A test for the mean vector with fewer observations than the dimension

Article *in* Journal of Multivariate Analysis · March 2008

DOI: 10.1016/j.jmva.2006.11.002 · Source: RePEc

M. S. Srivastava
University of Toronto

Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 386–402
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

A test for the mean vector with fewer observations than the dimension

Muni S. Srivastava, Meng Du*

Department of Statistics, University of Toronto, Toronto, Canada M5S 3G3

Received 30 April 2006
Available online 14 December 2006

Abstract

In this paper, we consider a test for the mean vector of independent and identically distributed multivariate normal random vectors where the dimension p is larger than or equal to the number of observations N. This test is invariant under scalar transformations of each component of the random vector. Theories and simulation results show that the proposed test is superior to other two tests available in the literature. Interest in such significance test for high-dimensional data is motivated by DNA microarrays. However, the methodology is valid for any application which involves high-dimensional data.

AMS 1991 subject classification: Primary, 62H15;; secondary 62F05

Keywords: Asymptotic distribution; DNA microarray; Multivariate normal; Power comparison; Significance test

1. Introduction

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be independent and identically distributed (*iid*) p-dimensional random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. We assume that both $\boldsymbol{\mu}$ and Σ are unknown. We also assume that Σ is positive definite (denoted as $\Sigma > 0$). We shall consider the hypothesis testing problem:

$$H : \boldsymbol{\mu} = \mathbf{0} \quad vs. \quad A : \boldsymbol{\mu} \neq \mathbf{0}, \tag{1.1}$$

when the sample size N is smaller than and equal to the dimension p, that is, $N \leqslant p$.

* Corresponding author. Fax: +1 416 978 5133.
E-mail addresses: srivasta@utstat.utoronto.ca (M.S. Srivastava), meng@utstat.utoronto.ca (M. Du).

0047-259X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2006.11.002

Interest in the testing problem (1.1) when $N \leqslant p$ arises from DNA microarrays, where thousands of gene expression levels are measured on relatively few subjects. However, when the number N of available observations is smaller than the dimension p of the observed vectors, traditional testing procedures for $N > p$ are no longer valid. For example, the traditional Hotelling T^2-test is based on the statistic

$$T^2 = N\bar{\boldsymbol{x}}' S^{-1} \bar{\boldsymbol{x}},$$

where the sample mean vector $\bar{\boldsymbol{x}}$ and the sample covariance matrix S are defined, respectively, by

$$\bar{\boldsymbol{x}} = N^{-1} \sum_{i=1}^{N} \boldsymbol{x}_i, \quad \text{and} \quad S = n^{-1} \sum_{i=1}^{N} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})' = (s_{ij}), \tag{1.2}$$

where

$$n = N - 1. \tag{1.3}$$

Since, when $N \leqslant p$, the $p \times p$ matrix S is singular with rank $n < p$, S^{-1} does not exist. Thus, Hotelling T^2-test is not valid when $N \leqslant p$. We therefore look for tests which do not require the nonsingularity of the sample covariance matrix S.

Define the diagonal matrix of sample variances by

$$D_S = diag(s_{11}, \ldots, s_{pp}), \tag{1.4}$$

where $s_{11}, \ldots, s_{pp}$ are the diagonal elements of S defined in (1.2). Recall that the sample correlation matrix is defined by

$$R = D_S^{-\frac{1}{2}} S D_S^{-\frac{1}{2}} = (r_{ij}), \tag{1.5}$$

where r_{ij} is the sample correlation between the ith and jth components of the random vector based on N observations and $r_{ii} = 1,\ i = 1, \ldots, p$.

We propose a test for the testing problem (1.1), which is based on the test statistic

$$T_1 = \frac{N\bar{\boldsymbol{x}}' D_S^{-1} \bar{\boldsymbol{x}} - \dfrac{np}{n-2}}{\sqrt{2\left(\operatorname{tr} R^2 - \dfrac{p^2}{n}\right) c_{p,n}}}, \tag{1.6}$$

where the adjustment coefficient $c_{p,n} \to 1$ in probability as $(n, p) \to \infty$ and is chosen in order to improve the convergence of T_1 to its asymptotic distribution $N(0, 1)$ under the null hypothesis. One particular choice of $c_{p,n}$ that we have found works is given by

$$c_{p,n} = 1 + \frac{\operatorname{tr} R^2}{p^{3/2}}, \tag{1.7}$$

where R is the sample correlation matrix defined in (1.5).

It is noted that the T_1 test is an invariant test under the group of scalar transformations $\boldsymbol{x} \to D\boldsymbol{x}$, where $D = diag(d_1, d_2, \ldots, d_p)$ and $d_1, d_2, \ldots, d_p$ are nonzero constants. It may also be noted that when $N \leqslant p$, for the testing problem (1.1), there are no affine invariant tests other than the test $\phi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \equiv \alpha$, where α is the level of the test, see Lehmann [10, p. 318]. That is, when $N \leqslant p$ there is no nontrivial test that is invariant under a nonsingular linear transformation $\boldsymbol{x} \to A\boldsymbol{x},\ A \in Gl_p$, where Gl_p is the group of $p \times p$ nonsingular matrices. It is therefore desired to find tests that are invariant under a smaller group of transformations. Other tests available in the literature for $N \leqslant p$ include the tests proposed by Dempster [4,5] and Bai and Saranadasa [3], respectively. These two tests are both invariant under an orthogonal transformation $\boldsymbol{x} \to c\Gamma\boldsymbol{x}$, where c is a nonzero constant and Γ is a $p \times p$ orthogonal matrix, and they are therefore not invariant under the group of scalar transformations.

The asymptotic distribution of the test statistic T_1 under the hypothesis as well as under the alternative is given in Section 2, where the multivariate normality is assumed for the distribution of the *iid* random vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$. The proofs of the asymptotic properties of T_1 is given in Section 3. In Section 4, the tests proposed by Dempster [4,5] and Bai and Saranadasa [3] are given, respectively. Two-sample versions of the three tests as well as their power functions are given in Section 5. In Section 6, the power comparison of the proposed test with the other two tests is carried out theoretically as well as through simulations. Theories and simulation results show that the proposed test is superior to the other two tests under certain circumstances. Two examples of DNA microarrays are analyzed in Section 7. We conclude in Section 8.

2. Asymptotic distributions of T_1

From now on, we shall assume that the *iid* random vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ follow a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, denoted as $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N \overset{iid}{\sim} N_p(\boldsymbol{\mu}, \Sigma)$. In this section, we will give the asymptotic distribution of the test statistic T_1 defined in (1.6) under the null hypothesis and under the alternative, respectively. The proofs are rather technical and are separately given in Section 3.

Let $\Sigma = (\sigma_{ij})$. Define the diagonal matrix of variances by

$$D_\sigma = diag(\sigma_{11}, \ldots, \sigma_{pp}). \tag{2.1}$$

Recall that the population correlation matrix is defined by

$$\mathcal{R} = D_\sigma^{-\frac{1}{2}} \Sigma D_\sigma^{-\frac{1}{2}} = (\rho_{ij}), \tag{2.2}$$

where ρ_{ij} is the correlation between the ith and jth components of the random vector and $\rho_{ii} = 1,\ i = 1, \ldots, p$. Let $\lambda_{ip},\ i = 1, \ldots, p$, be the eigenvalues of the correlation matrix $\mathcal{R}$. When there is no confusion, we shall drop the subscript p and write λ_{ip} as λ_i. We shall assume that

$$0 < \lim_{p\to\infty} \frac{\operatorname{tr} \mathcal{R}^i}{p} < \infty, \quad i = 1, 2, 3, 4, \tag{2.3}$$

and

$$\lim_{p\to\infty} \max_{1 \leqslant i \leqslant p} \frac{\lambda_{ip}}{\sqrt{p}} = 0. \tag{2.4}$$

It is mentioned in Section 1 that the test T_1 is invariant under scalar transformations. Therefore, the distribution of T_1 depends on Σ through $\mathcal{R}$ and will not be affected by the values of $\sigma_{11}, \ldots, \sigma_{pp}$. The following theorem gives the asymptotic distribution of T_1 under the null hypothesis.

Theorem 2.1. *Assume that the conditions* (2.3) *and* (2.4) *hold. Also assume that*

$$n = O(p^{\zeta}), \quad \frac{1}{2} < \zeta \leqslant 1. \tag{2.5}$$

Then under the hypothesis that $\boldsymbol{\mu} = \mathbf{0}$,

$$\lim_{(n,p)\to\infty} P_0(T_1 \leqslant z_{1-\alpha}) = \Phi(z_{1-\alpha}), \tag{2.6}$$

where P_0 *denotes that the probability is being computed under the null hypothesis and* $\Phi(\cdot)$ *is the standard normal cumulative distribution function.*

It is noted that the condition (2.5) includes both the cases that $n \leqslant p,\ n/p \to c,\ 0 \leqslant c \leqslant 1$ and that $n > p$, but $n/p \to c,\ 1 \leqslant c < \infty$.

When $\boldsymbol{\mu} \neq 0$, we consider the local alternative

$$\boldsymbol{\mu} = \left(\frac{1}{nN}\right)^{\frac{1}{2}} \boldsymbol{\delta}, \tag{2.7}$$

where $\boldsymbol{\delta}$ is a vector of constants. We assume that

$$\frac{\boldsymbol{\delta}' D_{\sigma}^{-1} \boldsymbol{\delta}}{p} \leqslant M \quad \forall p, \tag{2.8}$$

where the constant M does not depend on p. The following theorem gives the asymptotic distribution of T_1 under the local alternative (2.7).

Theorem 2.2. *Assume that the conditions* (2.3)–(2.5) *hold. Under the local alternative* (2.7) *and under the condition* (2.8),

$$\lim_{(n,p)\to\infty} \left[P_1(T_1 > z_{1-\alpha}) - \Phi\left(-z_{1-\alpha} + \frac{N \boldsymbol{\mu}' D_{\sigma}^{-1} \boldsymbol{\mu}}{\sqrt{2 \operatorname{tr} \mathcal{R}^2}} \right) \right] = 0, \tag{2.9}$$

where P_1 *denotes that the probability is being computed under the alternative.*

Therefore, under the local alternative (2.7) and the conditions (2.3)–(2.5) and (2.8), the asymptotic power of test T_1 as $(n, p) \to \infty$ is given by

$$\beta(T_1 | \boldsymbol{\mu}) \simeq \Phi\left(-z_{1-\alpha} + \frac{N \boldsymbol{\mu}' D_{\sigma}^{-1} \boldsymbol{\mu}}{\sqrt{2 \operatorname{tr} \mathcal{R}^2}} \right). \tag{2.10}$$

The proofs of Theorems 2.1 and 2.2 are given in the next section.

3. Proofs of Theorems 2.1 and 2.2

We start with stating two lemmas.

Lemma 3.1 (*Srivastava [13,14, Lemma 2.1]*). *Let* $\{a_{im}\}_{i=1}^{m}$ *be a sequence of constants such that*

$$\lim_{m\to\infty} \max_{1 \leqslant i \leqslant m} a_{im}^2 = 0, \quad \textit{and} \quad \lim_{m\to\infty} \sum_{i=1}^{m} a_{im}^2 = 1.$$

Then for any iid random variables $x_1, \ldots, x_m$ *with mean zero and variance one,*

$$\sum_{i=1}^{m} a_{im} x_i \xrightarrow{d} N(0,1), \quad m \to \infty,$$

where '$\xrightarrow{d}$' *denotes* '*converge in distribution*'.

Lemma 3.2. *If* $n = O(p^{\zeta})$, $0 < \zeta \leqslant 1$, *under the condition* (2.3), $(\operatorname{tr} R^2 - p^2/n)/p$ *converges to* $\lim_{p\to\infty} \operatorname{tr} \mathcal{R}^2/p$ *in probability as* $(n,p) \to \infty$ *and thus can be considered as a consistent estimator of* $\operatorname{tr} \mathcal{R}^2/p$ *as* $(n,p) \to \infty$.

The proof of Lemma 3.2 is given in Appendix.

From Lemma 3.2, we have that, under the conditions (2.3) and (2.5),

$$\frac{\operatorname{tr} R^2}{p^{3/2}} = \frac{1}{p^{3/2}} \left(\operatorname{tr} R^2 - \frac{p^2}{n} \right) + \frac{\sqrt{p}}{n} \xrightarrow{p} 0, \quad (n,p) \to \infty$$

where '$\xrightarrow{p}$' denotes 'converge in probability'. This justifies that the adjustment coefficient $c_{p,n}$ defined in (1.7) goes to one in probability as $(n,p) \to \infty$ under the conditions (2.3) and (2.5). And hence, under the conditions (2.3) and (2.5) and as $(n,p) \to \infty$,

$$T_1 = \frac{N\bar{\boldsymbol{x}}' D_S^{-1} \bar{\boldsymbol{x}} - \dfrac{np}{n-2}}{\sqrt{2\left(\operatorname{tr} R^2 - \frac{p^2}{n}\right) c_{p,n}}} \stackrel{p}{=} \frac{N\bar{\boldsymbol{x}}' D_S^{-1} \bar{\boldsymbol{x}} - p}{\sqrt{2 \operatorname{tr} \mathcal{R}^2}}, \tag{3.1}$$

where '$a \stackrel{p}{=} b$' denotes '$a - b \xrightarrow{p} 0$'. Next we shall give the proof of Theorem 2.1.

Proof of Theorem 2.1. Let $D_{S/\sigma} = diag(s_{11}/\sigma_{11}, \ldots, s_{pp}/\sigma_{pp})$. Since for each $i = 1, \ldots, p$, $\sqrt{n/2}\,(s_{ii}/\sigma_{ii} - 1) \to N(0,1)$ in distribution as $n \to \infty$, let

$$D_{\frac{S}{\sigma}} = I_p + n^{-\frac{1}{2}} D_u, \quad \text{and} \quad D_{\frac{S}{\sigma}}^{-1} = [I_p + n^{-\frac{1}{2}} D_u]^{-1} = I_p - n^{-\frac{1}{2}} D_u + C, \tag{3.2}$$

where $D_u = diag(u_1, \ldots, u_p)$ and $C = diag(c_1, \ldots, c_p)$ with

$$u_i = \sqrt{n} \left(\frac{s_{ii}}{\sigma_{ii}} - 1 \right), \quad \text{and} \quad c_i = \left(\frac{s_{ii}}{\sigma_{ii}} \right)^{-1} + \frac{s_{ii}}{\sigma_{ii}} - 2, \ \ i = 1, \ldots, p. \tag{3.3}$$

Let $\boldsymbol{z} = \sqrt{N} D_\sigma^{-\frac{1}{2}} \bar{\boldsymbol{x}}$ and $\boldsymbol{y} = \mathcal{R}^{-\frac{1}{2}} \boldsymbol{z}$, respectively. Since $\bar{\boldsymbol{x}}$ is independently distributed of S, $\boldsymbol{z}$ and $\boldsymbol{y}$ are also independent of S and thus independent of D_u and C. When $\boldsymbol{\mu} = \mathbf{0}$, we have

$$\boldsymbol{z} \sim N_p(\mathbf{0}, \mathcal{R}) \quad \text{and} \quad \boldsymbol{y} \sim N_p(\mathbf{0}, I_p).$$

Then from (3.2),

$$
\begin{aligned}
N\bar{\boldsymbol{x}}' D_S^{-1}\bar{\boldsymbol{x}} &= \boldsymbol{z}' D_{S/\sigma}^{-1}\boldsymbol{z} \\
&\stackrel{d}{=} \boldsymbol{y}'\mathcal{R}^{\frac{1}{2}} D_{\frac{S}{\sigma}}^{-1}\mathcal{R}^{\frac{1}{2}}\boldsymbol{y} \\
&= \boldsymbol{y}'\mathcal{R}\boldsymbol{y} - n^{-\frac{1}{2}}\boldsymbol{y}'\mathcal{R}^{\frac{1}{2}} D_u \mathcal{R}^{\frac{1}{2}}\boldsymbol{y} + \boldsymbol{y}'\mathcal{R}^{\frac{1}{2}} C \mathcal{R}^{\frac{1}{2}}\boldsymbol{y}.
\end{aligned} \tag{3.4}
$$

Hence from (3.1) and (3.4), under the conditions (2.3) and (2.5)

$$
\begin{aligned}
&\lim_{(n,p)\to\infty} P_0(T_1 > z_{1-\alpha}) \\
&= \lim_{(n,p)\to\infty} P_0\left(\frac{\boldsymbol{y}'\mathcal{R}\boldsymbol{y} - p}{\sqrt{2\operatorname{tr}\mathcal{R}^2}} - \frac{(np)^{-\frac{1}{2}}\boldsymbol{y}'\mathcal{R}^{\frac{1}{2}} D_u \mathcal{R}^{\frac{1}{2}}\boldsymbol{y}}{\sqrt{2\operatorname{tr}\mathcal{R}^2/p}} + \frac{\boldsymbol{y}'\mathcal{R}^{\frac{1}{2}} C \mathcal{R}^{\frac{1}{2}}\boldsymbol{y}/\sqrt{p}}{\sqrt{2\operatorname{tr}\mathcal{R}^2/p}} > z_{1-\alpha}\right), \\
&= \lim_{(n,p)\to\infty} P_0(I - II + III > z_{1-\alpha}).
\end{aligned} \tag{3.5}
$$

Let $\mathcal{R} = \Gamma'\Lambda\Gamma$, where $\Gamma\Gamma' = I_p$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$ with $\lambda_1, \ldots, \lambda_p$ being the eigenvalues of $\mathcal{R}$. Then $\operatorname{tr}\mathcal{R} = \sum_{i=1}^p \lambda_i = p$ and $\operatorname{tr}\mathcal{R}^2 = \sum_{i=1}^p \lambda_i^2$. And also let $\boldsymbol{w} = (w_1, \ldots, w_p)' = \Gamma\boldsymbol{y} \sim N_p(\boldsymbol{0}, I_p)$. Thus $w_1^2, \ldots, w_p^2 \stackrel{iid}{\sim} \chi_1^2$ with $E[w_i^2] = 1$ and $\operatorname{Var}(w_i^2) = 2$. Then from Lemma 3.1, under the conditions (2.3) and (2.4), we have

$$
\begin{aligned}
I &= \frac{\boldsymbol{y}'\mathcal{R}\boldsymbol{y} - p}{\sqrt{2\operatorname{tr}\mathcal{R}^2}} = \frac{\boldsymbol{w}'\Lambda\boldsymbol{w} - p}{\sqrt{2\sum_{i=1}^p \lambda_i^2}}, \\
&= \frac{\sum_{i=1}^p \lambda_i (w_i^2 - 1)}{\sqrt{2\sum_{i=1}^p \lambda_i^2}} \xrightarrow{d} N(0, 1), \quad p \to \infty.
\end{aligned} \tag{3.6}
$$

To show that $II \xrightarrow{p} 0$, we first note that $\lim_{p\to\infty} \operatorname{tr}\mathcal{R}^2/p > 0$ under the condition (2.3) and the denominator $\sqrt{2\operatorname{tr}\mathcal{R}^2/p}$ of the term II is thus bounded below by a positive number. Therefore, it suffices to show that

$$
(np)^{-\frac{1}{2}}\boldsymbol{y}'\mathcal{R}^{\frac{1}{2}} D_u \mathcal{R}^{\frac{1}{2}}\boldsymbol{y} \xrightarrow{p} 0, \quad (n, p) \to \infty. \tag{3.7}
$$

From (3.3) we have

$$
E[u_i] = \sqrt{n}\left(E\left[\frac{s_{ii}}{\sigma_{ii}}\right] - 1\right) = 0, \quad \operatorname{Var}(u_i) = n\operatorname{Var}\left(\frac{s_{ii}}{\sigma_{ii}}\right) = 2. \tag{3.8}
$$

According to Anderson [2, p. 39], we also have

$$
\operatorname{Cov}(u_i, u_j) = n(\sigma_{ii}\sigma_{jj})^{-1}\operatorname{Cov}(s_{ii}, s_{jj}) = 2\rho_{ij}^2. \tag{3.9}
$$

Since $\boldsymbol{y}$ is independent of D_u, then from (3.8) and (3.9),

$$
E[(np)^{-\frac{1}{2}}\boldsymbol{y}'\mathcal{R}^{\frac{1}{2}} D_u \mathcal{R}^{\frac{1}{2}}\boldsymbol{y}] = E[E[(np)^{-\frac{1}{2}}\boldsymbol{y}'\mathcal{R}^{\frac{1}{2}} D_u \mathcal{R}^{\frac{1}{2}}\boldsymbol{y}|\boldsymbol{y}]] = 0, \tag{3.10}
$$

and

$$
\begin{aligned}
\operatorname{Var}[(np)^{-\frac{1}{2}}\boldsymbol{y}'\mathcal{R}^{\frac{1}{2}}D_u\mathcal{R}^{\frac{1}{2}}\boldsymbol{y}] &= (np)^{-1}E[(\boldsymbol{y}'\mathcal{R}^{\frac{1}{2}}D_u\mathcal{R}^{\frac{1}{2}}\boldsymbol{y})^2] \\
&= (np)^{-1}E[2\operatorname{tr}(\mathcal{R}D_u)^2 + (\operatorname{tr}\mathcal{R}D_u)^2] = (np)^{-1}E\left[2\sum_{i,j=1}^{p}\rho_{ij}^2 u_i u_j + \left(\sum_{i=1}^{p} u_i\right)^2\right] \\
&= (np)^{-1}E\left[3\sum_{i=1}^{p} u_i^2 + \sum_{i\neq j}(1+2\rho_{ij}^2)u_i u_j\right] = (np)^{-1}\left[6p + 2\sum_{i\neq j}(1+2\rho_{ij}^2)\rho_{ij}^2\right] \\
&= \frac{4}{n} + \frac{2}{n}\frac{\operatorname{tr}\mathcal{R}^2}{p} + \frac{4}{np}\sum_{i\neq j}\rho_{ij}^4,
\end{aligned}
\tag{3.11}
$$

where the second equation follows from

$$
E[(\boldsymbol{x}'A\boldsymbol{x})^2] = E[2\operatorname{tr}A^2 + (\operatorname{tr}A)^2], \tag{3.12}
$$

with $\boldsymbol{x} \sim N_p(\mathbf{0}, I_p)$ and A a symmetric matrix independent of $\boldsymbol{x}$, see for example, Schott [11, p. 394 Theorem 9.21]. Note that

$$
\sum_{i\neq j}^{p}\rho_{ij}^4 \leqslant \sum_{i=1}^{p}\sum_{j=1}^{p}\rho_{ij}^4 \leqslant \sum_{i=1}^{p}\left(\sum_{j=1}^{p}\rho_{ij}^2\right)^2 \leqslant \sum_{k=1}^{p}\sum_{i=1}^{p}\left(\sum_{j=1}^{p}\rho_{ij}\rho_{kj}\right)^2 = \operatorname{tr}\mathcal{R}^4. \tag{3.13}
$$

And also note that $\lim_{p\to\infty}\operatorname{tr}\mathcal{R}^i/p < \infty,\ i = 1, 2, 3, 4$, under the condition (2.3). Then we have under the condition (2.3),

$$
\operatorname{Var}[(np)^{-\frac{1}{2}}\boldsymbol{y}'\mathcal{R}^{\frac{1}{2}}D_u\mathcal{R}^{\frac{1}{2}}\boldsymbol{y}] \longrightarrow 0 \quad \text{as } (n, p) \to \infty, \tag{3.14}
$$

which implies (3.7) and thus $II \xrightarrow{p} 0$, as $(n, p) \to \infty$.

Similarly, we can show that $III \xrightarrow{p} 0$, as $(n, p) \to \infty$. From Slutsky's Theorem, this completes the proof of Theorem 2.1. □

Proof of Theorem 2.2. It follows from the proof of Theorem 2.1 that, under the conditions (2.3)–(2.5),

$$
\frac{N(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'D_S^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) - p}{\sqrt{2\operatorname{tr}\mathcal{R}^2}} \xrightarrow{d} N(0, 1) \quad (n, p) \to \infty. \tag{3.15}
$$

Under the local alternative (2.7) that $\boldsymbol{\mu} = \left(\frac{1}{nN}\right)^{\frac{1}{2}}\boldsymbol{\delta}$, where $\boldsymbol{\delta}$ satisfies the condition (2.8), we have

$$
\begin{aligned}
&\frac{1}{\sqrt{p}}[N(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'D_S^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})] \\
&\quad = \frac{1}{\sqrt{p}}N\bar{\boldsymbol{x}}'D_S^{-1}\bar{\boldsymbol{x}} - 2\sqrt{\frac{N}{np}}\boldsymbol{\delta}'D_S^{-1}\bar{\boldsymbol{x}} + \frac{1}{n\sqrt{p}}\boldsymbol{\delta}'D_S^{-1}\boldsymbol{\delta}.
\end{aligned}
\tag{3.16}
$$

Since $\bar{\boldsymbol{x}} \xrightarrow{p} \boldsymbol{\mu} = \left(\frac{1}{nN}\right)^{1/2} \boldsymbol{\delta}$ and $D_S \xrightarrow{p} D_\sigma$, it follows that

$$\sqrt{\frac{N}{np}} \boldsymbol{\delta}' D_S^{-1} \bar{\boldsymbol{x}} \xrightarrow{p} \frac{\sqrt{p}}{n} \frac{\boldsymbol{\delta}' D_\sigma^{-1} \boldsymbol{\delta}}{p}, \quad (n, p) \to \infty, \tag{3.17}$$

under the condition (2.8) when $n = O(p^\zeta)$, $\frac{1}{2} < \zeta \leqslant 1$. Hence,

$$\begin{aligned} &\frac{1}{\sqrt{p}}[N(\bar{\boldsymbol{x}} - \boldsymbol{\mu})' D_S^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu})] \\ &\stackrel{p}{=} \frac{1}{\sqrt{p}} N \bar{\boldsymbol{x}}' D_S^{-1} \bar{\boldsymbol{x}} - \frac{1}{n\sqrt{p}} \boldsymbol{\delta}' D_\sigma^{-1} \boldsymbol{\delta}, \quad (n, p) \to \infty. \end{aligned} \tag{3.18}$$

Hence under the conditions (2.3)–(2.5) and (2.8),

$$\begin{aligned} \frac{N \bar{\boldsymbol{x}}' D_S^{-1} \bar{\boldsymbol{x}} - \frac{1}{n} \boldsymbol{\delta}' D_\sigma^{-1} \boldsymbol{\delta} - p}{\sqrt{2 \operatorname{tr} \mathcal{R}^2}} &\stackrel{p}{=} \frac{N(\bar{\boldsymbol{x}} - \boldsymbol{\mu})' D_S^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}) - p}{\sqrt{2 \operatorname{tr} \mathcal{R}^2}} \\ &\xrightarrow{d} N(0, 1), \quad (n, p) \to \infty. \end{aligned} \tag{3.19}$$

Thus, we have that under the local alternative (2.7) and under the conditions (2.3)–(2.5) and (2.8),

$$\begin{aligned} &\lim_{(n,p)\to\infty} P_1(T_1 > z_{1-\alpha}) \\ &= \lim_{(n,p)\to\infty} P_1\left(\frac{N \bar{\boldsymbol{x}}' D_S^{-1} \bar{\boldsymbol{x}} - p}{\sqrt{2 \operatorname{tr} \mathcal{R}^2}} > z_{1-\alpha}\right) \\ &= \lim_{(n,p)\to\infty} P_1\left(\frac{N \bar{\boldsymbol{x}}' D_S^{-1} \bar{\boldsymbol{x}} - \frac{1}{n} \boldsymbol{\delta}' D_\sigma^{-1} \boldsymbol{\delta} - p}{\sqrt{2 \operatorname{tr} \mathcal{R}^2}} > z_{1-\alpha} - \frac{\boldsymbol{\delta}' D_\sigma^{-1} \boldsymbol{\delta}}{n\sqrt{2 \operatorname{tr} \mathcal{R}^2}}\right) \\ &= \lim_{(n,p)\to\infty} \Phi\left(-z_{1-\alpha} + \frac{\boldsymbol{\delta}' D_\sigma^{-1} \boldsymbol{\delta}}{n\sqrt{2 \operatorname{tr} \mathcal{R}^2}}\right). \end{aligned} \tag{3.20}$$

This completes the proof of Theorem 2.2. □

4. Other competing tests

For the hypothesis testing problem (1.1) when $N \leqslant p$, another test has been proposed by Dempster [4,5]. This test is based on the statistic

$$T_2 = \frac{N \bar{\boldsymbol{x}}' \bar{\boldsymbol{x}}}{\operatorname{tr} S}, \tag{4.1}$$

where $\bar{\boldsymbol{x}}$ is the sample mean vector and S is the sample covariance matrix defined in (1.2). Using an orthogonal transformation, it can be shown that under the hypothesis that $\boldsymbol{\mu} = \mathbf{0}$,

$$T_2 = \frac{n Q_1}{Q_2 + \cdots + Q_N}, \tag{4.2}$$

where Q_i, $i = 1, \ldots, N$, are independent and identically distributed. Dempster [4] assumed that Q_i, $i = 1, \ldots, N$, are approximately distributed as $m\chi_r^2$, where $m > 0$ and χ_r^2 is the chi-square distribution with r degrees of freedom. By equating the first two moments of Q_i with those of $m\chi_r^2$, it can be shown that

$$r = \frac{(\operatorname{tr}\Sigma)^2}{\operatorname{tr}\Sigma^2} = p\frac{a_1^2}{a_2}, \quad \text{where } a_i = \operatorname{tr}\Sigma^i/p, \quad i = 1, 2. \tag{4.3}$$

Let

$$\hat{a}_1 = \frac{\operatorname{tr} S}{p}, \quad \hat{a}_2 = \frac{n^2}{(n-1)(n+2)}\frac{1}{p}\left[\operatorname{tr} S^2 - \frac{(\operatorname{tr} S)^2}{n}\right] \quad \text{and} \quad \hat{r} = p\frac{\hat{a}_1^2}{\hat{a}_2}. \tag{4.4}$$

It can be noted that $\hat{r}$ is a ratio consistent estimator of r under certain conditions, see Lemma A.1. Then T_2 has an approximate F-distribution with $[\hat{r}]$ and $[n\hat{r}]$ degrees of freedom, where $[a]$ denotes the largest integer less than or equal to a.

Bai and Saranadasa [3] proposed another test for testing the hypothesis (1.1), which is given by

$$T_3 = \frac{N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}} - \operatorname{tr} S}{\left[\frac{2n(n+1)}{(n-1)(n+2)}\left(\operatorname{tr} S^2 - \frac{(\operatorname{tr} S)^2}{n}\right)\right]^{\frac{1}{2}}}. \tag{4.5}$$

Bai and Saranadasa [3] showed that when $\boldsymbol{\mu} = \mathbf{0}$, T_3 is asymptotically distributed as $N(0, 1)$ under the conditions stated in their paper. They have also shown that the asymptotic powers of the test statistics T_2 and T_3 are the same and given by

$$\beta(T_2|\boldsymbol{\mu}) \simeq \beta(T_3|\boldsymbol{\mu}) \simeq \Phi\left(-z_{1-\alpha} + \frac{N\boldsymbol{\mu}'\boldsymbol{\mu}}{\sqrt{2\operatorname{tr}\Sigma^2}}\right). \tag{4.6}$$

It may be noted that the same asymptotic powers of T_2 and T_3 can also be obtained under conditions similar to (2.3) and (2.4) and the local alternative (2.7).

It is also noted that both T_2 and T_3 tests are invariant under the transformations $\boldsymbol{x}_i \to c\Gamma\boldsymbol{x}_i$, $c \neq 0,\ \Gamma\Gamma' = I_p,\ i = 1, \ldots, N$.

5. Two-sample case

Let $\boldsymbol{x}_{ij},\ j = 1, \ldots, N_i,\ i = 1, 2$, be independent p-dimensional multivariate normal random vectors with mean vectors $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{ip})',\ i = 1, 2$, and unknown common covariance matrix $\Sigma = (\sigma_{ij}) > 0$. We shall assume that $\lim_{N_1+N_2\to\infty} N_1/(N_1 + N_2) = c \in (0, 1)$. Let the sample mean vectors $\bar{\boldsymbol{x}}_i$ and the pooled sample covariance matrix $S = (s_{ij})$ be defined, respectively, by

$$\bar{\boldsymbol{x}}_i = N_i^{-1}\sum_{j=1}^{N_i}\boldsymbol{x}_{ij}, \quad i = 1, 2, \quad \text{and} \quad S = n^{-1}\sum_{i=1}^{2}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)', \tag{5.1}$$

where

$$n = N_1 + N_2 - 2. \tag{5.2}$$

The two-sample versions of the test statistics T_1, T_2 and T_3 for the hypothesis

$$H : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs.} \quad A : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \tag{5.3}$$

are, respectively, given by

$$T_1 = \frac{\frac{N_1 N_2}{N_1+N_2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' D_S^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - \frac{np}{n-2}}{\left[2\left(\operatorname{tr} R^2 - \frac{p^2}{n}\right) c_{p,n}\right]^{\frac{1}{2}}}, \tag{5.4}$$

$$T_2 = \frac{\frac{N_1 N_2}{N_1+N_2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)}{\operatorname{tr} S}, \tag{5.5}$$

and

$$T_3 = \frac{\frac{N_1 N_2}{N_1+N_2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - \operatorname{tr} S}{\left[\frac{2n(n+1)}{(n-1)(n+2)}\left(\operatorname{tr} S^2 - \frac{(\operatorname{tr} S)^2}{n}\right)\right]^{\frac{1}{2}}}, \tag{5.6}$$

where the quantities D_S, R and $c_{p,n}$ are defined as in (1.4), (1.5) and (1.7), respectively, with S being replaced by the pooled sample covariance matrix as in (5.1).

Under the local alternative

$$\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1 = \left(\frac{N_1+N_2}{nN_1N_2}\right)^{\frac{1}{2}} \boldsymbol{\delta}, \tag{5.7}$$

where $\boldsymbol{\delta}$ is a vector of constants and satisfies the condition (2.8), the power functions of the three tests are given, respectively, by,

$$\beta(T_1|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \simeq \Phi\left(-z_{1-\alpha} + \frac{N_1N_2}{N_1+N_2}\frac{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)' D_\sigma^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)}{\sqrt{2 \operatorname{tr} \mathcal{R}^2}}\right), \tag{5.8}$$

and

$$\beta(T_2|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \simeq \beta(T_3|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \simeq \Phi\left(-z_{1-\alpha} + \frac{N_1N_2}{N_1+N_2}\frac{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)'(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)}{\sqrt{2 \operatorname{tr} \Sigma^2}}\right). \tag{5.9}$$

Remark 5.1. It may be noted that Fujikoshi et al. [7] have generalized the T_2 and T_3 tests to MANOVA case when $p/n \to c,\ c \in (0, \infty)$. A generalization of T_1 test to MANOVA case is under investigation.

6. Power comparison

6.1. Theoretical power comparison for independence case

To compare the power of the proposed test T_1 with those of other two tests T_2 and T_3, we shall consider the case that all the components of the random vector are independent, that is, $\Sigma = diag(\sigma_{11}, \ldots, \sigma_{pp})$ and $\mathcal{R} = I_p$. Assume that the local alternative defined in (2.7) satisfies the condition that

$$0 < \lim_{p\to\infty} \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{p} = \lim_{p\to\infty} \frac{\boldsymbol{\delta}' D_\sigma^{-1}\boldsymbol{\delta}}{\operatorname{tr} D_\sigma^{-1}} < \infty. \tag{6.1}$$

It is noted that the condition (6.1) is satisfied if $\boldsymbol{\delta} = (\delta, \ldots, \delta)'$, $\delta \neq 0$, that is, all the components of the random vector have the same mean values. From the Cauchy–Schwartz inequality, we have

$$p \sum_{i=1}^{p} \sigma_{ii}^2 \geqslant \left(\sum_{i=1}^{p} \sigma_{ii} \right)^2, \quad \text{and} \quad \left(\sum_{i=1}^{p} \sigma_{ii} \right) \left(\sum_{i=1}^{p} \frac{1}{\sigma_{ii}} \right) \geqslant p^2, \tag{6.2}$$

with strict inequalities unless $\sigma_{11} = \cdots = \sigma_{pp}$. It then follows that

$$\left(\frac{1}{p} \sum_{i=1}^{p} \frac{1}{\sigma_{ii}} \right)^2 \geqslant \left(\frac{p}{\sum_{i=1}^{p} \sigma_{ii}} \right)^2 \geqslant \frac{p}{\sum_{i=1}^{p} \sigma_{ii}^2} \tag{6.3}$$

with strict inequality unless $\sigma_{11} = \cdots = \sigma_{pp}$. Then under the condition (6.1) and with $\Sigma = diag(\sigma_{11}, \ldots, \sigma_{pp})$ and $\mathcal{R} = I_p$,

$$\begin{aligned} \frac{\boldsymbol{\delta}' D_\sigma^{-1}\boldsymbol{\delta}}{\sqrt{\operatorname{tr} \mathcal{R}^2}} &= \frac{\boldsymbol{\delta}' D_\sigma^{-1}\boldsymbol{\delta}}{\operatorname{tr} D_\sigma^{-1}} \frac{\sum_{i=1}^{p} \sigma_{ii}^{-1}}{\sqrt{p}} \simeq \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{p} \frac{\sum_{i=1}^{p} \sigma_{ii}^{-1}}{\sqrt{p}}, \quad p \to \infty \\ &\geqslant \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{p} \frac{p}{\sqrt{\sum_{i=1}^{p} \sigma_{ii}^2}} = \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{\sqrt{\operatorname{tr} \Sigma^2}}, \end{aligned}$$

where '$\geqslant$' is strict unless $\sigma_{11} = \cdots = \sigma_{pp}$. From (2.10) and (4.6), we have Theorem 6.1.

Theorem 6.1. *Assume that* $\Sigma = diag(\sigma_{11}, \ldots, \sigma_{pp}) > 0$ *and the conditions* (2.3)–(2.5) *and* (6.1) *hold. Under the local alternative* (2.7),

$$\beta(T_1|\boldsymbol{\mu}) \geqslant \beta(T_2|\boldsymbol{\mu}) \simeq \beta(T_3|\boldsymbol{\mu}) \tag{6.4}$$

with strict inequality unless $\sigma_{11} = \cdots = \sigma_{pp}$.

Corollary 6.1. *Assume that* $\Sigma = diag(\sigma_{11}, \ldots, \sigma_{pp}) > 0$ *and the conditions* (2.3)–(2.5) *hold. Under the local alternative* (2.7) *with* $\boldsymbol{\delta} = (\delta, \ldots, \delta)'$, $\delta \neq 0$,

$$\beta(T_1|\boldsymbol{\mu}) \geqslant \beta(T_2|\boldsymbol{\mu}) \simeq \beta(T_3|\boldsymbol{\mu}) \tag{6.5}$$

with strict inequality unless $\sigma_{11} = \cdots = \sigma_{pp}$.

Thus when $\Sigma = \sigma^2 \mathcal{I}_p > 0$, the three tests have the same asymptotic powers under the local alternative (2.7), although T_2 test is the uniformly most powerful test among all the tests which are invariant under the transformation $\boldsymbol{x} \to c\Gamma\boldsymbol{x}$, $c \neq 0$, $\Gamma'\Gamma = I_p$ and whose powers depend on $\boldsymbol{\mu}'\boldsymbol{\mu}/\sigma^2$ alone, see Hsu [9] and Simaika [12]. When $\Sigma = diag(\sigma_{11}, \ldots, \sigma_{pp}) \neq \sigma^2 \mathcal{I}_p$, T_1 test has higher asymptotic power than the other two tests under the conditions (2.3)–(2.5) and (6.1).

Because of the difficulty in analyzing the power functions of the three tests when Σ is not diagonal, we shall compare their powers through simulations for both independence and non-independence cases next.

6.2. Simulation study

6.2.1. Attained significance level and empirical power

To compare the three tests, we need to define the attained significance levels and the empirical powers. Let $z_{1-\alpha}$ be the $100(1-\alpha)\%$ quantile of the asymptotic null distribution of the test statistic T. For example, if $T = T_1$, $z_{1-\alpha}$ is the $100(1-\alpha)\%$ quantile of $N(0,1)$. With m replications of the data set simulated under the null hypothesis, the attained significance level is computed as

$$\hat{\alpha} = \frac{(\#\text{ of } t_H \geqslant z_{1-\alpha})}{m}, \tag{6.6}$$

where t_H represents the values of the test statistic T based on the data sets simulated under the null hypothesis. In our study, we choose $m = 10,000$ as the number of replications and fix the nominal significance level $\alpha = 0.05$. And hence, under the null hypothesis, as $(n, p) \to \infty$, $\hat{\alpha}$ is approximately distributed as *Binomial*$(10000, 0.05)$ and has the standard deviation estimated by $\hat{se}(\hat{\alpha}) = \sqrt{0.05 \times 0.95/10,000} \simeq 0.0022$.

To compute the empirical powers, we shall use the empirical critical points. Specifically, we first simulate m replications of the data set under the null hypothesis, then select the $(m\alpha)$th largest value of the test statistic as the empirical critical point, denoted as $\hat{z}_{1-\alpha}$, that is, the $100(1-\alpha)\%$ quantile of the empirical null distribution of the test statistic obtained from the m replications. Then another m replications of the data set are simulated under the alternative with the given choice of $\boldsymbol{\mu} \neq \mathbf{0}$. The empirical power is calculated by

$$\hat{\beta} = \frac{(\#\text{ of } t_A \geqslant \hat{z}_{1-\alpha})}{m}, \tag{6.7}$$

where t_A represents the values of the test statistic T based on the data sets simulated from the alternative and m has been chosen as 10,000 in our simulations.

6.2.2. Parameter selection and simulation result

We consider both independent correlation structures $\mathcal{R} = \mathcal{I}_p = diag(1, 1, \ldots, 1)$ and equal correlation structure $\mathcal{R} = \mathcal{R}_1 = (\rho_{ij}) : \ \rho_{ij} = 0.25, \ i \neq j$. We also consider different scalar matrix $D_\sigma = diag(\sigma_{11}, \ldots, \sigma_{pp})$. We select $D_\sigma = \mathcal{I}_p$, $D_\sigma = D_{\sigma,1} : \ \sigma_{11}^{1/2}, \ldots, \sigma_{pp}^{1/2} \overset{iid}{\sim} Unif(2,3)$ and $D_\sigma = D_{\sigma,2} : \ \sigma_{11}, \ldots, \sigma_{pp} \overset{iid}{\sim} \chi_3^2$. For the alternative hypothesis, we choose $\boldsymbol{\mu} = \boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)' : \ \nu_{2k-1} = 0$ and $\nu_{2k} \overset{iid}{\sim} Unif(-1/2, 1/2), \ k = 1, \ldots, p/2$.

Since T_2 and T_3 tests have the same asymptotic powers as discussed previously, in this section we only compare T_1 and T_2 tests. The attained significance levels of T_1 and T_2 are given in Table 1. It shows that the attained significance levels of both tests approximate the nominal level $\alpha = 0.05$ reasonably well in all cases. When $\mathcal{R} = \mathcal{I}_p$, the powers of the two tests are very close to each other, as shown in Table 2. However, when $\mathcal{R} = \mathcal{R}_1$ and $D_\sigma = D_{\sigma,1}$ and $D_{\sigma,2}$, the powers of T_1

Table 1
Attained significance levels of T_1 and T_2 under the null hypothesis, when $\mathcal{R} = \mathcal{I}_p$ and $\mathcal{R} = \mathcal{R}_1$, respectively

		$D_\sigma = I_p$		$D_\sigma = D_{\sigma,1}$		$D_\sigma = D_{\sigma,2}$	
p	N	T_1	T_2	T_1	T_2	T_1	T_2
$\mathcal{R} = I_p$							
60	30	0.056	0.051	0.056	0.049	0.056	0.052
100	40	0.048	0.054	0.048	0.054	0.048	0.052
	60	0.050	0.053	0.050	0.053	0.050	0.051
	80	0.050	0.054	0.050	0.054	0.050	0.054
150	40	0.050	0.054	0.050	0.053	0.050	0.051
	60	0.048	0.049	0.048	0.048	0.048	0.053
	80	0.049	0.050	0.049	0.051	0.049	0.053
200	40	0.045	0.052	0.045	0.052	0.045	0.051
	60	0.048	0.052	0.048	0.051	0.048	0.051
	80	0.045	0.047	0.045	0.050	0.045	0.051
400	40	0.037	0.046	0.037	0.046	0.037	0.054
	60	0.035	0.050	0.035	0.051	0.035	0.053
	80	0.044	0.050	0.044	0.049	0.044	0.046
$\mathcal{R} = \mathcal{R}_1$							
60	30	0.058	0.058	0.058	0.057	0.0582	0.059
100	40	0.053	0.062	0.053	0.063	0.0526	0.061
	60	0.045	0.062	0.045	0.061	0.0450	0.057
	80	0.046	0.057	0.046	0.057	0.0463	0.060
150	40	0.049	0.065	0.049	0.065	0.0493	0.063
	60	0.050	0.064	0.050	0.063	0.0502	0.061
	80	0.044	0.059	0.044	0.059	0.0441	0.062
200	40	0.049	0.067	0.049	0.068	0.0485	0.064
	60	0.044	0.061	0.044	0.062	0.0441	0.060
	80	0.047	0.062	0.047	0.063	0.0469	0.063
400	40	0.045	0.063	0.045	0.063	0.0448	0.066
	60	0.041	0.064	0.041	0.063	0.0408	0.059
	80	0.036	0.062	0.036	0.063	0.0362	0.058

Nominal significance level $\alpha = 0.05$.

are substantially better than those of T_2, which can be explained by the invariance property of T_1 under the scalar transformations.

7. Two examples

In this section, we apply the proposed test and the other two tests to two examples of DNA microarrays. The data sets are described next.

- *Colon data*: 2000 (p) gene expression levels are available on 22 (N_1) normal colon tissues and 40 (N_2) tumor colon tissues. ($N = 62 < p$) (http://microarray.princeton.edu/oncology/affydata/index.html; [1]).

Table 2
Empirical powers of T_1 and T_2 under the alternative hypothesis, when $\mathcal{R} = \mathcal{I}_p$ and $\mathcal{R} = \mathcal{R}_1$, respectively

		$D_\sigma = I_p$		$D_\sigma = D_{\sigma,1}$		$D_\sigma = D_{\sigma,2}$	
p	N	T_1	T_2	T_1	T_2	T_1	T_2
$\mathcal{R} = I_p$							
60	30	0.999	1.000	0.287	0.291	0.976	0.542
100	40	1.000	1.000	0.622	0.593	1.000	0.939
	60	1.000	1.000	0.880	0.861	1.000	0.998
	80	1.000	1.000	0.973	0.962	1.000	1.000
150	40	1.000	1.000	0.698	0.661	1.000	0.962
	60	1.000	1.000	0.932	0.913	1.000	1.000
	80	1.000	1.000	0.991	0.983	1.000	1.000
200	40	1.000	1.000	0.831	0.789	1.000	0.992
	60	1.000	1.000	0.979	0.967	1.000	1.000
	80	1.000	1.000	0.999	0.998	1.000	1.000
400	40	1.000	1.000	0.926	0.903	1.000	1.000
	60	1.000	1.000	0.998	0.995	1.000	1.000
	80	1.000	1.000	1.000	1.000	1.000	1.000
$\mathcal{R} = \mathcal{R}_1$							
60	30	0.789	0.948	0.088	0.014	0.513	0.112
100	40	0.990	1.000	0.125	0.013	0.936	0.290
	60	1.000	1.000	0.203	0.067	1.000	0.813
	80	1.000	1.000	0.303	0.261	1.000	0.973
150	40	0.980	1.000	0.115	0.003	0.974	0.143
	60	1.000	1.000	0.143	0.024	1.000	0.681
	80	1.000	1.000	0.245	0.142	1.000	0.958
200	40	0.984	1.000	0.108	0.000	0.944	0.148
	60	1.000	1.000	0.157	0.016	1.000	0.792
	80	1.000	1.000	0.249	0.096	1.000	0.985
400	40	0.933	1.000	0.085	0.000	0.814	0.018
	60	1.000	1.000	0.118	0.000	0.995	0.589
	80	1.000	1.000	0.194	0.003	1.000	0.984

Nominal significance level $\alpha = 0.05$.

- *Leukemia data*: 3571 (p) genes expressions are available from 47 (N_1) patients suffering from acute lymphoblastic leukemia (ALL) and 25 (N_2) patients suffering from acute myeloid leukemia (AML) ($N = 72 < p$) (http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi; [6,8]).

We apply tests T_1, T_2 and T_3 to these two data sets. The results are shown in Table 3. For the colon data, the p-values of the three tests equal to $1.378268e-06$, $4.478795e-11$ and 0.00000, respectively. Thus, all the three tests lead to the rejection of the hypothesis that the tumor group have the same gene expression levels as the normal group. The p-values of the three tests for the leukemia data are all zero and lead to the rejection of the hypothesis that the ALL group and the AML group have identically expressed genes.

Table 3
Observed p-values for testing the equality of the gene expressions for colon data and leukemia data, respectively

	T_1	T_2	T_3
Colon data			
Statistic	4.6882	4.5198	11.4231
p-value	1.3783e-06	4.4788e-11	0.0000
Leukemia data			
Statistic	17.0758	8.3358	37.7236
p-value	0.0000	0.0000	0.0000

8. Conclusion

When the covariance matrix $\Sigma = \sigma^2 I_p,\ \sigma^2 > 0$, it is known that Dempster's test T_2 is the uniformly most powerful test among all the tests which are invariant under the transformation $\boldsymbol{x} \to c\Gamma\boldsymbol{x},\ c \neq 0,\ \Gamma'\Gamma = I_p$ and whose powers depend on $\boldsymbol{\mu}'\boldsymbol{\mu}/\sigma^2$ alone. Thus when $\Sigma = \sigma^2 I_p,\ \sigma^2 > 0$, Dempster's test is superior to the proposed test T_1. However, from the theoretical comparison of the asymptotic power functions, it follows that for large p and N, even when $\Sigma = \sigma^2 I_p,\ \sigma > 0$, asymptotically the proposed test has the same power as Dempster's test, which has also been confirmed from our simulation results that, except for small p and N, the two tests have almost the same powers. For all the other cases in our simulation, the proposed test has better powers than Dempster's test. As the T_3 test, proposed by Bai and Saranadasa [3], has the same asymptotic power function as Dempster's test, the proposed test is thus also superior to Bai and Saranadasa's test, which is confirmed by additional simulation results available with the authors. From all these considerations, the proposed test performs best among all the three tests under the circumstances we have considered in this paper.

Appendix. Proof of Lemma 3.2

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \overset{iid}{\sim} N_p(\boldsymbol{\mu}, \Sigma)$, where $\Sigma = (\sigma_{ij}) > 0$. And assume that

$$0 < \lim_{p\to\infty} \frac{\operatorname{tr}\Sigma^i}{p} < \infty, \quad i = 1, 2, 3, 4. \tag{A.1}$$

Let $n = N - 1$ and $S = (s_{ij})$ be the sample covariance matrix of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ defined in (1.2). Define the following statistics

$$\hat{a}_1 = \frac{\operatorname{tr} S}{p}, \quad \text{and} \quad \hat{a}_2 = \frac{n^2}{(n-1)(n+2)} \frac{1}{p} \left[\operatorname{tr} S^2 - \frac{(\operatorname{tr} S)^2}{n} \right]. \tag{A.2}$$

Then we have the following lemma.

Lemma A.1 (*Srivastava [15, Lemma 2.1]*). *Let $\hat{a}_1$ and $\hat{a}_2$ be defined in* (A.2). *Under the condition* (A.1), *we have*

1. $E[\hat{a}_i] = \operatorname{tr}\Sigma^i/p,\ i = 1, 2.$
2. $\hat{a}_i \xrightarrow{p} \operatorname{tr}\Sigma^i/p,\ i = 1, 2,$ *as* $n \to \infty$. *And therefore, when* $n = O(p^\zeta),\ 0 < \zeta \leqslant 1$, $(\hat{a}_i - \operatorname{tr}\Sigma^i/p) \xrightarrow{p} 0,\ i = 1, 2,$ *as* $(n, p) \to \infty$.

Since $(\text{tr}\, R^2 - p^2/n)/p$ is invariant under the scalar transformations of each component of $\boldsymbol{x}_i,\ i = 1, \ldots, N$, we may as well assume that $\Sigma = \mathcal{R}$. Then the condition (A.1) becomes equivalent to the condition (2.3). Let

$$w_i = 1 - 1/s_{ii}, \quad i = 1, \ldots, p.$$

Then

$$E(w_i) = 1 - \frac{n}{n-2} = -\frac{2}{n-2}, \quad \text{and} \quad \text{Var}(w_i) = \text{Var}(s_{ii}^{-1}) = \frac{2n^2}{(n-2)^2(n-4)} = O(n^{-1}).$$

Hence $\forall \varepsilon > 0,\ w_i/n^{-1/2+\varepsilon} \xrightarrow{p} 0,\ n \to \infty$, uniformly for $i = 1, \ldots, p$, that is, $w_i = O_p(n^{-\frac{1}{2}})$, uniformly for $i = 1, \ldots, p$, where the subscript 'p' denotes 'in probability'. Let

$$D_S^{-1} = I - D_w,$$

where $D_w = diag(w_1, \ldots, w_p)$. Then

$$\begin{aligned} \frac{1}{p}\left[\text{tr}\, R^2 - \frac{p^2}{n}\right] &= \frac{1}{p}\left[\text{tr}(D_S^{-1}S)^2 - \frac{(\text{tr}\, D_S^{-1}S)^2}{n}\right] \\ &= \frac{1}{p}\left[\text{tr}\, S^2 - \frac{(\text{tr}\, S)^2}{n}\right] - \frac{2}{p}\left[\text{tr}\, D_w S^2 - \frac{\text{tr}\, S \text{tr}\, D_w S}{n}\right] \\ &\quad + \frac{1}{p}\left[\text{tr}(D_w S)^2 - \frac{(\text{tr}\, D_w S)^2}{n}\right]. \end{aligned} \tag{A.3}$$

Under the condition (A.1) and when $n = O(p^{\zeta}),\ 0 < \zeta \leqslant 1$, from Lemma A.1, we have

$$\frac{1}{p}\left[\text{tr}\, S^2 - \frac{(\text{tr}\, S)^2}{n}\right] - \frac{\text{tr}\, \mathcal{R}^2}{p} \xrightarrow{p} 0, \quad (n, p) \to \infty, \tag{A.4}$$

and

$$\lim_{(n,p)\to\infty} \frac{1}{p}\left[\text{tr}\, D_w S^2 - \frac{\text{tr}\, S \text{tr}\, D_w S}{n}\right] \tag{A.5}$$

$$\begin{aligned} &\overset{p}{=} \lim_{(n,p)\to\infty} \frac{1}{p}\left[\sum_{i=1}^{p} w_i \sum_{j=1}^{p} \left(s_{ij}^2 - \frac{s_{ii}s_{jj}}{n}\right) I(r_{ij}^2 > 1/n)\right] \\ &\overset{p}{=} \lim_{(n,p)\to\infty} O_p(n^{-\frac{1}{2}}) \frac{1}{p}\left[\text{tr}\, S^2 - \frac{(\text{tr}\, S)^2}{n}\right] = 0, \end{aligned} \tag{A.6}$$

where the convergence holds, since $[\text{tr}\, S^2 - (\text{tr}\, S)^2/n]/p - \text{tr}\, \mathcal{R}^2/p \xrightarrow{p} 0$ and $\lim_{p\to\infty} \text{tr}\, \mathcal{R}^2/p < \infty$ under the condition (A.1). Similarly, when $n = O(p^{\zeta}),\ 0 < \zeta \leqslant 1$ and under the condition (A.1),

$$\begin{aligned} \frac{1}{p}\left[\text{tr}(D_w S)^2 - \frac{(\text{tr}\, D_w S)^2}{n}\right] &= O_p(n^{-1}) \frac{1}{p}\left[\text{tr}\, S^2 - \frac{(\text{tr}\, S)^2}{n}\right] \\ &\xrightarrow{p} 0, \quad (n, p) \to \infty. \end{aligned} \tag{A.7}$$

Thus, under the condition (A.1),

$$\frac{1}{p}\left[\operatorname{tr} R^2 - \frac{p^2}{n}\right] - \frac{\operatorname{tr} \mathcal{R}^2}{p} \xrightarrow{p} 0, \quad (n, p) \to \infty,\ n = O(p^{\zeta}),\ 0 < \zeta \leqslant 1.$$

This completes the proof of Lemma 3.2. □

References

[1] U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, A.J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 6745–6750.
[2] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1958.
[3] Z. Bai, H. Saranadasa, Effect of high dimension: an example of a two sample problem, Statist. Sinica 6 (1996) 311–329.
[4] A.P. Dempster, A high dimensional two sample significance test, Ann. Math. Statist. 29 (1958) 995–1010.
[5] A.P. Dempster, A significance test for the separation of two highly multivariate small samples, Biometrics 16 (1960) 41–50.
[6] S. Dudoit, J. Fridlyand, T.P. Speed, Comparison of discrimination methods for the classification of tremors using gene expression data, J. Amer. Statist. Assoc. 97 (2002) 77–87.
[7] Y. Fujikoshi, T. Himeno, H. Wakaki, Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size, J. Japan Statist. Soc. 34 (2004) 19–26.
[8] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, E.S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537.
[9] P.L. Hsu, Analysis of variance from the power function standpoint, Biometrika 32 (1941) 62–69.
[10] E.L. Lehmann, Testing Statistical Hypotheses, Wiley, New York, 1959.
[11] J.R. Schott, Matrix Analysis for Statistics, Wiley, New York, 1997.
[12] J.B. Simaika, On an optimum property of two important statistical tests, Biometrika 32 (1941) 70–80.
[13] M.S. Srivastava, On a class of non-parametric tests for regression parameters, Ann. Math. Statist. (Abstract) 39 (1968) 697.
[14] M.S. Srivastava, Asymptotically most powerful rank tests for regression parameters in MANOVA, Ann. Inst. Statist. Math. Tokyo 24 (2) (1972) 285–297.
[15] M.S. Srivastava, Some tests concerning the covariance matrix in high-dimensional data, J. Japan Statist. Soc. 35 (2005) 251–272.