diffusion tensor imaging, Geisser-Greenhouse sphericity estimator, Fractional anisotropy, Average diffusion coefficient.

1. INTRODUCTION

Diffusion tensor imaging (DTI) holds tremendous promise for improving our understanding of neural pathways, especially in the brain. 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. With water embedded in nonhomogeneous tissue, motion is expected to be anisotropic and might show preferred directions of mobility. DTI characterizes diffusion anisotropy locally in space, thus providing rich detail about white matter tissue microstructure. DTI allows tracking fibers in the brain, a result which has many potential applications in neuroscience and psychiatry. Combining fiber tracking with functional MRI may elucidate structure-function relationships. Due to the fact that MRI protocols are noninvasive and are deemed to provide essentially no risk to participants, longitudinal studies of both diseased and normal participants may be especially promising. DTI has already been used to show subtle white matter abnormalities in a variety of diseases; for example, stroke, multiple sclerosis, dyslexia, and schizophrenia (Le Bihan et al., 2001).

The current work was stimulated by a longitudinal study 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.

Definitions of credible statistical methods for analyzing DTI data are needed. Le Bihan et al. (2001) considered a diffusion tensor, $\boldsymbol{D}$, as a 3×3 estimated covariance matrix, $\widehat{\boldsymbol{\Sigma}}$, at the location of interest. To obtain an accurate evaluation of the probability distribution of diffusion in a region, one must use an orientation invariant measure for each tensor. 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. It will be 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. Using these transformed values will lead to outcomes in statistical models that avoid the "curse of dimensionality." Exact distributional results and a similar analysis for the average diffusion coefficient (ADC), volume ratio (VR) and relative anisotropy (RA) are also derived.

*Corresponding author.

This paper is organized as follows. Notation of the commonly used measures and the statistical results are in Section 2. Simulations and real data analysis are in Section 3. Finally, conclusions are stated in Section 4.

2. DTI MEASURES AND STATISTICAL PROPERTIES

2.1 DTI commonly used measures

The fact that diffusion can occur in three dimensions leads to using a diffusion tensor, $\boldsymbol{D}$, a 3×3 covariance matrix. The simplest possible covariance structure is $\sigma^2 \boldsymbol{I}$, referred to as *sphericity* due to the shape of the corresponding scattergram of data (and hence diffusion pattern, which is isotropic). The three most commonly used DTI summary measures are volume ratio, relative anisotropy, and fractional anisotropy. If $\{\lambda_3\}$ are the 3 eigenvalues of $\boldsymbol{D}$, then the measures are defined as follows.

Volume ratio, η

$$\eta = \frac{\prod_{k=1}^{3} \lambda_k}{\left(\sum_{j=1}^{3} \lambda_j/3\right)^3} = \frac{\mu_g^3}{{\mu_1'}^3}, \tag{1}$$

where μ_g^3 is the product of the 3 eigenvalues and μ_1' is the arithmetic mean of the eigenvalues. Here η expresses a relationship between the geometric and arithmetic mean of the variance of the diffusion.

Relative anisotropy, ζ

$$\zeta^2 = \frac{\sum_{k=1}^{3} \left(\lambda_k - \overline{\lambda}\right)^2}{\left(\sum_{k=1}^{3} \lambda_k\right)} = \mu_1' \cdot \frac{\mu_2' - (\mu_1')^2}{(\mu_1')^2}, \tag{2}$$

where μ_2' is the arithmetic mean of the squared eigenvalues. Hence ζ is a normalized standard deviation that measures the amount of anisotropy (non-sphericity) for the tensor.

Fractional anisotropy, ϕ

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

Here ϕ is a measure of the dispersion (variance) of the variances of the diffusion tensor.

2.2 Statistical properties

A covariance matrix always has a spectral decomposition with only positive or zero eigenvalues. If $\boldsymbol{\Upsilon}$ is the matrix of eigenvectors of the covariance matrix, then $\boldsymbol{\Sigma} = \boldsymbol{\Upsilon} \mathrm{Dg}(\boldsymbol{\lambda}) \boldsymbol{\Upsilon}'$ with $\boldsymbol{\Upsilon}'\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}' = \boldsymbol{I}_b$. Also if $\boldsymbol{Z} \sim \mathcal{N}_{\nu,b}(\boldsymbol{0}, \boldsymbol{I}_\nu, \boldsymbol{I}_b)$ and $\boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Phi}'$ where $\boldsymbol{\Phi} = \boldsymbol{\Upsilon} \mathrm{Dg}(\boldsymbol{\lambda})^{1/2}$, then $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Phi}' \sim \mathcal{N}_{\nu,b}(\boldsymbol{0}, \boldsymbol{I}_\nu, \boldsymbol{\Sigma})$ and

$$\boldsymbol{Y}'\boldsymbol{Y} \sim \mathcal{W}_b(\nu, \boldsymbol{\Sigma}), \tag{4}$$

where $\mathcal{W}_b$ denotes a Wishart distribution with ν the number of independent replicates used to find $\widehat{\boldsymbol{\Sigma}}$ and b the numbers of rows in $\widehat{\boldsymbol{\Sigma}}$ and $\mathcal{N}_{\nu,b}(\boldsymbol{0}, \boldsymbol{I}_\nu, \boldsymbol{I}_b)$ denotes a multivariate normal density with mean $\boldsymbol{0}$, ν representing the number of independent replicates and b representing the number of observations for each ν, and both $\boldsymbol{I}$ matrices representing the within and between subject variations, respectively (Muller and Stewart, 2006).

Box (1954a, b) defined the parameter ϵ as a measure of variance heterogeneity

$$\epsilon = \mathbf{tr}^2(\boldsymbol{\Sigma}) / \left[p\mathbf{tr}\left(\boldsymbol{\Sigma}^2\right)\right] = (\mu_1')^2/\mu_2', \tag{5}$$

where p is the rank of $\boldsymbol{\Sigma}$. It can also be interpreted as a measure of sphericity of the underlying principal components. The locally best invariant (LBI) test for testing sphericity ($H_0 : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p$) for unknown σ^2, against all alternatives, is to reject the null hypothesis for large values of $U = \mathbf{tr}(\widehat{\boldsymbol{\Sigma}}^2)/[\mathbf{tr}(\widehat{\boldsymbol{\Sigma}})]^2$ (John, 1971). Thus the maximum likelihood estimate (MLE), $\widehat{\epsilon}$, of the parameter ϵ is a one-to-one function of the LBI test for sphericity, $\widehat{\epsilon} = 1/(pU)$. In multivariate analysis of variance settings, $\widehat{\epsilon}$ is defined as the Geisser-Greenhouse sphericity estimator.

When $p = 3$, which is the case in diffusion tensors, the exact density function of U under the null and non-null hypotheses are known (Sugiura, 1995). Although analytic expression are known for both densities, the non-null density involves zonal polynomials, which makes computations difficult due to the need to evaluate an infinite series.

The null hypothesis for the likelihood ratio (LR) test for sphericity is to be rejected for sufficiently small values of

$$W = p^p |\widehat{\boldsymbol{\Sigma}}| \Big/ \left(\sum_{j=1}^{p} \widehat{\lambda}_j\right)^p = \widehat{\mu}_g^p \Big/ \widehat{\mu}_1'^p \tag{6}$$

(Khatri and Srivastava, 1971). The exact null and non-null densities of the LR test exist when $p = 3$ (Sugiura, 1995). For the remainder of this paper, $p = 3$ will be assumed. Although the non-null density exists, it involves infinite sums of infinite partitions; unless convergence occurs quickly, computations will be a problem. Power comparisons confirmed Grieve's (1984) conjecture that the LBI test is more powerful than the LR test if the population deviation from sphericity is large.

2.3 Statistical results

Given the assumption that the flow of water follows a Gaussian diffusion model arising from Brownian motion theory, $\boldsymbol{\Sigma}$ can be defined as the population covariance matrix (tensor) of the diffusion. Consequently $\nu\widehat{\boldsymbol{\Sigma}} \sim \mathcal{W}_p(\nu, \boldsymbol{\Sigma})$. The eigenvalues of $\widehat{\boldsymbol{\Sigma}}$, $\{\widehat{\lambda}_i\}$, are variance estimates for the 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, and thus of estimated variances.

First moment properties of eigenvalues (component variances)

Trace and ADC. Interpreting the eigenvalues, $\{\lambda_k\}$, as variances implies μ_1' is the average variance, or the arithmetic mean of the variances. When $\boldsymbol{S} = \nu\widehat{\boldsymbol{\Sigma}}$ is a Wishart, $\mathbf{tr}(\widehat{\boldsymbol{\Sigma}})$ is often called the generalized variance. Wishart properties imply the trace is a quadratic form in Gaussian variables (Glueck and Muller, 1998). Hence the average diffusion coefficient, ADC = $\mathbf{tr}(\widehat{\boldsymbol{\Sigma}})$, is distributed exactly as a weighted sum of central chi-square random variables. Kim, Gribbin, Muller, and Taylor (2005) provide a convenient review of exact and approximate calculations of probabilities for such quadratic forms. Both exact probabilities and excellent approximations are available. The following is the definition for an approximation for the quadratic form. If $Q = \sum_{k=1}^{K} \gamma_k X_k$ with $\gamma_k > 0$ and all mutually independent $X_k \sim \chi^2(\nu_k, \omega_k)$, then an approximation, $Q_* = \gamma_* X_*$, exists with $\gamma_* > 0$, $X_* > \chi^2(\nu_*, \omega_*)$ and

$$\gamma_* = \left(\sum_{k=1}^{K} \gamma_k^2\nu_k + 2\sum_{k=1}^{K} \gamma_k^2\omega_k\right) \Big/ \left(\sum_{k=1}^{K} \gamma_k\nu_k + 2\sum_{k=1}^{K} \gamma_k\omega_k\right) \tag{7}$$

$$\nu_* = \sum_{k=1}^{K} \gamma_k\nu_k \left(\sum_{k=1}^{K} \gamma_k\nu_k + 2\sum_{k=1}^{K} \gamma_k\omega_k\right) \Big/ \left(\sum_{k=1}^{K} \gamma_k^2\nu_k + 2\sum_{k=1}^{K} \gamma_k^2\omega_k\right) \tag{8}$$

$$\omega_* = \sum_{k=1}^{K} \gamma_k\omega_k \left(\sum_{k=1}^{K} \gamma_k\nu_k + 2\sum_{k=1}^{K} \gamma_k\omega_k\right) \Big/ \left(\sum_{k=1}^{K} \gamma_k^2\nu_k + 2\sum_{k=1}^{K} \gamma_k^2\omega_k\right) \tag{9}$$

Determinant. If the λ_i are thought of as measures of principal variation, then $\mu_g = (\mu_g^p)^{(1/p)}$ is the geometric mean of the variances. As an alternative to $\mathbf{tr}(\widehat{\boldsymbol{\Sigma}})$, the sample generalized variance is often defined as $|\widehat{\boldsymbol{\Sigma}}|$. It seems somewhat more natural to look at the geometric mean, $\widehat{\mu}_g = \sqrt[p]{|\widehat{\boldsymbol{\Sigma}}|}$. Gupta and Nagar (2000, Chapter 3) showed the following. If $\boldsymbol{S} \sim \mathcal{W}_p(\nu, \boldsymbol{\Sigma})$, $|\boldsymbol{S}|/|\boldsymbol{\Sigma}| \sim \prod_{i=1}^{p} u_i$, with independent $\{\mu_i\}$ and $u_i \sim \chi^2_{\nu-i+1}$, where $i \in \{1, \ldots, p\}$. Also,

$$\mathcal{E}\big(|\boldsymbol{S}|^h\big) = 2^{ph}|\boldsymbol{\Sigma}|^h \prod_{i=1}^{p} \left(\frac{\Gamma[(1/2)(\nu - i + 1) + h]}{\Gamma[(1/2)(\nu - i + 1)]}\right). \tag{10}$$

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 relatively the same in all three dimensions. The following discussion describes the very direct and simple connections among the DTI measures and Wishart distribution theory.

Volume ratio, $\boldsymbol{\eta}$

By Equation 1, $\widehat{\eta}$ is equivalent to W.

$$\widehat{\eta} = \frac{\widehat{\mu}_3'}{\widehat{\mu}_1'^{\,3}} = W. \tag{11}$$

Both η and $\widehat{\eta}$ are scale invariant with $\widehat{\eta}$ being exactly equal to the likelihood ratio test statistic for sphericity.

Relative anisotropy, $\boldsymbol{\zeta}$

By Equation 2,

$$\begin{aligned}\widehat{\zeta}^2 &= \frac{\widehat{\mu}_1'\widehat{\mu}_2'}{(\widehat{\mu}_1')^2} - \frac{\widehat{\mu}_1'\big[(\widehat{\mu}_1')^2\big]}{(\widehat{\mu}_1')^2} \\ &= \widehat{\mu}_1' \cdot \left(\frac{1}{\widehat{\epsilon}} - 1\right).\end{aligned} \tag{12}$$

It is straightforward to see $\widehat{\zeta}^2$ is not scale invariant. Relative anisotropy is a product of an invariant parameter, ϵ, and a non-invariant parameter, $\widehat{\mu}_1'$; thus $\widehat{\zeta}^2$ is not invariant to global scale. The lack of scale invariance makes RA an unappealing choice relative to VR or FA.

Fractional anisotropy, $\boldsymbol{\phi}$

By Equation 3,

$$\phi^2 = \frac{3 \cdot 3[\mu_2' - (\mu_1')^2]}{2 \cdot 3\mu_2'} = \frac{3}{2}(1 - \epsilon). \tag{13}$$

Hence $\widehat{\phi}^2$ is scale invariant and

$$\widehat{\epsilon} = 1 - \frac{2}{3}\widehat{\phi}^2. \tag{14}$$

Thus, a linear function of $\widehat{\phi}^2$ is a one-to-one function of a LBI test for sphericity. The LBI test for sphericity will be more powerful with values of ϵ near one (Sugiura, 1995). If care is taken to define essentially homogenous regions of tissue, DTI brain data can lead to values of $\widehat{\epsilon}$ that fit this case.

$\widehat{\boldsymbol{\epsilon}}$ and its relationship to a squared-beta distribution. It will be shown that $\widehat{\epsilon}$ can be approximated by a squared beta distribution. Thus, by equation 14, FA can also be approximated by a squared beta distribution. Approximately matching the first two moments of the $\widehat{\epsilon}$ to a squared beta random variable results in a simple approximate distribution. The fact that $1/p \le \widehat{\epsilon} \le 1$ allows concluding

$$0 \le \left(\widehat{\epsilon} - \frac{1}{p}\right)\left(1 - \frac{1}{p}\right)^{-1} \le 1. \tag{15}$$

This leads to defining

$$B^2 = \left(\widehat{\epsilon} - \frac{1}{p}\right)\left(1 - \frac{1}{p}\right)^{-1} = \widehat{\epsilon}\left(\frac{p}{p-1}\right) - \left(\frac{1}{p-1}\right). \tag{16}$$

If we let $c_1 = (p/(p-1))$ and $c_0 = (1/(p-1))$, then $\mathcal{E}B^2 = c_1\mathcal{E}\widehat{\epsilon} - c_0$ and $\mathcal{E}\widehat{\epsilon} = (\mathcal{E}B^2 + c_0)/c_1$. With $T_1 = \mathbf{tr}^2(\widehat{\boldsymbol{\Sigma}})$ and $T_2 = \mathbf{tr}(\widehat{\boldsymbol{\Sigma}}^2)$, it follows that $\widehat{\epsilon} = T_1/(pT_2)$. While it is known that the following assumptions are not precisely true, they were assumed to derive the approximate results. First, assume that T_1 and T_2 are independent. Second, assume that $\mathcal{E}(T_2^{-1}) = (\mathcal{E}(T_2))^{-1}$. Muller et al. (2007) reported that $\mathcal{E}(T_1) = 2\nu_e \sum_{k=1}^{p} \lambda_k^2 + \nu_e^2(\sum_{k=1}^{p} \lambda_k)^2$ and $\mathcal{E}(T_2) = \nu_e(\nu_e + 2)\sum_{k_1=1}^{p} \lambda_{k_1}^2 + 2\nu_e \sum_{k_1=2}^{p}\sum_{k_2=1}^{k_1-1} \lambda_{k_1}\lambda_{k_2}$. Also, from (16), it is known that $B^2 = (T_1/pT_2)[c_1] - [c_0]$; 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,

$$\mathcal{E}(\widehat{\epsilon}) = (\mathcal{E}(B^2) + c_0)/c_1 \approx \left(\frac{p-1}{p}\right)\mathcal{E}(B_*^2) + \left(\frac{1}{p}\right). \tag{17}$$

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

$$B_*^2 = \frac{\lambda_{*1}^2 X_1^2}{(\lambda_{*1}X_1 + \lambda_{*2}X_2)^2} = \frac{\lambda_{*1}^2 X_1^2}{\lambda_{*1}^2 X_1^2 + 2\lambda_{*1}\lambda_{*2}X_1X_2 + \lambda_{*2}^2 X_2^2} \tag{18}$$

with

$$\begin{aligned}\mathcal{E}(X_1^2 + 2X_1X_2 + X_2^2) &= \mathcal{E}X_1^2 + 2\mathcal{E}X_1\mathcal{E}X_2 + \mathcal{E}X_2^2 \\ &= \lambda_{*1}^2\nu_{*1}(\nu_{*1}+2) + 2\lambda_{*1}\lambda_{*2}\nu_{*1}\nu_{*2} \\ &\quad + \lambda_{*2}^2\nu_{*2}(\nu_{*2}+2).\end{aligned} \tag{19}$$

As a Beta random variable, $\mathcal{E}(B_*^2) = \nu_{*1}(\nu_{*1}+1)[(\nu_{*1} + \nu_{*2})(\nu_{*1} + \nu_{*2} + 1)]^{-1}$. Hence, by taking the expectation of the numerator and denominator separately,

$$\mathcal{E}(\widehat{\epsilon}) \approx \left(\frac{p-1}{p}\right)\frac{\nu_{*1}(\nu_{*1}+1)}{(\nu_{*1}+\nu_{*2})(\nu_{*1}+\nu_{*2}+1)} + \left(\frac{1}{p}\right). \tag{20}$$

Similarly, as a Beta random variable,

$$\begin{aligned}\mathcal{E}(B_*^4) = \nu_{*1}(\nu_{*1}+1)(\nu_{*1}+2)(\nu_{*1}+3)/ \\ (\nu_{*1}+\nu_{*2})(\nu_{*1}+\nu_{*2}+1) \\ \times (\nu_{*1}+\nu_{*2}+2)(\nu_{*1}+\nu_{*2}+3).\end{aligned} \tag{21}$$

Hence, by treating the numerator and denominator separately,

$$\begin{aligned}\mathrm{Var}(\widehat{\epsilon}) \approx \Big[((p-1)/p)^2 2\nu_{*1}\nu_{*2}(\nu_{*1}+1) \\ \times \left(2\nu_{*1}^2 + 2\nu_{*1}\nu_{*2} + 6\nu_{*1} + 3\nu_{*2} + 3\right)\Big]/ \\ \Big[(\nu_{*1}+\nu_{*2})^2(\nu_{*1}+\nu_{*2}+1)^2(\nu_{*1}+\nu_{*2}+2) \\ \times (\nu_{*1}+\nu_{*2}+3)\Big].\end{aligned} \tag{22}$$

Without the $(p-1)/p$ term, this would be what one would expect for the variance of a squared beta distribution. In more general cases, the approximation depends on p such that as p increases, the approximation gets better since $((p-1)/p) \to 1$ as $p \to \infty$. The simulations below show that the approximations work well when $p = 3$. However, if larger dimensional tensors were used in imaging analysis (for example, High Angular Resolution Diffusion Imaging (HARDI)), then these approximations would only improve.

3. SIMULATIONS AND DATA ANALYSIS

Simulations were conducted to evaluate the accuracy of the approximations, using one million replications. Using a known ϵ, pseudo random Wishart matrices were generated by first creating a random Gaussian Z matrix. Throughout $\Upsilon = \boldsymbol{I}$ without loss of generality because the parameters and estimates of interest are invariant to Υ (from trace and determinant invariance properties). With ϵ a function of eigenvalues, $\mathrm{Dg}(\lambda)$ was chosen to achieve the ϵ of choice and $Y = Z\Upsilon\mathrm{Dg}(\lambda)^{1/2} = Z\mathrm{Dg}(\lambda)^{1/2}$ and $\Sigma = \mathrm{Dg}(\lambda)$. The following entries comprised the matrix, $\mathrm{Dg}(\lambda)$: $(0.80, 0.09, 0.10)$ for $\epsilon = 0.496$ and $(1.00, 0.55, 0.55)$ for $\epsilon = 0.889$. By (4), $Y'Y$ was the resulting random Wishart matrix. For each $Y'Y$, $\widehat{\epsilon}$ was then computed. The values were then transformed into B values using the square root of (13). The distribution of the transformed values as well as the corresponding beta distribution were plotted.

Figures 1a and 1b show the accuracy of the approximations when $p = 3$. The solid line represents the approximations based on a Beta random variable, while the histogram is from the sample values from the simulation described above. Since so many replicates were used, a p-value for a goodness of fit test does not make sense because if one were used, the fit would need to be perfect in order to get a non-significant result due to the large sample size. However, the Kolmogorov-Smirnov test statistic, $KS = \max|\widehat{F}(x) - F_{\mathrm{obs}}(x)|$, was calculated as a measure of discrepancy.

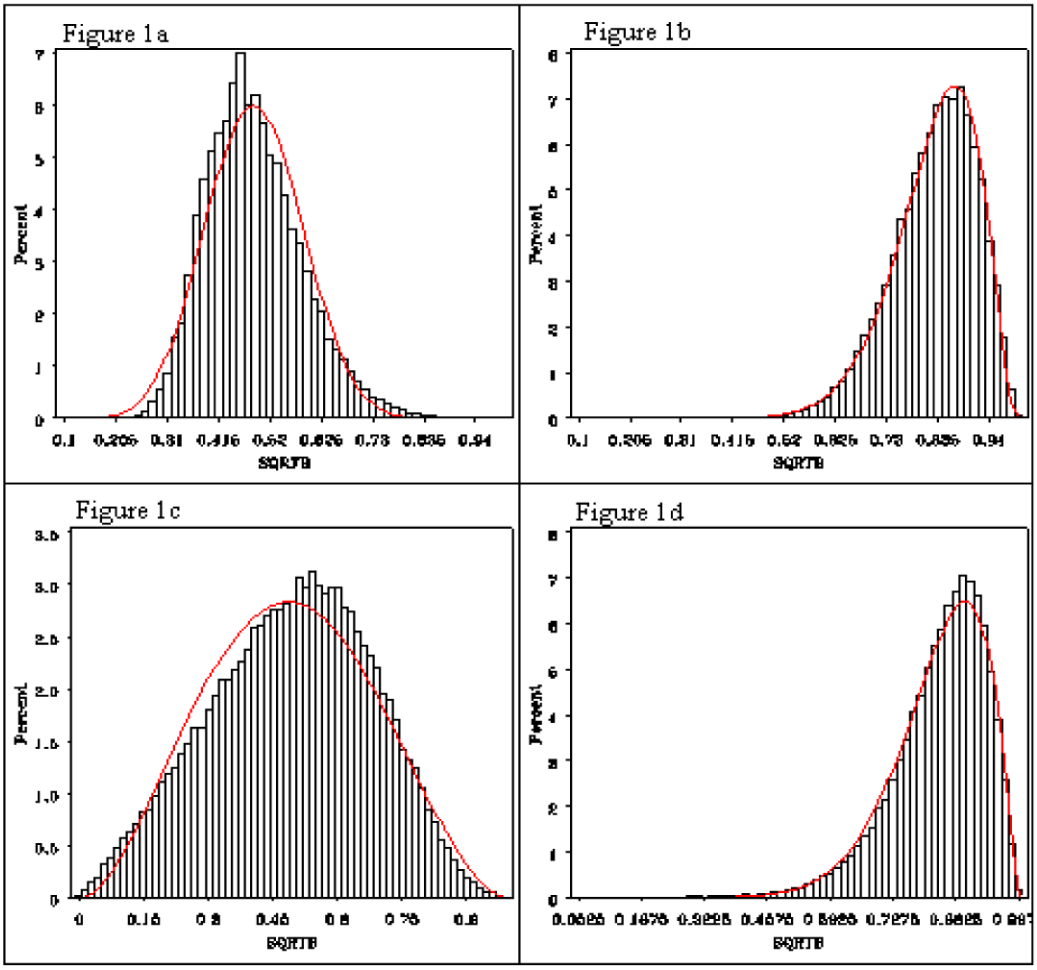

Figure 1. Approximate (line) and simulated (histogram) densities of $\widehat{\epsilon}$ when $p = 3$. 1a: $\epsilon \approx 0.496$ ($KS = 0.045$); 1b: $\epsilon \approx 0.889$ ($KS = 0.011$); 1c: $\epsilon \approx 0.496$ and SNR $= 5$ ($KS = 0.029$); 1d: $\epsilon \approx 0.889$ and SNR $= 5$ ($KS = 0.018$).

Figure 1a summarizes the simulation when $\epsilon \approx 0.496$. In Figure 1a, the fit is good, but not perfect ($KS = 0.045$). Given that the LBI test is more powerful for large values of ϵ, we would expect better fits as $\epsilon \to 1$. Figure 1b summarizs the simulation when $\epsilon \approx 0.889$. The figure shows that this distribution works extremely well for larger values of epsilon ($KS = 0.011$). This again was expected, as the LBI test is better test for large ϵ. For imaging analysis, the definition of large can be arbitrary to define. The techniques below (see Data Example section) will allow one to decide whether or not this approach would be appropriate.

When dealing with DTI data, random noise must be accounted for. Thus, an additional simulation was performed that created random $\widehat{\boldsymbol{D}}$ by using the WLS algorithm defined in Zhu et al. (2007) with 100,000 replicates. Simulated diffusion-weighted images were generated as follows: S_o, the signal intensity in the absence of a diffusion-sensitizing field gradient, was fixed at 1,500, values of σ_o were varied to provide differing signal-to-noise ratios (SNR $= S_o/\sigma_o$) of 5, 10, 15, 20, 25 and 30. Similar to the simulations in Zhu et al. (2007) paper, an imaging acquisition scheme $\{(b_i, r_i) : i = 1, \ldots, 30\}$ was defined with $\boldsymbol{r}_i$ a 1×3 vector that represents the ith direction of the diffusion gradient such that $\boldsymbol{r}_i'\boldsymbol{r}_i = 1$, and b_i is the corresponding b factor of each i^{th} DW MRI. The scheme included a total 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 $\boldsymbol{r}_i$ equivalent to the matrix provided in the Hardin (1994) web site for $m = 30$. For a given diffusion tensor, $\boldsymbol{D}$, x_i, and y_i were generated from a Gaussian random number generator with mean zero and standard deviation σ_0. Similar to the simulations above, $\boldsymbol{D}$ was defined with the diagonal elements as follows (units: 10^{-1}mm/s): $(0.80, 0.09, 0.1)$, for $\epsilon = 0.496$ and $(1.00, 0.55, 0.55)$ for $\epsilon = 0.889$.

Finally, the resulting i^{th} acquisition of the diffusion-weighted data was calculated by S_i =

$\sqrt{[S_o \exp(-b_i r_i' D r_i) + x_i]^2 + y_i^2}$. The $\widehat{\epsilon}$ values were then calculated for each $\widehat{\boldsymbol{D}}$ computed by the WLS algorithm defined in Zhu et al. (2007) with $k = 5$. Similar to the first simulation, the $\widehat{\epsilon}$ values were then transformed into B values using the square root of (13) and plotted. The distribution of the transformed values as well as the corresponding beta distribution were plotted.

Figures 1c and 1d show the accuracy of the approximations when the simulation accounts for random noise. The solid 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 1c summarizes the simulation when $\epsilon \approx 0.496$. In Figure 1c, the fit works reasonably well ($KS = 0.029$). This result is consistent with the results from the first simulation. A main reason why the addition of the noise did not hinder the fits was that the noise added was Gaussian, which helps homogenize the data. Beta random variables can be expressed as functions of chi-square variables, which are functions of Gaussian variables. Figure 1d summarizes the simulation when $\epsilon \approx 0.889$. The figure shows that this distribution works well for larger values of epsilon ($KS = 0.018$). The reason for the fit not being as good as Figure 1b is that this simulation was with SNR = 5, which means a larger σ_o was used to define x_i and y_i. As SNR increases, the fits also improve.

Thus, it has been shown that $\widehat{\epsilon}$ can be approximated by a beta-squared random variable and that the approximation holds in the DTI setting when random noise is incorporated into the diffusion matrix. For region of interest analysis of DTI images this is extremely useful, as there is now an approximate statistical distribution that can be associated with the FA values coming from a given region. As noted earlier, even better approximations or exact results are available for ADC.

Data example. The UNC Neurodevelopmental Disorders Research Center performed a study to identify if differences in the brain exist between autistic, developmentally delayed and typical children. This study was approved by both the University of North Carolina at Chapel Hill and Duke Institutional Review Boards and parental assent was obtained for all participants. Data was provided for the 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 $128 \times 128 \times 130$ image matrix at 1.875 mm $\times$ 1.875 mm $\times$ 3.8 mm resolution with a 0.4 mm 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. 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 (13). 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 performed. The data were ordered from smallest to largest for each subject.

Having the ordered data allowed for simple computation of the empirical quantile function to be used. QQ-envelope plots were used on the ordered data to see how the fit relates to other samples from the same beta distribution. Programs supplied by Marron (2007) delved into an approach called QQ-envelope 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.

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 2 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 the fits seem to work well; however, we will still look at the diagnostic tests.

In QQ-envelope, the data are plotted on the y-axis and the values from the assumed distribution are plotted on the x-axis. The red bold solid line is the Beta Q-Q plot for the transformed FA data; the green dashed line is theoretically what would occur if the data were drawn with no sampling variation from the beta distribution; and the blue dot-dash lines are the result of resampling the exact beta distribution. If the results fit, then the red line will be encompassed by the blue lines.

Figures 3 and 4 show the worst and best fits from Figure 2, respectively. The Q-Q plot for the transformed FA data (red solid line) in the QQ-envelope plots can deviate quite a bit from the theoretical beta distribution (green dashed line) and is not always encompassed in the 1,000 resamplings of the beta distributions (blue dot-dash lines). In Figure 3, it appears that the data in the lower 5^{th} percentile deviates from the estimated distribution. In Figure 4, the red line is encompassed by the blue lines at all points except a small region around 0.55 on the x-axis. The occurrence of multimodality was considered, but did not exist.

Thus, the beta distribution approximation works well in the best case and not as well for other subjects in the right cerebellum region of interest. The fits for the right cerebellum appear to work well enough to be used in analyses, as

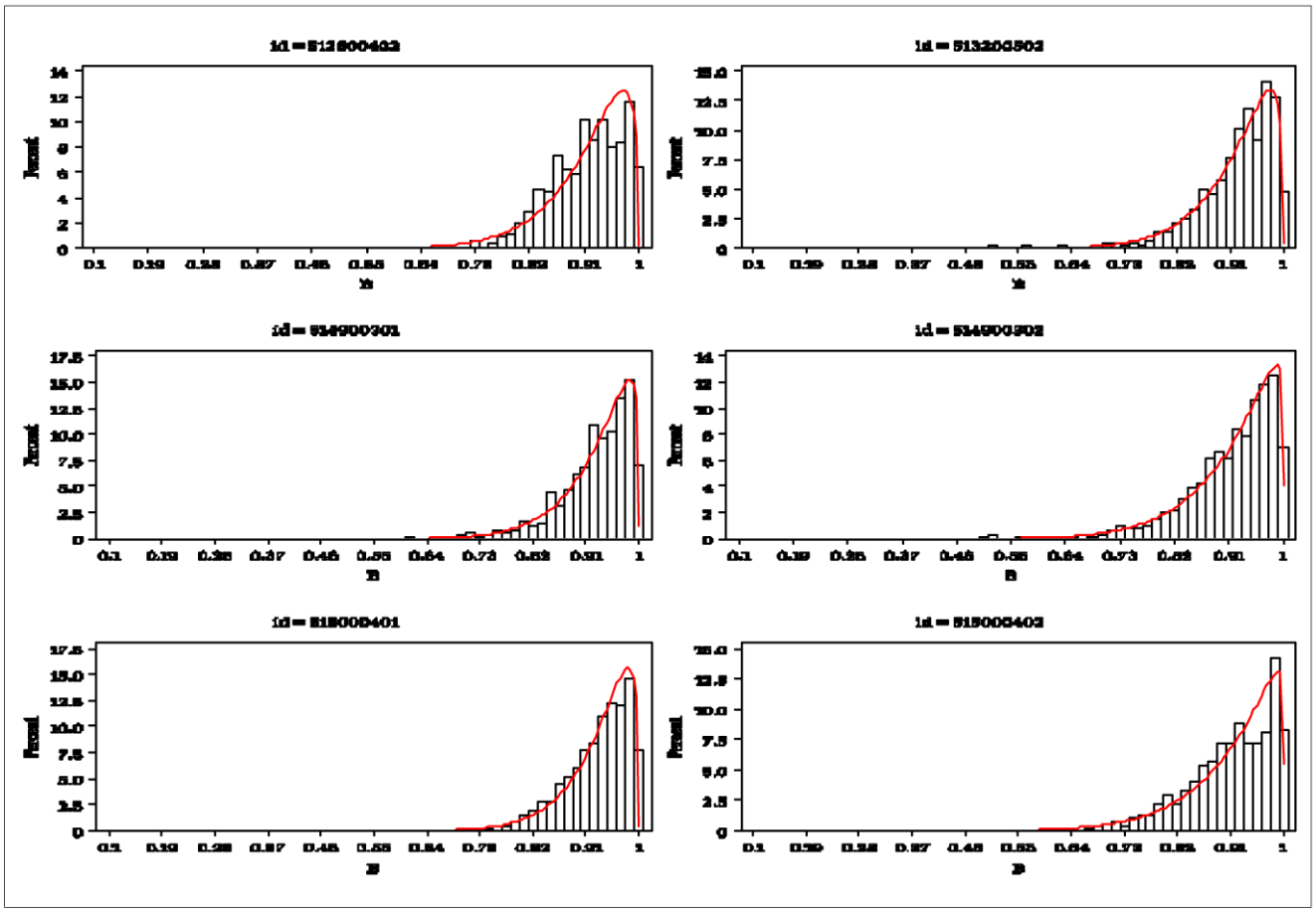

Figure 2. Six histograms of the fits of the transformed right cerebellum data; actual values (histogram) and approximate beta distribution (line).

the fit did work for all but a small fraction of the data in the worst case, mainly the left tail.

Hence, the data from a region for each subject can be represented by a Beta distribution with subject and region specific parameters. Note that the F-distribution is a one-to-one function of the Beta. Using the F distribution yields a non-bounded distribution with scale-free random variables, with several simple statistical properties. Unlike the Beta, the F varies in the same direction as FA, which simplifies understanding of the analysis of a novel outcome variable and its relationship to the physical property that it represents. Each subject can be identified with a single summary measure, $\widehat{\delta} = \text{mean} + \text{standard deviation}$ of the F distribution.

A general linear mixed model assuming Gaussian errors was fitted with $\widehat{\delta}$ as the response, and fixed effects of age, group, gender, age by group interaction. The analysis was used to look at differences in the right cerebellum between developmentally delayed children and typical children across time. Figure 5 shows the least square means of each group at given ages. Note that at less than 24 months the difference between subjects is greater than that at 60 months. Even with these differences, there was no significant difference between the two groups of children in the right cerebellum (p-value = 0.16).

4. CONCLUSIONS AND FUTURE RESEARCH

A multivariate Gaussian has a sample covariance following a Wishart distribution. This corresponds with DTI analysis in simple ways. In the absence of sphericity, Box (1954) proposed quantifying the deviation from sphericity with ϵ a function of the trace of the covariance matrix and the trace of the matrix squared. The parameter corresponds to the locally best invariant (LBI) test statistic. 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.

Common DTI measures can be expressed as functions of the LBI and LR tests. Both FA and RA essentially only de-

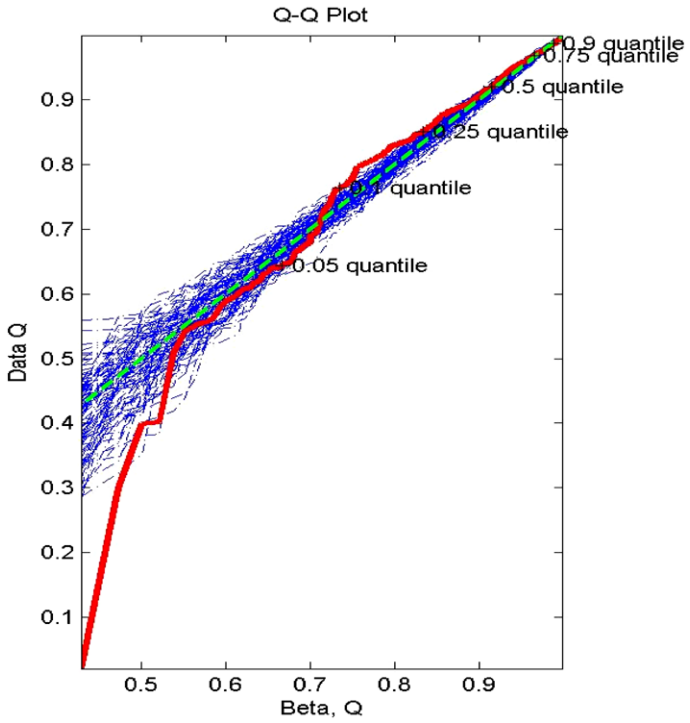

Figure 3. QQ-envelop plot of the subject (ID = 515000402) with the worst fit from Figure 2 (plot on third row, second column). Q-Q plot for the data (red solid lines), Q-Q plot for the theoretical distribution (green dashed), and Q-Q plot for the resampled data (blue dot-dashed) are displayed. Fit works well except for the left tail ($< 5^{th}$ percentile).

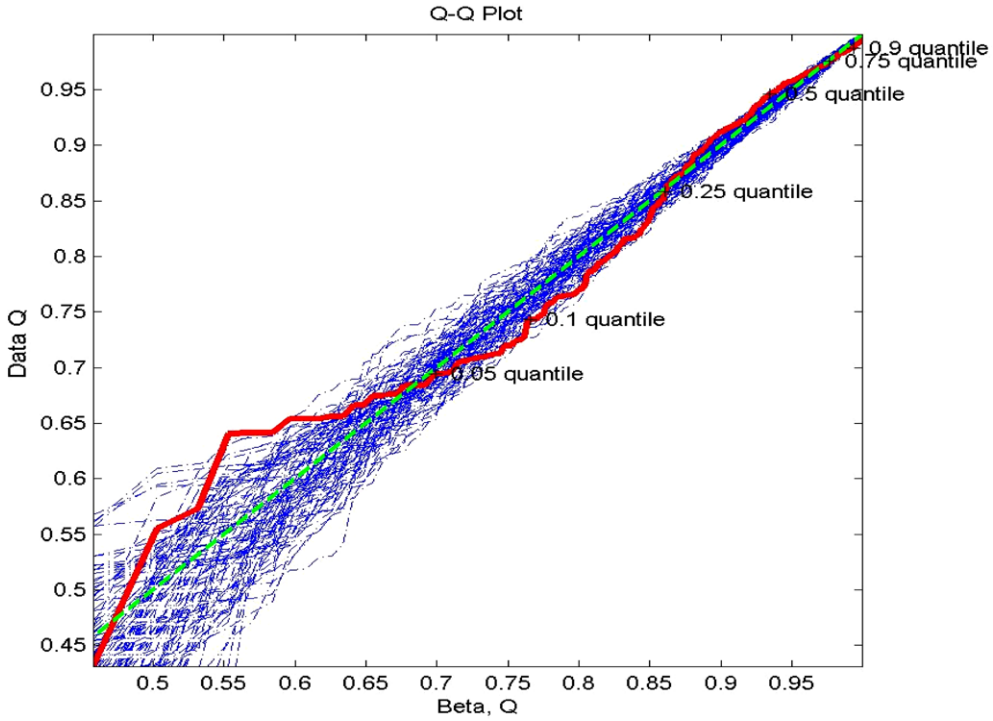

Figure 4. QQ-envelop plot of the subject (ID = 514900302) with the best fit from Figure 2 (plot on second row, second column). Q-Q plot for the data (red solid lines), Q-Q plot for the theoretical distribution (green dashed), and Q-Q plot for the resampled data (blue dot-dash) are displayed. The fit works well except for a small region around 0.55 on the x-axis.

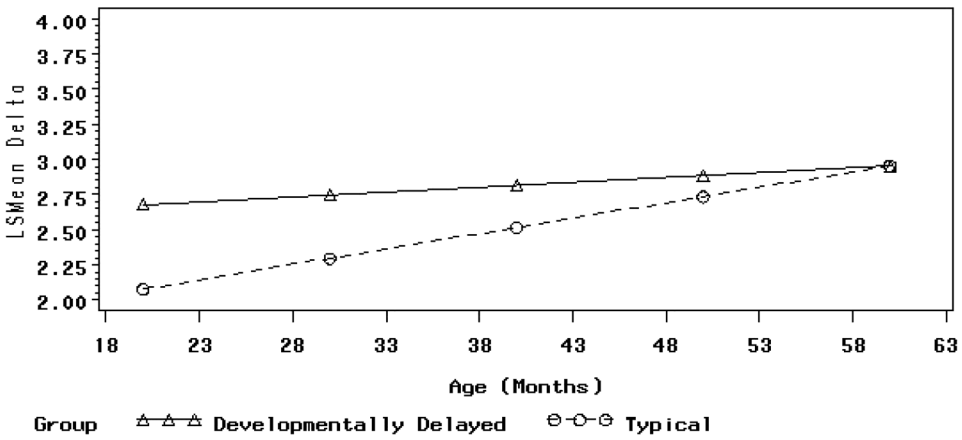

Figure 5. Least square means of $\widehat{\delta}$ across time by group.

pend on the LBI test statistic for sphericity. However, FA is scale-free; thus, the use of RA should be dismissed. Choosing between FA and VR should reflect the performance of the corresponding tests for sphericity. The LBI test for sphericity (test using FA) will be more powerful with values of ϵ near one. Symmetrically, the LR test for sphericity (test using VR) will be more powerful with small values of ϵ. 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. Experience with a limited range of MRI data for brains, as in the Autism study, shows that ϵ is typically close to one. Thus, FA is the more sensible measure.

Two types of simulations were performed to show that the approximation of FA by a squared Beta distribution is valid. 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, $\widehat{\epsilon}$ 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. This approximation was then shown to work on the UNC Neurodevelopmental data; however, it is necessary to perform diagnostic tests.

Received 31 July 2009

REFERENCES

Box, G. E. P. (1954). "Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and of correlation between errors in the two-way classification." *Annals of Mathematical Statistics* **25** 484–498. MR0064361

Glueck, D. H. and Muller, K. E. (1998). "On the trace of a Wishart." *Communications in Statistics: Theory and Methods* **27** 2137–2141. MR1646629

Grieve, A. P. (1984). "Test of sphericity of normal distributions and the analysis of repeated measures designs." *Psychometrika* **49** 257–267.

Gupta, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions*. Boca Raton, Fla.: Chapman and Hall. MR1738933

Hardin, R. H., Sloane, N. J. A., and Smith, W. D. (1994). "Minimal energy arrangements of points of a sphere with minimal 1/r potential," available at *http://www.research.att.com/∼njas/electrons/*

Hernández-Campos, F., Marron, J. S., Samorodnitsky, G., Smith, F. D. (2004). "Variable heavy tails in Internet traffic." *Performance Evaluation* **58** 261–284.

John, S. (1971). "Some optimal multivariate tests." *Biometrika* **58** 123–127. MR0275568

John, S. (1972). "The distribution of a statistic used for testing sphericity of normal distributions." *Biometrika* **59** 169–173. MR0312619

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, 2nd ed. New York, NY: John Wiley & Sons, Inc. MR1326603

Khatri, C. G. and Srivastava, M. S. (1971). "On exact non-null distributions of likelihood ratio criteria for sphericity test and equality of two covariance matrices." *Sankhyā Ser. A.* **33** 201–206. MR0323027

Kim, H. Y., Gribbin, M. J., Muller, K. E., and Taylor, D. J. (2006). "Analytic, computational, and approximate forms for the ratio of a non-central and central Gaussian quadratic forms." *Journal of Computational and Graphical Statistics* **15** 443–459. MR2256153

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions Volume 1: Models and Applications*, 2nd ed. New York, NY: John Wiley & Sons, Inc. MR1788152

Le Bihan, D., Mangin, J. F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., and Chabriat, H. (2001). "Diffusion tensor imaging: concepts and applications." *Journal of Magnetic Resonance Imaging* **13** 534–546.

Marron, J. S. (2007). 22 Sept. 2004. "qqSM Matlab Software." *http://www.unc.edu/∼marron/marron_software.html* 28 Aug. 2007.

Muller, K. E., Edwards, L. J., Simpson, S. L., and Taylor, D. J. (2007). "Statistical tests with accurate size and power for balanced linear mixed models." *Statistics in Medicine* **26** 3639–3660. MR2393741

Muller, K. E. and Stewart, P. W. (2006). *Linear Model Theory; Univariate, Multivariate and Mixed Models*. New York: Wiley. MR2242366

Puff, D. T., Pisano, E. D., Muller, K. E., Johnston, R. E., Hemminger, B. M., Burbeck, C. A., McLelland, R., and Pizer, S. M. (1994). "A method for determination of optimal image enhancement for the detection of mammographic abnormalities." *Journal of Digital Imaging* **7** 161–171.

Sugiura, N. (1972). "Locally best invariant test for sphericity and the limiting distributions." *The Annals of Mathematical Statistics* **43** 1312–1316. MR0311032

Sugiura, N. (1995). "Exact non-null distributions of sphericity tests for trivariate normal population with power comparisons." *American Journal of Mathematical and Management Sciences* **15** 355–374. MR1397513

Westin, C. F., Maier, S. E., Khidhir, B., Everett, P., Jolesz, F. A., and Kikinis, R. (1999). "Image processing for Diffusion Tensor Magnetic Resonance Imaging." In proceedings of *Second International Conference on Medical Image Computing and Computer-assisted Interventions (MICCAI'99)*, 441–452.

Zhu, H. T., Zhang, H. P., Ibrahim, J. G., and Peterson, B. (2007). "Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data (with discussion)." *Journal of the American Statistical Association* **102** 1085–1112. MR2412530

Meagan E. Clement-Spychala
Rho Inc.
6330 Quadrangle Dr.
Chapel Hill, NC 27510
E-mail address: Meagan_Spychala@rhoworld.com

David Couper
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599

Hongtu Zhu
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599

Keith E. Muller
Division of Biostatistics
University of Florida
Gainesville, FL 32611