Testing the structure of the covariance matrix with fewer observations than the dimension☆

Muni S. Srivastava, N. Reid

Department of Statistics
100 St. George St.
Toronto Canada M5S 3G3

Abstract

We consider two hypothesis testing problems with N independent observations on a single m-vector, when $m > N$, and the N observations on the random m-vector are independently and identically distributed as multivariate normal with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, both unknown. In the first problem, the m-vector is partitioned into two subvectors of dimensions m_1 and m_2, respectively, and we propose two tests for the independence of the two sub-vectors that are valid as $(m, N) \to \infty$. The asymptotic distribution of the test statistics under the hypothesis of independence is shown to be standard normal, and the power examined by simulations. The proposed tests perform better than the likelihood ratio test, although the latter can only be used when m is smaller than N. The second problem addressed is that of testing the hypothesis that the covariance matrix Σ is of the intraclass correlation structure. A statistic for testing this is proposed, and assessed via simulations; again the proposed test statistic compares favorably with the likelihood ratio test.

Keywords: attained significance level, false discovery rate, high dimensional data, independence of sub-vectors, intraclass correlation, multivariate normal distribution, $p > n$
2000 MSC: 62H15, 62P10

☆This research was supported by the Natural Sciences and Engineering Research Council of Canada. We wish to thank Jingchunzi Shi and Xin Wang for computational assistance.

Email address: `srivasta,reid@utstat.toronto.edu` (Muni S. Srivastava, N. Reid)

1. Introduction

Recent advances in technology to obtain DNA microarrays have made it possible to measure quantitatively the expressions of thousands of genes. These expression levels within subjects may, however, be expected to be correlated. Since the number of subjects, N, is usually quite small compared to the number of genes, m, multivariate theory for the situation when $m >> N$ needs to be developed: classical asymptotic theory requires m fixed so that $m/N \to 0$. Alternatively, some authors such as Dudoit et al. (2002) have suggested ordering the m genes by, for example, their sample means and selecting a very small number of them, much smaller than N, so that the usual asymptotic theory can be applied. The implicit assumption is that the remaining genes have mean zero and thus should not have much effect on the analysis. But unless the selected set is distributed independently of the remaining set of variables for which the mean is zero, this remaining set can provide significant information about the mean vector of the selected set; see, Srivastava and Khatri (1979, p. 115–118).

For example, consider the problem of classifying an individual with m-dimensional observed vector $\mathbf{x}$, into two known groups with mean vectors $\boldsymbol{\mu}$, $\boldsymbol{\mu}+\boldsymbol{\delta}$, and common positive definite covariance matrix Σ. Using Fisher's linear discriminant rule, both errors of misclassification are equal and given by $\Phi(-\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta}/2)$, where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable. If we use only the first m_1 components $\mathbf{x}_1$ of $\mathbf{x}$, then the errors of misclassification are equal and given by $\Phi(-\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{\delta}_1/2)$, where $\boldsymbol{\delta}$ and Σ have been partitioned according to the partitioning of $\mathbf{x}$. Since

$$\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta} = \boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{\delta}_1 + (\boldsymbol{\delta}_2 - \boldsymbol{\beta}\boldsymbol{\delta}_1)'\Sigma_{2.1}^{-1}(\boldsymbol{\delta}_2 - \beta\boldsymbol{\delta}_1),$$

where $\boldsymbol{\beta} = \Sigma_{22}^{-1}\Sigma_{12}'$ and $\Sigma_{2.1} = \Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12}$, we have $\boldsymbol{\delta}_1'\Sigma^{-1}\boldsymbol{\delta}_1 \leq \boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta}$, even when $\boldsymbol{\delta}_2 = 0$: equality holds if both $\boldsymbol{\delta}_2 = 0$ and $\boldsymbol{\beta} = 0$, or $\boldsymbol{\delta}_2 = \boldsymbol{\beta}\boldsymbol{\delta}_1$. That is, unless the two sub-vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent, dropping $\mathbf{x}_2$ loses efficiency even when the mean is the same in both groups.

Another problem of importance for the analysis of microarray data is that of testing the hypothesis that the covariance matrix has an intraclass correlation structure when $N \leq m$. Such a test is needed to select the differentially expressed genes using Benjamini and Hochberg's (1995) procedure to control the false discovery rate at a specified level. It is shown in Benjamini and Yekutieli (2001) that the false discovery rate can be controlled at a specified level if either the m genes are independently distributed, or the covariance matrix of the m genes has an intraclass correlation structure with positive correlation. Verification of intraclass correlation structure is very important, because if it fails the overall level will be $\alpha\Sigma_{j=1}^{m}(1/j)$ rather than α and adjustment for this will lead to a considerably less powerful procedure.

Tests for complete independence of all m genes can be obtained by testing the diagonality of the covariance matrix, under the assumption of normality. Such tests have been proposed by Schott (2005) and Srivastava (2005, 2006). Tests for independence that do not require normality are proposed by Szekely et al. (2009). The null distribution of these tests is based on simulation from the permutation distribution.

In this article tests for independence of two sub-vectors and for intraclass correlation structure are proposed. Both tests apply whether $N \leq m$ or $N > m$.

For the development of these tests we assume the response vector $\mathbf{x}$ follows an m-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ, and that we have a sample of N independent and identically distributed observations $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from this distribution. The sufficient statistics for $\boldsymbol{\mu}$ and Σ are

$$\begin{aligned} \bar{\mathbf{x}} &= N^{-1}\sum_{i=1}^{N}\mathbf{x}_i \ , \\ nS &= V = \sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \ , \text{ where } n = N-1 \ . \end{aligned} \tag{1}$$

Since the testing problem described above remains invariant under the additive group of transformations, $\mathbf{x} \rightarrow \mathbf{x} + \mathbf{c}, \mathbf{c} \neq 0$, we shall base our test on S, or equivalently, V.

To test the hypothesis of independence of two subvectors, we partition $\mathbf{x}$ as $(\mathbf{x}_1, \mathbf{x}_2)$, of length m_1, m_2, respectively, and consider

$$H_1 : \Sigma = \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix} \quad \text{versus} \quad A_1 : \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} ,$$

or equivalently

$$H_1 : \Sigma_{12} = \mathbf{0} \text{ versus } A_1 : \Sigma_{12} \neq \mathbf{0} \ , \tag{2}$$

where Σ is partitioned compatibly with $\mathbf{x}$.

The hypothesis that the covariance matrix Σ is of the intraclass correlation structure is

$$H_2 : \Sigma = \tau^2 \left[(1-\rho)I_m + \rho \mathbf{1}_m \mathbf{1}_m'\right] \text{ vs } A_2 : \Sigma > 0, \tag{3}$$

where I_m is the $m \times m$ identity matrix and $\mathbf{1}_m = (1, \ldots, 1)'$ is an m-vector of 1's. For convenience and simplicity, instead of V, we consider the $m \times m$ random matrix

$$W = GVG' \sim W_m(\Omega, n), \ \Omega = G\Sigma G' , \tag{4}$$

where G is a known $m \times m$ orthogonal matrix, $GG' = G'G = I_m$, of Helmert form. The first column is $(\mathbf{1}_m/\sqrt{m})'$, and the remaining columns $G_2 = (\boldsymbol{g}_2, \ldots, \boldsymbol{g}_m)$ are given by

$$\boldsymbol{g}_i = \left(\frac{1}{\sqrt{i(i-1)}}, \ldots, \frac{1}{\sqrt{i(i-1)}}, -\frac{i-1}{\sqrt{i(i-1)}}, 0, \ldots, 0\right)' . \tag{5}$$

In §2 we propose two test statistics, T_1 and T_1^*, for the problem of testing independence of two sub-vectors, (2). We show that the limiting distribution of T_1 and T_1^* are standard normal under H_1, when $(m, N) \to \infty$, and study the finite sample performance by simulations. In §3 we propose a test statistic, T_2, for testing the hypothesis (3) that the covariance matrix Σ is of the intraclass correlation structure, show that T_2 is asymptotically standard normal under H_2, and study its finite sample performance through simulations. We compare these test statistics to the relevant likelihood ratio tests, which are only valid for $m < N$, and show that the performance of the proposed tests is generally better than that of the likelihood ratio test. The methods of proof are similar in the two cases, and use results on invariance and asymptotic normality that are outlined in the Appendix. In §4, we illustrate the proposed tests on a microarray dataset.

2. Testing the independence of two sub-vectors

2.1. The proposed test statistics

Our proposed test statistics are based on consistent estimates of two parametric measures of distance δ_1^2, and δ_2^2 which we now introduce. As shown in the Appendix, for $n < m$ no invariant test exists under the group of non-singular linear transformations. We consider tests that are invariant under a smaller group of transformations

$$\mathbf{x} \to \begin{pmatrix} c_1\Gamma_1 & 0 \\ 0 & c_2\Gamma_2 \end{pmatrix} \mathbf{x} \,, \tag{6}$$

where $c_i > 0, i = 1, 2$, and Γ_1 and Γ_2 are orthogonal matrices. A distance function between the hypothesis H_1 and the alternative A_1, invariant under this group of transformations, is

$$\delta_1^2 = \frac{1}{2m\sqrt{2}} \operatorname{tr} \left[D^{-1} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} - D^{-1} \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right]^2 ,$$

where D is a diagonal matrix in which the first m_1 diagonal elements are $a_{2(1)}^{1/2}$ and the remaining m_2 diagonal elements are $a_{2(2)}^{1/2}$, and

$$a_{2(1)} = \operatorname{tr}(\Sigma_{11}^2)/m, \quad a_{2(2)} = \operatorname{tr}(\Sigma_{22}^2)/m \,, \quad a_{(1,2)} = \operatorname{tr}(\Sigma_{12}\Sigma_{12}')/m. \tag{7}$$

It can be easily seen that

$$\delta^2 = \frac{1}{2m\sqrt{2}} \operatorname{tr} \begin{pmatrix} 0 & a_{2(1)}^{-1/2}\Sigma_{12} \\ a_{2(2)}^{-1/2}\Sigma_{12}' & 0 \end{pmatrix}^2 = \frac{a_{(1,2)}}{\sqrt{2a_{2(1)}a_{2(2)}}} \,. \tag{8}$$

Note that $a_{(1,2)} = 0$ if and only if $\Sigma_{12} = 0$ and $a_{(1,2)} > 0$, otherwise.

Let

$$\hat{a}_{(1,2)} = \frac{n^2}{(n-1)(n+2)m} \left[\operatorname{tr}(S_{12}S_{12}') - \frac{1}{n}\operatorname{tr}(S_{11})\operatorname{tr}(S_{22}) \right] , \tag{9}$$

$$\hat{a}_{2(i)} = \frac{n^2}{(n-1)(n+2)m} \left[\operatorname{tr}(S_{ii}^2) - \frac{1}{n}\{\operatorname{tr}(S_{ii})\}^2 \right], \quad i = 1, 2 \,, \tag{10}$$

where S, defined at (1), is partitioned compatibly with Σ:

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{pmatrix} . \tag{11}$$

Our first test statistic for H_1 is

$$T_1 = n\frac{\hat{a}_{(1,2)}}{\sqrt{2\hat{a}_{2(1)}\hat{a}_{2(2)}}}. \tag{12}$$

A smaller group of transformations is given by the group of $m \times m$ nonsingular diagonal matrices

$$\mathbf{x} \to D^*\mathbf{x} = \begin{pmatrix} D_1^* & 0 \\ 0 & D_2^* \end{pmatrix} \mathbf{x} \ , \tag{13}$$

where $D^* = \mathrm{diag}(d_1^*, \ldots, d_m^*)$, with $D_1^* = \mathrm{diag}(d_1^*, \ldots, d_{m_1}^*)$, and D_2^* the remaining components, where we assume $0 < d_i^* < \infty, i = 1, \ldots, m$. Let

$$\begin{aligned} R_{11} &= D_1^{*-1/2}\Sigma_{11}D_1^{*-1/2}, \quad R_{22} = D_2^{*-1/2}\Sigma_{22}D_2^{*-1/2}, \\ R_{12} &= D_1^{*-1/2}\Sigma_{12}D_2^{*-1/2}, \quad a_{2(1)}^* = \mathrm{tr}(R_{11}^2)/m, \quad a_{2(2)}^* = \mathrm{tr}(R_{22}^2)/m. \end{aligned}$$

We choose

$$d_i^* = (\sigma_{ii}/a_{2(1)}^*)^{1/2}, i = 1, \ldots, m_1; \quad d_i^* = (\sigma_{ii}/a_{2(2)}^*)^{1/2}, \quad , i = m_1 + 1, \ldots, m,$$

and consider the distance measure between the hypothesis H_1 and the alternative A_1 as

$$\begin{aligned} \delta^{*2} &= \frac{1}{2m\sqrt{2}}\mathrm{tr}\left[D^{*-1}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} - D^{*-1}\begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right]^2 \\ &= \frac{a_{(1,2)}^*}{\sqrt{2a_{2(1)}^* a_{2(2)}^*}}. \end{aligned} \tag{14}$$

Thus we need to obtain consistent estimators of $a_{(1,2)}^*$, $a_{2(1)}^*$, and $a_{2(2)}^*$. Since $\mathrm{diag}(s_{11}, \ldots, s_{mm})$ is a consistent estimator of $(\sigma_{11}, \ldots, \sigma_{mm})$, it follows that consistent estimators are given respectively by

$$\hat{a}_{(1,2)}^* = \frac{1}{m}\{\mathrm{tr}(R_{12}R_{12}') - \frac{m_1 m_2}{m}\}, \tag{15}$$

$$\hat{a}_{2(1)}^* = \frac{1}{m}\{\mathrm{tr}(R_{11}^2) - \frac{m_1^2}{m}\}, \tag{16}$$

$$\hat{a}_{2(2)}^* = \frac{1}{m}\{\mathrm{tr}(R_{22}^2) - \frac{m_2^2}{m}\}, \tag{17}$$

where

$$\begin{aligned} R &= \begin{pmatrix} R_{11} & R_{12} \\ R_{12'} & R_{22} \end{pmatrix} = D_s^{*-1/2} S D_s^{*-1/2}, \\ D_s^* &= \mathrm{diag}(s_{11}, \ldots, s_{mm}). \end{aligned}$$

Thus another test statistic T_1^* is given by

$$T_1^* = n \frac{\hat{a}^*_{(1,2)}}{\sqrt{2\hat{a}^*_{2(1)}\hat{a}^*_{2(2)}}}. \tag{18}$$

In the next subsection we show that T_1 is asymptotically normally distributed with mean 0 and variance 1 under $H_1 : \Sigma_{12} = 0$. From this result it also follows that T_1^* is asymptotically distributed as a standard normal under H_1: this is stated in Corollary 2.1. We require the following assumption, writing $a_i = \mathrm{tr}(\Sigma^i)/m$:

Assumption A:

(i) $0 < a_i^0 = \lim_{m\to\infty} a_i < \infty, \lim_{m\to\infty} m^{-1}a_4 \to 0,\ i = 1, 2$

(ii) $0 < \lim_{m\to\infty} (m_j/m) = c_j < \infty,\ j = 1, 2,$

(iii) $n = O(m^\delta),\ \delta > 0$.

The following lemma is proved in Srivastava (2005, p.252, Lemma 2.1):

Lemma 2.1. *Let $V \sim W_m(\Sigma, n)$ and $a_i = tr(\Sigma^i)/m, \quad i = 1, \ldots, 4$. Then under Assumption A, unbiased and consistent estimators of a_1 and a_2 as $(n, m) \to \infty$ are given by*

$$\hat{a}_1 = \frac{tr(V)}{nm}\ , \quad \hat{a}_2 = \frac{1}{(n-1)(n+2)m}\left[tr(V^2) - \frac{1}{n}\{tr(V)\}^2\right]. \tag{19}$$

2.2. Asymptotic Distribution of the Test Statistic T_1

The proposed test statistic is based on consistent estimator of δ^2, for which we need consistent estimators of $a_{2(1)}$, $a_{2(2)}$ and $a_{(1,2)}$. Note that

$$a_2 = \frac{1}{m}\mathrm{tr}(\Sigma^2) = \frac{1}{m}\mathrm{tr}\left[\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{pmatrix}\right] = a_{2(1)} + a_{2(2)} + 2a_{(1,2)}\ ,$$

where $a_{2(i)}, i = 1, 2$, and $a_{(1,2)}$ are defined in (7). From the definition of $\hat{a}_{(1,2)}$ in (9), and $\hat{a}_{2(i)}$ in (10), we can write

$$\begin{aligned}\hat{a}_2 &= \frac{n^2}{(n-1)(n+2)m}\left[\mathrm{tr}(S^2) - \frac{1}{n}\{\mathrm{tr}(S)\}^2\right]\\ &= \hat{a}_{2(1)} + \hat{a}_{2(2)} + 2\hat{a}_{(1,2)}\ .\end{aligned}$$

Since

$$\frac{n^2}{(n-1)(n+2)m}\left[\mathrm{tr}(S_{ii}^2) - \frac{1}{n}\{\mathrm{tr}(S_{ii})\}^2\right], \quad i = 1, 2 \ ,$$

are consistent and unbiased estimators of $(1/m)\mathrm{tr}(\Sigma_{ii}^2), i = 1, 2$, by Lemma 2.1 $\hat{a}_{(1,2)}$ is a consistent and unbiased estimator of $a_{(1,2)}$, under Assumption A. In the next theorem, we give an expression for the asymptotic variance of $\hat{a}_{(1,2)}$.

Theorem 2.1. *Let $\hat{a}_{(1,2)}$ be as defined in (9). Then the variance of $\hat{a}_{(1,2)}$ under the hypothesis H_1 and assumption A is given by*

$$\mathrm{Var}(\hat{a}_{(1,2)}) = \frac{2}{n^2} a_{2(1)} a_{2(2)} + O(\frac{1}{n^3}) \ .$$

Proof: Since $V \sim W_n(\Sigma, n)$, we can write $V = nS = YY'$, where $Y = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$, and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are independent and identically distributed as $N_m(\mathbf{0}, \Sigma)$, where we write Σ_0 for the covariance matrix under $H_1 : \Sigma_{12} = 0$. Let Γ be an $m \times m$ orthogonal matrix given by $\Gamma = \mathrm{diag}(\Gamma_1, \Gamma_2)$, where Γ_1 is $m_1 \times m_1$ and Γ_2 is $m_2 \times m_2$, and

$$\Gamma \Sigma_0 \Gamma' = \begin{pmatrix} \Gamma_1 \Sigma_{11} \Gamma_1' & 0 \\ 0 & \Gamma_2 \Sigma_{22} \Gamma_2' \end{pmatrix} = \begin{pmatrix} D_{\lambda_{(1)}} & 0 \\ 0 & D_{\lambda_{(2)}} \end{pmatrix} \ ,$$

where $D_{\lambda_{(1)}} = \mathrm{diag}(\lambda_{(1)1}, \ldots, \lambda_{(1)m_1})$ and $D_{\lambda_{(2)}} = \mathrm{diag}(\lambda_{(2)1}, \ldots, \lambda_{(2)m_2})$ are diagonal matrices composed of the eigenvalues of Σ_{11} and Σ_{22}. Thus, with $U = (U^{(1)'}, U^{(2)'})' = \Gamma' \Sigma_0^{-\frac{1}{2}} Y$,

$$V = \Gamma \left(\begin{array}{cc} D_{\lambda(1)}^{\frac{1}{2}} & 0 \\ 0 & D_{\lambda(2)}^{\frac{1}{2}} \end{array} \right) \left(\begin{array}{c} U^{(1)} \\ U^{(2)} \end{array} \right) \left(U^{(1)'} \ , \ U^{(2)'} \right) \left(\begin{array}{cc} D_{\lambda(1)}^{\frac{1}{2}} & 0 \\ 0 & D_{\lambda(2)}^{\frac{1}{2}} \end{array} \right) \Gamma' \ .$$

The n columns of U are independently distributed as $N_m(\mathbf{0}, I_m)$, $U^{(1)}$ and $U^{(2)}$ are independently distributed under H_1, and the n columns of $U^{(i)}$ are independently distributed as $N_{m_i}(\mathbf{0}, \ I_{m_i})$, $i = 1, 2$. Writing

$$U^{(1)'} = \left(\boldsymbol{u}_1^{(1)}, \ldots, \boldsymbol{u}_{m_1}^{(1)} \right), \ U^{(2)'} = \left(\boldsymbol{u}_1^{(2)}, \ldots, \boldsymbol{u}_{m_2}^{(2)} \right), \qquad (20)$$

then $\boldsymbol{u}_1^{(1)}, \ldots, \boldsymbol{u}_{m_1}^{(1)}, \boldsymbol{u}_1^{(2)}, \ldots, \boldsymbol{u}_{m_2}^{(2)}$ are independent and identically distributed as $N_n(\mathbf{0}, I)$ under $H_1 : \Sigma_{12} = 0$. Using (9) we have

$$\begin{aligned}\hat{a}_{(1,2)} &= \frac{1}{m(n-1)(n+2)} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \lambda_{(1)i}\lambda_{(2)j} \left[\left(\boldsymbol{u}_i^{(1)'} \boldsymbol{u}_j^{(2)} \right)^2 - \frac{1}{n} (\boldsymbol{u}_i^{(1)'} \boldsymbol{u}_i^{(1)})(\boldsymbol{u}_j^{(2)'} \boldsymbol{u}_j^{(2)}) \right], \\ &\simeq \frac{1}{mn^2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \lambda_{(1)i}\lambda_{(2)j} z_{ij}, \qquad (21)\end{aligned}$$

where

$$\begin{aligned} z_{ij} &= \left(\boldsymbol{u}_i^{(1)'} \boldsymbol{u}_j^{(2)} \right)^2 - \frac{1}{n} (\boldsymbol{u}_i^{(1)'} \boldsymbol{u}_i^{(1)})(\boldsymbol{u}_j^{(2)'} \boldsymbol{u}_j^{(2)}), \\ &= (w_{ij}^2 - n) - \frac{1}{n}(w_{ii}^{(1)} w_{jj}^{(2)} - n^2), \qquad (22)\end{aligned}$$

where $w_{ij} = \boldsymbol{u}_i^{(1)'} \boldsymbol{u}_j^{(2)}$, and $w_{ii}^{(1)} = \boldsymbol{u}_i^{(1)'} \boldsymbol{u}_i^{(1)}$ and $w_{jj}^{(2)} = \boldsymbol{u}_j^{(2)'} \boldsymbol{u}_j^{(2)}$ are independently and identically distributed under H_1 for all i, j as χ_n^2 random variables. Hence, under H_1, $\mathrm{E}(z_{ij}) = 0$, $\mathrm{Cov}\,(z_{ij}, z_{kl}) = 0$ for all distinct (j, ℓ) or (i, k) and $\mathrm{Var}\,(z_{ij}) = 2(n+2)(n-1)$. Hence, under H_1, $\mathrm{E}(\hat{a}_{(1,2)}) = 0$ and

$$\mathrm{Var}(\hat{a}_{(1,2)}) \simeq \frac{1}{m^2 n^4} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \lambda_{(1)i}^2 \lambda_{(2)j}^2 \mathrm{Var}(z_{ij}) = \frac{2}{n^2} a_{2(1)} a_{2(2)} \ ,$$

neglecting terms of $O(n^{-3})$.

Theorem 2.2. *Let $\hat{a}_{(1,2)}$ and $\hat{a}_{2(i)}$ be defined as in (9) and (10). Then T_1 defined in (12) is asymptotically normally distributed as $(m, n) \to \infty$ under the hypothesis H_1 and Assumption A; i.e.,*

$$\lim_{(m,n)\to\infty} P_0(T_1 \leq z) = \Phi(z)$$

where $\Phi(\cdot)$ is the distribution function of a standard normal random variable and P_0 denotes the distribution under the null hypothesis.

Proof: As noted above $\hat{a}_{2(1)}$ and $\hat{a}_{2(2)}$ are consistent estimators of $a_{2(1)}$ and $a_{2(2)}$ respectively. Thus, we need to find the asymptotic distribution of $n\hat{a}_{(1,2)}$ where we use the asymptotic expression for $\hat{a}_{(1,2)}$ given at (21).

We note that

$$\mathrm{Var}\left(\frac{1}{mn^3} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \lambda_{(1)i}\lambda_{(2)j} w_{ii}^{(1)} w_{jj}^{(2)} \right) = \frac{4}{n^4} a_{2(1)} a_{2(2)} = O(n^{-4}) \ .$$

Since this is of order $O(n^{-4})$, the second term of $n\hat{a}_{(1,2)}$ converges in probability to its expectation. Thus

$$\hat{a}_{(1,2)} \overset{d}{=} \frac{1}{mn} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \lambda_{(1)i} \lambda_{(2)j} \left[(w_{ij}^2/n) - 1 \right] \ ,$$

and the asymptotic distribution of $n\hat{a}_{(1,2)}$ as $(m,n) \to \infty$, is the same as that of

$$\left(\frac{m_1 m_2}{m^2} \right)^{\frac{1}{2}} \frac{1}{(m_1 m_2)^{\frac{1}{2}}} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \lambda_{(1)i} \lambda_{(2)j} \left[\left(\eta_{ij}^2 \nu_i^2 / n \right) - 1 \right]$$

where

$$\nu_i^2 = \boldsymbol{u}_i^{(1)'} \boldsymbol{u}_i^{(1)}, \text{ and } \eta_{ij} = \boldsymbol{u}_i^{(1)'} \boldsymbol{u}_j^{(2)} / \nu_i .$$

Given $\boldsymbol{u}_i^{(1)}, \eta_{ij}$ has a normal distribution with mean 0 and variance 1 which does not depend on $\boldsymbol{u}_i^{(1)}$; hence η_{ij} are independently distributed of ν_i for all i, j. Noting that $\nu_i^2/n = 1 + O_p(n^{-\frac{1}{2}})$, we find that the asymptotic distribution of $n\hat{a}_{(1,2)}$ is the same as that of $[(m_1 m_2)/m^2]^{\frac{1}{2}} Q$, where

$$Q = \frac{1}{(m_1 m_2)^{\frac{1}{2}}} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \lambda_{(1)i} \lambda_{(2)j} (\eta_{ij}^2 - 1) \ .$$

Then

$$\frac{1}{(m_1 m_2)} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \lambda_{(1)i}^2 \lambda_{(2)j}^2 \int\limits_{|\gamma| > \varepsilon \sqrt{m_1 m_2}} \gamma^2 \, \mathrm{dF}(\gamma) \ \leq \left[\frac{m^2}{(m_1 m_2)} \right] a_{2(1)} a_{2(2)} \left[\frac{1}{\varepsilon^2 m_1 m_2} \right] E\left(\eta_{ij}^4 \right) \ .$$

which goes to zero, as $(m_1, m_2) \to \infty$. Hence, from the Lyapunov central limit theorem, it follows that under H_1

$$\frac{1}{m} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \lambda_{(1)i} \lambda_{(2)j} \left[\left(\boldsymbol{u}_i^{(1)'} \boldsymbol{u}_j^{(2)} \right)^2 - 1 \right] \to N(0, 2a_{2(1)} a_{2(2)}) \ .$$

This proves Theorem 2.2. An alternative proof can be obtained by using Lemma A.1 of the Appendix.

Corollary 2.1. *Let* $\hat{a}^*_{(1,2)}$ *and* $\hat{a}^*_{2(i)}$ *be defined as in (15, 16, 17), respectively. Then* T_1^* *defined in (18) is asymptotically normally distributed as* $(m,n) \to \infty$ *under the hypothesis* H_1 *and Assumption A; i.e.,*

$$\lim_{(m,n)\to\infty} P_0(T_1^* \leq z) = \Phi(z)$$

where $\Phi(\cdot)$ *is the distribution function of a standard normal random variable and* P_0 *denotes the distribution under the null hypothesis.*

It may be noted that following Srivastava (2005), where a test of independence of all components of $\mathbf{x}$ is given, another test can be proposed based on the distance function

$$\delta^{*2} = \left[\frac{\mathrm{tr}(\Sigma^2)}{\mathrm{tr}(\Sigma_{11}^2) + \mathrm{tr}(\Sigma_{22}^2)} - 1\right] = \left[\frac{a_2}{a_{2(1)} + a_{2(2)}} - 1\right] ,$$

which takes the value zero if and only if $\Sigma_{12} = 0$; otherwise $\delta^{*2} > 0$. A test based on a consistent estimator of δ^*, namely

$$\begin{aligned} T_{1A} =& \frac{\hat{a}_2}{\hat{a}_{2(1)} + \hat{a}_{2(2)}} - 1 \\ =& \frac{\hat{a}_{2(1)} + \hat{a}_{2(2)} + 2\hat{a}_{(1,2)}}{\hat{a}_{2(1)} + \hat{a}_{2(2)}} - 1 \\ =& \frac{2\hat{a}_{(1,2)}}{\hat{a}_{2(1)} + \hat{a}_{2(2)}}, \end{aligned}$$

can also be proposed. However this test is also based on $\hat{a}_{(1,2)}$, hence asymptotically equivalent to the proposed test statistic T_1, and thus needs no further consideration.

2.3. Power of the Test of Independence and its Attained Significance Level

In this section we consider the performance of the test statistics T_1 and T_1^* in finite samples by simulation. We first examine the attained significance level of the test statistic compared to the nominal value $\alpha = 0.05$. We use $\Sigma = DRD, D = \mathrm{diag}(d_1, \ldots, d_m),\ R = (r_{ij}), r_{ii} = 1, r_{ij} = (-1)^{i+j}(\rho)^{|i-j|^{0.1}}, i \neq j, i, j = 1, \ldots, m$; and report results for the choices $d_i = 2+(m-i+1)/m$ and $D-i$ independently distributed as χ_3^2. For the hypothesis, we make $\Sigma_{12} = 0$ by taking $\Sigma = \mathrm{diag}(\Sigma_{11}, \Sigma_{22})$, where $\Sigma_{11} = D_1R_1D_1$, $\Sigma_{22} = D_2R_2D_2$, D_1

and R_1 are the corresponding sub-matrices of D and R, and D_2 and R_2 are similarly defined.

The attained significance level (ASL) is $\hat{\alpha}_T = \#(T_{1H} > z_{1-\alpha})/r$ where T_{1H} are values of the test statistic T_1 (or T_1^*) computed from data simulated under H_1, r is the number of replications and $z_{\alpha/2}$ is the $100(1-\alpha)\%$ point of the standard normal distribution. The ASL assesses how close the null distribution of T_1 (or T_1^*) is to its limiting null distribution. From the same simulation, we also obtain $\hat{z}_\alpha$ as the $100(1-\alpha)\%$ point of the empirical null distribution, and define the attained power by $\hat{\beta}_T = \#(T_{1A} > \hat{z}_{1-\alpha})/r$, where T_{1A} are values of the T_1 (or T_1^*) computed from data simulated under A_1.

In Table 1 we compare the proposed tests T_1 and T_1^* with the likelihood ratio test, when $m < N$. We use two approximations to the distribution of the likelihood ratio statistic

$$\lambda^* = |S|/(|S_{11}||S_{22}|) .$$

Under H_1, $-g \log \lambda^*$ is asymptotically distributed as $\chi^2_{m_1 m_2}$, where $g = N - 3 - m/2, \gamma = m_1 m_2 (m_1^2 + m_2^2 - 5)/48, f = m_1 m_2$ (Srivastava and Khatri, 1979, p.222). The test based on this approximation will be denoted LR_1. Another approximation, which may have better performance when m is close to n is

$$LR_2 = (-g \log \lambda^* - f)/(2f)^{1/2};$$

this is asymptotically distributed as $N(0,1)$ under H_1, as $n \to \infty$. The results in Table 1 show that even for small m and large n, the tests based on T_1 and T_1^* perform better than both approximations to the distribution of the likelihood ratio test, and the test based on T_1^* is better than that based on T_1, which is to be expected since our simulations are consistent with the invariance structureof (13).

It may be noted that irrespective of the ASL of any statistic, the power has been computed when all the statistics in the comparison have the same specified significance level as the cut off points have been obtained by simulation. Thus the empirical powers for LR_1 and LR_2 are the same; only one is shown. The ASL gives an idea as to how close it is to the specified significance level. If it is not close, the only choice left is to obtain it from simulation, not from the asymptotic distribution. It is common in practice, although not recommended, to depend on the asymptotic distribution, rather than relying on simulations to determine the ASL.

Szekely et al. (2009) proposed a nonparametric test for independence; the p-value for their test statistic is estimated by using the permutation distribution. Limited simulations, not shown here, indicated that compared to the test based on T_1 or T_1^*, their test has size closer to nominal, although slightly less power, for $N < m$, and much lower power for $N > m$.

3. Testing intraclass correlation

3.1. The test statistic

In this section, we consider the problem of testing that the covariance matrix Σ has the intraclass correlation structure,

$$\Sigma = \tau^2[(1-\rho)I_m + \rho\mathbf{1}_m\mathbf{1}_m'], \quad -1/(m-1) < \rho < 1 \ . \tag{23}$$

When Σ is of the form (3.1), from (1.5) we can write

$$\Omega = \begin{pmatrix} \Omega_{11} & \boldsymbol{\Omega}_{12}' \\ \boldsymbol{\Omega}_{12} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} \lambda^2 & 0 \\ 0 & \sigma^2 I_{m-1} \end{pmatrix}, \tag{24}$$

where $\lambda^2 = \tau^2[1+(m-1)\rho] > 0$, and $\sigma^2 = \tau^2(1-\rho)$. Thus, we can re-express H_2 as

$$H_2 : \Omega_{11} = \lambda^2, \quad \boldsymbol{\Omega}_{12} = \mathbf{0}, \quad \Omega_{22} = \sigma^2 I_{m-1}, \quad \sigma^2 > 0.$$

When $n > m$, the maximum likelihood estimate of Ω_{11} under A_2 remains the same as the maximum likelihood estimate of λ^2 under H_2, since both Ω_{11} and λ^2 are unknown scalars. Thus H_2 is equivalent to

$$H_2 : \boldsymbol{\Omega}_{12} = 0, \ \Omega_{22} = \sigma^2 I_{m-1}, \ \sigma^2 > 0 \ ,$$

with Ω_{11} a nuisance parameter in both H_2 and A_2. Under H_2 we note that

$$\sigma^2 = \mathrm{tr}(\Omega_{22})/(m-1) \equiv a_{1(2)}^* \ .$$

We also define

$$a_{2(1)}^* = \frac{\Omega_{11}^2}{m-1}, \ a_{2(2)}^* = \frac{\mathrm{tr}(\Omega_{22}^2)}{m-1}, \ a_{(1,2)}^* = \frac{\boldsymbol{\Omega}_{12}'\boldsymbol{\Omega}_{12}}{m-1} \ , \tag{25}$$

and make the following assumption:
Assumption B:

Table 1: Attained significance level and attained power of the tests of $\Sigma_{12} = 0$ based on T_1 and T_1^* given in (12) and (18), compared to two versions of the likelihood ratio test. The covariance matrix is constructed from $D = \text{diag}(d_i)$ where $d_i = 2+(m-i+1)/m$.The likelihood ratio test can only be used when $m < N$. These tables are based on 1,000 simulations; additional runs with 10,000 simulations for several cases gave very similar results.

			ASL under H_1				Power ($\rho = 0.2$)			
N	m_1	m_2	T_1	T_1^*	LR_1	LR_2	T_1	T_1^*	LR_1	LR_2
	2	3	0.064	0.056	0.024	0.033	0.169	0.191	0.060	0.076
	5	5	0.075	0.064	0.019	0.027	0.235	0.217	0.034	0.042
	10	15	0.054	0.055	—	—	0.373	0.347	—	—
15	50	50	0.056	0.054	—	—	0.612	0.595	—	—
	50	100	0.051	0.054	—	—	0.651	0.606	—	—
	100	200	0.049	0.047	—	—	0.704	0.675	—	—
	200	300	0.055	0.059	—	—	0.710	0.671	—	—
	400	600	0.047	0.047	—	—	0.745	0.733	—	—
	2	3	0.059	0.054	0.034	0.059	0.315	0.343	0.151	0.193
	5	5	0.069	0.069	0.028	0.069	0.389	0.362	0.081	0.102
	10	15	0.067	0.066	—	—	0.626	0.597	—	—
	50	50	0.057	0.050	—	—	0.845	0.852	—	—
25	50	100	0.051	0.044	—	—	0.891	0.882	—	—
	100	200	0.071	0.066	—	—	0.917	0.913	—	—
	200	300	0.061	0.058	—	—	0.909	0.904	—	—
	400	600	0.066	0.067	—	—	0.916	0.914	—	—
	2	3	0.061	0.060	0.037	0.056	0.530	0.528	0.298	0.356
	5	5	0.054	0.056	0.035	0.042	0.780	0.782	0.324	0.353
	10	15	0.058	0.061	0.061	0.037	0.929	0.921	0.135	0.148
50	50	50	0.048	0.0571	—	—	0.994	0.996	—	—
	50	100	0.042	0.049	—	—	0.999	0.998	—	—
	100	200	0.059	0.058	—	—	0.999	0.999	—	—
	200	300	0.068	0.065	—	—	0.999	0.999	—	—
	400	600	0.061	0.059	—	—	0.999	0.998	—	—
	2	3	0.068	0.056	0.045	0.065	0.848	0.841	0.705	0.750
	5	5	0.058	0.064	0.039	0.055	0.972	0.972	0.746	0.773
	10	15	0.061	0.060	0.036	0.045	0.998	0.998	0.518	0.542
	50	50	0.051	0.045	—	—	1	1	—	—
100	50	100	0.064	0.061	—	—	1	1	—	—
	100	200	0.044	0.044	—	—	1	1	—	—
	200	300	0.060	0.059	—	—	1	1	—	—
	400	600	0.059	0.059	—	—	1	1	—	—

Table 2: Attained significance level and attained power of the tests of $\Sigma_{12} = 0$ based on T_1 and T_1^* given in (12) and (18), compared to two versions of the likelihood ratio test. The covariance matrix is constructed from $D = \text{diag}(d_i)$ where $d_i \approx \chi_3^2$. The likelihood ratio test can only be used when $m < N$. These tables are based on 1,000 simulations.

			ASL under H_1				Power ($\rho = 0.2$)			
N	m_1	m_2	T_1	T_1^*	LR_1	LR_2	T_1	T_1^*	LR_1	LR_2
15	2	3	0.074	0.075	0.032	0.047	0.096	0.149	0.058	0.081
	5	5	0.055	0.047	0.020	0.024	0.124	0.276	0.035	0.045
	10	15	0.060	0.056	—	—	0.141	0.385	—	—
	50	50	0.063	0.047	—	—	0.188	0.585	—	—
	50	100	0.064	0.047	—	—	0.201	0.640	—	—
	100	200	0.059	0.061	—	—	0.258	0.625	—	—
	200	300	0.038	0.051	—	—	0.454	0.677	—	—
	400	600	0.058	0.050	—	—	0.480	0.712	—	—
25	2	3	0.065	0.048	0.028	0.038	0.202	0.312	0.129	0.168
	5	5	0.080	0.054	0.024	0.031	0.234	0.464	0.102	0.121
	10	15	0.072	0.052	—	—	0.229	0.613	—	—
	50	50	0.069	0.051	—	—	0.344	0.844	—	—
	50	100	0.060	0.050	—	—	0.479	0.858	—	—
	100	200	0.054	0.056	—	—	0.608	0.899	—	—
	200	300	0.060	0.0548	—	—	0.685	0.935	—	—
	400	600	0.072	0.060	—	—	0.682	0.910	—	—
50	2	3	0.066	0.066	0.044	0.055	0.250	0.504	0.325	0.390
	5	5	0.059	0.063	0.039	0.046	0.487	0.768	0.322	0.366
	10	15	0.057	0.058	0.035	0.039	0.782	0.931	0.152	0.164
	50	50	0.062	0.056	—	—	0.631	0.993	—	—
	50	100	0.054	0.066	—	—	0.837	0.996	—	—
	100	200	0.052	0.062	—	—	0.956	0.996	—	—
	200	300	0.050	0.053	—	—	0.969	0.999	—	—
	400	600	0.055	0.055	—	—	0.964	0.998	—	—
100	2	3	0.073	0.069	0.047	0.060	0.662	0.826	0.700	0.750
	5	5	0.073	0.060	0.049	0.060	0.704	0.974	0.732	0.768
	10	15	0.061	0.060	0.036	0.045	0.739	0.999	0.532	0.550
	50	50	0.070	0.062	—	—	0.997	1	—	—
	50	100	0.068	0.057	—	—	0.992	1	—	—
	100	200	0.060	0.055	—	—	0.997	1	—	—
	200	300	0.069	0.064	—	—	1	1	—	—
	400	600	0.068	0.067	—	—	1	1	—	—

(i) $0 < \lim_{m\to\infty} a^*_{i(2)} < \infty, i = 1, 2,$

(ii) $0 \leq \lim_{m\to\infty} a^*_{(1,2)} < \infty,$

(iii) $n = O(m^\delta), \quad \delta > 0.$

The parameters

$$F_1 = \frac{a^*_{(1,2)}}{\sqrt{2a^*_{2(1)}a^*_{2(2)}}} \text{ and } F_2 = \frac{1}{2}\left(1 - \frac{a^{*2}_{1(2)}}{a^*_{2(2)}}\right) , \tag{26}$$

are invariant under the group of transformations

$$\mathbf{x} \to \begin{pmatrix} c_1 & \mathbf{0}' \\ \mathbf{0} & c_2 G_{m-1} \end{pmatrix} \mathbf{x} , \tag{27}$$

where G_{m-1} is orthogonal and $c_i > 0, \ i = 1, 2$. We consider a distance function that measures the difference between the hypothesis H_2 and the alternative hypothesis $A_2 : \Sigma > 0$. Let D be an $m \times m$ diagonal matrix given by

$$\begin{aligned} D &= \operatorname{diag}\left[\frac{1}{2}(2a^*_{2(1)})^{-\frac{1}{2}}, \ \frac{1}{2}(a^*_{2(2)})^{-\frac{1}{2}} I_{m-1}\right] \\ &= \operatorname{diag}\left(d_1, d_2 I_{m-1}\right) . \end{aligned}$$

We define a distance that measures the difference between the hypothesis H_2 and A_2 by

$$\begin{aligned} \eta^2 &= \frac{1}{(m-1)} \operatorname{tr} \left[D \begin{pmatrix} \Omega_{11} & \mathbf{\Omega}'_{12} \\ \mathbf{\Omega}_{12} & \Omega_{22} \end{pmatrix} - D \begin{pmatrix} \Omega_{11} & 0 \\ 0 & a^*_{1(2)} I_{m-1} \end{pmatrix}\right]^2 \\ &= \frac{1}{(m-1)} \operatorname{tr} \left[D \begin{pmatrix} 0 & \mathbf{\Omega}'_{12} \\ \mathbf{\Omega}_{12} & \left(\Omega_{22} - a^*_{1(2)} I_{m-1}\right) \end{pmatrix}\right]^2 \\ &= \frac{1}{(m-1)} \operatorname{tr} \begin{pmatrix} 0 & d_1 \mathbf{\Omega}'_{12} \\ d_2 \mathbf{\Omega}_{12} & d_2\left(\Omega_{22} - a^*_{1(2)} I_{m-1}\right) \end{pmatrix}^2 \\ &= \frac{{a^*}_{(1,2)}}{\sqrt{2a^*_{2(1)}a^*_{2(2)}}} + \frac{1}{2}\left[\frac{1 - a^{*2}_{1(2)}}{a^*_{2(2)}}\right] \\ &= F_1 + F_2 \end{aligned}$$

It may be noted that $\eta^2 = 0$ if and only if H_2 holds, otherwise $\eta^2 > 0$. Thus, a test statistic based on a consistent estimator of η^2 can be proposed.

We consider tests based on the sample covariance matrix $S = n^{-1}V$, or equivalently on the $m \times m$ matrix $W = GVG' \sim W_m(\Omega, n)$, $\Omega = G\Sigma G'$, where G has the Helmert form described at (4), and W is partitioned to conform with the partition of Ω at (24).

The following results from Srivastava and Khatri (1979, p.80) hold whether $n < m$ or $n \geq m$:

$$\begin{aligned} (i) \quad & W_{2.1} = W_{22} - W_{11}^{-1}\boldsymbol{W}_{12}\boldsymbol{W}'_{12} \sim W_{m-1}(\Omega_{2.1}, n-1) \\ & \quad \text{is independently distributed of } (\boldsymbol{W}_{12}, W_{11}) \\ (ii) \quad & \boldsymbol{W}_{12} \text{ given } W_{11} \sim N_{m-1}(\boldsymbol{\beta} W_{11}, W_{11}\Omega_{2.1}) \ , \\ (iii) \quad & W_{11} \sim \Omega_{11}\chi_n^2 \ , \end{aligned}$$

where

$$\boldsymbol{\beta} = \Omega_{11}^{-1}\boldsymbol{\Omega}_{12}, \ \text{and} \ \Omega_{2.1} = \Omega_{22} - \Omega_{11}^{-1}\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}'_{12} \ .$$

We define

$$\hat{a}^*_{(1,2)} = \frac{1}{(n-1)(n+2)(m-1)}\left[\boldsymbol{W}'_{12}\boldsymbol{W}_{12} - \frac{1}{n}W_{11} \ \mathrm{tr}(W_{22})\right], \tag{28}$$

$$\hat{a}^*_{1(2)} = \frac{\mathrm{tr}(W_{22})}{n(m-1)}, \ \hat{a}^*_{1(1)} = \frac{W_{11}}{n(m-1)}, \tag{29}$$

$$\hat{a}^*_{2(1)} = \frac{W_{11}^2}{(n-1)(n+2)(m-1)}, \tag{30}$$

$$\hat{a}^*_{2(2)} = \frac{1}{(n-1)(n+2)(m-1)}\left[\mathrm{tr}(W_{22}^2) - \frac{1}{n}\{\mathrm{tr}(W_{22})\}^2\right] \ . \tag{31}$$

We propose the statistic

$$T_2 = \frac{n}{\sqrt{2}}\left(\hat{F}_1 + \hat{F}_2\right), \tag{32}$$

where

$$\hat{F}_1 = \frac{\hat{a}^*_{(1,2)}}{\sqrt{2\hat{a}^*_{2(1)}\hat{a}^*_{2(2)}}} = [(n-1)(n+2)(m-1)]^{1/2}\,\frac{\hat{a}^*_{(1,2)}}{\sqrt{2\hat{a}^*_{2(2)}W_{11}}} \tag{33}$$

$$,\hat{F}_2 = \frac{1}{2}\left(1 - \frac{\hat{a}^{*2}_{1(2)}}{\hat{a}^*_{2(2)}}\right) \ , \tag{34}$$

for testing the hypothesis H_2 against the alternative A_2. The statistic T_2 is invariant under the transformation:

$$W \to \begin{pmatrix} c_1 & \mathbf{0} \\ \mathbf{0} & c_2 I_{m-1} \end{pmatrix} W \begin{pmatrix} c_1 & \mathbf{0} \\ \mathbf{0} & c_2 I_{m-1} \end{pmatrix}.$$

Hence, without any loss of generality, we may assume that the matrix $\Omega = I$ when obtaining the distribution of T_2 under the hypothesis H_2 and calculating its average significance level (ASL) or power; see the discussion of Tables 3 and 4 below.

3.2. Asymptotic null distribution of T_2

Under $H_2, W \sim W_m(\Omega, n)$, where $\Omega = \mathrm{diag}(\lambda^2, \sigma^2 I_{m-1})$. Hence, we can write $W = (\boldsymbol{z}_1, Z_2)'(\boldsymbol{z}_1, Z_2)$, and

$$W = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m)'(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m) = \begin{pmatrix} W_{11} & \boldsymbol{W}'_{12} \\ \boldsymbol{W}_{12} & W_{22} \end{pmatrix}, \tag{35}$$

where $\boldsymbol{z}_i$ are independently distributed with $\boldsymbol{z}_1 \sim N_n(\mathbf{0}, \lambda I_n)$ and $\boldsymbol{z}_2, \ldots, \boldsymbol{z}_m \sim N_n(\mathbf{0}, \sigma^2 I_n)$. Also $W_{11} = \boldsymbol{z}'_1\boldsymbol{z}_1$, $\boldsymbol{W}'_{12} = \boldsymbol{z}'_1 Z_2$, $W_{22} = Z'_2 Z_2$. Hence,

$$\begin{aligned} \frac{n\hat{a}^*_{(1,2)}}{\sqrt{\hat{a}^*_{2(1)}}} &= \frac{1}{(m-1)^{1/2}(\boldsymbol{z}'_1\boldsymbol{z}_1)}\left[\boldsymbol{z}'_1 Z_2 Z'_2 \boldsymbol{z}_1 - \frac{1}{n}(\boldsymbol{z}'_1\boldsymbol{z}_1)\mathrm{tr}(Z_2 Z'_2)\right] \\ &= \frac{\sigma^2}{(m-1)^{1/2}}\sum_{j=2}^{m}\left[\frac{(\boldsymbol{z}'_1\boldsymbol{z}_j)^2}{\sigma^2(\boldsymbol{z}'_1\boldsymbol{z}_1)} - \frac{1}{n\sigma^2}(\boldsymbol{z}'_j\boldsymbol{z}_j)\right] \end{aligned}$$

By the law of large numbers, $(n\sigma^2)^{-1}(\boldsymbol{z}'_j\boldsymbol{z}_j) \xrightarrow{p} 1$ as $n \to \infty$. Given $\boldsymbol{z}_1, \boldsymbol{z}'_1\boldsymbol{z}_j/\sigma(\boldsymbol{z}'_1\boldsymbol{z}_1)^{1/2}$ is standard normal, so

$$\left[(\boldsymbol{z}'_1\boldsymbol{z}_j)^2/\sigma^2(\boldsymbol{z}'_1\boldsymbol{z}_1)\right]$$

is distributed as χ^2_1, independently of $\boldsymbol{z}_1$. From Slutzky's theorem and the central limit theorem,

$$\begin{aligned} \frac{n\hat{a}^*_{(1,2)}}{\sigma^2\sqrt{2\hat{a}^*_{2(1)}}} &= \frac{1}{(m-1)^{1/2}}\sum_{j=2}^{m}\frac{1}{\sqrt{2}}\left[\frac{(\boldsymbol{z}'_1\boldsymbol{z}_j)^2}{\sigma^2(\boldsymbol{z}'_1\boldsymbol{z}_1)} - \frac{1}{n\sigma^2}(\boldsymbol{z}'_j\boldsymbol{z}_j)\right] \\ &\to N(0,1)\ , \end{aligned} \tag{36}$$

as $(m, n) \to \infty$. A consistent estimator of σ^2 is given by $\hat{a}^{*1/2}_{2(2)}$. Hence, we get the following theorem.

Theorem 3.1. *Let $W \sim W_m(\Omega, n)$, where $\boldsymbol{\Omega}_{12} = \mathbf{0}$, $\Omega_{22} = \sigma^2 I_{m-1}$, and Ω is partitioned as in (24). Then, under the hypothesis H_2 and the assumption (B), $n\hat{F}_1$ defined in (33) is asymptotically distributed as $N(0,1)$ as $(m,n) \to \infty$:*

$$\lim_{(m,n)\to\infty} P_0(n\hat{F}_1 \le f_1) = \Phi(f_1) \ ,$$

where P_0 denotes the distribution under the hypothesis H_2.

Since $\lim_{m\to\infty} \lambda^2/m = \tau^2\rho < \infty$, we have the following Corollary.

Corollary 3.1. *As $(m,n) \to \infty$, the limiting distribution of $n\hat{a}^*_{(1,2)}$ under H_2 is $N(0, 2a^*_{2(1)}a^*_{2(2)})$.*

Next, we obtain the asymptotic normality of $\hat{F}_2$. It may be noted that $\hat{F}_2$ is invariant under scale transformation of the observation vectors and thus we shall assume without loss of generality that $\boldsymbol{z}_2, \dots, \boldsymbol{z}_m$ are iid $N_n(\mathbf{0}, I_n)$. Now, from the definition of $\hat{a}^*_{2(2)}$, we have

$$\begin{aligned} \hat{a}^*_{2(2)} &= \frac{1}{(n-1)(n+2)(m-1)}\left[\mathrm{tr}\ (W_{22}^2) - \frac{1}{n}\{\mathrm{tr}(W_{22})\}^2\right] \qquad (37) \\ &= \frac{1}{(n-1)(n+2)(m-1)}\left[\sum_{j=2}^m (\boldsymbol{z}_j'\boldsymbol{z}_j)^2 + 2\sum_{2\le k<l}^m (\boldsymbol{z}_k'\boldsymbol{z}_l)^2 - \right. \\ &\qquad \left. -\frac{1}{n}\sum_{j=2}^m (\boldsymbol{z}_j'\boldsymbol{z}_j)^2 - \frac{2}{n}\sum_{2\le k<l}^m (\boldsymbol{z}_k'\boldsymbol{z}_k)(\boldsymbol{z}_l'\boldsymbol{z}_l)\right] = Q_1 + Q_2, \text{say,} \end{aligned}$$

where

$$Q_1 = \frac{n-1}{n(n-1)(n+2)(m-1)}\sum_{j=2}^m \left(\boldsymbol{z}_j'\boldsymbol{z}_j\right)^2, \qquad (38)$$

$$Q_2 = \frac{2}{(n-1)(n+2)(m-1)}\sum_{2\le k<l}^m \left[(\boldsymbol{z}_k'\boldsymbol{z}_l)^2 - \frac{1}{n}(\boldsymbol{z}_k'\boldsymbol{z}_k)(\boldsymbol{z}_l'\boldsymbol{z}_l)\right], \qquad (39)$$

and

$$\begin{aligned} \mathrm{E}(Q_1) &= 1, \quad \mathrm{Var}(Q_1) \simeq 8/(nm), \\ \mathrm{E}(Q_2) &= 0, \quad \mathrm{Var}(Q_2) \simeq 4/n^2. \end{aligned}$$

It follows from the central limit theorem that as $(m,n) \to \infty$

$$\sqrt{mn}\left(\frac{Q_1}{\sigma^4} - 1\right) \xrightarrow{d} N(0,8),$$

where now we give the result for general σ^2.

To find the distribution of Q_2, let

$$\eta_j = \frac{2}{n(m-1)} \sum_{i=2}^{j-1} \left[(\boldsymbol{z}_i' \boldsymbol{z}_j)^2 - \frac{1}{n} (\boldsymbol{z}_i' \boldsymbol{z}_i)(\boldsymbol{z}_j' \boldsymbol{z}_j) \right], \; j = 3, \ldots, m-1 \tag{40}$$

Then

$$E(\eta_j | \mathcal{F}_{j-1}) = 0, \text{and } \mathrm{E}(\eta_j^2 | \mathcal{F}_{j-1}) < \infty \;.$$

where $\mathcal{F}_j$ is the σ-algebra generated by the random vectors $\boldsymbol{z}_2, \ldots, \boldsymbol{z}_j$. Letting $\boldsymbol{z}_1 = \mathbf{0}$, and $\mathcal{F}_1 = (\emptyset, \mathcal{X})$, where Φ is the empty set, and $\mathcal{X}$ is the whole space, we find that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset, \ldots, \subset \mathcal{F}_m \subset \mathcal{F}$, and $\{\eta_j, \mathcal{F}_j\}$ is a sequence of integrable martingale differences. We note that

$$nQ_2 \simeq \sum_{j=3}^{m} \eta_j \;. \tag{41}$$

We need to show that the Lindeberg condition

$$L = \sum_{j=3}^{m} E\left[\eta_j^2 I(|\eta_j| > \varepsilon) \mid \mathcal{F}_{j-1}\right] \xrightarrow{p} 0$$

is satisfied. From Markov's inequality and the Cauchy-Schwarz inequality, as in the Appendix, we have

$$P(L > \xi) \leq \sum_{j=3}^{m} E\left(\eta_j^4\right) / \varepsilon^2 \xi.$$

As in §2, write

$$u_{ij} = (\boldsymbol{z}_i' \boldsymbol{z}_j)^2 - \frac{1}{n} (\boldsymbol{z}_i' \boldsymbol{z}_i) \left(\boldsymbol{z}_j' \boldsymbol{z}_j\right).$$

Then, it can be shown that

$$n^4(m-1)^4 \sum_{j=3}^{m} E\left(\eta_j^4\right) = 16 \sum_{j=3}^{m} E\left(\sum_{i=2}^{j-1} u_{ij}^4 + 6 \sum_{2 \leq k < l}^{j-1} u_{kj}^2 u_{il}^2 \right) = O\left(m^3 n^4\right).$$

Thus, the Lindeberg condition is satisfied. We now show that

$$M = \sum_{j=3}^{m} E(\eta_j^2 | \mathcal{F}_{j-1}) \xrightarrow{p} 4, \quad \text{and } \mathrm{Var}(M) \to 0.$$

The variance of M is

$$\mathcal{V}^2 = \mathrm{Var}\left[\frac{4}{n^2(m-1)^2} \sum_{j=3}^{m} \left(\sum_{i=2}^{j-1} b_{in}^{(j)} + 2 \sum_{2 \le k < l}^{j-1} c(j)_{kln} \right)\right] ,$$

where

$$b_{in}^{(j)} = E(u_{ij}^2 \mid \mathcal{F}_{j-1}),\ c_{kln}^{(j)} = E(u_{kl} u_{lj} \mid \mathcal{F}_{j-1}).$$

It can be shown that

$$E\left[\sum_{j=3}^{m} \mathrm{E}(\eta_j^2 \mid \mathcal{F}_{j-1})\right] = \sum_{j=3}^{m} \mathrm{E}(\eta_j^2) \simeq 4\ .$$

As well,

$$\mathrm{Var}\left[\frac{4}{n^2(m-1)^2} \sum_{j=3}^{m} \sum_{i=2}^{m-1} b_{in}^{(j)}\right] = O(m^{-1}n^{-2})\ ,\ \text{and}$$

$$\mathrm{Var}\left[\frac{8}{n^2(m-1)^2} \sum_{j=3}^{m} \sum_{2 \le k < l} c_{kln}^{(j)}\right] = O(m^{-1}n^{-2}),\ \text{so } \mathcal{V}^2 \to 0.$$

Hence, from Theorem 4 of Shiryayev (1984), as $(m,n) \to \infty$, the limiting distribution of nQ_2 is $N(0,4)$. Next, we consider the joint distribution of $\hat{a}_{1(2)}^*$ and Q_1, where

$$\hat{a}_{1(2)}^* = \frac{\sum_{j=2}^{m} (\boldsymbol{z}_j' \boldsymbol{z}_j)}{n(m-1)} \quad \text{and} \quad Q_1 \stackrel{d}{=} \frac{n-1}{n^2(m-1)} \sum_{j=2}^{m} (\boldsymbol{z}_j' \boldsymbol{z}_j)^2\ .$$

As before, σ^2 will be assumed to be one. Let $\varepsilon_{1i} = (\boldsymbol{z}_i' \boldsymbol{z}_i - n)\ / \sqrt{n}, \varepsilon_{2i} = [(\boldsymbol{z}_i' \boldsymbol{z}_i)^2 - n(n+2)]\ /\ \sqrt{n(n+2)(n+3)}\ ,\ i = 2, ..., m$. Then $\mathrm{E}(\varepsilon_{1i}) = 0$, $\mathrm{Var}(\varepsilon_{1i}) = 1$, $\mathrm{E}(\varepsilon_{2i}) = 0$, $\mathrm{Var}(\varepsilon_{2i}) = 1$, and $\mathrm{Cov}(\varepsilon_{1i}, \varepsilon_{2i}) = 4\delta_n$, $\delta_n = \sqrt{(n+2)/(n+3)}$. The bivariate random vectors $(\varepsilon_{1i}, \varepsilon_{2i})'$ are independent and identically distributed with mean vector $\mathbf{0}$, and finite covariance matrix,

$i = 2, \ldots, m$. Hence, from the multivariate central limit theorem, it follows that as $(m, n) \to \infty$, in any manner,

$$\sqrt{mn} \begin{pmatrix} \hat{a}^*_{1(2)} \\ Q_1 \end{pmatrix} \xrightarrow{d} N_2 \left[\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix} \right]$$

It can easily be shown that $\mathrm{Cov}(\hat{a}^*_{1(2)}, Q_2) = 0$. Now we apply Lemma A.1 in the Appendix to conclude that $\hat{a}^*_{(1,2)}, \hat{a}^*_{1(2)}$and $\hat{a}^*_{2(2)}$ defined in (3.6) – (3.9) are jointly normal. From this, it follows that $\hat{a}^*_{(1,2)}$ and $(\hat{a}^*_{1(2)}, \hat{a}^*_{2(2)})$ are asymptotically independently distributed under H_2. Since $\hat{a}^*_{2(2)} \xrightarrow{p} \sigma^4$ and $\hat{a}^*_{2(1)} \xrightarrow{p} \lambda^2$, it follows that $\hat{F}_1$ and $\hat{F}_2$ are asymptotically independently distributed. To find the distribution of $\hat{F}_2$, we apply the delta method to the joint distribution of $\hat{a}^*_{1(2)}$ and $\hat{a}^*_{2(2)}$, using

$$\frac{\partial F_2}{\partial \hat{a}^*_{1(2)}} = \frac{2\hat{a}^*_{1(2)}}{\hat{a}^*_{2(2)}} \ , \text{ and } \frac{\partial F_2}{\partial \hat{a}^*_{2(2)}} = -\frac{\hat{a}^{*2}_{1(2)}}{\hat{a}^{*2}_{2(2)}} \ ,$$

and

$$(2, -1) \begin{pmatrix} \frac{2}{nm} & \frac{4}{nm} \\ \frac{4}{nm} & \frac{8}{nm} + \frac{4}{n^2} \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \left(0, -\frac{4}{n^2}\right) \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \frac{4}{n^2}.$$

Hence, as $(m, n) \to \infty$, $(n/2)\hat{F}_2 \xrightarrow{d} N(0, 1)$, and we have the following:

Theorem 3.2. *Let $W \sim W_m(\Omega, n)$, where $\mathbf{\Omega}_{12} = 0$, $\Omega_{22} = \sigma^2 I_{m-1}$, $\Omega_{11} = \lambda^2$, and $\lim_{m\to\infty}(\lambda^2/m) < \infty$. Then under H_2 and assumption B, $T_2 \xrightarrow{d} N(0, 1)$, as m and $n \to \infty$.*

3.3. Power of the test T_2 and its attained significance level

As in §2, we examine attained significance level (ASL) first. Since the statistic T_2 is invariant under scale transformations of the first component and the remaining $(m - 1)$ components, we shall assume without loss of generality that $\Omega = G\Sigma G' = I_m$. For the alternative, we consider $\Omega = DRD$, $D = \mathrm{diag}(d_1, \ldots, d_m), d_i = 2 + (m - i + 1)/m$, $R = (r_{ij})$, where $r_{ii} = 1, r_{ij} = (-1)^{i+j}(\rho)^{|i-j|^{0.1}}$. . The ASL and power are defined in the same manner as in §2.3.

We compare the performance of T_2 with that of the likelihood ratio statistic

$$\lambda^* = \frac{|W_{2.1}|}{[\mathrm{tr}(W_{22})/(m-1)]^{m-1}} \ ,$$

given by Wilks (1946). The asymptotic distribution as $n \to \infty$ can be obtained from a general result of Box (1949). Let

$$\tilde{Q} = -[(n-1) - m(m+1)^2(2m-3)/\{6(m-1)(m^2+m-4)\}] \log \lambda^*.$$

Then

$$LR_1 = \tilde{Q} \sim \chi^2_g \ , \quad g = \frac{1}{2}m(m+1) - 2,$$

and

$$LR_2 = \frac{\tilde{Q} - g}{\sqrt{2g}} \xrightarrow{d} N(0,1) \ .$$

The likelihood ratio statistic is not invariant under the group of transformations (27), although it is invariant under the smaller group of transformations

$$\mathbf{x} \to \begin{pmatrix} c & \mathbf{0}' \\ \mathbf{0} & cG_{m-1} \end{pmatrix} \mathbf{x} \ .$$

The test based on T_2 has better ASL and power than the likelihood ratio test, even when $m < N$. In Table 4 we computed the percentage points by simulation, as in Table 3, but with $\lambda^2 = 10$ and $\sigma^2 = 2$, to demonstrate the fact noted at the end of §3.1 that the results are the same whether or not we impose the assumption $\Omega = I$.

4. Example

For illustration we applied the proposed test statistics to a microarray dataset, which has expression levels for 6500 human genes, for 40 samples of colon tumour tissue and 22 samples of normal colon tissue. A selection of 2000 genes with highest minimal intensity across the samples was made in Alon et al. (1999), and we use these 2000 genes. Thus $m = 2000$ and there are 60 degrees of freedom for estimating the covariance matrix. These data are publicly available at `http://www.molbio.princeton.edu/colondata`. The expression levels have been transformed by $\log_{10}$ transformation.

The description of the datasets and preprocessing are due to Dettling and Buhlmann (2002), except that we do not standardize each tissue sample to have zero mean and unit variance across genes, as it may invalidate our normality assumptions, and is not necessary. The preprocessed datasets were obtained from Professor Tatsuya at `http://www.tatsuya.e.u-tokyo.ac.jp/data1/colon_xtr`.

Table 3: Attained significance level and attained power of the test of intraclass correlation based on T_2 given in (32), compared to two versions of the likelihood ratio test. The covariance matrix under H_2 is the identity. The likelihood ratio test can only be used when $m < N$. The number of simulations is 10,000.

		ASL Under H			Power (ρ=0.4)	
N	m	T_2	LR_1	LR_2	T_2	LR_1
15	5	0.0408	0.0287	0.0262	0.6209	0.4866
	20	0.0457	—	—	0.9513	—
	50	0.0450	—	—	0.9934	—
	75	0.0464	—	—	0.9971	—
	100	0.0464	—	—	0.9988	—
	200	0.0467	—	—	0.9999	—
25	5	0.0404	0.0382	0.0345	0.8985	0.8174
	20	0.0469	0.1683	0.1251	0.9988	0.9370
	50	0.0470	—	—	0.9999	—
	75	0.0447	—	—	1	—
	100	0.0448	—	—	1	—
	200	0.0468	—	—	1	—
50	5	0.0419	0.0445	0.0398	0.9975	0.9944
	20	0.0533	0.0425	0.0504	1	1
	50	0.0492	—	—	1	—
	75	0.0461	—	—	1	—
	100	0.0493	—	—	1	—
	200	0.0474	—	—	1	—
100	5	0.0455	0.0457	0.0412	1	1
	20	0.0503	0.0415	0.0456	1	1
	50	0.0469	0.0922	0.0663	1	1
	75	0.0486	0.7705	0.6772	1	1
	100	0.0487	—	—	1	—
	200	0.0501	—	—	1	—

Table 4: Attained significance level and attained power of the test of intraclass correlation based on T_2 given in (32), compared to two versions of the likelihood ratio test. The covariance matrix under H_2 is the matrix at (24) with $\lambda^2 = 10$ and $\sigma^2 = 2$. The likelihood ratio test can only be used when $m < N$. The number of simulations is 10,000.

		ASL Under H			Power (ρ=0.2)	
N	m	T_2	LR_1	LR_2	T_2	LR_1
15	5	0.0368	0.0289	0.0261	0.1679	0.1245
	20	0.0446	—	—	0.4546	—
	50	0.0447	—	—	0.6695	—
	75	0.0429	—	—	0.7741	—
	100	0.0449	—	—	0.8163	—
	200	0.0474	—	—	0.9111	—
25	5	0.0380	0.0385	0.0350	0.3116	0.2456
	20	0.0447	0.1637	0.1288	0.7645	0.3122
	50	0.0483	—	—	0.9334	—
	75	0.0472	—	—	0.9708	—
	100	0.0438	—	—	0.9876	—
	200	0.0463	—	—	0.9975	—
50	5	0.0382	0.0442	0.0392	0.6664	0.5972
	20	0.0447	0.0409	0.0475	0.9912	0.9554
	50	0.0495	—	—	1	—
	75	0.0449	—	—	1	—
	100	0.0492	—	—	1	—
	200	0.0453	—	—	1	—
100	5	0.0412	0.0502	0.0434	0.9635	0.9449
	20	0.0506	0.0409	0.0479	1	1
	50	0.0507	0.0863	0.0615	1	1
	75	0.0492	0.7640	0.6741	1	1
	100	0.0494	—	—	1	—
	200	0.0451	—	—	1	—

Table 5: Tests of independence for the colon data set, based on T_1 defined at (12), for various values of m. Results based on T_1^* (not shown) were more extreme. The associated p-values are all essentially 0, since $T_1 \dot\sim N(0,1)$.

m_1	25	50	100	200	1000	1500	1900
T_1	24.958	26.402	30.098	32.883	39.613	36.655	28.730

The tests developed in §2 and §3 are for a sample from the same normal distribution, whereas the colon dataset has two sub-samples, from normal distributions with potentially different means. To accommodate this we use the pooled estimate of the covariance matrix

$$\hat\Sigma = (n_1 S_1 + n_2 S_2)/n,$$

where S_i is the sample covariance matrix of the ith group. The implicit assumption of a common covariance matrix was tested using the method given in Srivastava and Yanagihara (2010), and there was no evidence that the covariance matrices differed ($p = 0.5$). Consistent with the suggestion in Dudoit et al. (2002), we re-ordered the genes according the magnitude of the t-statistic for comparing the two groups. We then tested the independence of the first m_1, and the remaining m_2, genes: under independence there is no loss of power in retaining only the set of m_1 corresponding to the largest values of the t-statistic.

Table 5 shows the results of applying the test of independence, based on T_1. There is strong evidence against the hypothesis of independences of the first m_1 genes from the remaining $m_2 = m - m_1$, for a range of values of m_1. This implies that the second set of variables cannot be omitted, without losing power in testing, or the probability of correct classification in a discriminant analysis. Results obtained by applying T_1 separately to the tumor and normal classes are consistent with the conclusions of Table 5; the sub-vectors of differentially-expressed genes are not independent of the remaining set.

We also applied the test for intraclass correlation structure, based on T_2, to this dataset, both before, and after, re-ordering according to the magnitude of the m two-sample t-statistics. The test statistic took the values 26.5 before re-ordering, and 27.7 after re-ordering; thus there is strong evidence that the intraclass correlation model does not hold, and the method of false discovery rates should no be applied for this dataset.

5. Concluding Remarks

In this paper, we propose test statistics for testing independence, as well as for testing intraclass correlation structure, based on consistent estimators of the distance function between the hypothesis and the alternative. We have compared the attained significance level with the nominal level $\alpha = 0.05$. It seems that the asymptotic null distributions provide good approximations to the significance level, and the power of the tests are excellent. It may be noted that the proposed tests are valid for both $m < N$ and $m > N$, and can thus be recommended over the likelihood ratio test, which can only be used when $m < N$. Particularly when m is close to N, results in Tables 1 and 2 indicate that the likelihood ratio test can have very poor power.

Appendix

In §2 and §3 we used invariance arguments and a central limit theorem for independent but not identically distributed random variables. In this appendix we present these results in general notation.

Assume that the sample of n observations $\mathbf{x}, i = 1, \ldots, n$ are independent and identically distributed with mean $\mathbf{0}$ and positive definite covariance matrix Σ. Since $n < m$, the sample covariance matrix S as well as $V = nS$ are singular. Consider two sample points $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $X^* = (\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*)$ and let

$$Z = (X, X_1) \text{ and } Z^* = (X^*, X_1^*)$$

where X_1 and X_1^* are both $m \times (m - n)$ matrices of arbitrary values so that the $m \times m$ matrices Z and Z^* are nonsingular. Let I_r denote the $r \times r$ identity matrix. Then,

$$\begin{aligned} I_m &= (Z^*)^{-1} Z^* = (Z^*)^{-1}(X^*, X_1^*) \\ &= \left[(Z^*)^{-1} X^*, (Z^*)^{-1} X_1^*\right] = \begin{pmatrix} I_n & 0 \\ 0 & I_{m-n} \end{pmatrix} . \end{aligned}$$

Hence,

$$(Z^*)^{-1} X^* = \begin{pmatrix} I_n \\ 0 \end{pmatrix} , \text{ and } Z(Z^*)^{-1} X^* = (X, X_1) \begin{pmatrix} I_n \\ 0 \end{pmatrix} = X,$$

and $X = AX^*$, $A = Z(Z^*)^{-1}$, where A is nonsingular. Thus for any two points, there exists a nonsingular matrix taking one to the other; i.e. the whole space is a single orbit. This implies that the group of nonsingular transformations is transitive, and no invariant statistic exists.

For example, for testing the independence of two subvectors $\mathbf{x}_1$ and $\mathbf{x}_2$ where $\mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2')$, no invariant test exists under the nonsingular group of transformation

$$\mathbf{x} \to \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \mathbf{x} \ ,$$

where A_1 and A_2 are $m_1 \times m_1$ and $m_2 \times m_2$, $m_1 + m_2 = m$ are non singular matrices. For this reason we consider in §2 and §3 the more restricted group of transformations given at (2.1) and (3.5).

We now give a lemma to establish the joint asymptotic normality of the k statistics

$$t_{i,m}^{(n)} = \sum_{j=1}^{m} x_{ij}^{(n)}, \quad i = 1, ..., k.$$

where $x_{ij}^{(n)}$ is a sequence of random variables which may depend on n. We consider an arbitrary linear combination of these k statistics, namely,

$$t_m^{(n)} = c_1 t_{1,m}^{(n)} + ... + c_k t_{k,m}^{(n)} = \sum_{j=1}^{m} \sum_{i=1}^{k} c_i x_{ij}^{(n)} \equiv \sum_{j=1}^{m} y_j^{(n)}$$

where without any loss of generality, we assume that $c_1^2 + ... + c_k^2 = 1$. From the definition of multivariate normality, see Srivastava and Khatri (1979, p. 43), joint normality of $t_{im}^{(n)}, i = 1, ..., k$, will follow if the normality of $t_m^{(n)}$ is established for all $c_1, \ldots, c_k$. Let $\mathcal{F}_\ell^{(n)}$ be the σ-algebra generated by the random variables $(x_{1j}^{(n)}, \ldots, x_{kj}^{(n)})$, $j = 1, ..., \ell$, $\ell = 1, ..., m$. Then $\mathcal{F}_0 \subset \mathcal{F}_1^{(n)} \subset ... \subset \mathcal{F}_m^{(n)} \subset \mathcal{F}$, where $(\Omega, \mathcal{F}, P)$ is the probability space, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\emptyset$ is the null set and Ω is the whole space.

Lemma A.1 Let $x_{ij}^{(n)}$ be a sequence of random variables, $y_j^{(n)} = \sum_{i=1}^{k} c_i x_{ij}^{(n)}$, $i = 1, \ldots, k$, $j = 1, \ldots, m$, and $n = O(m^\delta)$, $\delta > 0$. We assume that

(i) $\mathrm{E}(y_j^{(n)} \mid \mathcal{F}_{j-1}^{(n)}) = 0$,

(ii) $\lim_{(n,m)\to\infty} E[(y_j^{(n)})^2] < \infty$,

(iii) $\sum_{j=0}^{m} E[(y_j^{(n)})^2 \mid \mathcal{F}_{j-1}^{(n)}] \xrightarrow{p} \sigma_0^2, \text{as}(n,m) \to \infty,$

(iv) $L \equiv \sum_{j=0}^{m} E[(y_j^{(n)})^2 \; I(|y_j^{(n)}| > \epsilon) \mid \mathcal{F}_{j-1}^{(n)}] \xrightarrow{p} 0, \text{ as } \; (n,m) \to \infty,$

Then

$$t_m^{(n)} = \sum_{j=1}^{m} y_j^{(n)} \xrightarrow{d} N(0, \sigma_0^2), \text{ as } (n,m) \to \infty.$$

The proof of this lemma follows from Theorem 4 of Shiryayev (1984, p. 511), since the first two conditions imply that $\{x_j^{(n)}, \mathcal{F}_j^{(n)}\}$ forms a sequence of integrable martingale differences. Condition (iv) is known as Lindeberg's condition. To verify this condition, we note that from the Markov and Cauchy-Schwarz inequalities

$$\begin{aligned} P[L > \delta] \;\; &\leq \;\; \sum_{j=0}^{m} E[(y_j^{(n)})^2 I(|y_j^{(n)}| > \epsilon]/\delta \\ &\leq \;\; \sum_{j=0}^{m} E[(y_j^{(n)})^4]/\delta\epsilon^2 \; . \end{aligned}$$

We also know that

$$E[(y_j^{(n)})^4] \leq k^3 \sum_{i=1}^{k} c_i^4 E[(x_{ij}^{(n)})^4].$$

Hence, if $\sum_{j=0}^{m} E[(x_{ij}^{(n)})^4] \to 0$, for all $i = 1, ..., k$, the Lindeberg condition is satisfied. It is rather simple to evaluate σ_0^2 in most cases.

References

Alon, U., Barkai, N. Motterman, D. Gish, K., Mack, S. and Levine, 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.* **96**, 6745–6750.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and power approach to multiple testing. *J. R. Statist. Soc.* B **57**, 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165–1181.

Box, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36**, 317–346.

Dettling, M. and Buhlmann, P. (2002). Boosting for tumor classification with gene expression data. *Bioinformatics* **19**, 1061–1069.

Dudoit, S., Fridlyand, J., and Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Statist. Assoc.* **97**, 77–87.

Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M. and Downing, J. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **296**, 531–537.

Schott, J.R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92**, 951–956.

Shiryayev, A.N. (1984). *Probability*. Springer-Verlag, New York.

Srivastava, M.S. (2005). Some tests concerning the covariance matrix in high-dimensional data. *J. Japan Stat. Soc.* **35**, 251–272.

Srivastava, M.S. (2006). Some test criteria for the covariance matrix with fewer observations than the dimension. *Acta et Commentations Universitatis Tartuensis de Mathematica* **10**, 77–93.

Srivastava, M.S., and Khatri, C.G. (1979). *An Introduction to Multivariate Statistics.* North-Holland, New York.

Srivastava, M.S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *J. Mult. Anal.* **101**, 1310–1329.

Székely, G.J. and Rizzo, M.L. (2009). Brownian distance covariance. *Ann. Appl. Statist.* **3**, 1236–1265.

Wilks, S.S.(1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Ann. Statist.* **17**, 257–281.