Inference for eigenvalues and eigenvectors in exponential families of random symmetric matrices

Han Na Lee, Armin Schwartzman

PII: S0047-259X(17)30544-4
DOI: http://dx.doi.org/10.1016/j.jmva.2017.08.006
Reference: YJMVA 4283

To appear in: *Journal of Multivariate Analysis*

Received date : 10 March 2016

Inference for eigenvalues and eigenvectors in exponential families of random symmetric matrices

Han Na Lee[a], Armin Schwartzman[b]

[a] *Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC 27695-8203, USA*
[b] *Division of Biostatistics, University of California, San Diego, 9500 Gilman Drive, MC0631, La Jolla, CA 92093-0631, USA*

Abstract

Diffusion tensor imaging (DTI) data consist of a 3×3 positive definite random matrix at every voxel. Motivated by the anatomical interpretation of the data, we define a matrix-variate exponential family of distributions for random positive definite matrices and develop estimation and testing procedures for the eigenstructure of the mean parameter. The exponential family includes the spherical Gaussian and matrix-Gamma distributions as special cases. Maximum likelihood estimation and likelihood ratio testing procedures are carried out both in the one-sample and two-sample problems. In addition to their large-sample behavior, their non-asymptotic performance is evaluated via simulations. The methods are illustrated in a real DTI data example.

Keywords: Diffusion Tensor Imaging (DTI), matrix-variate Gamma distribution, statistics on manifolds, symmetric positive definite matrices, Wishart distribution.

1. Introduction

There are several situations where inference based on random symmetric matrix data is of interest. In Diffusion Tensor Imaging (DTI), the random movement of water molecules at each voxel (3D pixel) is captured by a 3×3 symmetric positive definite matrix, called "diffusion tensor" [4]. A diffusion tensor can be thought of as the covariance matrix of a Gaussian distribution, which models the physical Brownian motion of the water molecules in the voxel. Thus, the set of diffusion tensors from different subjects at a given voxel can be considered as a random sample from the population corresponding to that voxel. In addition, the eigenvalues of the diffusion tensor represent information about the type of tissue and its condition, while the eigenvectors provide information about the spatial orientation of neural fibers [4]. Thus, statistical inference for the eigenstructure of random symmetric matrices can be useful in the analysis of DTI data [27–29, 33].

In addition to DTI, inference for symmetric positive definite matrices is relevant in the analysis of covariance matrices or inverse covariance matrices representing networks, such as in brain connectivity [14, 19, 26] and stock portfolios [11]. In these applications, the number of rows (or columns) of the matrices corresponds to the number of nodes in the network, which can be in the order of several items to tens or even hundreds.

More generally, consider a $p \times p$ random symmetric matrix Y from some probability density function $f(\cdot; M)$, where M is the mean parameter of this density. Here, Y and M are in the set of $p \times p$ symmetric matrices, denoted by $\mathbb{S}_p$. Our goal is to make inference about the mean parameter M, with some assumptions on the distribution of the random matrix Y and restrictions on the parameter space, i.e., M lies in a specific subset of $\mathbb{S}_p$. Our work is focused on estimating and testing the eigenvalues and eigenvectors of M, when we have restrictions on the eigenstructure of the mean parameter matrix M. In addition, we consider the comparison of two group means, by estimating and testing the eigenvalues and eigenvectors of each group mean parameter M_1 and M_2. For the two-sample case, we are interested in comparing the eigenstructure of the two group means under restrictions on the eigenvalues and eigenvectors, such as the existence of particular multiplicities in eigenvalues, common eigenvalues, or common eigenvectors.

The problem of inference for matrix-variate data has been treated before in the multivariate statistics literature, although not so much from the angle of their eigenstructure. The statistical properties and moments of specific parametric matrix-variate distributions are provided in several books and papers. For the Wishart distribution, see [21, 32]; for the matrix-variate Gaussian distribution on rectangular random matrices, the matrix-Gamma distribution, as well as spherical and elliptically contoured distributions, refer to [13]. Our work focuses on making

Email addresses: hlee20@ncsu.edu (Han Na Lee), armins@ucsd.edu (Armin Schwartzman)

inference on random symmetric matrices by suggesting a distribution family that can embrace some commonly used distributions for symmetric matrix data, such as the spherical Gaussian, Wishart and matrix-Gamma distributions.

There are a few works about inference on the eigenstructure of $p \times p$ random symmetric matrices when the data are assumed to be from a specific distribution. Mallows [20] presented work about linear hypotheses involved in testing the eigenvectors of a single matrix in the case of a Gaussian distribution with orthogonally invariant covariance structure. This reference also provides the concept of orthogonally invariant covariance structure, which means that the distribution of a random matrix Y is the same as that of $QYQ^\top$ for any orthogonal matrix Q. In addition, maximum likelihood estimation (MLE) and a likelihood ratio test (LRT) for eigenvalues and eigenvectors of M under the assumption of a symmetric matrix-variate Gaussian distribution with orthogonally invariant covariance are derived in [27, 29]. However, these works do not model positive definite matrix data directly. Instead, a one-to-one matrix log transformation is applied to remove the positive-definite constraint before data analysis [12, 25]. We attempt to include the positive definite constraint in the model so that we can include distributions such as the Wishart and matrix-Gamma.

In hypothesis testing problems, the LRT is one of the most commonly used methods. The distribution of the log likelihood ratio (LLR) depends on the geometry of the null and alternative parameter sets. It is well known that if the null set M_0 is nested in the alternative set M_a, i.e., the null set M_0 is a subset of M_a and both sets are algebraic subsets of Euclidean space (such as affine subspaces or differentiable manifolds), then as the sample size $n \to \infty$, the LLR follows a χ^2 whose degrees of freedom d equal the difference between the dimension of the alternative set and that of the null set [21]. Thus, by embedding the matrices of interest in Euclidean space we can also perform a LRT for the parameter matrix. For example, the LLR for the eigenvalues and eigenvectors of M are asymptotically χ^2 distributed when the null and alternative sets are closed embedded submanifolds corresponding to hypotheses about the eigenvalues and eigenvectors of M [27, 29]. We use this device as well, as we consider the same hypotheses tests but different distributions.

We have assumed a parametric distribution for the matrix-variate data to make inferences for the eigenstructure of symmetric random matrix using MLE and LRT, since this assumption has all the convenience of parametric models in terms of analytic expressions and can be efficiently applied to each of tens of thousands of voxels in DTI data. One motivation of our work is that the Gaussian model, which is described in [29], may not be appropriate for the distribution of diffusion tensors in DTI. Previously, the comparison of two different groups in DTI was usually performed under the normal assumption [5], but some literature suggests that the distribution of diffusion tensors may be Gamma, rather than Gaussian [16, 24]. So, we propose to call on a class of distributions that includes several commonly used models.

In this paper, we consider an exponential family, which includes many common distributions such as the matrix-variate Gaussian, exponential, matrix-Gamma, and Wishart distribution. Exponential families for univariate and multivariate data and inference methods for them are well known [9, 21]. Analogous to other multivariate exponential families, we define a matrix-variate version of an exponential family. Analytical calculations are possible if the distribution is assumed to be equivariant under orthogonal transformations. With this exponential family form, we can estimate and test the eigenstructure of M in the general exponential family using MLE and LRT. Then we can apply the result to each specific distribution.

In our problem, the parameter space is, in general, a submanifold of Euclidean space. For instance, requiring the eigenvalue matrix to be diagonal creates a linear submanifold in the Euclidean space of square matrices of size p. In contrast, the constraint that the eigenvector matrix should be orthogonal defines an algebraic curved submanifold. For this reason, we cannot use the usual method for calculating MLEs in a Euclidean space. Instead, we obtain the critical points by finding the score function in two ways: (i) using the projection of the unrestricted gradient onto the manifold or (ii) tracing a curve directly on the manifold. Once we have the score function by either method, we obtain critical points by setting the score function equal to 0. These points maximize the likelihood because of convexity, so we can get the MLEs. Using the MLEs derived from this procedure, we perform LRTs for the hypotheses related to the constraints on the eigenstructure of the parameter matrix M.

The organization of the paper is as follows. We first demonstrate the definition and properties of the matrix-variate exponential family and give expressions for the moments of this family. Then we derive the MLE and LRT for the eigenvalues and eigenvectors of the mean parameter M with some restrictions on the eigenstructures, based on the orthogonally equivariant (OE) property of the exponential family. The inference problems are essentially the same as those considered by [29], but the solutions are given in greater generality, being applicable in a more general distribution family. Numerical studies are also provided to check the validity of our derived estimators and testing procedures under non-asymptotic conditions. The methods are illustrated with a DTI data set also analyzed in [28]. The goals of the analysis are to compare brain images and find regions of anatomic difference between two groups. Here we use our derived estimators and testing procedures instead of those derived under the

Gaussian assumption. We show that changing the modeling assumptions has a substantial impact on the results.

Proofs of all the theorems, propositions and properties in this article can be found in the Online Supplement. All the simulations and data analysis in this article were performed in R.

2. Exponential family of symmetric variate random matrices

2.1. Matrix-variate exponential family

2.1.1. Definition

Suppose the probability density of a $p \times p$ symmetric random matrix Y can be written in the form

$$f(Y;\Theta) = H(Y)e^{[\mathrm{tr}\{\eta(\Theta)T(Y)\}-K\{\eta(\Theta)\}]}, \tag{1}$$

where Θ is an original parameter matrix, the $p \times p$ symmetric matrix $T(Y)$ is referred to as a sufficient statistic, and η is called a natural parameter. Here $H(Y)$ is a scalar function that does not depend on the value of Θ. The function $K(\eta)$ is the logarithm of the normalization factor, and we assume that $K(\eta)$ is twice differentiable. Then, Y is said to belong to an exponential family of distributions, and we call (1) the canonical form of the exponential family.

In this paper, we focus on estimating and testing the mean parameter $M = \mathrm{E}\{T(Y)\}$, where Y, M, $T(Y)$, and $\eta(M)$ are symmetric. Any other parameters, such as the variance or the scale and shape parameters are considered as nuisance parameters. Therefore, we rewrite (1) as a function of M, which has the form

$$f(Y;M) = H(Y)e^{[\mathrm{tr}\{\eta(M)T(Y)\}-K\{\eta(M)\}]}. \tag{2}$$

Example 1. The probability density function (pdf) of the symmetric matrix-variate Gaussian distribution with spherical covariance in [29], denoted as $Y \sim \mathcal{N}_{pp}(M, \sigma^2 I_q)$, is given by

$$f(Y;M) = \frac{1}{(2\pi)^{q/2}\sigma^q} e^{-\frac{1}{2\sigma^2}\mathrm{tr}\{(Y-M)^2\}} = H(Y)e^{\frac{1}{2\sigma^2}\mathrm{tr}(2YM-M^2)},$$

where $q = p(p+1)/2$. As it can be written in the canonical form, it is an exponential family. In this case, we have

$$H(Y) = \frac{1}{(2\pi)^{q/2}\sigma^q}\, e^{-\frac{1}{2\sigma^2}\mathrm{tr}(Y^2)}, \quad \eta(M) = \frac{M}{\sigma^2}, \quad T(Y) = Y, \quad K\{\eta(M)\} = \frac{1}{2\sigma^2}\,\mathrm{tr}(M^2).$$

Example 2. The pdf of the matrix-Gamma distribution is given by

$$f(Y;M) = \frac{|M|^{-\alpha}}{\alpha^{-\alpha p}\Gamma_p(\alpha)}|Y|^{\alpha-\frac{p+1}{2}}e^{\alpha\mathrm{tr}\{-(M^{-1}Y)\}} = H(Y)e^{\alpha\{-\mathrm{tr}(M^{-1}Y)-\ln|M|\}},$$

where $\Gamma_p(\alpha)$ is the multivariate gamma function defined, for any positive definite matrix S, by

$$\Gamma_p(\alpha) = \pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma\left(\alpha + \frac{1-j}{2}\right).$$

As it can be rewritten in the canonical form of an exponential family, it is an exponential family. We have

$$H(Y) = \frac{|Y|^{\alpha-\frac{p+1}{2}}}{\alpha^{-\alpha p}\Gamma_p(\alpha)}, \quad \eta(M) = -\alpha M^{-1}, \quad T(Y) = Y, \quad K\{\eta(M)\} = \alpha\ln|M|.$$

Other examples include the Inverse-Gamma distribution (including the Inverse-Wishart distribution) and the matrix-variate Beta (both type I and II) distribution [13].

2.1.2. Moments

To obtain the general form of the moments of $T(Y)$ in the exponential family, we will use a vectorization operator for any symmetric matrix X defined in [29], viz.

$$\mathbf{vecd}(X) = (\mathrm{diag}(X)^\top,\ \sqrt{2}\,\mathrm{offdiag}(X)^\top)^\top,$$

where the diag(X) operator is defined as the vector whose elements are the diagonal elements of X, and offdiag(X) is defined as the vector that is a vectorization of the upper (or lower) triangular part of X. Using this operator, the correlation between diagonal elements and off-diagonal elements of a matrix is seen more clearly. The fact that all off-diagonal elements are multiplied by $\sqrt{2}$ implies that the equation $\mathbf{vecd}(A)^\top \mathbf{vecd}(B) = \mathrm{tr}(AB)$ is satisfied, so we can easily present the trace of a matrix product as an inner product of two vectors. We can also connect this definition to **vec** operator, by choosing a duplication matrix $\mathcal{D}$ such that $\mathbf{vec}(Y) = \mathcal{D}\mathbf{vecd}(Y)$. Such a duplication matrix always exists for any p, although it is generally not unique.

Theorem 1. *The expected value and covariance matrix of $T(Y)$ are as follows.*

(i) $M = \mathrm{E}\{T(Y)\} = \partial K\{\eta(M)\}/\partial\eta(M)$.

(ii) $$\mathrm{cov}[\mathbf{vecd}\{T(Y)\}] = \frac{\partial^2 K\{\eta(M)\}}{\partial\mathbf{vecd}\{\eta(M)\}\partial\mathbf{vecd}\{\eta(M)\}^\top}.$$

The proof of Theorem 1 is given in the Online Supplement. The basic idea consists of differentiating the moment generating function of $T(Y)$ with respect to a matrix argument.

2.2. Orthogonal equivariance of the exponential family

Orthogonal equivariance is an additional property imposed on the exponential family that makes it easier to obtain analytical expressions for the MLEs and LLRs considered in this work. To make this explicit, some definitions and assumptions are given first.

Definition 1. A square matrix function $a(X)$ of square matrix argument X is analytic if $a(X)$ can be represented in the form of a convergent power series such that $a(X) = \sum_{k=-\infty}^{\infty} c_k X^k$ for some coefficients c_k. In particular, an analytic function is infinitely differentiable.

Definition 2. Let X be a square matrix.

(i) A function $g(X)$ is orthogonally invariant if $g(X)$ satisfies $g(RXR^\top) = g(X)$ for any orthogonal matrix R.

(ii) A squared matrix function $g(X)$ is orthogonally equivariant (OE) if $g(X)$ satisfies $g(RXR^\top) = Rg(X)R^\top$ for any orthogonal matrix R.

Note that an analytic function $a(X)$ is OE, since an analytic function satisfies $a(PXP^{-1}) = Pa(X)P^{-1}$ for any invertible matrix P and an orthogonal matrix R satisfies $R^{-1} = R^\top$. Another way to understand the OE property from the point of view of representation theory is as follows. The special orthogonal group $SO(p)$ is a group under matrix multiplication and acts on the set of $p\times p$ symmetric positive definite matrices by a similarity transformation [18, 27]. The function g is OE if it commutes with this group action.

Definition 3. If a pdf $f(\cdot; M)$ satisfies $f(Y; M) = f(RYR^\top; RMR^\top)$ for any orthogonal matrix R, the pdf is then said to be an OE family.

Assumption 1. For a density function $f(\cdot; M)$ that has the form (2), we assume that (i) K and H are orthogonally invariant functions, η is an analytic function, and T is an OE function.

Proposition 1. *Let $f(\cdot; M)$ have the form* (2) *and suppose that Assumption 1 holds. Then, the density of Y is an OE family.*

It can be seen that both the spherical Gaussian and matrix-Gamma distributions of Examples 1 and 2 have OE density functions.

Corollary 1. *Let $W = RT(Y)R^\top$ and $\tilde{M} = RMR^\top$ for any orthogonal matrix R. Suppose the pdf of Y is an OE family. Then,*

(i) $\mathrm{E}(W) = \partial K\{\eta(\tilde{M})\}/\partial\eta(\tilde{M}) = \tilde{M}$;

(ii) $$\mathrm{cov}\{\mathbf{vecd}(W)\} = \frac{\partial^2 K\{\eta(\tilde{M})\}}{\partial\mathbf{vecd}\{\eta(\tilde{M})\}\partial\mathbf{vecd}\{\eta(\tilde{M})\}^\top}.$$

2.3. Eigenvalue parameterization

Now, consider the eigenstructure of M, denoted as $M = UDU^\top$, where D is a diagonal matrix whose elements are eigenvalues of M and U is an eigenvector matrix corresponding to D. Then, we can rewrite the pdf as a function of U and D as

$$f(Y; U, D) = H(Y)e^{[(\mathrm{tr}\{U\eta(D)U^\top T(Y)\} - K\{\eta(D)\}]}.$$

The following useful lemma establishes the moments of the random matrix $T(Y)$ rotated by the eigenvectors.

Lemma 1. *Suppose all the assumptions in Corollary 1 are satisfied and $W = RT(Y)R^\top$. The following statements then hold.*

(i) $\mathrm{E}(W) = \partial K\{\eta(\tilde{M})\}/\partial\eta(\tilde{M})\big|_{R=U^\top} = D$*;*

(ii) $\mathrm{cov}\{\mathbf{vecd}(W)\} = \left.\dfrac{\partial^2 K\{\eta(\tilde{M})\}}{\partial\mathbf{vecd}\{\eta(\tilde{M})\}\partial\mathbf{vecd}\{\eta(\tilde{M})\}^\top}\right|_{R=U^\top} = \mathcal{V}$ *is a diagonal matrix, i.e., all the elements in* $W = U^\top T(Y)U$ *are mutually uncorrelated.*

3. Inferences for eigenstructures in the one-sample problem

3.1. Likelihood for estimating M

Let $Y_1, \ldots, Y_n$ be a random sample of size n from the density (2). The log likelihood function with respect to M is then

$$\ell(M; Y_1, \ldots, Y_n) \propto \mathrm{tr}\left\{\eta(M)\sum_{i=1}^{n} T(Y_i)\right\} - nK\{\eta(M)\} = n\mathrm{tr}\{\eta(M)\bar{T}(Y)\} - nK\{\eta(M)\},$$

where $\bar{T}(Y) = \{T(Y_1) + \cdots + T(Y_n)\}/n$. In this paper, we are interested in the eigenstructure of the mean parameter. Setting $M = UDU^\top$, we see that the likelihood with respect to U and D is

$$\ell(M; Y_1, \ldots, Y_n) \propto n\mathrm{tr}\{\eta(UDU^\top)\bar{T}(Y)\} - nK\{\eta(UDU^\top)\} = n\mathrm{tr}\{U\eta(D)U^\top\bar{T}(Y)\} - nK\{\eta(D)\},$$

where the simplified expression in the second row follows from Assumption 1. Note that the last term does not depend on U. Some matrix properties that are necessary to derive the following estimates of M, U, and D are given in the Online Supplement.

3.2. Inference for M in the unrestricted case

To fix ideas, we first present the inference for the mean parameter M, without any restriction on the eigenstructure. Using the central limit theorem, we can obtain the following result.

Theorem 2. *Let $Y_1, \ldots, Y_n$ be a random sample from the pdf $f(\cdot; M)$ in (2). If there is no restriction for estimating M, then (i) the MLE of M is $\hat{M} = \bar{T}(Y)$; and (ii) the asymptotic distribution of $\hat{M}$ is*

$$\sqrt{n}\,\{\mathbf{vecd}(\hat{M}) - \mathbf{vecd}(M)\} \rightsquigarrow \mathcal{N}[0, \mathrm{cov}[\mathbf{vecd}\{T(Y)\}]],$$

where $\mathrm{cov}[\mathbf{vecd}\{T(Y)\}]$ *is the covariance matrix of* $\mathbf{vecd}\{T(Y)\}$ *in the exponential family given in Theorem 1.*

The LRT statistic for the hypothesis test of $\mathcal{H}_0 : M = M_0$ vs. $\mathcal{H}_a : M \neq M_0$ is

$$2(\ell_a - \ell_0) = 2n[\mathrm{tr}[[\eta\{\bar{T}(Y)\} - \eta(M_0)]\bar{T}(Y)] + K\{\eta(M_0)\} - K[\eta\{\bar{T}(Y)\}]],$$

and it follows a χ^2_q distribution asymptotically as $n \to \infty$. This is obtained by noticing that under $\mathcal{H}_0$, the MLE for M is just $\hat{M} = M_0$ while under $\mathcal{H}_a$, the MLE of M is $T(Y)$. This test shows whether the mean parameter has a specific value.

3.3. Inference for eigenvectors

Now, we focus on the inference for the eigenstructure. According to the restrictions for the structure of the mean parameter $M = UDU^\top$, we have different estimates. The following result gives the MLE of the eigenvectors and its asymptotic distribution.

Theorem 3. *Let $Y_1, \ldots, Y_n$ be a random sample from the pdf $f(\cdot; M)$ in (2) and assume the pdf of Y is an OE family. Suppose that D is fixed at D_0, with diagonal entries in non-increasing order, with k distinct eigenvalues and corresponding multiplicities $m_1, \ldots, m_k$.*

(i) Let the decomposition $\bar{T}(Y) = V\Lambda V^\top$ be chosen so that Λ and D_0 have their diagonal entries in the same rank order. Then, defining $\mathbb{O}_p$ as a set of $p \times p$ orthogonal matrices, the MLE of U is $\hat{U} = VQ$, where $Q \in \mathbb{O}_p$ satisfies the condition that $QD_0Q^\top = D_0$, i.e., Q is a block diagonal matrix with orthogonal blocks of size m_j. If the diagonal entries of D_0 are distinct, then Q has diagonal entries ± 1.

(ii) Suppose all the elements of D_0 are distinct so M has unique eigenvectors U up to sign. Let $\hat{U}$ be chosen to minimize the norm $\|U\hat{U}^\top - I_p\|$ and let $\hat{A} = \ln(U^\top \hat{U}) \in \mathbb{A}_p$, where $\mathbb{A}_p$ denotes a set of $p \times p$ antisymmetric matrices. Then, as the sample size n gets larger, the off-diagonal entries of $\hat{A}$ denoted by $(\hat{a}_{ij})_{i\neq j}$ are asymptotically independent, and for any $i \neq j$ one has, as $n \to \infty$,

$$\frac{\sqrt{n}\,\hat{a}_{ij}}{2\sqrt{[\{\eta(D_0)_i - \eta(D_0)_j\}^2 \mathrm{var}(w_{ij})]^{-1}}} \rightsquigarrow \mathcal{N}(0,1),$$

where w_{ij} is the (i, j)th element of $W = U^\top T(Y)U$.

Note that the variance of each off-diagonal entry (i, j) of A increases without bound as $\eta(D_0)_i$ and $\eta(D_0)_j$ get closer to one another. In the limit, when $\eta(D_0)_i = \eta(D_0)_j$, the variance goes to infinity and $\eta(D_0)_i$ and $\eta(D_0)_j$ can no longer be estimated separately because the two parameters become unidentifiable. In fact, in that situation, the dimension of the parameter space goes down by 1.

In part (ii) of Theorem 3 above, we assume that eigenvalues are known and distinct, so they can be ordered. Thus, the problem of testing whether a set of eigenvectors is equal to U_0 is formulated as a test of whether the columns of U_0 are the eigenvectors corresponding to the ordered eigenvalues in D_0. Let $\mathbb{M}_{D_0}$ be $\mathbb{M}_{D_0} = \{M = UD_0U^\top : U \in \mathbb{O}_p\}$ and $\bar{T}(Y) = V\Lambda V^\top$ be an eigendecomposition where the diagonal elements of Λ are in non-increasing order. Using the MLE in Theorem 3, the LRT statistic for testing $\mathcal{H}_0 : M = U_0D_0{U_0}^\top$ vs. $\mathcal{H}_a : M \in \mathbb{M}_{D_0}$ is

$$2(\ell_a - \ell_0) = 2n[\mathrm{tr}\{\eta(D_0)(\Lambda - {U_0}^\top \bar{T}(Y)U_0)\}],$$

which is asymptotically χ^2 with $q - \sum_{i=1}^k m_i(m_i + 1)/2$ degrees of freedom, where for each $i \in \{1, \ldots, k\}$, m_i is the multiplicity corresponding to ith distinct eigenvalue of M [29].

3.4. Inference for eigenvalues

We now consider inference for the eigenvalues of mean parameter.

Theorem 4. *Let $Y_1, \ldots, Y_n$ be a random sample from (2) and suppose that Assumption 1 is satisfied. Then the following statements hold true.*

(i) Let $\bar{W} = U^\top \bar{T}(Y)U$. If we assume that $U = U_0$ is fixed, then the MLE of D is $\hat{D} = \mathrm{diag}\{{U_0}^\top \bar{T}(Y)U_0\} = \mathrm{diag}(\bar{W})$, where $\mathrm{diag}(X)$ denotes the diagonal matrix whose diagonal elements are the diagonal entries of the matrix X.

(ii) If we assume that $U = U_0$ is fixed, then the asymptotic distribution of diagonal entries of $\hat{D} = \mathrm{diag}(\bar{W})$ denoted by $(\hat{d}_i)$ is

$$\frac{\sqrt{n}\,(\hat{d}_i - d_i)}{\sqrt{\mathrm{var}(w_{ii})}} \rightsquigarrow \mathcal{N}(0,1),$$

where w_{ii} is the ith diagonal element of $W = U^\top YU$.

An associated test is whether the eigenvalues of the mean parameter M are different from some fixed values while the eigenvectors are fixed. Let $\mathbb{M}_{U_0} = \{M = U_0DU_0{}^\top : D \in \mathbb{D}_p\}$, where $\mathbb{D}_p$ denotes the set of $p \times p$ diagonal matrices. Then, under the hypothesis that $\mathcal{H}_a : M \in \mathbb{M}_{U_0}$, the MLE of D is $\hat{D} = \mathrm{diag}\{U_0{}^\top T(Y)U_0\}$, using the result of Theorem 4 with restriction $U = U_0$. The LRT statistic for testing $\mathcal{H}_0 : M = U_0 D_0 U_0{}^\top$ vs. $\mathcal{H}_a : M \in \mathbb{M}_{U_0}$ is

$$2(\ell_a - \ell_0) = 2n[\mathrm{tr}[U_0\{\eta(\hat{D}) - \eta(D_0)\}U_0{}^\top \bar{T}(Y)] - K\{\eta(\hat{D})\} + K\{\eta(D_0)\}]$$

and it asymptotically follows χ^2_p, since the number of parameters tested is p.

3.5. MLE of eigenvectors and eigenvalues with known multiplicities

In this section, we assume that we do not know the eigenvalues or the eigenvectors, but the order and multiplicity of the eigenvalues are known. To state the results, we first need to define how to perform block averages.

Definition 4. Define the *block average* of a diagonal matrix B according to the multiplicities $m_1, \ldots, m_k$, as the block diagonal matrix formed by partitioning B into k diagonal blocks of sizes $m_1, \ldots, m_k$ and replacing the diagonal entries in each block by the mean of the diagonal entries in each block. Denote this term operator as $\mathrm{blk}_{m_1,\ldots,m_k}(B)$. If we denote the original diagonal blocks as $B_{11}, \ldots, B_{kk}$, then the jth diagonal block of $\mathrm{blk}_{m_1,\ldots,m_k}(B)$ has the form $(1/m_j)\mathrm{tr}(B_{jj})I_{m_j}$.

Definition 5. For any square matrix B, define $\mathbf{Blk}_{m_1,\ldots,m_k}(B) = \mathrm{blk}_{m_1,\ldots,m_k}\{\mathrm{diag}(B)\}$. Denote the diagonal block of B with respect to the multiplicities $m_1, \ldots, m_k$ as $B_{11}, \ldots, B_{kk}$. Then, each diagonal block of $\mathbf{Blk}_{m_1,\ldots,m_k}(B)$ has the form $(1/m_j)\mathrm{tr}(B_{jj})I_{m_j}$. If B is a diagonal matrix, then $\mathbf{Blk}_{m_1,\ldots,m_k}(B) = \mathrm{blk}_{m_1,\ldots,m_k}(B)$.

Theorem 5. *Let $\mathbb{M}_{m_1,\ldots,m_k}$ be the set of symmetric matrices which have k distinct eigenvalues with multiplicities $m_1, \ldots, m_k$, i.e., $\mathbb{M}_{m_1,\ldots,m_k} = \{M = UDU^\top : U \in \mathbb{O}_p, D \in \mathbb{D}_p, d_1 \geq \cdots \geq d_k, \text{mult. } m_1, \ldots, m_k\}$. Let $\bar{T}(Y) = V\Lambda V^\top$ be an eigendecomposition, where the diagonal elements of Λ are in non-increasing order. Suppose that D is unknown but the order and multiplicities of D are known. Then the MLE of M is $\hat{M} = \hat{U}\hat{D}\hat{U}^\top$, where $\hat{D} = \mathbf{Blk}_{m_1,\ldots,m_k}(\Lambda)$ and $\hat{U}$ is any matrix that has the form $\hat{U} = VQ$ and $Q \in \mathbb{O}_p$ satisfies the condition that $Q\hat{D}Q^\top$ is a diagonal matrix. Here, Q must be a block diagonal matrix with orthogonal blocks of size $m_1, \ldots, m_k$, in that order.*

Using the MLE values obtained from Theorem 5, we can test whether the eigenvalues of the matrix M have particular multiplicities. Using the notation $\mathbb{M}_{m_1,\ldots,m_k} = \{M = UDU^\top : U \in \mathbb{O}_p, D \in \mathbb{D}_p, d_1 \geq \cdots \geq d_k, \text{mult. } m_1, \ldots, m_k\}$, the test compares $\mathcal{H}_0 : M = UDU^\top \in \mathbb{M}_{m_1,\ldots,m_k}$ to $\mathcal{H}_a : M \notin \mathbb{M}_{m_1,\ldots,m_k}$. The test statistic is

$$2(\ell_a - \ell_0) = 2n[\mathrm{tr}\{(\eta(\Lambda) - Q\eta(\hat{D})Q^\top)\Lambda\} - K\{\eta(\Lambda)\} + K\{\eta(\hat{D})\}],$$

where $\hat{D}$ is the MLE under the restriction, which is obtained from Theorem 5, and $\hat{M} = \bar{T}(Y) = V\Lambda V^\top$ is the MLE from Theorem 2, which maximizes the likelihood without any restriction. It is asymptotically χ^2 with $\sum_{i=1}^k m_i(m_i + 1)/2 - k$ degrees of freedom. The number of degrees of freedom is the difference between the dimension of alternative parameter space q and the dimension of $\mathbb{M}_{m_1,\ldots,m_k}$, which is equal to $k + q - \sum_{i=1}^k m_i(m_i + 1)/2$.

4. Inferences for eigenstructures in the two-sample case

In the previous section, we only considered problems for one-sample cases. Now, we will look at two-sample problems. Let $Y_1, \ldots, Y_{n_1}$ and $Y_{n_1+1}, \ldots, Y_n$, $n = n_1 + n_2$ be two independent iid samples from two exponential family distributions $f(Y; M_1)$ and $f(Y; M_2)$, and let M_1, M_2 be their respective means. Here, we are interested in estimating and testing the pair of means $M = (M_1, M_2)$. In general, the inference depends on other nuisance parameters, such as variance parameters and shape parameters.

4.1. Estimation in the two-sample case under the same nuisance parameters

In order to obtain analytical expressions, we assume here that the nuisance parameters are the same in both distributions. We start from the joint likelihood function of the two samples. With the assumption that $\eta_1 = \eta_2 = \eta$, the form of the joint likelihood function is

$$L = \prod_{i=1}^{n_1} H(Y_i) e^{[\mathrm{tr}\{\eta(M_1)T_1(Y_i)\} - K\{\eta(M_1)\}]} \prod_{j=n_1+1}^{n} H(Y_j) e^{[\mathrm{tr}\{\eta(M_2)T_2(Y_j)\} - K\{\eta(M_2)\}]}.$$

Now, assuming the eigendecompositions $M_1 = U_1 D U_1^\top$ and $M_2 = U_2 D U_2^\top$ with common eigenvalue matrix D, the log likelihood can be written as

$$\begin{aligned}\ell(M; Y_1, \ldots, Y_n) &\propto n_1[\mathrm{tr}\{\eta(M_1)\bar{T}_1(Y)\} - K\{\eta(M_1)\}] + n_2[\mathrm{tr}\{\eta(M_2)\bar{T}_2(Y)\} - K\{\eta(M_2)\}] \\ &\propto n_1 \mathrm{tr}\{U_1 \eta(D) U_1^\top \bar{T}_1(Y)\} + n_2 \mathrm{tr}\{U_2 \eta(D) U_2^\top \bar{T}_2(Y)\} - (n_1 + n_2) K\{\eta(D)\}.\end{aligned}$$

Theorem 6. *Define the set $\mathbb{M}_{2,D} = \{(M_1, M_2) : M_1 = U_1 D U_1{}^\top, M_2 = U_2 D U_2{}^\top, U_1, U_2 \in \mathbb{O}_p, D \in \mathbb{D}_p$, mult. $m_1, \ldots, m_k$, diag. entries of D are in non-increasing order$\}$ and let $\bar{Y}_1 = V_1 \Lambda_1 V_1{}^\top$, $\bar{Y}_2 = V_2 \Lambda_2 V_2{}^\top$ be eigendecompositions of the two sample means, where the diagonal elements of Λ_1, Λ_2 are in non-increasing order. Suppose that D is unknown but the order and multiplicities of D are known and η is the same for the two groups. Then, the MLEs of M_1 and M_2 are $\hat{M}_1 = \hat{U}_1 \hat{D} \hat{U}_1^\top$ and $\hat{M}_2 = \hat{U}_2 \hat{D} \hat{U}_2^\top$, respectively, where*

(i) $\hat{U}_1$ and $\hat{U}_2$ are any matrices that have the form $\hat{U}_1 = V_1 Q_1$ and $\hat{U}_2 = V_2 Q_2$, where $Q_1, Q_2 \in \mathbb{O}_p$ satisfy the condition that $Q_1 \hat{D} Q_1^\top = \hat{D}$, $Q_2 \hat{D} Q_2^\top = \hat{D}$;

(ii) $\hat{D} = \boldsymbol{Blk}_{m_1,\ldots,m_k}\left(\dfrac{n_1 \Lambda_1 + n_2 \Lambda_2}{n_1 + n_2}\right)$.

Next, we assume that $\eta_1(D) \neq \eta_2(D)$. Then, the log likelihood function is

$$\ell(M; Y_1, \ldots, Y_n) \propto n_1 \mathrm{tr}\{U_1 \eta_1(D) U_1^\top \bar{T}_1(Y)\} + n_2 \mathrm{tr}\{U_2 \eta_2(D) U_2^\top \bar{T}_2(Y)\} - n_1 K\{\eta_1(D)\} - n_2 K\{\eta_2(D)\}.$$

Theorem 7. *Let $\bar{Y}_1 = V_1 \Lambda_1 V_1{}^\top$, $\bar{Y}_2 = V_2 \Lambda_2 V_2{}^\top$ be eigendecompositions of two sample means, where the diagonal elements of Λ_1, Λ_2 are in non-increasing order. Suppose that D has the multiplicities $m_1, \ldots, m_k$ and is the same for two groups but that the η's are different. Then, the MLEs of M_1 and M_2 are $\hat{M}_1 = \hat{U}_1 \hat{D} \hat{U}_1^\top$ and $\hat{M}_2 = \hat{U}_2 \hat{D} \hat{U}_2^\top$, respectively, where*

(i) $\hat{U}_1 = V_1 Q_1$ and $\hat{U}_2 = V_2 Q_2$, where $Q_1, Q_2 \in \mathbb{O}_p$ satisfy the condition that $Q_1 \hat{D} Q_1^\top = \hat{D}$, $Q_2 \hat{D} Q_2^\top = \hat{D}$.

(ii) $\hat{D}_{ii} = \boldsymbol{Blk}_{m_1,\ldots,m_k}\left\{\dfrac{n_1 (\Lambda_1)_{ii} \frac{\partial \eta_1(D)_{ii}}{\partial D_{ii}} + n_2 (\Lambda_2)_{ii} \frac{\partial \eta_2(D)_{ii}}{\partial D_{ii}}}{n_1 \frac{\partial \eta_1(D)_{ii}}{\partial D_{ii}} + n_2 \frac{\partial \eta_2(D)_{ii}}{\partial D_{ii}}}\right\}$.

4.2. Testing in the two-sample case under the same nuisance parameters

4.2.1. Full matrix test

First we start with the test of $\mathcal{H}_0 : M_1 = M_2$ vs. $\mathcal{H}_a : M_1 \neq M_2$. This is a test of whether the means of two groups are equal. Under $\mathcal{H}_0$, the MLE for M_1, M_2 is just the MLE of the one-sample case with sample size $n = n_1 + n_2$. By setting $\bar{T}(Y) = \{n_1 \bar{T}_1(Y) + n_2 \bar{T}_2(Y)\}/n$, we have $\hat{M}_1 = \hat{M}_2 = \bar{T}(Y)$. Then, the LRT statistic for this hypothesis is

$$\begin{aligned}2(\ell_a - \ell_0) = 2[&n_1 \mathrm{tr}\{\eta(\bar{T}_1(Y))\bar{T}_1(Y)\} + n_2 \mathrm{tr}\{\eta(\bar{T}_2(Y))\bar{T}_2(Y)\} - n_1 K\{\eta(\bar{T}_1(Y))\} \\ &- n_2 K\{\eta(\bar{T}_2(Y))\} - n \mathrm{tr}\{\eta(\bar{T}(Y))\bar{T}(Y)\} + n K\{\eta(\bar{T}(Y))\}],\end{aligned}$$

and it follows a χ_q^2 distribution asymptotically, since the number of parameters tested is q.

4.2.2. Testing for the eigenvectors

We move to the testing for the eigenstructure of M_1 and M_2. A test of eigenvectors can be performed by testing the hypothesis that $\mathcal{H}_0 : M_1 = M_2$ vs. $\mathcal{H}_a : (M_1, M_2) \in \mathbb{M}_{2,D}$. This is the test of equal sets of eigenvectors when the two mean parameters have the same eigenvalues. Under $\mathcal{H}_0$, the MLE for M_1, M_2 is just the MLE in the one-sample case with sample size $n = n_1 + n_2$. Setting $\bar{T}(Y) = \{n_1 \bar{T}_1(Y) + n_2 \bar{T}_2(Y)\}/n = V \Lambda V^\top$, we have $M_1 = M_2 = \hat{U} \hat{D}^* \hat{U}^\top$, where $\hat{U} = VQ$ and $\hat{D}^* = \mathbf{Blk}_{m_1,\ldots,m_k}(\Lambda)$ are calculated by the one-sample case formula in Theorem 5, with the sample size $n = n_1 + n_2$. Then, the test statistic is

$$\begin{aligned}2(\ell_a - \ell_0) &= 2[n_1 \mathrm{tr}\{\hat{U}_1 \eta(\hat{D}) \hat{U}_1^\top \bar{T}_1(Y)\} + n_2 \mathrm{tr}\{\hat{U}_2 \eta(\hat{D}) \hat{U}_2^\top \bar{T}_2(Y)\} - n K\{\eta(\hat{D})\} - n \mathrm{tr}\{\hat{U} \eta(\hat{D}^*) \hat{U}^\top \bar{T}(Y)\} + n K\{\eta(\hat{D}^*)\}] \\ &= 2[n_1 \mathrm{tr}\{\eta(\hat{D})\Lambda_1\} + n_2 \mathrm{tr}\{\eta(\hat{D})\Lambda_2\} - n \mathrm{tr}\{\eta(\hat{D}^*)\Lambda\} + n\{K(\eta(\hat{D}^*)) - K(\eta(\hat{D}))\}],\end{aligned}$$

and it follows the χ^2 distribution asymptotically. The degrees of freedom are the difference between the dimension of the alternative parameter set and that of the null parameter set, viz.

$$\left\{k + 2q - \sum_{i=1}^{k} m_i(m_i+1)\right\} - \left\{q + k - \sum_{i=1}^{k} m_i(m_i+1)/2\right\} = q - \sum_{i=1}^{k} m_i(m_i+1)/2.$$

4.2.3. Testing for the eigenvalues

The next test is a test of whether the means of the two groups have the same eigenvalues, while differing on their eigenvectors. The test is $\mathcal{H}_0 : (M_1, M_2) \in \mathbb{M}_{2,D}$ vs. $\mathcal{H}_a : (M_1, M_2) \notin \mathbb{M}_{2,D}$. The MLE for M_1 and M_2 under the alternative hypothesis are just $\hat{M}_1 = \bar{T}_1(Y)$, $\hat{M}_2 = \bar{T}_2(Y)$. Therefore, the test statistic is

$$\begin{aligned} 2(\ell_a - \ell_0) &= 2[n_1 \mathrm{tr}\{\eta(\bar{T}_1(Y))\bar{T}_1(Y)\} + n_2 \mathrm{tr}\{\eta(\bar{T}_2(Y))\bar{T}_2(Y)\} - n_1 K\{\eta(\bar{T}_1(Y))\} \\ &\quad - n_2 K\{\eta(\bar{T}_2(Y))\} - n_1 \mathrm{tr}\{\hat{U}_1 \eta(\hat{D})\hat{U}_1^\top \bar{T}_1(Y)\} - n_2 \mathrm{tr}\{\hat{U}_2 \eta(\hat{D})\hat{U}_2^\top \bar{T}_2(Y)\} - nK\{\eta(\hat{D})\}] \\ &= 2[n_1 \mathrm{tr}\{\eta(\Lambda_1)\Lambda_1\} + n_2 \mathrm{tr}\{\eta(\Lambda_2)\Lambda_2\} - n_1 K\{\eta(\Lambda_1)\} - n_2 K\{\eta(\Lambda_2)\} \\ &\quad - n_1 \mathrm{tr}\{\eta(\hat{D})\Lambda_1\} - n_2 \mathrm{tr}\{\eta(\hat{D})\Lambda_2\} + K\{\eta(\hat{D})\}] \end{aligned}$$

and it is asymptotically χ^2. The degrees of freedom are the difference between the dimension of the alternative parameter set and that of the null parameter set, viz.

$$2q - \left\{k + 2q - \sum_{i=1}^{k} m_i(m_i+1)\right\} = \sum_{i=1}^{k} m_i(m_i+1) - k.$$

4.3. Extension to multiple groups

We can extend the idea of estimation and hypothesis testing described in Sections 4.1. and 4.2. to the multiple-group case. For each $i \in \{1, \dots, g\}$, let $Y_{i1}, \dots, Y_{in_i}$ be g independent iid samples from g exponential family distributions $f(Y; M_i)$. Let also $n = n_1 + \cdots + n_g$. Here the g distributions have the same form but different mean parameter $M_1, \dots, M_g$. We can write the joint likelihood of g groups as follows:

$$L = \prod_{i=1}^{g} \prod_{j=1}^{n_i} H(Y_{ij}) e^{[\mathrm{tr}\{\eta_i(M_i)T_i(Y_{ij})\} - K_i\{\eta_i(M_i)\}]}.$$

Using the eigendecompositions $M_i = U_i D_i U_i^\top$, the log likelihood of g group data can be written as

$$\ell(M_i; Y_{11}, \dots, Y_{gn_g}) \propto \sum_{i=1}^{g} n_i[\mathrm{tr}\{\eta_i(M_i)\bar{T}_i(Y)\} - K_i\{\eta_i(M_i)\}] \propto \sum_{i=1}^{g} n_i[\mathrm{tr}\{U_i \eta_i(D_i) U_i^\top \bar{T}_i(Y)\} - K_i\{\eta_i(D_i)\}].$$

Then, we can estimate the MLE of M_i, U_i and D_i using this joint log- likelihood. In particular, estimates may be obtained under ANOVA-type assumptions such as $M_1 = \cdots = M_g$ or $D_1 = \cdots = D_g$ with given eigenvalue multiplicities. In the latter case, multi-group versions of Theorems 6–7 hold for common and different η, respectively. The eigenvector estimates have the same forms as the ones there and the eigenvalue estimates contain the sums of g terms instead of only two. Given these estimates, the log-likelihood ratio test statistics can be written down similar to those in Sections 4.2.1, 4.2.2 and 4.2.3 (with indices ranging from 1 to g instead of 1 to 2) in order to test equality of means, equality of eigenvectors and equality of eigenvalues, respectively. The number of degrees of freedom can be easily adjusted accordingly.

5. Estimating nuisance parameters

In addition to estimating M and calculating related test statistics, in practice, we need to estimate other parameters such as variance components, scale parameters, and shape parameters. Thus, we need to estimate these parameters under the restrictions we have made for the mean parameter.

5.1. Gaussian distribution with spherical covariance matrix

5.1.1. Test of sphericity of covariance matrices

In the Gaussian distribution example, we assume that the covariance matrix is spherical. Thus, estimation and tests for M and σ^2 are derived based on this assumption. In data analysis, however, we need to test whether this assumption holds. Here, we use a special case of the test of orthogonal invariance described in [29]. The test statistic considered there is

$$T_o = nq \ln(\hat{\sigma}^2) - n \ln(1 - p\hat{\tau}) - n \ln |S|,$$

where S is the sample covariance matrix and τ is an additional parameter required by the orthogonally invariant Gaussian distribution. This test statistic follows a χ^2 distribution with $df = q(q+1)/2 - 2$, since the dimension of the parameter space under $\mathcal{H}_a$ is $q(q+1)/2$ and the dimension of null parameter space is 2.

To test sphericity, we set $\tau = 0$ in the above test. Then, the null hypothesis for this test is that the covariance matrix is spherical. As $\tau = 0$, the test statistic for the sphericity of a covariance matrix is simplified as

$$T_s = nq \ln(\hat{\sigma}^2) - n \ln |S|,$$

which follows the χ^2 distribution with $q(q+1)/2 - 1$ degrees of freedom. The number of degrees of freedom changes because the dimension of the null parameter space is reduced to 1.

5.1.2. Estimating covariance parameters in the one-sample case

If the null hypothesis of the sphericity test is not rejected, the next step is to find an appropriate way to estimate the variance parameter. In the case of Gaussian distribution with a spherical covariance matrix that has the likelihood

$$\ell(M, \sigma^2; Y) = -\frac{nq}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \mathrm{tr}\{(Y_i - M)^2\},$$

we have the variance parameter σ^2. To find the distribution of $\hat{M}$, we also need to estimate σ^2. Here, we use the MLE for σ^2 described in [29], viz.

$$\hat{\sigma}^2 = \frac{1}{qn} \sum_{i=1}^{n} \mathrm{tr}(Y_i - \bar{Y})^2 + \frac{1}{q} \mathrm{tr}(\bar{Y} - \hat{M})^2, \tag{3}$$

where $\hat{M}$ is the MLE under any of the eigenvalue or eigenvector restrictions. This form is obtained by taking the derivative with respect to σ^2 and setting it equal to 0.

Proposition 2. *Let $\hat{\sigma}^2$ be given by* (3)*. Then $qn\hat{\sigma}^2/\sigma^2 \sim \chi^2_{q(n-1)+df_r}$, where df_r is the difference between the dimension of $\bar{Y}$ and the dimension of $\hat{M}$.*

5.1.3. Estimating covariance parameters in the two-sample case

For the two-sample Gaussian case, we have two options: equal or unequal covariance matrix assumption. In general, the equal covariance matrix assumption is preferred because it makes it possible to use Hotelling's T^2 test when we compare two group means, which is simpler and has an exact distribution. If the covariance matrices are not equal, we should use other alternative tests such as the Behrens–Fisher test, which is more complicated and its distribution is only approximate.

If we assume that the covariance matrices of the two groups are spherical and equal, then we use a pooled estimate $\hat{\sigma}^2_{12}$ of the common variance. The estimate has the form

$$\hat{\sigma}^2_{12} = s^2_{12} + \frac{1}{q(n_1 + n_2)} \{n_1 \mathrm{tr}(\bar{Y}_1 - \hat{M}_1)^2 + n_2 \mathrm{tr}(\bar{Y}_2 - \hat{M}_2)^2\},$$

where

$$s^2_{12} = \frac{1}{q(n_1 + n_2)} \left\{ \sum_{i=1}^{n_1} \mathrm{tr}(Y_i - \bar{Y}_1)^2 + \sum_{i=1}^{n_2} \mathrm{tr}(Y_i - \bar{Y}_2)^2 \right\}$$

and $\hat{M}_1, \hat{M}_2$ are the MLEs under the relevant eigenvalue and eigenvector restrictions [29]. Like the one-sample case, $\{n_1 \mathrm{tr}(\bar{Y}_1 - \hat{M}_1)^2 + n_2 \mathrm{tr}(\bar{Y}_2 - \hat{M}_2)^2\}/\sigma^2$ also follows the χ^2 distribution, with degrees of freedom that depend on the differences between the dimension of $(\bar{Y}_1, \bar{Y}_2)$ and that of $(\hat{M}_1, \hat{M}_2)$ (denoted by df_{r2}).

If we assume that two variance parameters are different, we can estimate two variance parameters separately. For $i \in \{1, 2\}$, we have

$$\hat{\sigma}_i{}^2 = \frac{1}{qn_i} \sum_{j=1}^{n_i} \mathrm{tr}(Y_{ij} - \bar{Y}_i)^2 + \frac{1}{q} \mathrm{tr}(\bar{Y}_i - \hat{M}_i)^2,$$

where $\hat{M}_i$s are the MLE of M_i under the corresponding restrictions. The procedure is the same as the one-sample case described in Proposition 2, since each group can be treated as one sample.

In addition, using the property that

$$s_i^2 = 1/qn_i \sum_{i=1}^{n_i} \mathrm{tr}(Y_i - \bar{Y}_i)^2$$

has the distribution $qn_i s_i^2/\sigma^2 \sim \chi^2_{q(n_i-1)}$ for $i \in \{1,2\}$, as mentioned in the proof of Proposition 5.1, $n_i \mathrm{tr}(\bar{Y}_i - \hat{M}_i)^2/\sigma^2$ also follows the χ^2 distribution, with degrees of freedom which is the difference between the dimension of $\bar{Y}_i$ and the dimension of $\hat{M}_i$ (denoted by df_{r_i}). As described in Proposition 2, note that $\hat{\sigma}_i^2$ is equal to the sum of ${s_i}^2$ and $2(\ell_a - \ell_0)$ and the two are independent. Since the sum of two independent χ^2-distributed statistics also follows a χ^2 distribution, $qn_i\hat{\sigma}_i^2/\sigma^2$ follows a χ^2 distribution with degrees of freedom $q(n_i-1) + df_{r_i}$.

5.2. Matrix-Gamma distribution

5.2.1. Estimating shape parameters in the one-sample case

Recall that the pdf of matrix-Gamma distribution described in Example 2 is

$$f(Y;\alpha, M) = \frac{|M|^{-\alpha}}{\alpha^{-\alpha p}\Gamma_p(\alpha)} |Y|^{\alpha - \frac{p+1}{2}} e^{\alpha \mathrm{tr}\{-(M^{-1}Y)\}},$$

where α is a shape parameter. In one-sample cases, α can be estimated from the first derivative of the log likelihood. Starting from the log likelihood with respect to α and M, we have

$$\ell(M,\alpha) \propto n\left[-\ln\Gamma_p(\alpha) + \alpha\left\{p\ln\alpha - \ln|M| + \frac{1}{n}\sum_{i=1}^{n}\ln|Y_i| - \mathrm{tr}(M^{-1}\bar{Y})\right\}\right].$$

Taking the derivative with respect to α, we have

$$\frac{\partial \ell}{\partial\alpha} = -n\ln|M| - n\frac{\partial \ln\Gamma_p(\alpha)}{\partial\alpha} + np\ln\alpha + np + \sum_{i=1}^{n}\ln|Y_i| - n\mathrm{tr}(M^{-1}\bar{Y}).$$

Thus, the MLE of α should satisfy the condition that the first derivative of likelihood evaluated at $M = \hat{M}$, the MLE under each restriction, is equal to 0. Therefore, the MLE of α satisfies the equation

$$p\ln\alpha - \frac{\partial\ln\Gamma_p(\alpha)}{\partial\alpha} = \ln|\hat{M}| + \mathrm{tr}(\hat{M}^{-1}\bar{Y}) - \frac{1}{n}\sum_{i=1}^{n}\ln|Y_i| - p.$$

Since there is no analytical expression for multivariate Gamma function, there is no closed-form solution for α. However, the equation for α presented above is numerically well behaved. Therefore, if a numerical solution is desired, it can be found using an approximation method. Thus, we estimated $\hat{\alpha}$ using this formula and numerical computation via Newton's method.

As $\hat{\alpha}$ is the MLE, we can find the asymptotical distribution of the estimated value, using the property of MLE that the asymptotic variance is the inverse of Fisher's information number.

Proposition 3. *Let $\hat{\alpha}$ be the MLE of α. Then,*

$$\sqrt{n}\,(\hat{\alpha} - \alpha) \rightsquigarrow \mathcal{N}\left[0, \left\{\frac{\partial^2\ln\Gamma_p(\alpha)}{\partial\alpha^2} - \frac{p}{\alpha}\right\}^{-1}\right].$$

5.2.2. Estimating shape parameters in the two-sample case

For the two-sample case, α_1 and α_2 can be estimated in the same manner as in the one-sample case. We have the likelihood

$$\ell(M_1, M_2, \alpha_1^2, \alpha_2^2) \propto n_1\left[\ln\Gamma_p(\alpha_1) + \alpha_1\left\{p\ln\alpha_1 - \ln|M_1| + \frac{1}{n_1}\sum_{i=1}^{n_1}\ln|Y_i| - \mathrm{tr}(M_1^{-T}\bar{Y}_1)\right\}\right]$$
$$+ n_2\left[\ln\Gamma_p(\alpha_2) + \alpha_2\left\{p\ln\alpha_2 - \ln|M_2| + \frac{1}{n_2}\sum_{j=n_1+1}^{n_1+n_2}\ln|Y_j| - \mathrm{tr}(M_2^{-T}\bar{Y}_2)\right\}\right].$$

Taking the derivative with respect to each α_i, we have

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha_1} &= -n_1 \ln|M_1| - n_1 \frac{\partial \ln \Gamma_p(\alpha_1)}{\partial \alpha_1} + n_1 p \ln \alpha_1 + n_1 p + \sum_{i=1}^{n_1} \ln|Y_i| - n_1 \mathrm{tr}(M_1^{-1}\bar{Y}_1) \\ \frac{\partial \ell}{\partial \alpha_2} &= -n_2 \ln|M_2| - n_2 \frac{\partial \ln \Gamma_p(\alpha_2)}{\partial \alpha_2} + n_2 p \ln \alpha_2 + n_2 p + \sum_{i=1}^{n_2} \ln|Y_i| - n_2 \mathrm{tr}(M_2^{-1}\bar{Y}_2).\end{aligned}$$

The value of α_i should satisfy the condition that the first derivative of likelihood evaluated at $M_i = \hat{M}_i$ is equal to 0. However, since $\hat{M}_i$ can be a function of both α_1 and α_2, depending on the restriction, estimates of α_i cannot be obtained separately. We can get the solution for α_1 and α_2 simultaneously by solving the above system of equations.

As in the one-sample case, no closed-form solution for α_i is available. Instead, the system of equations for α_i presented above can be solved using an approximation method such as a multivariate version of Newton's method. Thus, we estimate $\hat{\alpha}_i$ using this formula and numerical computation.

6. Numerical studies

6.1. Simulation setting

We performed simulation studies for the scenarios presented in Sections 3–4 to evaluate the distributions of the estimators and the test statistics for finite sample sizes in comparison with their theoretical asymptotic distributions, and evaluate their statistical power. The dimensions of the data matrices were chosen as $p = 3$ or $p = 4$, depending on the scenarios.

In the case of the Gamma distribution, the shape parameter α should be greater than p. Since we do simulations with $p = 3$ or $p = 4$ in this section, we set $\alpha = 4$ for $p = 3$ and $\alpha = 5$ for $p = 4$. In addition, since we ought to use M/α for generation of matrix-Gamma data in R, we use $M = \alpha V$ for a $p \times p$ matrix V. We chose the same value of M for the Gaussian distribution, and picked the covariance parameter as $\sigma^2 = 4^2 = 16$ for $p = 3$, and $\sigma^2 = 5^2 = 25$ for $p = 4$.

6.1.1. Null distribution of test statistics

For the one-sample case, we performed the four tests listed below:

1. Test of means (Full Matrix test): $\mathcal{H}_0 : M = U_0 D_0 U_0^\top$ vs. $\mathcal{H}_a : M$ is unrestricted.
2. Test of eigenvectors with fixed eigenvalues: $\mathcal{H}_0 : M = U_0 D_0 U_0^\top$ vs. $\mathcal{H}_a : M \in \mathbb{M}_{D_0}$.
3. Test of eigenvalues with fixed eigenvectors: $\mathcal{H}_0 : M = U_0 D_0 U_0^\top$ vs. $\mathcal{H}_a : M \in \mathbb{M}_{U_0}$.
4. Test of multiplicities: $\mathcal{H}_0 : M \in \mathbb{M}_{m_1,\dots,m_k}$ vs. $\mathcal{H}_a : M$ is unrestricted.

Taking the dimension of the data matrices as $p = 3$, $n = 100$ iid samples were generated from the following distributions:

1. Gaussian with $M = \mathrm{diag}(8, 8, 4)$ and $\mathrm{cov}\{\mathbf{vecd}(Y)\} = 4^2 I_6$.
2. Wishart distribution, with parameters $\alpha = 4$ and $M = \mathrm{diag}(8, 8, 4)$.

With this dataset, we estimated the eigenvalues and eigenvectors of the mean parameter matrix using the formulas derived in Section 3. The likelihood ratio test statistics and p-values were also computed for each hypothesis test. This process was repeated $r = 10{,}000$ times, so 10,000 test statistics and p-values were computed for each scenario.

In the two-sample case, we used the following three tests:

1. Test of means (Full Matrix test): $\mathcal{H}_0 : M_1 = M_2$ vs. $\mathcal{H}_a : M_1 \neq M_2$.
2. Test of equality of two eigenvectors (eigenvector test): $\mathcal{H}_0 : M_1 = M_2$ vs. $\mathcal{H}_a : (M_1, M_2) \in \mathbb{M}_{2,D}$.
3. Test of eigenvalue structure (eigenvalue test) : $\mathcal{H}_0 : (M_1, M_2) \in \mathbb{M}_{2,D}$ vs. $\mathcal{H}_a : (M_1, M_2) \notin \mathbb{M}_{2,D}$.

As in the one-sample case, we took the dimension of the data matrices to be $p = 3$, and $n_1 = n_2 = 100$ iid samples were generated from the following distributions:

1. Gaussian with M = Diag(12,8,4) and cov{$\mathbf{vecd}(Y)$} = $4^2 I_6$.

2. Wishart distribution, with parameters $\alpha = 4$ and $M = \mathrm{diag}(12, 8, 4)$.

With these two groups of data, we estimated the eigenvalues and eigenvectors of the mean parameter matrices using the formulas derived in Section 5. Then, using the estimated eigenvalues and eigenvectors, the likelihood ratio test statistics and p-values were also computed for each hypothesis. This process was repeated r = 10,000 times, so 10,000 test statistics and p-values were computed for each scenario.

For $p = 4$, we performed the same process as in the case $p = 3$ with different parameters. In the one-sample case, $n = 100$ iid samples were generated from the following distributions:

1. Gaussian with $M = \mathrm{diag}(15, 15, 10, 5)$ and cov{$\mathbf{vecd}(Y)$} = $5^2 I_{10}$.

2. Wishart distribution, with parameters $\alpha = 5$ and $M = \mathrm{diag}(15, 15, 10, 5)$.

For the two-sample case, $n_1 = n_2 = 100$ iid samples were generated from the distributions:

1. Gaussian with $M = \mathrm{diag}(20, 15, 10, 5)$ and cov{$\mathbf{vecd}(Y)$} = $5^2 I_{10}$.

2. Wishart distribution, with parameters $\alpha = 5$ and $M = \mathrm{diag}(20, 15, 10, 5)$.

6.1.2. Statistical power of tests

The simulation scenarios were based on [28], with dimension $p = 3$. In the one-sample case, under the null hypothesis, $n = 50$ iid samples were generated from the following distributions:

1. Gaussian with M = Diag(12,8,4) and cov{$\mathbf{vecd}(Y)$} = $4^2 I_6$.

2. Wishart distribution, with parameters $\alpha = 4$ and $M = \mathrm{diag}(12, 8, 4)$.

We considered two sets of alternatives to assess the power of the proposed test statistics. The first set of alternatives was defined in terms of changes in the eigenvalues of M_0. These alternatives were defined as $M_a = M_0 + U_0(\Delta D)U_0^\top$, where ΔD is a diagonal matrix that represents changes in eigenvalues. This reduces to $M_a = M_0 + \Delta D$ in our simulation, since M_0 is diagonal and U_0 is the identity matrix. Four values of ΔD were chosen from easier to harder to detect: (i) $\Delta D = (0.8, 0.8, 0.4)$; (ii) $\Delta D = (0.8, -0.8, 0.4)$, (iii) $\Delta D = (-0.4, -0.4, 0.4)$; (iv) $\Delta D = (0.2, 0.2, 0.2)$. It gets harder to detect as ΔD gets smaller.

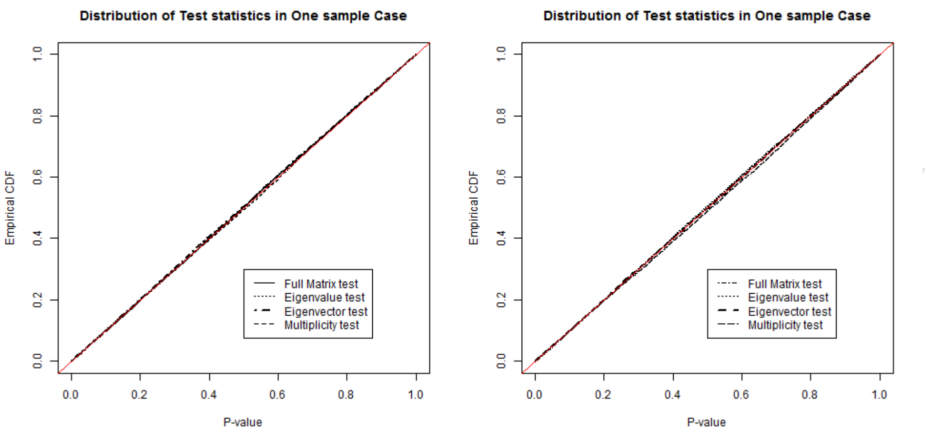
The second set of alternatives was defined by changes in eigenvectors. Using Rodrigues' rotation formula, a rotation matrix $Q \in O_3$ is given as

$$Q = \exp\left[\theta \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}\right],$$

where $\mathbf{a} = (a_1, a_2, a_3)$ is a unit vector representing the axis of rotation and θ is the rotation angle around this axis. With this rotation matrix, the alternative is formulated as $M = QM_0Q^\top$. Four unit vectors were chosen to make the rotation matrix, with a fixed angle $\theta = \pi/12$ for all cases: (i) $\mathbf{a} = 1/\sqrt{3}(-1, -1, -1)$; (ii) $\mathbf{a} = (1, 0, 0)$; (iii) $\mathbf{a} = 1/\sqrt{2}(0, 1, 1)$; (iv) $\mathbf{a} = 1/\sqrt{2}(1, 0, -1)$.

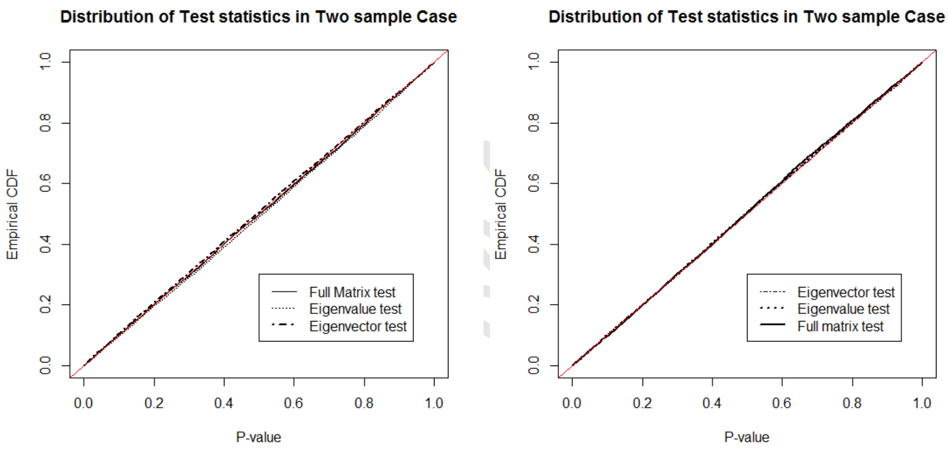
With these datasets, we estimated eigenvalues and eigenvectors of the mean parameter matrix using the formulas derived in Section 3, and the likelihood ratio test statistics were computed for each of the four scenarios (Full Matrix Test, Eigenvalue test, Eigenvector test, and Multiplicity test). With these test statistics, the statistical power was computed for each set of alternatives. This process is repeated r = 10,000 times.

In the two-sample case, one sample group was generated from the same process as the null hypothesis setting in the one-sample case, with the same parameters, while the other group was generated under the alternative hypothesis setting as the one-sample case (eigenvalue change, eigenvector change). The sample size was $n_1 = n_2 = 50$ for each hypothesis set. With these two groups of data, we estimated eigenvalues and eigenvectors of the mean parameter matrix for each group by using the formula derived in the Section 4. These were performed for both Gaussian and Wishart distribution samples. Likelihood ratio test statistics and statistical powe were computed for each of the three scenarios (Full Matrix Test, Eigenvalue test, Eigenvector test). This process was repeated r = 10,000 times.



(a) Gaussian (b) Matrix-Gamma

Figure 1: Distribution of p-values for hypothesis testing scenarios: one-sample case, with $n = 100$.



(a) Gaussian (b) Matrix-Gamma

Figure 2: Distribution of p-values for hypothesis testing scenarios: two-sample case, with $n_1 = n_2 = 100$.

6.2. Simulation results

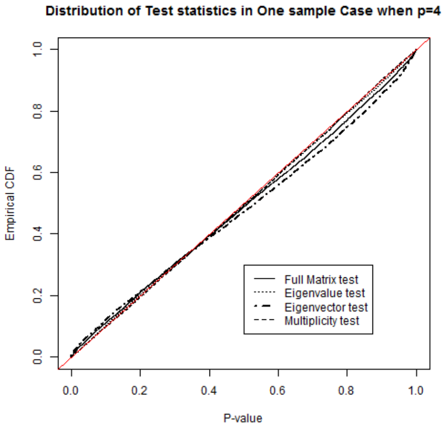
6.2.1. Distribution of p-values

We evaluate the accuracy of the approximate distributions of test statistics by examining the pattern of empirical distribution of plots. Figure 1 shows the distribution of the p-values for test statistics of each hypothesis described above in the one-sample case. The closeness of the empirical cumulative distribution function of p-values to the 45-degree line indicates the goodness of the fit. When we look at the plot, all tests show good performances since these empirical distributions of p-values are near the 45-degree red line.

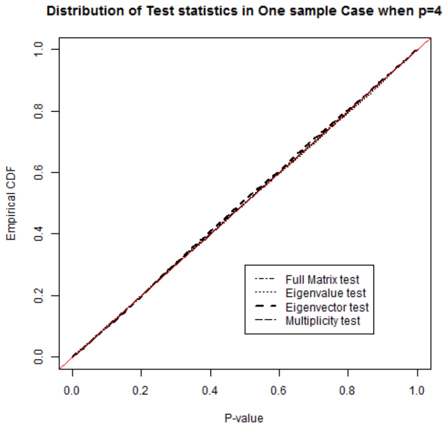
Figure 2 represents the distribution of the p-values for test statistics in the two-sample case. Like the one-sample case, the closeness of the empirical cdf of p-values to the 45-degree line indicates the goodness of the fit. When we examine the plots, we see that all tests show good performances since these empirical distributions of p-values are close to the 45-degree red line. Thus, the derived distributions of test statistics are accurate.

Moving to the simulations for $p = 4$, Figure 3 shows the distribution of the p-values for test statistics of each hypothesis described above in the one-sample case. Similar to the result of $p = 3$, all tests exhibit good performance since these empirical distributions of p-values are near the 45-degree red line.

Figure 4 presents the distribution of the p-values for test statistics in the two-sample case when $p = 4$. Like the one-sample case, all tests show good performances since these empirical distributions of p-values are close to

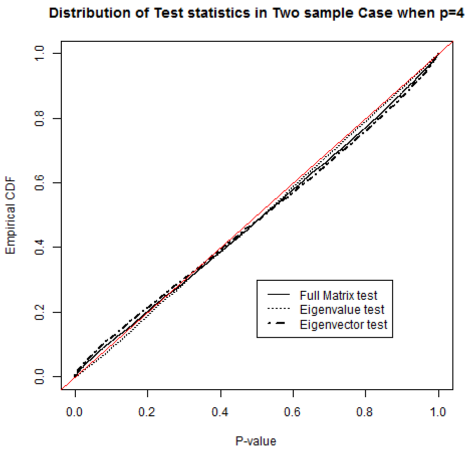


(a) Gaussian

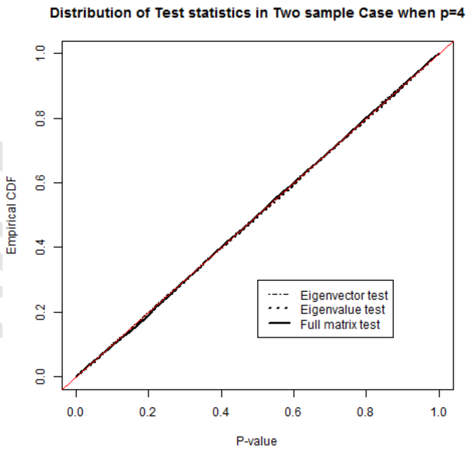


(b) Matrix-Gamma

Figure 3: Distribution of p-values for hypothesis testing scenarios: $p = 4$, one-sample case, with $n = 100$.



(a) Gaussian



(b) Matrix-Gamma

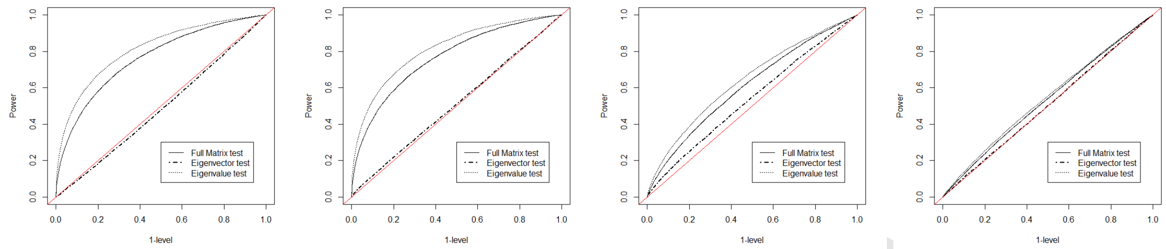
Figure 4: Distribution of p-values for hypothesis testing scenarios: $p = 4$, two-sample case, with $n_1 = n_2 = 100$.

the 45-degree red line. Thus, we can verify that the derived distributions of test statistics are accurate.

6.2.2. Statistical power of tests

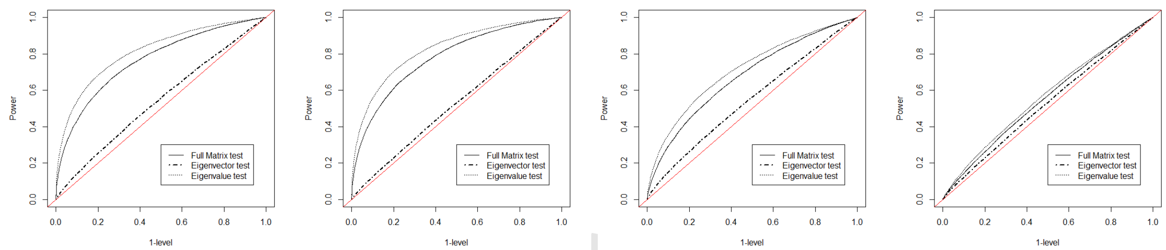
Figures 5 and 6 show the results for various alternatives related to eigenvalue changes in the one-sample case. When the eigenvalues are closer to each other, it is harder to detect the difference. By design, the eigenvalue tests are the most powerful, with the full matrix tests not far behind. The eigenvector tests have no power since they are not designed to capture changes in eigenvalues. Figures 7–8 show the results for various alternatives related to the eigenvector changes in the one-sample case. For all alternatives, eigenvector tests are the most powerful, while the eigenvalue tests have very little power.

Figures 9–10 represent the results for various alternatives related to the eigenvalue changes in the two-sample case. Here, similar to the one-sample case, eigenvalue tests are the most powerful for all the alternatives and the full matrix test is also powerful in both distributions. Still, eigenvector tests have no power since they are not designed to capture changes in the eigenvalues. Figures 11–12 present the results for various alternatives related to the eigenvector changes in the two-sample case. As in the one-sample case, eigenvector tests are the most powerful in both distributions, while the eigenvalue tests have very little power.



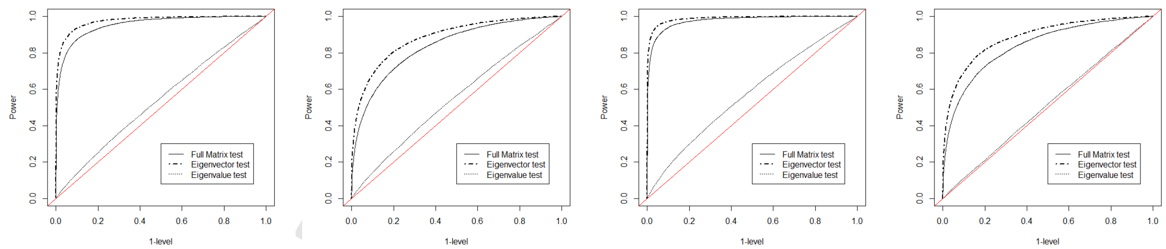
(a) (b) (c) (d)

Figure 5: ROC curves for eigenvalue changes: one-sample case, with $n = 50$, Gaussian sample. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \text{diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \text{diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \text{diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \text{diag}(0.2, 0.2, 0.2)$.



(a) (b) (c) (d)

Figure 6: ROC curves for eigenvalue changes: one-sample case, with $n = 50$, Wishart sample. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \text{diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \text{diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \text{diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \text{diag}(0.2, 0.2, 0.2)$.



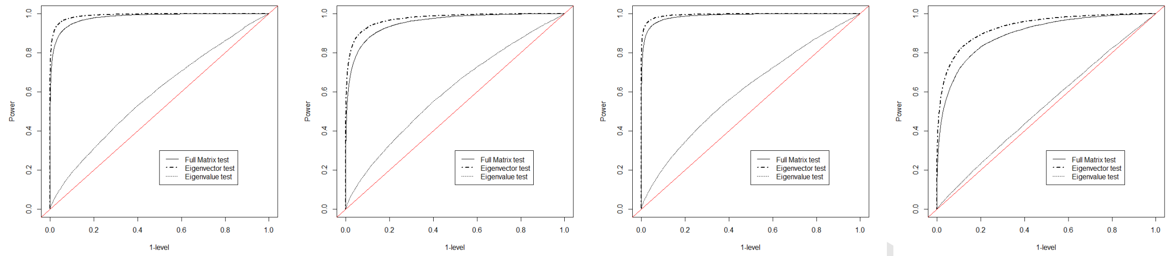
(a) (b) (c) (d)

Figure 7: ROC curves for eigenvector changes: one-sample case, with $n = 50$, Gaussian sample. Four unit vectors for rotation matrices are chosen as (a) $\mathbf{a} = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, (b) $\mathbf{a} = (1, 0, 0)$, (c) $\mathbf{a} = (0, 1/\sqrt{2}, 1/\sqrt{2})$, (d) $\mathbf{a} = (1/\sqrt{2}, 0, -1/\sqrt{2})$. The angle $\theta = \pi/12$ is fixed for all cases.

7. Real data example

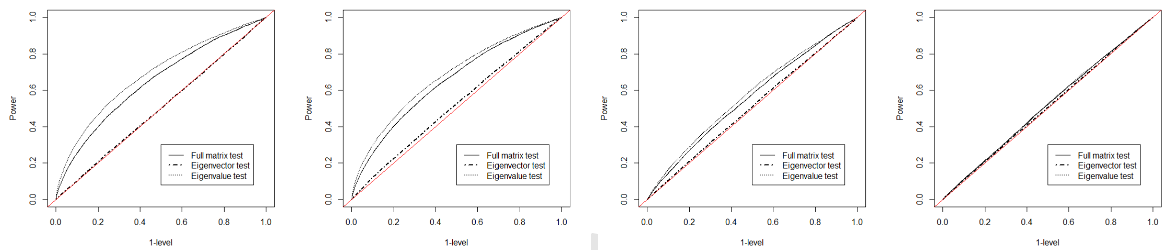
7.1. Data description

We analyze a dataset from an observational study of brain images of children [28]. The dataset consists of 34 brain DTI images of 10-year old children, where 12 images are from boys and 22 images are from girls. The goal of this data analysis is to see whether there exist regions that present anatomic differences between genders.



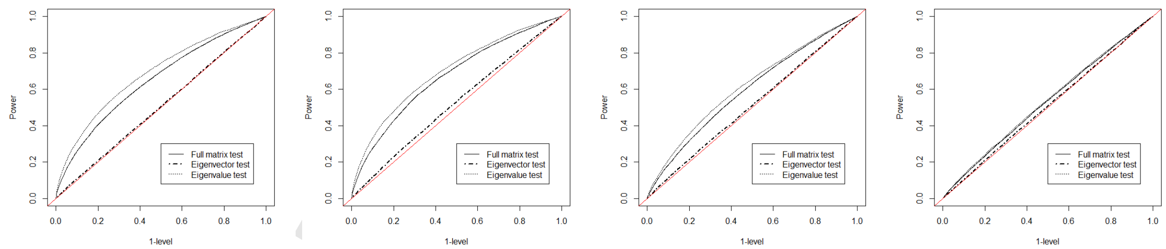
(a) (b) (c) (d)

Figure 8: ROC curves for eigenvector changes: one-sample case, with $n = 50$, Wishart sample. Four unit vectors for rotation matrices are chosen as (a) $\mathbf{a} = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, (b) $\mathbf{a} = (1, 0, 0)$, (c) $\mathbf{a} = (0, 1/\sqrt{2}, 1/\sqrt{2})$, (d) $\mathbf{a} = (1/\sqrt{2}, 0, -1/\sqrt{2})$. The angle $\theta = \pi/12$ is fixed for all cases.



(a) (b) (c) (d)

Figure 9: ROC curves for eigenvalue changes: two-sample Gaussian case, with $n_1 = n_2 = 50$. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \mathrm{diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \mathrm{diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \mathrm{diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \mathrm{diag}(0.2, 0.2, 0.2)$.

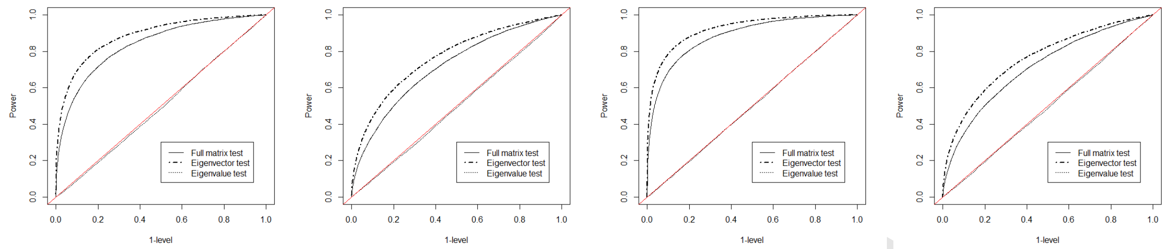


(a) (b) (c) (d)

Figure 10: ROC curves for eigenvalue changes: two-sample Wishart case, with $n_1 = n_2 = 50$. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \mathrm{diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \mathrm{diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \mathrm{diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \mathrm{diag}(0.2, 0.2, 0.2)$.

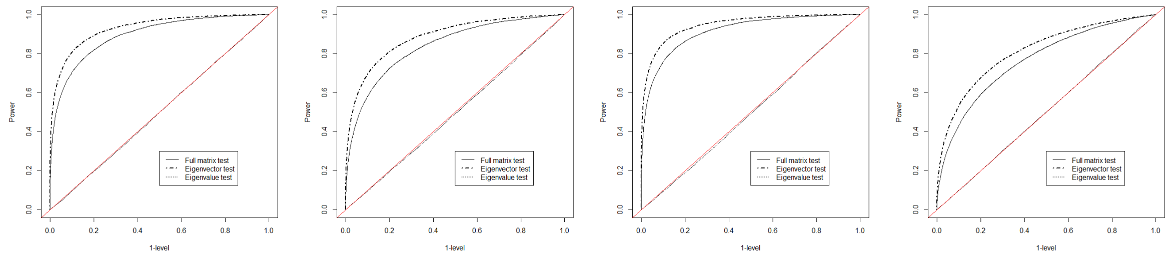
7.2. Analysis

To analyze the data, we assumed three possible distributions for the DT : Gaussian distribution, log-normal distribution, and matrix-Gamma distribution. Our data consist of diffusion tensors, and the DTs in each voxel are all positive definite by definition. So, the positive definite constraint for the data is satisfied in the log-normal and the matrix-Gamma distributions. For the Gaussian distribution, we assume that the probability that matrices are not positive definite is small. At each voxel, the two-sample comparisons are performed, by computing the test statistics for each hypothesis test (eigenvector test, eigenvalue test, full matrix test) under each distribution



(a) (b) (c) (d)

Figure 11: ROC curves for eigenvector changes: two-sample Gaussian case, with $n_1 = n_2 = 50$. Four unit vectors for rotation matrices are chosen as (a) $\mathbf{a} = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, (b) $\mathbf{a} = (1, 0, 0)$, (c) $\mathbf{a} = (0, 1/\sqrt{2}, 1/\sqrt{2})$, (d) $\mathbf{a} = (1/\sqrt{2}, 0, -1/\sqrt{2})$. The angle $\theta = \pi/12$ is fixed for all cases.



(a) (b) (c) (d)

Figure 12: ROC curves for eigenvector changes: two-sample Wishart case, with $n_1 = n_2 = 50$. Four unit vectors for rotation matrices are chosen as (a) $\mathbf{a} = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, (b) $\mathbf{a} = (1, 0, 0)$, (c) $\mathbf{a} = (0, 1/\sqrt{2}, 1/\sqrt{2})$, (d) $\mathbf{a} = (1/\sqrt{2}, 0, -1/\sqrt{2})$. The angle $\theta = \pi/12$ is fixed for all cases.

assumption. With these test statistics, the corresponding p-values are also computed using their approximate null distributions that were described in previous sections. For our data, $n_1 = 12$, $n_2 = 22$, and $p = 3$ at each voxel, and we have $v = 105{,}822$ voxels in total inside the brain.

7.2.1. *Model checking for the Gaussian model*

Figure 13(a) presents the p-value distribution of the LRT and Box's M test, which tests whether the covariance matrices of the two groups are equal. The LRT statistic for this hypothesis test is

$$T = (n_1 + n_2)\ln|S| - n_1 \ln|S_1| - n_2 \ln|S_2|,$$

where S_i is a covariance matrix of each group and $S = (n_1\Sigma_1 + n_2\Sigma_2)/(n_1 + n_2)$ is a pooled covariance matrix [21]. For a case where the sample size is small, Box's M test [8] is also provided; one has $M = \gamma(S_{u1} + S_{u2})$, where

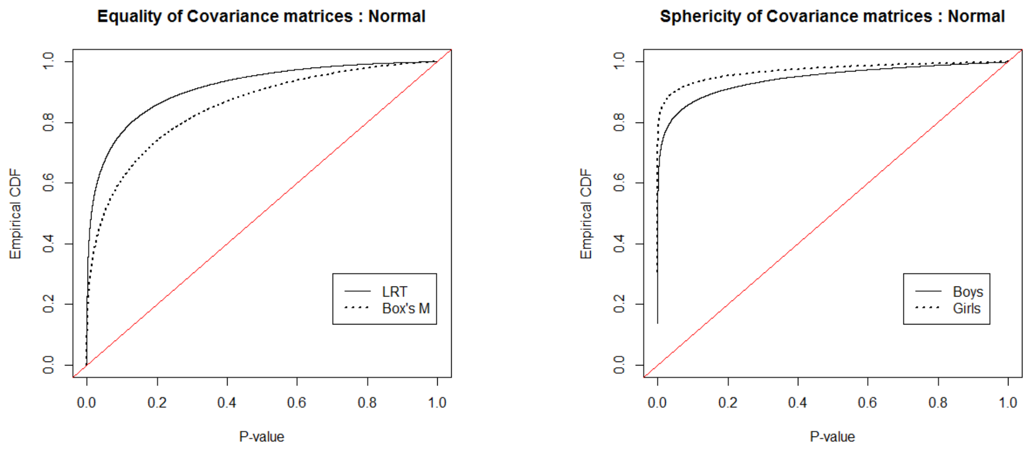
$$\gamma = 1 - \frac{(2p^2 + 3p - 1)}{6(p+1)}\{(n_1 - 1)^{-1} + (n_2 - 1)^{-1} - 2(n_1 + n_2 - 2)^{-1}\}$$

and

$$S_{ui} = (n_i - 1)\ln\left|\left(\frac{n_i}{n_i - 1}S_i\right)^{-1} \frac{(n_1 + n_2)}{(n_1 + n_2 - 2)}S\right|.$$

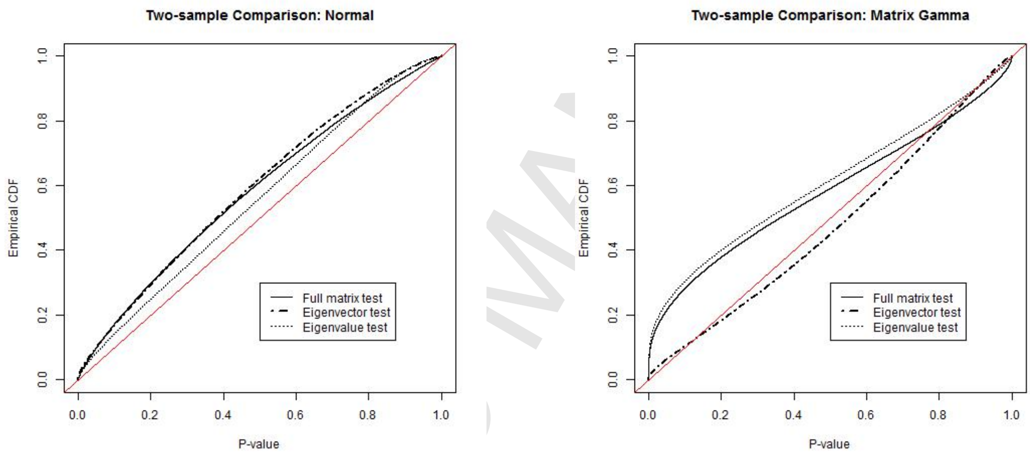
Both the LRT and Box's M test statistics follow the χ^2 distribution with degrees of freedom $q(q+1)/2$ asymptotically. Since the plot shows that two lines of p-value distributions are far from 45-degree lines, we need to make an assumption that the covariance matrices are different in most voxels.

Figure 13(b) shows the test of sphericity of covariance matrices for each voxel, as described in Section 5.1.1. We can see that the distribution of p-values for boys and girls are far from the 45-degree line. Thus, in most voxels, we reject the null hypothesis that the covariance matrices are spherical. Based on this result, we need to make an arbitrary covariance assumption for the model.



(a) Test for the equality of covariance | (b) The test of spherical covariance

Figure 13: Model checking tests for Gaussian model



(a) Gaussian distribution | (b) Wishart distribution assumption

Figure 14: The distribution of p-values:

7.2.2. Two-sample tests of the mean parameter

Under the Gaussian distribution assumption, we used an arbitrary covariance assumption, based on the result of equality and the sphericity test in the previous section. The form of the test statistic and its distribution are described in [28]. We used a sample covariance matrix which, for $i \in \{1, 2\}$, has the form

$$S_i = \widehat{\mathrm{cov}}\{\mathbf{vecd}(Y_i)\} = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \mathbf{vecd}(Y_{ij} - \bar{Y}_i)\{\mathbf{vecd}(Y_{ij} - \bar{Y}_i)\}^\top.$$

Figures 14(a) shows the empirical distribution of the p-values for the three tests. The full matrix test and eigenvector test are not so different from each other, but both are better than the eigenvalue test.

The matrix-Gamma distribution is defined for positive definite matrix data. Thus, we can also assume that the original data follows the matrix-Gamma distribution but with different shape parameters, estimated as described in Section 5.2. Figure 14(b) shows the empirical distribution of the p-values for the three tests with the matrix-Gamma distribution assumption. Three tests are entangled in the high p-value area, but the eigenvector test is the closest to the 45-degree line and the eigenvalue test is better than the other two tests. Therefore, to detect the difference between these two groups, the eigenvalue test seems to be better.

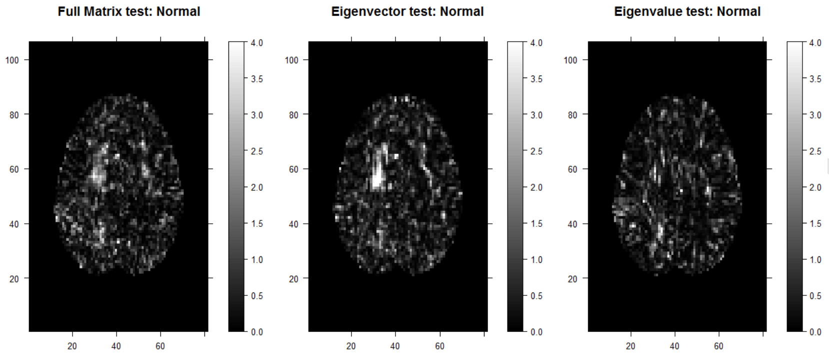


Figure 15: The p-value level plot in scale $-\log_{10}(p)$ at slice $z = 49$: Gaussian distribution

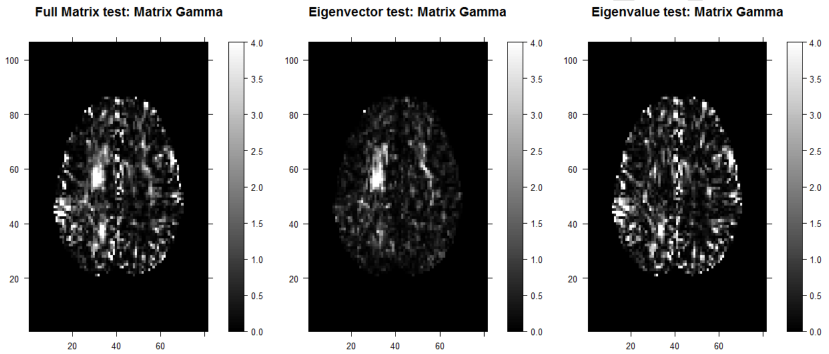


Figure 16: The p-value level plot in scale $-\log_{10}(p)$ at slice $z = 49$: matrix-Gamma distribution

7.2.3. p-value maps

Figure 15 shows maps of the p-values on the $-\log_{10}$ scale under the Gaussian distribution assumption. Some of the most significant differences between boys and girls are presented as white in the plot. We can see that the eigenvector test shows the largest area of differences between the two groups, highlighted in white. Figure 16 shows maps of the p-values in $-\log_{10}$ scale under the matrix-Gamma distribution assumption. We can see that the eigenvalue test and full matrix test show similar size of highlighted area.

While the eigenvalue and full matrix tests in the matrix-Gamma case detect more voxels that are different between the two groups, the eigenvector test in the matrix-Gamma case shows the same anatomical pattern as in the Gaussian and log-normal cases. Therefore, the various modeling assumptions do not give contradictory results in terms of the eigenvector test. Yet, the eigenvalue test in the matrix-Gamma case is sensitive to differences in eigenvalues in the gray matter that are not picked up by the Gaussian model.

8. Summary and discussion

This paper has described inferences for the eigenstructure of the mean matrix of symmetric positive definite matrix-variate data. We have studied an exponential family of positive definite symmetric matrices, including the derivation of MLEs and LLR test statistics for the eigenstructure of the mean parameter M under the exponential family assumption. These procedures were applied to both one-sample and two-sample problems. Through simulations and data analysis, we showed that the exponential family assumption could be appropriate to analyze the matrix-variate data.

We defined an exponential family model for symmetric matrix-variate data to find a general way to make inferences for the mean matrix. This model flexibly included both the spherical Gaussian and matrix-variate Gamma distributions. The matrix-variate normal distribution with orthogonally invariant covariance (OI covariance) as defined in [29] does not conform to this model because it requires another parameter τ in addition to the variance parameter σ^2, which affects the norm of the matrix $Y - M$, and therefore it cannot be written in the exponential family form we defined. Our research could be extended to find a more general form of distribution family that handles this problem.

In the data analysis, we found that the sphericity of covariance matrices assumption was invalid. To handle this in the Gaussian case, we used the approximation in [28] for the distribution of the LRT statistics assuming an arbitrary covariance structure. The problem did not occur in the case of the matrix-Gamma distribution because the covariance matrix depends on the mean matrix, which is arbitrary. However, the covariance cannot be specified separately, as in the Gaussian case.

In the future, research could be extended to generalized linear models (GLMs) for matrix-variate data. In this work, we focused on looking for differences in eigenvalues and eigenvectors between two or more groups under the exponential family assumption. In a GLM formulation it could be assumed that each outcome of the dependent variables Y is generated from a particular distribution in the exponential family with $\mathrm{E}(Y) = M = \psi^{-1}(X, B)$, B is an (unknown) coefficient matrix, X is a suitably defined independent variable matrix, and ψ^{-1} is a link function. Such a GLM could be used to include covariates that can affect the eigenstructure of mean parameter.

Supplementary Material

The Online Supplement contains the proofs for all the theoretical results mentioned in this paper.

References

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, 3rd Ed., Wiley, New York, 2003.
[2] J. Barnard, R. McCulloch, X. Meng, Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage, Statistica Sinica 10 (2000) 1281–1312.
[3] M.S. Bartlett, The statistical conception of mental factors, British J. Psychol. 28 (1937) 97–104.
[4] P.J. Basser, J. Mattiello, D. LeBihan, MR diffusion tensor spectroscopy and imaging, Biophysical J. 66 (1994) 259–267.
[5] P.J. Basser, S.A. Pajevic, Normal distribution for tensor-valued random variables: Applications to diffusion tensor MRI, IEEE Trans Med Imaging 22 (2003) 785–794.
[6] S. Basu, T. Fletcher, S. Whitaker, Rician noise removal in diffusion tensor MRI, in: MICCAI 2006, LNCS, vol. 4190 (2006), Springer, Heidelberg, pp. 117–125.
[7] R.J. Boik, Spectral models for covariance matrices, Biometrika 89 (2002) 583–639.
[8] G.E.P. Box, A general distribution theory for a class of likelihood criteria, Biometrika 36 (1949) 317–346.
[9] G. Casella, R.L. Berger, Statistical Inference, Duxbury, London, 2002.
[10] Z. Chen, D. Dunson, Random effects selection in linear mixed models, Biometrics 59 (2003) 762–769.
[11] J. Fan, Y. Liao, H. Liu, An overview of the estimation of large covariance and precision matrices, Econometrics J. 19 (2016) 1368–1423.
[12] P.T. Fletchera, S. Joshi, Riemannian geometry for the statistical analysis of diffusion tensor data, Signal Processing 87 (2007) 250–262.
[13] A.K. Gupta, D.K. Nagar, Matrix Variate Distribution, Chapman & Hall/CRC, Boca Raton, FL, 1999.
[14] M. Ingalhalikar, A. Smith, D. Parker, T.D. Satterthwaite, M.A. Elliott, K. Ruparel, H. Hakonarson, R.E. Gur, R.C. Gur, R. Verma, Sex differences in the structural connectome of the human brain, PNAS 111 (2013) 823–828.
[15] A. Izenman, Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning, Springer, New York, 2008.
[16] B. Jian, B.C. Vemuri, E. Ozarslan, P.R. Carney, T.H. Mareci, A novel tensor distribution model for the diffusion-weighted MR signal, Neuroimage 37(2007) 164–176.
[17] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, 6th Ed., Pearson Prentice Hall, Upper Saddle River, NJ, 2007.
[18] S. Lang, Fundamentals of Differential Geometry, Springer, New York, 1999.
[19] Y.-Y. Lee, S. Hsieh, Classifying different emotional states by means of EEG-based functional connectivity patterns, PLoS ONE 9(2014) e95415.
[20] C.L. Mallows, Latent vectors of random symmetric matrices, Biometrika 48 (1961) 133–149.
[21] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, San Diego, CA, 1979.
[22] A.M. Mathai, Jacobians of Matrix Transformations and Functions of Matrix Argument, World Scientific Publishing, New York, 1997.
[23] K. Pearson, Method of moments and method of maximum likelihood, Biometrika 28 (1936) 34–59.
[24] B. Scherrer, A. Schwartzman, M. Taquet, M. Sahin, S.P. Prabhu, S.K. Warfield, Characterizing brain tissue by assessment of the distribution of anisotropic microstructural environments in diffusion-compartment imaging (DIAMOND), Magnetic Resonance in Medicine 76 (2016) 963–977.
[25] J.R. Schott, Matrix Analysis for Statistics, Wiley, New York, 2005.
[26] Schurz et al. (2015) Resting-state and task-based functional brain connectivity in developmental dyslexia, Cerebral Cortex 25 (2015) 3502–3514.
[27] A. Schwartzman, Random Ellipsoids and False Discovery Rates: Statistics for Diffusion Tensor Imaging Data, PhD dissertation, Stanford University, Stanford, CA, 2006.
[28] A. Schwartzman, R.F. Dougherty, J.E. Taylor, Group comparison of eigenvalues and eigenvectors of diffusion tensors, J. Amer. Statist. Assoc. 105 (2010) 588–599.
[29] A. Schwartzman, W. Mascarenhas, J.E. Taylor, Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices, Ann. Statist. 36 (2008) 2886–2919.
[30] M. Smith, R. Kohn, Parsimonious covariance matrix estimation for longitudinal data, J. Amer. Statist. Assoc. 97 (2002) 1141–1153.
[31] E. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, Ann. Math. 62 (1955) 548–564.
[32] J. Wishart, The generalised product moment distribution in samples from a normal multivariate population, Biometrika 20A (1928) 32–52.
[33] H. Zhu, H. Zhang, J.G. Ibrahim, B.G. Peterson, Statistical analysis of diffusion tensors in diffusion-weighted magnetic image resonance data (with discussion), J. Amer. Statist. Assoc. 102 (2007) 1085–1102.