Journal of Statistical Planning and Inference 142 (2012) 312–326

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size

Thomas J. Fisher*

Department of Mathematics and Statistics, University of Missouri-Kansas City, Kansas City, MO 64110, USA

ARTICLE INFO

Article history:
Received 19 January 2011
Received in revised form
19 July 2011
Accepted 22 July 2011
Available online 29 July 2011

Keywords:
Covariance matrix
High-dimensional data analysis
Hypothesis testing
Identity matrix

ABSTRACT

This article explores the problem of testing the hypothesis that the covariance matrix is an identity matrix when the dimensionality is equal to the sample size or larger. Two new test statistics are proposed under comparable assumptions to those statistics in the literature. The asymptotic distribution of the proposed test statistics are found and are shown to be consistent in the general asymptotic framework. An extensive simulation study shows the newly proposed tests are comparable to, and in some cases more powerful than, the tests for an identity covariance matrix currently in the literature.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

With the development of new and advanced technologies, many applications of multivariate analysis involve a large number of variables. For example, Microarray DNA analysis may entail the measure of thousands of gene expressions on a small group of individuals. In stock market analysis, a portfolio can include many companies. Advancements in computer technology have made the analysis of these large dimensional data sets feasible; however many of the classical multivariate methods fail when the dimensionality (the number of variables measured) equals or exceeds the sample size. The classical techniques rely on the asymptotic behavior of the sample size. In particular, it is assumed that the dimensionality, p, is fixed, and the sample size, N, exceeds p and grows large.

In this paper, the problem of testing whether a covariance matrix is an identity matrix is considered when the dimensionality is equal to or exceeds the sample size. That is, if $\mathbf{\Sigma}$ denotes the $p \times p$ dimensional *true* covariance matrix, we are interested in the hypothesis test:

$$H_0 : \mathbf{\Sigma} = \mathbf{I} \quad \text{versus} \quad H_A : \mathbf{\Sigma} \neq \mathbf{I}. \tag{1}$$

Classically, a statistic based on the unbiased modified likelihood ratio criterion is used; see Anderson (2003). This test statistic is derived from the geometric mean of the characteristic roots, or eigenvalues, of the maximum likelihood estimator of $\mathbf{\Sigma}$. When the dimensionality is larger than the sample size, it is well known that the maximum likelihood estimator has zero-eigenvalues, and hence the likelihood ratio criterion is degenerate. Furthermore, if $p \approx N$, the maximum likelihood estimator can become ill-conditioned or near-singular, resulting in a degenerate test.

This article tackles the problem of testing the hypothesis in (1) when the dimensionality is at least as large as the sample size by building on the work of Schott (2005, 2006, 2007), Srivastava (2006, 2009), and Ledoit and Wolf (2002). Specifically, a proposal in Srivastava (2005) is generalized and the work of Fisher et al. (2010) has been extended to testing the identity

* Tel.: +1 816 235 2853.
E-mail address: fishertho@umkc.edu

0378-3758/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2011.07.019

hypothesis. Section 2 provides the necessary preliminary results to develop two new statistics for testing the hypothesis in the general asymptotic framework (i.e., $(N,p)\to\infty$). A new estimator for the third arithmetic mean of the eigenvalues of $\boldsymbol{\Sigma}$ is proposed. A theorem showing estimators for the first four arithmetic means of the characteristic roots of $\boldsymbol{\Sigma}$ are asymptotically normally distributed for large (N,p) is provided. Section 3 provides an overview of the statistics available in the literature and introduces two new statistics. Section 4 provides a comprehensive simulation study demonstrating the effectiveness of the new statistics and comparing them with the methods available in the literature. A brief data analysis is provided in Section 5. Concluding remarks are included in Section 6. Appendix A highlights the technical results associated with the new findings. For simplicity, the boldface notation of vectors and matrices is dropped in the remainder of the article.

2. Preliminaries

Let $x_1,x_2,\ldots,x_N$ be independent and identically distributed p-dimensional multivariate normal random variables with unknown mean vector, μ, and covariance matrix Σ. Let $\overline{x}$ and S be the sample mean and sample covariance matrix defined in the typical fashion:

$$\overline{x}=\frac{1}{N}\sum_{i=1}^{N}x_i$$

and

$$S=\frac{1}{n}\sum_{i=1}^{N}(x_i-\overline{x})(x_i-\overline{x})',$$

where $N=n+1$ and A' denotes the transpose operation on the matrix A.

Let $\lambda_1,\lambda_2,\ldots,\lambda_p$ be the characteristic roots of Σ and define

$$a_i=\frac{1}{p}\sum_{j=1}^{p}\lambda_j^i=\frac{1}{p}\,\mathrm{tr}\,\Sigma^i$$

as the ith arithmetic mean of the eigenvalues of Σ.

Many of the test statistics designed for high-dimensional data analysis, both from the literature and introduced in this article, are based on arithmetic means of the eigenvalues of Σ and are designed to work in the general asymptotic framework. It should be noted that unlike the likelihood ratio criterion, a zero eigenvalue will not cause a test based on the a_is to become degenerate.

Under the assumptions that

(1a): As $p\to\infty$, $a_i\to a_i^0$, $0<a_i^0<\infty$ for $i=1,\ldots,4$,
(2a): $n=O(p^{\delta})$ for $0<\delta\le 1$

Srivastava (2005) provides

$$\hat{a}_1=\frac{1}{p}\mathrm{tr}\,S \tag{2}$$

and

$$\hat{a}_2=\frac{n^2}{(n-1)(n+2)p}\left[\mathrm{tr}\,S^2-\frac{1}{n}(\mathrm{tr}\,S)^2\right] \tag{3}$$

as unbiased and (n,p)-consistent estimators for a_1 and a_2, respectively.

Recently, Fisher et al. (2010) provided an estimator for a_4:

$$\hat{a}_4=\frac{\gamma}{p}\left[\mathrm{tr}\,S^4-\frac{4}{n}\mathrm{tr}\,S^3\,\mathrm{tr}\,S-\frac{2n^2+3n-6}{n(n^2+n+2)}(\mathrm{tr}\,S^2)^2+\frac{2(5n+6)}{n(n^2+n+2)}\mathrm{tr}\,S^2(\mathrm{tr}\,S)^2-\frac{5n+6}{n^2(n^2+n+2)}(\mathrm{tr}\,S)^4\right], \tag{4}$$

where

$$\gamma=\frac{n^5(n^2+n+2)}{(n+1)(n+2)(n+4)(n+6)(n-1)(n-2)(n-3)}.$$

Under the assumptions that

(1b): As $p\to\infty$, $a_i\to a_i^0$, $0<a_i^0<\infty$ for $i=1,\ldots,8$,
(2b): As $(n,p)\to\infty$, $p/n\to c$ with $1\le c<\infty$,

where c is known as the concentration, they show $\hat{a}_4$ is unbiased and (n,p)-consistent for a_4.

Remark. Assumption (2b) can be relaxed to include the case of $0<c<1$ and the asymptotic arguments will hold. However, when testing the hypothesis in (1), if $0<c<1$, then $p<n$ and the likelihood ratio criterion should be considered.

Note that for an (n,p)-consistent estimator of a_i, it is required that the $(2i)$th arithmetic mean to converge. The similar assumption

(1c): As $p\to\infty$, $a_i\to a_i^0$, $0<a_i^0<\infty$ for $i=1,\dots,6$

along with (2b) allows the following:

Theorem 1. *An unbiased and* (n,p)*-consistent estimator for* a_3 *is provided by*

$$\hat{a}_3=\frac{\tau}{p}\left(\mathrm{tr}\,S^3-\frac{3}{n}\mathrm{tr}\,S^2\mathrm{tr}\,S+\frac{2}{n^2}(\mathrm{tr}\,S)^3\right), \tag{5}$$

where

$$\tau=\frac{n^4}{(n-1)(n-2)(n+2)(n+4)}.$$

Proof. From Proposition 1 in Appendix A, $\hat{a}_3$ is unbiased for a_3. Use the asymptotic behavior of the variance terms from A.1.2 and Chebyshev's inequality to complete the result:

$$P(|\hat{a}_3-a_3|>\varepsilon)\le\frac{1}{\varepsilon^2}V[\hat{a}_3]\simeq\frac{1}{\varepsilon^2}\left(\frac{18}{np}a_6+\frac{18}{n^2}(a_4a_2+a_3^2)+\frac{6}{n^2}ca_2^3\right)\to 0\quad\text{as }(n,p)\to\infty.\quad\square$$

Unbiased and (n,p)-consistent estimators are available for the first four arithmetic means of the eigenvalues of Σ; however the test statistics are based on their asymptotic distribution. Consider the stronger first assumption:

(1d): As $p\to\infty$, $a_i\to a_i^0$ for $0<a_i^0<\infty$, $i=1,\dots,16$

and assumption (2b). Assumption (1d) requires convergence up to the sixteenth arithmetic mean. This is necessary because each $\hat{a}_i$ requires up to the $(4i)$th arithmetic mean for asymptotic normality. Assumption (2b) is required for an asymptotic finite variance for the estimators.

Theorem 2. *Under assumptions* (1d) *and* (2b), *the estimators* $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$ *and* $\hat{a}_4$ *are asymptotically distributed multivariate normal, i.e.,*

$$\sqrt{np}\begin{pmatrix}\hat{a}_1-a_1\\ \hat{a}_2-a_2\\ \hat{a}_3-a_3\\ \hat{a}_4-a_4\end{pmatrix}\xrightarrow{D}N_4\left(\begin{pmatrix}0\\0\\0\\0\end{pmatrix},\begin{pmatrix}\sigma_{11}&\sigma_{12}&\sigma_{13}&\sigma_{14}\\ \sigma_{12}&\sigma_{22}&\sigma_{23}&\sigma_{24}\\ \sigma_{13}&\sigma_{23}&\sigma_{33}&\sigma_{34}\\ \sigma_{14}&\sigma_{24}&\sigma_{34}&\sigma_{44}\end{pmatrix}\right),$$

where

$$\sigma_{11}=2a_2,\quad\sigma_{12}=4a_3,\ \sigma_{13}=6a_4,\ \sigma_{14}=8a_5,$$

$$\sigma_{22}=8a_4+4ca_2^2,\ \sigma_{23}=12a_5+12ca_3a_2,\ \sigma_{24}=16a_6+16ca_4a_2+8ca_3^2,$$

$$\sigma_{33}=18a_6+18ca_4a_2+18ca_3^2+6c^2a_2^3,\ \sigma_{34}=24a_7+24ca_5a_2+48ca_4a_3+24c^2a_3a_2^2,$$

$$\sigma_{44}=32a_8+48ca_4^2+32ca_6a_2+64ca_5a_3+32c^2a_4a_2^2+64c^2a_3^2a_2+8c^3a_2^4.$$

Proof. The proof follows from Proposition 2 in appendix by performing an elementary linear combination. $\square$

3. Testing for an identity covariance matrix

The aim of this article is to analyze the methods from the literature and introduce new statistics for testing hypothesis (1). As aforementioned, in the classical scenario, the modified likelihood ratio statistic is typically used as it is unbiased and has a monotone power function. However, when the dimensionality exceeds the sample size, it becomes degenerate since it is based on the geometric mean of the characteristic roots of S. Srivastava (2006) proposes a statistic based on the modified likelihood ratio where the p has been replaced by n and the n by p. Specifically he suggest using the first n non-zero eigenvalues of the matrix nS. Let $l_1,l_2,\dots,l_n$, be those eigenvalues and define

$$L=\left(\frac{e}{p}\right)^{pn/2}\left(\prod_{i=1}^{n}l_i\right)^{p/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}l_i\right)$$

and

$$m = 1 - \frac{2n^2+3n+1}{6p(n+1)}.$$

He shows

$$Q_2 = -2m \log L \tag{6}$$

is a statistic for the hypothesis in (1) and its asymptotic null distribution is chi-squared with $\frac{1}{2}p(p+1)$ degrees of freedom. Much like the classical likelihood ratio test the distribution can be further expanded; see Srivastava (2006) for details. His simulations indicate that under H_0 the statistic Q_2 follows the theoretical asymptotic distribution quite well when n and p are not close to one-another.

Nagao (1973) suggested a criterion based on the first and second arithmetic means of the eigenvalues of S. Ledoit and Wolf (2002) show the statistic suggested in Nagao (1973) has poor properties when $p > n$ and in the general asymptotic setting. They introduce the modified statistic

$$W = \frac{1}{p}\mathrm{tr}[(S-I)^2] - \frac{p}{n}\left[\frac{1}{p}\mathrm{tr}\, S\right]^2 + \frac{p}{n}$$

and show that under H_0,

$$T_W = \frac{nW-p-1}{2} \xrightarrow{D} N(0,1) \tag{7}$$

making similar assumptions as (1a) and (2b).

Srivastava (2005) proposes a similar test using the first and second arithmetic means. The derivation follows the work of Nagao (1973), but unlike Ledoit and Wolf (2002), Srivastava (2005) proposes a different test statistic using unbiased and (n,p)-consistent estimators for the components of the equation. To summarize, note that under H_0, each eigenvalue of Σ takes on the value one. Furthermore,

$$\frac{1}{p}\sum_{i=1}^{p}(\lambda_i-1)^2 \geq 0 \tag{8}$$

and equal if and only if each $\lambda_i = 1$. Recalling the definition of a_i and a simple algebraic expansion results in

$$a_2 - 2a_1 + 1 \geq 0.$$

Utilizing the unbiased and (n,p)-consistent estimators in (2) and (3) he shows the statistic

$$\frac{n}{2}(\hat{a}_2 - 2\hat{a}_1 - a_2 + 2a_1) \xrightarrow{D} N\left(0, \frac{2}{c}(a_2 - 2a_3 + a_4) + a_2^2\right)$$

in general, and

$$T_S = \frac{n}{2}(\hat{a}_2 - 2\hat{a} + 1) \xrightarrow{D} N(0,1) \tag{9}$$

under the null hypothesis that $\Sigma = I$.

Recently, Fisher et al. (2010) introduced a statistic for testing the sphericity hypothesis (i.e., $\Sigma = \sigma^2 I$) using higher powers of the arithmetic means. Their simulations indicate that you can gain power when the covariance matrix is nearly spherical. That is, if the covariance matrix had just a few deviations from satisfying the null hypothesis, you can gain power by utilizing a statistic using higher arithmetic means. That work is the motivating factor for two new statistics based on the generalized form of the inequality in (8):

$$\frac{1}{p}\sum_{i=1}^{p}(\lambda_i^r - 1)^{2s} \geq 0. \tag{10}$$

The statistic T_S in (9) from Srivastava (2005) is the case of $r=1$ and $s=1$. The cases of $r=1$, $s=2$ and $r=2$, $s=1$ are explored here.

When $r=1$, $s=2$ the inequality in (10) becomes

$$a_4 - 4a_3 + 6a_2 - 4a_1 + 1 \geq 0.$$

When $r=2$, $s=1$, the inequality in (10) becomes

$$a_4 - 2a_2 + 1 \geq 0.$$

Theorem 2 indicates that for large (n,p), the estimators $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$ and $\hat{a}_4$ are approximately jointly normal. A simple linear transformation on the vector in Theorem 2 and a standardization argument suggest the statistics:

$$T_1 = \frac{n}{c\sqrt{8}}(\hat{a}_4 - 4\hat{a}_3 + 6\hat{a}_2 - 4\hat{a}_1 + 1) \tag{11}$$

and

$$T_2 = \frac{n}{\sqrt{8(c^2+12c+8)}}(\hat{a}_4 - 2\hat{a}_2 + 1) \tag{12}$$

can be utilized in testing the identity hypothesis.

The following theorems and corollaries provide the distribution for T_1 and T_2.

Theorem 3. *Under assumptions* (1d) *and* (2b), *as* $(n,p)\to\infty$,

$$\frac{n}{c\sqrt{8}}(\hat{a}_4 - 4\hat{a}_3 + 6\hat{a}_2 - 4\hat{a}_1 - a_4 + 4a_3 - 6a_2 + 4a_1) \xrightarrow{D} N(0,\tau_1),$$

where

$$\tau_1 = \frac{1}{c^2}\left(60a_4a_2 + 18a_2^2 + 48a_3^2 + 6a_4^2 + 4a_6a_2 + 8a_5a_3 - 72a_3a_2 - 24a_5a_2 - 48a_4a_3 + 12ca_2^3 + 4ca_4a_2^2 + 8ca_3^2a_2 \right.$$
$$\left. -24ca_3a_2^2 + \frac{4}{c}a_8 + \frac{4}{c}a_2 + \frac{60}{c}a_4 + \frac{60}{c}a_6 - \frac{24}{c}a_7 - \frac{80}{c}a_5 - \frac{24}{c}a_3 + c^2a_2^4\right).$$

Corollary 1. *Under* $H_0 : \Sigma = I$ *and assumptions* (1d) *and* (2b), *as* $(n,p)\to\infty$,

$$T_1 = \frac{n}{c\sqrt{8}}(\hat{a}_4 - 4\hat{a}_3 + 6\hat{a}_2 - 4\hat{a}_1 + 1) \xrightarrow{D} N(0,1).$$

Proof. Under $H_0 : \Sigma = I$, each $a_i = 1$ for all i, hence $\tau_1 = 1$. □

Theorem 4. *Under assumptions* (1d) *and* (2b), *as* $(n,p)\to\infty$,

$$\frac{n}{\sqrt{8(c^2+12c+8)}}(\hat{a}_4 - 2\hat{a}_2 - a_4 + 2a_2) \xrightarrow{D} N(0,\tau_2),$$

where

$$\tau_2 = \frac{1}{c^2+12c+8}\left(\frac{4}{c}a_4 + \frac{4}{c}a_8 - \frac{8}{c}a_6 + 2a_2^2 + 6a_4^2 + 4a_6a_2 + 8a_5a_2 - 8a_4a_2 - 4a_3^2 + 4ca_4a_2^2 + 8ca_3^2a_2 + c^2a_2^4\right).$$

Corollary 2. *Under* $H_0 : \Sigma = I$ *and assumptions* (1d) *and* (2b), *as* $(n,p)\to\infty$,

$$T_2 = \frac{n}{\sqrt{8(c^2+12c+8)}}(\hat{a}_4 - 2\hat{a}_2 + 1) \xrightarrow{D} N(0,1).$$

Proof. Under H_0, each $a_i = 1$, hence $\tau_2 = 1$. □

Theorem 5. *Under assumptions* (1d) *and* (2b), *as* $(n,p)\to\infty$ *the test statistics* T_1 *in* (11) *and* T_2 *in* (12) *are* (n,p)*-consistent.*

Proof. The proof is analogous to Theorem 3 in Fisher et al. (2010). □

4. Simulation study

The purpose of the simulation study is to (i) show the effectiveness of the newly defined test statistics T_1 and T_2 and (ii) to perform a comparative study on the available statistics in testing the hypothesis that the covariance matrix is the identity.

Remark. The simulation study conducted in this article involves computation on large matrices, i.e., the fourth power of S is needed for $\hat{a}_4$ and S is potentially as large as a 2560×2560 matrix. As such, much of the study was performed in a parallelized framework utilizing the `multicore` package in the GNU licensed `R-Project`. All simulations included were run on a Dell Optiplex 980 with 16 GB of RAM running Ubuntu 10.10 Linux with kernel 2.6.35–23.

4.1. Normality and size study

To demonstrate the effectiveness of the new test statistics begin by confirming the results of Corollaries 1 and 2 which state that for large (n,p), the statistics T_1 and T_2 will behave like standard normal random variables under the null hypothesis. To validate the results, consider the simple study of exploring QQ-Plots. The true covariance matrix Σ is set to be a 512×512 identity matrix. A sample of size $N=257$ (recall, $N=n+1$, hence $n=256$) is simulated from a multivariate normal distribution with a mean zero vector and $\Sigma = I$. This sample is used to calculate T_1 and T_2; the process is repeated 1000 times. Fig. 1 shows that the newly introduced statistics behave like standard normal random variables when the sample size and dimensionality are large. The results from Fig. 1 indicate that for sufficiently large n and p, Corollaries 1 and 2 hold. Table 1 provides the simulated size, or attained significance level, of the two proposed test statistics for

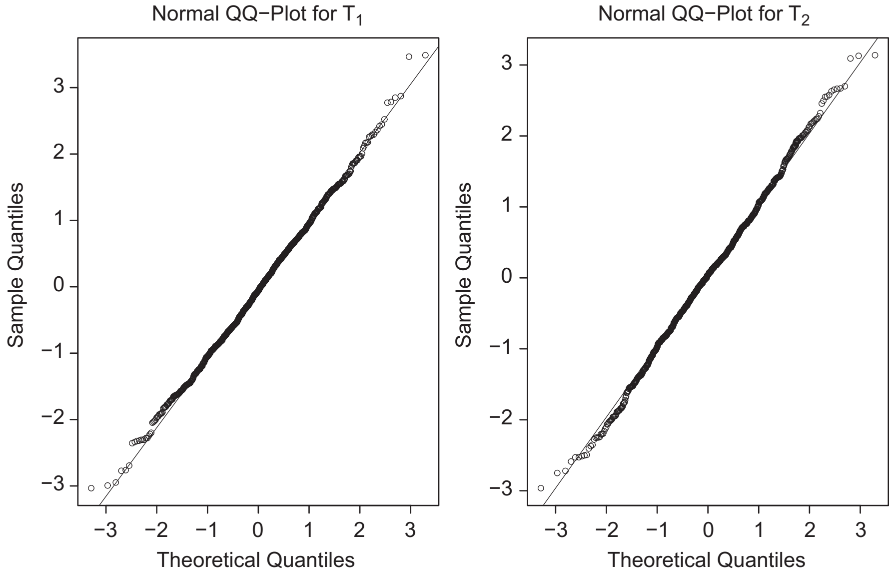

Fig. 1. Normal QQ-plots for T_1 and T_2, $n=256$, $p=512$ under H_0.

Table 1
Empirical size at $\alpha=0.05$ and $p=cn$ for T_1 and T_2.

n	T_1 in (11)					T_2 in (12)					n
	$c=1$	$c=2$	$c=3$	$c=4$	$c=5$	$c=1$	$c=2$	$c=3$	$c=4$	$c=5$	
16	0.039	0.044	0.039	0.041	0.042	0.052	0.057	0.046	**0.071**	0.051	16
32	0.049	0.059	**0.036**	0.040	0.057	0.061	0.057	**0.069**	0.059	**0.065**	32
64	0.054	0.050	0.053	0.044	0.054	0.057	0.055	0.052	0.052	**0.068**	64
128	0.061	0.060	0.049	0.050	0.057	0.058	0.060	0.050	0.061	0.058	128
256	0.061	0.054	0.056	0.047	0.044	0.062	0.051	0.051	0.055	0.052	256
512	0.049	0.045	0.054	0.060	0.060	0.049	0.061	0.052	0.061	0.059	512

a variety of n and p combinations when sampled under the null hypothesis $\Sigma=I$. For any particular combination of n and p, 1000 instances of T_1 and T_2 are calculated from corresponding samples under H_0. The simulated size is found by calculating $(\#T_i>z_\alpha)/1000$ for $i=1,2$ and where z_α is the standard normal critical value at $\alpha=0.05$. Table 1 shows that overall the newly proposed statistics appear to satisfy the normality results. Only in a few cases (see boldface in Table 1) does the simulated size significantly differ from the theoretical size of $\alpha=0.05$, and these typically occur when n is relatively small compared to p. As (n,p)-increase the simulated size appears to become more accurate.

4.2. Power study

The second goal of the simulation study is to compare the newly suggested statistics and those from the literature by looking at their power functions. The statistic Q_2 (6) from Srivastava (2006) is designed for the case when n is fixed and p increases. His simulations show that Q_2 is most effective when n and p are very different. This article explores the general asymptotic framework, when $(n,p)\to\infty$, so it is not appropriate to compare Q_2 with the other available methods. The simulated power of the statistics, T_1, T_2, T_S and T_W from (11), (12), (9) and (7), respectively are compared. For a given n and p, a sample is taken under the null hypothesis, $\Sigma=I$, and the statistics are computed. This process is repeated 1000 times. The empirical critical value, at $\alpha=0.05$, is found for each statistic. Then a sample is taken under the alternative hypothesis, the test statistics are found, and compared to their empirical critical value. The proportion of times the statistic exceeds the empirical critical value is the simulated power. The setup of this study is analogous to that found in Srivastava (2006) and Fisher et al. (2010). However, the simulation experiments in those articles is fairly limited in terms of large matrices. Both articles only explore matrices as large as 400×400. In this study, the power study is performed on an assortment of concentrations; as large as $c=5$, with a sample size of $n=512$ resulting in a 2560×2560 covariance matrix. For each covariance matrix under the alternative hypothesis, we would expect the newly introduced statistics to demonstrate a decrease in power as the concentration increases since it influences the convergence rate. This phenomenon can be seen in many of the results in this article as typically two concentrations are explored for each covariance matrix studied. The statistic with the largest power for each n has been put in boldface to assist the reader.

4.2.1. Nearly identity covariance matrices

Fisher et al. (2010) demonstrated in their simulations that by utilizing higher powers of arithmetic means of the eigenvalues of the empirical covariance matrix you can gain power when only a few elements of the covariance matrix deviate from the null hypothesis. The first series of simulations study the analogous case where the true covariance is of the form:

$$\Sigma = \begin{pmatrix} \Theta & 0^T \\ 0 & I \end{pmatrix}, \tag{13}$$

where Θ is a $k \times k$ diagonal matrix, $k<p$, with all elements $\theta_i \neq 1$, $i=1,\ldots,k$. I is a $(p-k)\times(p-k)$ identity matrix and 0 is a $(p-k)\times(p-k)$ matrix of zeros. k is chosen to be small, so the covariance matrix will be nearly the identity with the exception of a few elements.

The first example is in Table 2 with Θ defined to be a diagonal matrix of size $c \times c$ with diagonal elements taking on the value 3.5; concentrations $c=3$ and 4 are explored.

We see that the newly suggested statistic T_2 appears to perform the best. Although the power function for each, T_1, T_S and T_W, appear to be increasing, all seem to have a slower convergence rate compared to T_2. Also note the overall appearance of a slight decrease in power between the two concentrations for the newly proposed statistics. Each row of Table 2 corresponds to the same proportion of non-unity diagonal elements and there is a decrease in power for the newly introduced statistics. The statistics T_S and T_W do not seem affected by the concentration.

Table 3 provides a study when $\Theta = \text{diag}(4,3,2)$ with concentrations $c=2$ and 3. The results are similar to that of Table 2. Again, it appears the statistic T_2 outperforms the others.

Based on the variance terms (see τ_1 in Theorem 3 and τ_2 in Theorem 4) of the two newly proposed statistics, it appears T_2 will be better since its variance is smaller. This was seen in the previous two examples as T_1 outperformed T_S and T_W asymptotically, but appeared to be dominated by T_2. In general, the power studies conducted support this claim, however T_1 can be comparable to T_2 in terms of power. Consider the case where $\Theta = \text{diag}(4.404)$ or diag(4.40425), both with a concentration $c=1$. Table 4 provides the results demonstrating that the statistic T_1 appears to be equivalent to, or better than, T_2 for larger n. Both T_1 and T_2 outperform the other statistics available.

4.2.2. Other covariance matrices

When the true covariance greatly differs from the identity (e.g., $\Sigma = \text{diag}(Unif(1,10))$, $\Sigma = \text{diag}(5,\ldots,5)$, etc.), each statistic performs very well. These simulations are omitted from this article as they are not very interesting.

However, when the true covariance is similar to the identity, we see a difference in performance between the statistics. Consider the sphericity model $\Sigma = \sigma^2 I$ with σ^2 relatively close to the value 1. In Table 5 the results for $\sigma^2 = 1.5$ with concentrations $c=3$ and 5 are provided. Table 6 provides the results for $\sigma^2 = 0.5$ with $c=1$ and 3. In each case, the newly suggested statistics are dominated by the statistics in Srivastava (2005) or Ledoit and Wolf (2002). In fact, when n is fairly small, the performance of the new statistics are quite poor. However, even with their poor performance for small n, the simulations show the statistics are consistent as n and p increase.

Table 2
Empirical power at $\alpha=0.05$ for T_1, T_2, T_S, T_W, $\Theta = \text{diag}_c(3.5,\ldots,3.5)$.

n	Dimensionality $p=3n$				Dimensionality $p=4n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	0.471	0.655	0.732	**0.746**	0.417	0.590	0.698	**0.712**	16
32	0.625	**0.796**	0.762	0.772	0.575	**0.774**	0.799	0.813	32
64	0.744	**0.908**	0.838	0.846	0.656	**0.866**	0.847	0.851	64
128	0.873	**0.964**	0.893	0.896	0.772	**0.963**	0.877	0.880	128
256	0.945	**0.993**	0.914	0.914	0.859	**0.986**	0.936	0.938	256
512	0.981	**1.000**	0.938	0.940	0.919	**0.998**	0.928	0.928	512

Table 3
Empirical power at $\alpha=0.05$ for T_1, T_2, T_S, T_W, $\Theta = \text{diag}(4,3,2)$.

n	Dimensionality $p=2n$				Dimensionality $p=3n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	0.499	0.653	0.695	**0.712**	0.399	0.529	0.567	**0.583**	16
32	0.708	**0.838**	0.785	0.801	0.517	0.632	0.639	**0.657**	32
64	0.839	**0.925**	0.861	0.865	0.632	**0.773**	0.648	0.661	64
128	0.956	**0.983**	0.926	0.929	0.774	**0.912**	0.707	0.711	128
256	0.993	**0.999**	0.941	0.941	0.830	**0.961**	0.676	0.677	256
512	0.999	**1.000**	0.948	0.948	0.918	**0.978**	0.764	0.764	512

Table 4
Empirical power at $\alpha=0.05$ for T_1, T_2, T_S, T_W, $c=1$.

n	$\Theta=\text{diag}(4.404)$				$\Theta=\text{diag}(4.40425)$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	0.651	**0.713**	0.677	0.696	0.629	**0.744**	0.702	0.707	16
32	0.808	**0.877**	0.797	0.800	0.824	**0.866**	0.807	0.814	32
64	0.964	**0.974**	0.909	0.911	0.968	**0.979**	0.922	0.925	64
128	**0.996**	**0.996**	0.959	0.959	**0.997**	0.996	0.955	0.956	128
256	**1.000**	**1.000**	0.984	0.984	**1.000**	**1.000**	0.982	0.982	256
512	**1.000**	**1.000**	0.990	0.990	**1.000**	**1.000**	0.991	0.992	512

Table 5
Empirical power at $\alpha=0.05$ for T_1, T_2, T_S, T_W, $\Sigma=\text{diag}(1.5,\ldots,1.5)$.

n	Dimensionality $p=3n$				Dimensionality $p=5n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	0.297	0.341	0.554	**0.624**	0.300	0.300	0.522	**0.582**	16
32	0.332	0.473	0.863	**0.903**	0.346	0.441	0.851	**0.901**	32
64	0.395	0.719	**0.999**	**0.999**	0.369	0.631	0.997	**0.998**	64
128	0.404	0.938	**1.000**	**1.000**	0.436	0.875	**1.000**	**1.000**	128
256	0.503	**1.000**	**1.000**	**1.000**	0.445	0.998	**1.000**	**1.000**	256
512	0.610	**1.000**	**1.000**	**1.000**	0.528	**1.000**	**1.000**	**1.000**	512

Table 6
Empirical power at $\alpha=0.05$ for T_1, T_2, T_S, T_W, $\Sigma=\text{diag}(0.5,\ldots,0.5)$.

n	Dimensionality $p=n$				Dimensionality $p=3n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	0.001	0.000	**0.732**	0.347	0.000	0.000	**0.837**	0.369	16
32	0.013	0.003	**1.000**	**1.000**	0.000	0.000	**1.000**	**1.000**	32
64	0.402	**1.000**	**1.000**	**1.000**	0.000	0.102	**1.000**	**1.000**	64
128	**1.000**	**1.000**	**1.000**	**1.000**	0.001	**1.000**	**1.000**	**1.000**	128
256	**1.000**	**1.000**	**1.000**	**1.000**	0.954	**1.000**	**1.000**	**1.000**	256
512	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	512

Table 7
Empirical power at $\alpha=0.05$ for T_1, T_2, T_S, T_W, $\Sigma=\text{diag}(\text{Unif}(0.5,1.5))$.

n	Dimensionality $p=2n$				Dimensionality $p=2.5n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	0.113	0.155	0.196	**0.204**	0.136	0.164	0.236	**0.257**	16
32	0.022	0.053	**0.232**	0.213	0.095	0.135	**0.320**	0.310	32
64	0.089	0.304	**0.806**	0.797	0.108	0.308	**0.814**	0.806	64
128	0.161	0.749	**1.000**	**1.000**	0.162	0.742	**1.000**	**1.000**	128
256	0.214	0.991	**1.000**	**1.000**	0.178	0.980	**1.000**	**1.000**	256
512	0.365	**1.000**	**1.000**	**1.000**	0.309	**1.000**	**1.000**	**1.000**	512

Another example where the covariance matrix is similar to the identity in structure is when it is diagonal with elements distributed uniformly over the interval (0.5, 1.5). Table 7 shows the results for concentrations $c=2$ and 2.5. The statistics from Srivastava (2005) and Ledoit and Wolf (2002) appear to outperform the newly proposed statistics. In particular, the statistic T_1 appears to perform very poorly under this alternative. Its power function has a very slow convergence rate.

Lastly, consider a matrix similar to the sphericity model. Let Σ be a diagonal matrix with half of the elements taking on value 0.5, and the other half taking on the value 1.5. The idea is that the average eigenvalues will still sum to one but the covariance matrix is clearly not an identity matrix. Table 8 provides the results with concentrations $c=1.5$ and 2. Each statistic appears to be consistent with T_S performing the best overall.

Table 8
Empirical power at $\alpha=0.05$ for T_1, T_2, T_S, T_W, $\Sigma=\text{diag}(0.5,\ldots,0.5,1.5,\ldots,1.5)$.

n	Dimensionality $p=1.5n$				Dimensionality $p=2n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	0.118	0.210	0.538	**0.536**	0.137	0.252	**0.530**	0.517	16
32	0.199	0.468	**0.942**	0.941	0.191	0.419	**0.972**	**0.972**	32
64	0.306	0.891	**1.000**	**1.000**	0.279	0.826	**1.000**	**1.000**	64
128	0.487	**1.000**	**1.000**	**1.000**	0.440	**1.000**	**1.000**	**1.000**	128
256	0.807	**1.000**	**1.000**	**1.000**	0.686	**1.000**	**1.000**	**1.000**	256
512	0.998	**1.000**	**1.000**	**1.000**	0.962	**1.000**	**1.000**	**1.000**	512

Table 9
Empirical size under uniform distribution on interval $(-\sqrt{3},\sqrt{3})$.

n	Dimensionality $p=2n$				Dimensionality $p=4n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	**0.017**	**0.029**	**0.016**	**0.018**	**0.022**	**0.028**	**0.016**	**0.016**	16
32	**0.032**	**0.023**	**0.019**	**0.019**	**0.033**	**0.034**	**0.010**	**0.012**	32
64	0.047	**0.036**	**0.019**	**0.019**	**0.066**	0.053	**0.012**	**0.013**	64
128	0.048	**0.027**	**0.013**	**0.013**	0.042	**0.024**	**0.018**	**0.018**	128
256	0.049	**0.024**	**0.011**	**0.012**	0.043	**0.022**	**0.006**	**0.006**	256
512	0.054	**0.032**	**0.008**	**0.008**	**0.070**	0.044	**0.007**	**0.007**	512

Table 10
Empirical power at $\alpha=0.05$, $\Theta=\text{diag}_c(3.5,\ldots,3.5)$, data uniformly distributed.

n	Dimensionality $p=2n$				Dimensionality $p=4n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	0.454	0.704	0.742	**0.763**	0.357	0.582	0.699	**0.730**	16
32	0.674	**0.871**	0.789	0.809	0.507	0.794	0.809	**0.833**	32
64	0.864	**0.960**	0.864	0.871	0.659	**0.906**	0.869	0.875	64
128	0.995	**0.991**	0.914	0.916	0.800	**0.971**	0.900	0.900	128
256	0.990	**0.999**	0.921	0.922	0.867	**0.996**	0.901	0.901	256
512	**1.000**	**1.000**	0.931	0.932	0.901	**0.996**	0.942	0.943	512

4.3. Robustness

In many modern applications it is imperative to statistical practitioners that their test statistics be fairly robust; i.e., if we have some violations from our underlying assumptions, the method still works reasonably well and will have use in practical applications. In this section we study the robustness of the statistics T_1, T_2, T_S and T_W. In particular, we look at the behavior of the statistics when the normality assumption does not hold. The behavior of the statistics is explored when the data comes from a light tailed distribution, a heavy tailed distribution and a skewed distribution. The study is analogous to the empirical size study in Section 4.1 except the normality assumption is relaxed. Any empirical sizes that significantly differ from the predicted size of $\alpha=0.05$ are in boldface. Two concentrations, $c=2$ and 4, are studied for each statistic.

In the first example in Table 9, consider a random sample from a multivariate uniform distribution under the null hypothesis; i.e., each r.v. $x_{ij}\sim Unif(-\sqrt{3},\sqrt{3})$ where x_{ij} is the jth measurement of the ith observation. Under H_0, the true covariance matrix will be the identity, however we are relaxing the underlying assumption of normality with a light tailed (no tailed) distribution.

The results in Table 9 indicate that the statistic T_1 may be the most robust under H_0 with data from a light tailed distribution. Overall it has the most accurate empirical size at $\alpha=0.05$ while the others appear conservative. The statistics T_S and T_W appear to be very conservative. As a follow up, a power study is performed analogous to Table 2 from Section 4.2.1. The first c columns of the data matrix are uniform random variables over the interval $(-\sqrt{10.5},\sqrt{10.5})$, and the remaining $p-c$ columns are uniformly distributed over the interval $(-\sqrt{3},\sqrt{3})$. The true covariance matrix will follow the model (13) with Θ a $c\times c$ diagonal matrix with entries 3.5. Table 10 shows that the statistics T_2, T_S and T_W do not appear to be hindered by their conservativeness under the null hypothesis. We see that T_2 still appears to be the most powerful even though its empirical size was fairly conservative.

In the second result we look at a heavy tailed distribution. When simulating under the null distribution with each x_{ij} distributed as a scaled Students' t-random variable with small degrees of freedom, all of the statistics perform poorly.

Table 11
Empirical size under scaled t_8 distribution.

n	Dimensionality $p=2n$				Dimensionality $p=4n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	0.060	**0.111**	**0.181**	**0.188**	**0.068**	**0.137**	**0.199**	**0.203**	16
32	**0.071**	**0.138**	**0.177**	**0.178**	**0.064**	**0.121**	**0.183**	**0.191**	32
64	0.061	**0.157**	**0.187**	**0.189**	**0.077**	**0.124**	**0.181**	**0.190**	64
128	0.062	**0.123**	**0.181**	**0.182**	0.052	**0.092**	**0.174**	**0.176**	128
256	0.055	**0.117**	**0.192**	**0.193**	0.039	**0.093**	**0.200**	**0.200**	256
512	0.051	**0.102**	**0.184**	**0.184**	0.047	**0.093**	**0.169**	**0.170**	512

Table 12
Empirical size under scaled χ_7^2 distribution.

n	Dimensionality $p=2n$				Dimensionality $p=4n$				n
	T_1	T_2	T_S	T_W	T_1	T_2	T_S	T_W	
16	**0.075**	**0.127**	**0.176**	**0.186**	**0.071**	**0.123**	**0.195**	**0.200**	16
32	**0.066**	**0.139**	**0.220**	**0.227**	**0.072**	**0.114**	**0.186**	**0.187**	32
64	**0.067**	**0.134**	**0.217**	**0.218**	**0.072**	**0.125**	**0.231**	**0.235**	64
128	0.052	**0.122**	**0.218**	**0.220**	0.047	**0.115**	**0.217**	**0.220**	128
256	**0.074**	**0.121**	**0.229**	**0.229**	0.058	**0.102**	**0.232**	**0.232**	256
512	0.057	**0.121**	**0.221**	**0.222**	0.060	**0.078**	**0.215**	**0.215**	512

However, once the degrees of freedom is as large as 7 or 8, a difference in empirical size is noticed. We include the study where each x_{ij} is iid $(\sqrt{6/8})t_8$, so the true covariance matrix will be the identity but the data comes from a heavy tailed distribution. The results in Table 11 indicate that the statistic T_1 may be the most robust in handling a heavy tailed distribution under H_0. The statistics T_2, T_S and T_W have very poor empirical size in this situation. As the degrees of freedom increase, the empirical size improves as the t-distribution approaches normality. Since the empirical size of T_2, T_S and T_W is so liberal, any follow up power study would provide inflated results for these statistics.

Lastly, a skewed distribution is explored. Data generated from a chi-squared distribution is considered. Similar to the results from the t-distribution above, when the degrees of freedom are small, all the test statistics have poor Type I error performance. However, when the data is generated from a scaled chi-squared distribution with 7 (and larger) degrees of freedom a difference in performance is observed. Each x_{ij} is distributed as $\sqrt{1/14}\chi_7^2$ so the $p\times p$ covariance will be an identity matrix. Much like the results from the scaled t-distribution, Table 12 indicates that the statistic T_1 may be the most robust when the underlying distribution is skewed. The statistics T_2, T_S and T_W have very liberal size and their power performance will be inflated when the data comes from a skewed distribution.

4.4. Simulation conclusions

From the results of the simulation study, particularly the power study, we make the following conclusions: In general, T_2 and T_1 are comparable with T_2 generally performing better. Similarly, the statistics based on lower powers, T_S and T_W, are comparable. T_2 performs the best when the true covariance matrix is nearly an identity matrix. The statistics using lower powers perform better when the true covariance greatly differs from the identity. In some cases the statistics using higher arithmetic means perform very poorly while T_S and T_W always appear to be consistent. Lastly, from the robustness study, it appears that the statistic T_1 may be the most robust when the data comes from a skewed or heavy tailed distribution.

5. Data analysis

The statistics discussed are applied on two classic datasets: the colon data of Alon et al. (1999) and the leukemia data of Golub et al. (1999). Both datasets are preprocessed following the protocol of Dettling and Bühlmann (2003) and are publicly available on the website of Tatsuya Kubokawa: http://www.e.u-tokyo.ac.jp/~tatsuya/index.html (last accessed: 11 January 2011). The results in Fisher et al. (2010) and Srivastava (2006) demonstrate the covariance matrix from neither dataset fits the sphericity model, whence will not be the identity. However, it is interesting to note that when the statistics derived for testing the identity hypothesis are applied to the two datasets, the computed values are $T_1=6062.642$, $T_2=5666.707$, $T_S=180.929$ and $T_W=183.095$ for the colon data, and $T_1=6955.651$, $T_2=6640.174$, $T_S=198.442$ and $T_W=200.483$ for the leukemia data. When testing the sphericity hypothesis, the statistic using high powers (derived in

Fisher et al., 2010) had a substantially lower observed Z-score, but here they greatly exceed the values of the statistics based on lower powers. For each dataset, all test statistics have a p-value of zero indicating any assumption of an identity covariance matrix to be false.

6. Conclusions and discussion

This article introduced an unbiased and (n,p)-consistent estimator for the third arithmetic mean of the eigenvalues of Σ. The algebraic rationale on testing for an identity covariance matrix in Srivastava (2005) has been generalized. The result in Fisher et al. (2010) has been extended to testing for an identity covariance matrix. Two new statistics for testing the identity hypothesis have been introduced. Simulations indicate that these statistics appear to perform better in cases when just a few elements deviate from the identity matrix. Simulations also indicate the statistics in the literature perform better when many elements deviate from the identity but take on values relatively close to one.

The newly proposed statistics can be hindered by their large variance. In Section 5.1 of Fisher et al. (2010), they provide a discussion about the impact on using higher arithmetic means. The results in this article suffer the same potential drawbacks.

Acknowledgement

The author would like to thank the reviewers for their comments and suggestions to improve this article, particularly their suggestions that have improved the simulation study.

Appendix A. Technical details and proofs

A.1. Derivation of $\hat{a}_3$

The derivation of the estimator $\hat{a}_3$ follows the same methodology as Srivastava (2005) and Fisher et al. (2010).

Let $nS=YY'\sim W_p(\Sigma,n)$, where $Y=(y_1,y_2,\ldots,y_n)$ and each $y_i\sim N_p(0,\Sigma)$ and independent. By orthogonal decomposition, $\Sigma=\Gamma'\Lambda\Gamma$, where $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_p)$ with λ_i being the ith eigenvalue of Σ and Γ is an orthogonal matrix. Define $U=(u_1,u_2,\ldots,u_n)$, where u_i are iid $N_p(0,I)$ and we can write $Y=\Sigma^{1/2}U$ where $\Sigma^{1/2}\Sigma^{1/2}=\Sigma$. Define $W'=(w_1,w_2,\ldots,w_p)=U'\Gamma'$ and each w_i is iid $N_n(0,I)$. Thus, $n\mathrm{tr}\,S=\mathrm{tr}\,W'\Lambda W$.

Define $v_{ii}=w_i'w_i$ and $v_{ij}=w_i'w_j$. It is easy to see that each v_{ii} is iid chi-squared random variables with n degrees of freedom. Each v_{ij} is a sum of the product of independent standard normals. This leads to:

$$n^3\mathrm{tr}\,S^3=\sum_{i=1}^{p}\lambda_i^3v_{ii}^3+3\sum_{i\neq j}^{p}\lambda_i^2\lambda_jv_{ij}^2v_{ii}+6\sum_{i<j<k}^{p}\lambda_i\lambda_j\lambda_kv_{ij}v_{ik}v_{jk},$$

$$n^3\mathrm{tr}\,S^2\mathrm{tr}\,S=\sum_{i=1}^{p}\lambda_i^3v_{ii}^3+\sum_{i\neq j}^{p}\lambda_i^2\lambda_j(2v_{ij}^2v_{ii}+v_{ii}^2v_{jj})+2\sum_{i<j<k}^{p}\lambda_i\lambda_j\lambda_k(v_{ij}^2v_{kk}+v_{ik}^2v_{jj}+v_{jk}^2v_{ii})$$

and

$$n^3(\mathrm{tr}\,S)^3=\sum_{i=1}^{p}\lambda_i^3v_{ii}^3+3\sum_{i\neq j}^{p}\lambda_i^2\lambda_jv_{ii}^2v_{jj}+6\sum_{i<j<k}^{p}\lambda_i\lambda_j\lambda_kv_{ii}v_{jj}v_{kk}.$$

Then

$$\frac{1}{p}\left(\mathrm{tr}\,S^3-\frac{3}{n}\mathrm{tr}\,S^2\mathrm{tr}\,S+\frac{2}{n^2}(\mathrm{tr}\,S)^3\right)=\psi_1+\psi_2+\psi_3,$$

where

$$\psi_1=\frac{(n-1)(n-2)}{n^2}\frac{1}{n^3p}\sum_{i=1}^{p}\lambda_i^3v_{ii}^3, \tag{A.1}$$

$$\psi_2=\frac{3(n-2)}{n^2}\frac{1}{n^3p}\sum_{i\neq j}^{p}\lambda_i^2\lambda_j(nv_{ij}^2v_{ii}-v_{ii}^2v_{jj}) \tag{A.2}$$

and

$$\psi_3=\frac{6}{n^2}\frac{1}{n^3p}\sum_{i<j<k}^{p}\lambda_i\lambda_j\lambda_k(n^2v_{ij}v_{ik}v_{jk}-n(v_{ij}^2v_{kk}+v_{ik}^2v_{jj}+v_{jk}^2v_{ii})+2v_{ii}v_{jj}v_{kk}). \tag{A.3}$$

A.1.1. Expectations

Note the following expectations from the moments of chi-squared random variables:

Lemma 1. *For $v_{ii}=(w_i'w_i)$ and $v_{ij}=(w_i'w_j)$ for any $i\neq j$,*

$$E[v_{ii}^3]=n(n+2)(n+4),\quad E[v_{ij}^2v_{ii}]=n(n+2),$$

$$E[v_{ii}^2v_{jj}]=n^2(n+2),\quad E[v_{ij}v_{ik}v_{jk}]=n,$$

$$E[v_{ij}^2v_{kk}]=n^2,\quad E[v_{ii}v_{jj}v_{kk}]=n^3,$$

whence

Lemma 2.

$$E[\psi_1]=\frac{n(n+2)(n+4)(n-1)(n-2)}{n^5p}\sum_{i=1}^p\lambda_i^3=\frac{(n+2)(n+4)(n-1)(n-2)}{n^4}a_3$$

and

Lemma 3.

$$E[\psi_2]=E[\psi_3]=0.$$

Proposition 1. *An unbiased estimator for a_3 is provided by*

$$\hat{a}_3=\frac{\tau}{p}\left(\mathrm{tr}\,S^3-\frac{3}{n}\mathrm{tr}\,S^2\mathrm{tr}\,S+\frac{2}{n^2}(\mathrm{tr}\,S)^3\right),$$

where

$$\tau=\frac{n^4}{(n-1)(n-2)(n+2)(n+4)}.$$

Proof. This follows from Lemmas 1, 2 and 3. □

A.1.2. Variance

The variance of ψ_1 is relatively easy to find. First note that

$$V[v_{ii}^3]=6n(n+2)(n+4)(3n^2+30n+80)$$

and by independence

$$V[\psi_1]=\frac{(n-1)^2(n-2)^2}{n^{10}p^2}\sum_{i=1}^pV[v_{ii}^3]=\frac{6n(n+2)(n+4)(n-1)^2(n-2)^2(3n^2+30n+80)}{n^{10}p}a_6\simeq\frac{18}{np}a_6.$$

The variance of ψ_2 is slightly more complicated. Following the methodology of Section A.3.2 from Fisher et al. (2010) provides the result. Important results are highlighted here.

Define $V_2(i,j)=nv_{ij}^2v_{ii}-v_{ii}^2v_{jj}$ and ψ_2 can be expressed as

$$\psi_2=\frac{3(n-2)}{n^5p}\sum_{i<j}^p\lambda_i^2\lambda_jV_2(i,j)+\lambda_i\lambda_j^2V_2(j,i).$$

Since ψ_2 has expectation zero, we can look at its second moment. The following cross product terms each have expectation zero:

$$E[V_2(i,j)V_2(i,k)]=E[V_2(i,j)V_2(k,i)]=E[V_2(i,j)V_2(j,k)]=E[V_2(i,j)V_2(k,j)]=0$$

and terms of the form $V_2(i,j)V_2(k,l)$ have expectation zero by independence. The only terms contributing to the variance of ψ_2 are

$$V_2(i,j)^2=n^2v_{ij}^4v_{ii}^2-2nv_{ii}^3v_{ij}^2v_{jj}+v_{ii}^4v_{jj}^2$$

and

$$V_2(i,j)V_2(j,i)=n^2v_{ij}^4v_{ii}v_{jj}-2nv_{ij}^2v_{ii}^2v_{jj}^2+v_{ii}^3v_{jj}^3.$$

Using Lemma 4 in Fisher et al. (2010) we find

$$E[\psi_2^2]=\frac{18n^2(n-2)^2(n+2)(n+4)}{n^{10}p}\times((n-1)(n+6)(pa_4a_2-a_6)+(n-2)(n+2)(pa_3^2-a_6))\simeq\frac{18}{n^2}(a_4a_2+a_3^2).$$

Similarly, the variance of ψ_3 is found to be

$$E[\psi_3^2] = \frac{6n^3(n+2)(n^3+n^2+26n+8)}{n^{10}p}(p^2a_2^3 - 3pa_4a_2 + 2a_6) \simeq \frac{6}{n^2}ca_2^3.$$

Similar to the argument in A.3.4 from Fisher et al. (2010), the covariance terms between ψ_1, ψ_2 and ψ_3 are all zero.

A.2. *Covariance with $\hat{a}_1$, $\hat{a}_2$ and $\hat{a}_4$*

As described in Srivastava (2005) and Fisher et al. (2010), $\hat{a}_2$ and $\hat{a}_4$ can be written as components much like $\hat{a}_3 = \psi_1 + \psi_2 + \psi_3$. Let $\hat{a}_2 = q_1 + q_2$ where q_1 and q_2 are defined in Eq. (6.8) of Srivastava (2005). Let $\hat{a}_4 = \eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5$ where η_1, η_2, η_3, η_4 and η_5 are defined in Eqs. (6)–(10), respectively, in Fisher et al. (2010). The estimator for a_1 can be written as

$$\hat{a}_1 = \frac{1}{np}\sum_{i=1}^{p} \lambda_i v_{ii}.$$

Using the methodology above and that from Srivastava (2005) and Fisher et al. (2010), we can find the covariance between all the terms. Srivastava (2005) provides the covariance for the terms of $\hat{a}_1$, q_1 and q_2. Fisher et al. (2010) provide the covariance for the terms of η_i and q_j for $i = 1, \ldots, 5$ and $j=1,2$. The asymptotic covariance terms for $\hat{a}_3$ with the respective components are found using the same methodology. They are outlined below:

$$Cov(\psi_1, \hat{a}_1) \simeq \frac{6}{np}a_4,$$

$$Cov(\psi_1, q_1) \simeq \frac{12}{np}a_5,$$

$$Cov(\psi_1, \eta_1) \simeq \frac{24}{np}a_7,$$

$$Cov(\psi_1, q_2) = Cov(\psi_1, \eta_j) = 0 \quad \text{for } j = 2,3,4,5,$$

$$Cov(\psi_2, \hat{a}_1) = Cov(\psi_2, q_1) = Cov(\psi_2, \eta_j) = 0 \quad \text{for } j = 1,4,5,$$

$$Cov(\psi_2, q_2) \simeq \frac{12}{n^2}a_3a_2 = \frac{12}{np}ca_3a_2,$$

$$Cov(\psi_2, \eta_2) \simeq \frac{24}{n^2}a_5a_2 = \frac{24}{np}ca_5a_2,$$

$$Cov(\psi_2, \eta_3) \simeq \frac{48}{n^2}a_4a_3 = \frac{48}{np}ca_4a_3,$$

$$Cov(\psi_3, \hat{a}_1) = Cov(\psi_3, q_i) = Cov(\psi_3, \eta_j) = 0 \quad \text{for } i = 1,2, \quad \text{and} \quad j = 1,2,3,5$$

and

$$Cov(\psi_3, \eta_4) \simeq \frac{24}{n^2}ca_3a_2^2 = \frac{24}{np}c^2a_3a_2^2.$$

Each term was found using the methodology from Appendix A.1.2 and that of A.3.2, A.4, and A.5 in Fisher et al. (2010).

A.3. *Asymptotic normality*

The asymptotic distribution of the estimators for a_4, a_3, a_2 and a_1 can be found using the Central Limit Theorem for Martingale-Differences; see Durrett (1996), Shiryaev (1996) or Billingsley (1995).

Lemma 4. *Let $X_{n,p}$ be a sequence of random variables with $\mathcal{F}_{n,p}$ the σ-field generated by the random variables $(w_1, \ldots, w_p)$, then $\mathcal{F}_{n.0} \subset \mathcal{F}_{n,1} \subset \cdots \subset \mathcal{F}_{n,p}$. If $E[X_{n,p}|\mathcal{F}_{n,p-1}] = 0$ a.s. then $(X_{n,p}, \mathcal{F}_{n,p})$ is known as a martingale-difference array. If*

(1) $\sum_{j=0}^{p} E[(X_{n,j})^2|\mathcal{F}_{n,j-1}] \xrightarrow{p} \sigma^2$ *as* $(n,p) \to \infty$

(2) $\sum_{j=0}^{p} E[X_{n,j}^2 I(X_{n,j} > \varepsilon)|\mathcal{F}_{n,j-1}] \xrightarrow{p} 0$

then $Y_{n,p} = \sum_{j=0}^{p} X_{n,p} \xrightarrow{D} N(0, \sigma^2)$.

The second condition is known as the Lindeberg condition but can be satisfied with the stronger Lyapounov type condition:

$$\sum_{j=0}^{p} E[X_{n,j}^4 | \mathcal{F}_{n,j-1}] \xrightarrow{p} 0.$$

Proposition 2. *Under assumptions* (1d) *and* (2b), *as* $(n,p)\to\infty$

$$\sqrt{np}\dot{a} = \sqrt{np}\begin{pmatrix} \hat{a}_1-a_1 \\ \hat{a}_2-a_2 \\ \hat{a_4}-a_4 \\ \psi_1-a_3 \\ \psi_2 \\ \psi_3 \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0\\0\\0\\0\\0\\0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{14} & \sigma_{a_1\psi_1} & 0 & 0 \\ \sigma_{12} & \sigma_{22} & \sigma_{24} & \sigma_{a_2\psi_1} & \sigma_{a2\psi_2} & 0 \\ \sigma_{14} & \sigma_{24} & \sigma_{44} & \sigma_{a_4\psi_1} & \sigma_{a4\psi_2} & \sigma_{a4\psi_3} \\ \sigma_{a_1\psi_1} & \sigma_{a_2\psi_1} & \sigma_{a_4\psi_1} & \sigma^2_{\psi_1} & 0 & 0 \\ 0 & \sigma_{a2\psi_2} & \sigma_{a4\psi_2} & 0 & \sigma^2_{\psi_2} & 0 \\ 0 & 0 & \sigma_{a4\psi_3} & 0 & 0 & \sigma^2_{\psi_3} \end{pmatrix}\right),$$

where σ_{11}, σ_{12}, σ_{14}, σ_{22}, σ_{24}, *and* σ_{44} *are defined in Theorem* 2, *and* $\sigma^2_{\psi_1}$, $\sigma^2_{\psi_3}$, $\sigma^2_{\psi_3}$, $\sigma_{a_1\psi_1}$, $\sigma_{a_2\psi_1}$, $\sigma_{a_4\psi_1}$, $\sigma_{a_2\psi_2}$, $\sigma_{a_4\psi_2}$ *and* $\sigma_{a_4\psi_3}$ *are the asymptotic variances and covariances of* ψ_1, ψ_2, ψ_3, $\hat{a}_1$, $\hat{a}_2$ *and* $\hat{a}_4$ *with respect to the convergence rate of* $\sqrt{np}$. *These are easily found by taking a linear combination of the results from* Appendix A.2 *and factoring out an* (np).

Proof. The proof follows Proposition 1 in Fisher et al. (2010). Begin by considering an arbitrary non-zero linear combination of the components of our vector $\dot{a}$. With respect to the increasing set of σ-algebras, $\mathcal{F}_{n,l} = \sigma\{w_1,\dots,w_l\}$, note using the inequalities in Fisher et al. (2010) that $\dot{a}$ will satisfy the conditions of Lemma 4 if each term satisfies the requirements. The terms $(\hat{a}_1-a_1)$, $(\hat{a}_2-a_2)$, and $(\hat{a}_4-a_4)$ have been shown to satisfy the requirements of Lemma 4 in the literature. The term ψ_1 is comprised of iid components and will satisfy the requirements for the standard central limit theorem, hence it must satisfy the requirements for martingale-difference central limit theorem. Many of the details showing the asymptotic normality of the ψ_2 and ψ_3 terms follow the procedure in Fisher et al. (2010). Some details are provided outlining the proof.

As stated in Appendix A.1.2, consider writing ψ_2 as

$$\psi_2 = \frac{3(n-2)}{n^5 p}\sum_{i<j}^{p} \lambda_i^2\lambda_j V_2(i,j) + \lambda_i\lambda_j^2 V_2(j,i),$$

where $V_2(i,j) = nv_{ij}^2 v_{ii} - v_{ii}^2 v_{jj}$. Define the σ-algebra generated by the random variables $\mathcal{F}_{n,j} = \sigma\{w_1,\dots,w_j\}$ and note the following conditional expectations:

$$E[v_{ii}v_{ij}^2|\mathcal{F}_{n,j-1}] = v_{ii}^2, \quad E[v_{ii}^2 v_{jj}|\mathcal{F}_{n,j-1}] = nv_{ii}^3,$$

$$E[v_{jj}v_{ij}^2|\mathcal{F}_{n,j-1}] = (n+2)v_{ii}, \quad E[v_{jj}^2 v_{ii}|\mathcal{F}_{n,j-1}] = n(n+2)v_{ii},$$

hence $E[V_2(i,j)|\mathcal{F}_{n,j-1}] = 0$ a.s. and $E[V_2(j,i)|\mathcal{F}_{n,j-1}] = 0$ a.s. Therefore ψ_2 can be shown to be a martingale-difference.

Likewise, with respect to the σ-algebra generated as $\mathcal{F}_{n,k} = \sigma\{w_1,\dots,w_k\}$ we obtain the following conditional expectations:

$$E[v_{ij}v_{ik}v_{jk}|\mathcal{F}_{n,k-1}] = v_{ij}^2, \quad E[v_{ij}^2 v_{kk}|\mathcal{F}_{n,k-1}] = nv_{ij}^2,$$

$$E[v_{ik}^2 v_{jj}|\mathcal{F}_{n,k-1}] = v_{ii}v_{jj}, \quad E[v_{jk}^2 v_{ii}|\mathcal{F}_{n,k-1}] = v_{ii}v_{jj},$$

$$E[v_{ii}v_{jj}v_{kk}|\mathcal{F}_{n,k-1}] = v_{ii}v_{jj},$$

hence it can be shown that $E[\psi_3|\mathcal{F}_{n,k-1}] = 0$ a.s. and ψ_3 can be expressed as a martingale-difference.

To show that ψ_2 satisfies the two conditions of Lemma 4 the argument is analogous to that of the proof of Proposition 1 in Fisher et al. (2010). To demonstrate that ψ_3 satisfies the two conditions, assumption (2b) must be utilized to handle the double summation that arises when rewriting ψ_3. This is analogous to the terms η_4 and η_5 in Fisher et al. (2010). □

References

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., Levine, A.J., 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Nat Acad. Sci. USA 96(12), 6745–6750.

Anderson, T.W., 2003. An introduction to multivariate statistical analysis, Wiley Series in Probability and Statistics, third ed. Wiley-Interscience John Wiley & Sons, Hoboken, NJ.

Billingsley, P., 1995. Probability and measure, Wiley Series in Probability and Mathematical Statistics, third ed. John Wiley & Sons Inc., New York, Wiley-Interscience Publication.

Dettling, M., Bühlmann, P., 2003. Boosting for tumor classification with gene expression data. Bioinformatics 19 (9), 1061–1069.

Durrett, R., 1996. Probability: Theory and Examples, second ed. Duxbury Press, Belmont, CA.

Fisher, T.J., Sun, X., Gallagher, C.M., 2010. A new test for sphericity of the covariance matrix for high dimensional data. J. Multivariate Anal. 101 (10), 2554–2570.
Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M., Bloomfield, C., Lander, E., 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531–537.
Ledoit, O., Wolf, M., 2002. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Ann. Statist. 30 (4), 1081–1102.
Nagao, H., 1973. On some test criteria for covariance matrix. Ann. Statist. 1, 700–709.
Schott, J.R., 2005. Testing for complete independence in high dimensions. Biometrika 92 (4), 951–956.
Schott, J.R., 2006. A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. J. Multivariate Anal. 97 (4), 827–843.
Schott, J.R., 2007. A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. Comput. Statist. Data Anal. 51 (12), 6535–6542.
Shiryaev, A.N., 1996. Probability, Graduate Texts in Mathematics, vol. 95, second ed. Springer-Verlag, New York (translated from the first (1980) Russian edition by R.P. Boas).
Srivastava, M.S., 2005. Some tests concerning the covariance matrix in high dimensional data. J. Japan Statist. Soc. 35 (2), 251–272.
Srivastava, M.S., 2006. Some tests criteria for the covariance matrix with fewer observations than the dimension. Acta Comment. Univ. Tartu. Math. 10, 77–93.
Srivastava, M.S., 2009. A review of multivariate theory for high dimensional data with fewer observations. In: Advances in Multivariate Statistical Methods. Statistics Science and Interdisciplinary Research, vol. 4. World Sci. Publ., Hackensack, NJ, pp. 25–51.