Statistica Sinica **6**(1996), 311-329

EFFECT OF HIGH DIMENSION: BY AN EXAMPLE OF A TWO SAMPLE PROBLEM

Zhidong Bai and Hewa Saranadasa

National Sun Yat-sen University

曾銮杰2020：
非精确检验

Abstract. With the rapid development of modern computing techniques, statisticians are dealing with data with much higher dimension. Consequently, due to their loss of accuracy or power, some classical statistical inferences are being challenged by non-exact approaches. The purpose of this paper is to point out and briefly analyze such a phenomenon and to encourage statisticians to reexamine classical statistical approaches when they are dealing with high dimensional data. As an example, we derive the asymptotic power of the classical Hotelling's T^2 test and Dempster's non-exact test for a two-sample problem. Also, an asymptotically normally distributed test statistic is proposed. Our results show that both Dempster's non-exact test and the new test have higher power than Hotelling's test when the data dimension is proportionally close to the within sample degrees of freedom. Although our new test has an asymptotic power function similar to Dempster's, it does not rely on the normality assumption. Some simulation results are presented which show that the non-exact tests are more powerful than Hotelling's test even for moderately large dimension and sample sizes.

Key words and phrases: Edgeworth expansion, Hotelling T^2 test, hypothesis test, power function, significance test, χ^2 approximation.

1. Introduction

Modern computation techniques make it possible to deal with high dimensional data. Some recent examples of interest in dealing with high dimensional data can be found in Narayanaswamy and Raghavarao (1991) and Saranadasa (1991, 1993). Examples may also be found in applied statistical inference handling samples of many measurements on individuals. For example, in a clinical trial of pharmaceutical studies, many blood chemistry measurements are measured on each individual. In some studies the number of variables is comparable to or even exceeds the total sample size. The purpose of this article is to raise the following questions: What's new in high dimensional statistical inference and what should be done? The difference of high dimensional statistical inference from that in classical statistical inference will be referred to as the "**Effect of High Dimension**" (EHD).

There are two aspects of the EHD. The first, there are too many interesting or nuisance parameters in the model. For example, in M-estimation in linear models, the number of regression parameters may be proportional to the sample size. This problem remains unsolved. The best results are due to Huber's work (1973) in which the consistency of estimation is proved under the assumption that $p^2/n \to 0$ and the asymptotic normality under $p^3/n \to 0$, where n and p are the sample size and the dimension of regression coefficent vector. Althogh these requirements on the ratio of the dimention to the sample size were reduced, very strong assumptions were made on the design sequence. References are made to Portnoy (1984,1985). Another example is the model of Error in Variables in which the true regressor variables can be considered as nuisance parameters whose number is np (while the number of observations is $n(p+1)$). In these cases, either the estimation is very poor or it is impossible to get an unbiased or consistent estimator. The second case is that the dimension itself of the data is very high. Of course, the number of parameters to be estimated must be very large. An example is the detection of the signal number in omni-directional signal processing. When the number of sensors are increased, the detection accuracy is supposed to be better. However, the simulation results show the opposite when the traditional method (the MUSIC method) is used if the number of sensors is 10 or more. We believe that the reason is that the number of elements of the covariance matrix (parametrs to be estimated) becomes very large ($2p^2$ and 200 if $p = 10$). Some references in this direction are Bai, Krishnaiah and Zhao (1989) and Zhao, Krishinaiah and Bai (1986a,b).

Although the EHD has been noticed in many different directions of multivariate statistical inferences, the problem has not yet been clearly stated in the literature and no appropriate methods have been proposed to deal with the EHD. To this end, we shall analyze these problems through the two sample problem, as an example to show how and why the EHD affects inferences and how the EHD can be reduced.

A classical method to deal with this problem is the famous Hotelling's T^2 test. Its advantages include: it is invariant under linear transformation, its exact distribution is known under the null hypothesis and it is powerful when the dimension of data is sufficiently small, compared with the sample sizes. However, Hotelling's test has the serious defect that the T^2 statistic is undefined when the dimension of data is greater than the within sample degrees of freedom.

Seeking remedies, Chung and Fraser (1958) proposed a nonparametric test and Dempster (1958, 1960) discussed the so-called "non-exact" significance test. Dempster (1960) also considered the so-called randomization test. These works seek alternatives to Hotelling's test in situations when the latter does not apply. Not only being a remedy when the T^2 is undefined, we show that even it is well defined, the non-exact test is more powerful than the T^2 test when the dimension is proportionally "close to" (more discussion on the ratio will be given in Section 5) the sample degrees of freedom.

Both the T^2 test and Dempster's non-exact test strongly rely on the normality assumption. Moreover, Dempster's non-exact test statistic involves a complicated estimation of r, the "degrees of freedom" for the chi-square approximation. To simplify the testing procedure, a new method is proposed in Section 4. It is proven in Sections 3 and 4 that the asymptotic power of the new test is equivalent to that of Dempster's test. Simulation results further show that our new approach is slightly more powerful than Dempster's. We believe that the estimation of r and its rounding to an integer in Dempster's procedure may cause an error of order $O(1/n)$. This might indicate that the new approach is superior to Dempster's test in the second order term in some Edgeworth-type expansions. We shall not discuss this in detail in this paper but hope to address it in future work. Some simulation results and discussions are presented in Section 5 and some technical proofs are given in the Appendix.

2. Asymptotic Power of Hotelling's Test

In this section, we derive the asymptotic power functions of the T^2 test for the two sample problem. The model described here is the same as the one in Dempster's test given in the next section.

Suppose that $\mathbf{x}_{i,j} \sim N_p(\mu_i, \Sigma),\ j = 1, \ldots, N_i,\ i = 1, 2$ are two independent samples. To test the hypothesis $H_0 : \mu_1 = \mu_2$, vs $H_1 : \mu_1 \neq \mu_2$, traditionally one uses Hotelling's famous T^2 test which is defined by

$$T^2 = \eta(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)' A^{-1} (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2), \tag{2.1}$$

where $\overline{\mathbf{x}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} \mathbf{x}_{i,j},\ i = 1, 2,\ A = \sum_{i=1}^{2}\sum_{j=1}^{N_i} (\mathbf{x}_{i,j} - \overline{\mathbf{x}}_i)(\mathbf{x}_{i,j} - \overline{\mathbf{x}}_i)'$, and $\eta = n\frac{N_1 N_2}{N_1+N_2}$ with $n = N_1 + N_2 - 2$. The purpose of this section is to investigate the power function of Hotelling's test when $p/n \to y \in (0, 1)$ for guaranteeing the existence of the T^2 statistic, and to compare it with other non-exact tests given in later sections.

To derive the asymptotic power of Hotelling's test, we first derive an asymptotic expression for the threshold of the test. It is well known that under the null hypothesis, $\frac{n-p+1}{np}T^2$ has an F-distribution with degrees of freedom p and $n - p + 1$. Let the significance level be chosen as α and the threshold be denoted by $F_\alpha(p,\ n - p + 1)$. We have the following lemma.

Lemma 2.1. $\frac{p}{n-p+1}F_\alpha(p, n-p+1) = \frac{y_n}{1-y_n} + \sqrt{\frac{2y}{(1-y)^3 n}}\ \xi_\alpha + o(\frac{1}{\sqrt{n}})$, *where* $y_n = \frac{p}{n}$, $\lim_{n\to\infty} y_n = y \in (0, 1)$ *and* ξ_α *is the* $1 - \alpha$ *quantile of standard normal distribution.*

Proof. Under the null hypothesis, by the Central Limit Theorem,

$$\sqrt{\frac{(1-y)^3 n}{2y}}\Big(\frac{T^2}{n}-\frac{y_n}{1-y_n}\Big)\to\mathcal{N}(0,1),\quad \text{as } n\to\infty,$$

from which the result follows immediately.

Now, we consider the behavior of T^2/n under H_1. In this case, its distribution is the same as

$$(\mathbf{w}+\tau^{-1/2}\delta)'U^{-1}(\mathbf{w}+\tau^{-1/2}\delta), \tag{2.2}$$

where $\delta=\Sigma^{-\frac{1}{2}}(\mu_1-\mu_2)$, $U=\sum_{i=1}^n \mathbf{u}_i\mathbf{u}_i'$, $\mathbf{w}=(w_1,\ldots,w_p)'$ and $\mathbf{u}_i,\ i=1,\ldots,n$ are i.i.d. $N(\mathbf{0},I_p)$ random vectors and $\tau=(N_1+N_2)/N_1N_2$. Denote the spectral decomposition of U^{-1} by $O'\text{diag}[d_1,\ldots,d_p]O$ with eigenvalues $d_1\geq\cdots\geq d_p>0$. Then, (2.2) becomes

$$(O\mathbf{w}+\tau^{-1/2}\|\delta\|\mathbf{v})'\text{diag}[d_1,\ldots,d_p](O\mathbf{w}+\tau^{-1/2}\|\delta\|\mathbf{v}), \tag{2.3}$$

where $\mathbf{v}=O\delta/\|\delta\|$. Since U has the Wishart distribution $W(n,I_p)$, the orthogonal matrix O has the Haar distribution on the group of all orthogonal p-matrices, and hence the vector $\mathbf{v}$ is uniformly distributed on the unit p-sphere. Note that the conditional distribution of $O\mathbf{w}$ given O is $\mathcal{N}(0,I_p)$, the same as that of $\mathbf{w}$ which is independent of O. This shows that $O\mathbf{w}$ is independent of $\mathbf{v}$. Therefore, replacing $O\mathbf{w}$ in (2.3) by $\mathbf{w}$ does not change the joint distribution of $O\mathbf{w}$, $\mathbf{v}$ and the d_i's. Consequently, $\frac{T^2}{n}$ has the same distribution as

$$\Omega_n=\sum_{i=1}^p(w_i^2+2w_iv_i\ \tau^{-1/2}\|\delta\|+\tau^{-1}\|\delta\|^2v_i^2)d_i, \tag{2.4}$$

where $\mathbf{v}=(v_1,\ldots,v_p)'$ is uniformly distributed on the unit sphere of R^p and is independent of $\mathbf{w}$ and the d_i's.

Lemma 2.2. *Using the above notation, we have* $\sqrt{n}\Big(\sum_{i=1}^p d_i-\frac{y_n}{1-y_n}\Big)\to 0$, *and* $n\sum_{i=1}^p d_i^2\to\frac{y}{(1-y)^3}$ *in probability.*

Proof. Recalling (2.4) with $\delta=\mathbf{0}$ under the null hypothesis and applying the Central Limit Theorem with $D_n=\{d_1,\ldots,d_p\}$ given, we have

$$\begin{aligned}
&P\Big(\frac{T^2}{n}\leq\frac{1-y_n}{y_n}+\sqrt{\frac{2y}{(1-y)^3n}}\ x\Big)\\
&=E\Big[P\Big(\frac{\sum_{i=1}^p(w_i^2-1)d_i}{\sqrt{2\sum_{i=1}^p d_i^2}}\leq\frac{\sqrt{n}(\frac{1-y_n}{y_n}-\sum_{i=1}^p d_i)+\sqrt{\frac{2y}{(1-y)^3}}\ x}{\sqrt{2n\sum_{i=1}^p d_i^2}}\ \Big|\ D_n\Big)\Big]\\
&=E\Big[\Phi\Big(\frac{\sqrt{n}(\frac{1-y_n}{y_n}-\sum_{i=1}^p d_i)+\sqrt{\frac{y}{(1-y)^3}}\ x}{\sqrt{n\sum_{i=1}^p d_i^2}}\Big)+o(1)\Big],
\end{aligned} \tag{2.5}$$

where Φ is the distribution function of a standard normal variable. On the other hand, as shown in the proof of Lemma 2.1, the Central Limit Theorem implies that the above quantity tends to $\Phi(x)$, for all x. Hence, by the type-convergence theorem (see Page 216 of Loéve (1977)), the lemma is proved.

Now we are in position to derive an approximation of the power function of Hotelling's test.

Theorem 2.1. *If* $y_n = \frac{p}{n} \to y \in (0,1)$, $N_1/(N_1+N_2) \to \kappa \in (0,1)$ *and* $\|\delta\|^2 = o(1)$, *then*

$$\beta_{\mathrm{H}}(\delta) - \Phi\Big(-\xi_\alpha + \sqrt{\frac{n(1-y)}{2y}}\ \kappa(1-\kappa)\|\delta\|^2\Big) \to 0, \tag{2.6}$$

where $\beta_{\mathrm{H}}(\delta)$ *is the power function of Hotelling's test.*

Remark 2.1. The usual consideration of the alternative hypothesis in limiting theorems is to assume that $\sqrt{n}\|\delta\|^2 \to a > 0$. Under this additional assumption, it follows from (2.6) that the limiting power of Hotelling's test is given by $\Phi(-\xi_\alpha + ((1-y)/2y)^{1/2}\ \kappa(1-\kappa)a)$. This formula shows that the limiting power of Hotelling's test is slowly increasing for y close to 1, as the non-central parameter (namely a) increases.

Proof. Write $D_n = (d_1, \ldots, d_n)$. Using the facts $Ev_1^2 = 1/p$, $Ev_1^4 = 3/[p(p+2)]$ and $Ev_1^2 v_2^2 = 1/[p(p+2)]$ and then applying Lemma 2.2, one easily obtains

$$nE\Big[\Big(\sum_{i=1}^p w_i v_i d_i \tau^{-1/2}\|\delta\|\Big)^2 \Big| D_n\Big] = n\sum_{i=1}^p d_i^2\ \frac{\|\delta\|^2}{\tau p} \to 0, \text{ in Pr.}, \tag{2.7}$$

$$\begin{aligned} &nE\Big[\Big(\sum_{i=1}^p (v_i^2 - Ev_i^2)\tau^{-1}\|\delta\|^2 d_i\Big)^2 \Big| D_n\Big] \\ &= n\tau^{-2}\|\delta\|^4\Big[\frac{2}{p(p+2)}\sum_{i=1}^p d_i^2\ - \frac{2}{p^2(p+2)}\Big(\sum_{i=1}^p d_i\Big)^2\Big] \to 0 \text{ in Pr.} \end{aligned} \tag{2.8}$$

and

$$\sum_{i=1}^p (Ev_i^2)\ \tau^{-1}\|\delta\|^2 d_i = \frac{1}{\tau p}\ \|\delta\|^2 \sum_{i=1}^p d_i = \frac{y_n\ \|\delta\|^2}{\tau p(1-y_n)}(1 + o_p(\frac{1}{\sqrt{n}})). \tag{2.9}$$

Thus, by the above and Lemma 2.1, we have

$$\beta_{\mathrm{H}}(\delta) = P\Big(\sum_{i=1}^p w_i^2 d_i \ge \frac{y_n}{1-y_n} + \sqrt{\frac{2y^2}{(1-y)^3 p}}\xi_\alpha - \frac{y_n\ \|\delta\|^2}{\tau p(1-y_n)} + o(\frac{1}{\sqrt{n}})\Big)$$

$$= E\Big[P\Big(\frac{\sum_{i=1}^p (w_i^2-1)d_i}{\sqrt{\sum_{i=1}^p 2d_i^2}} \geq \frac{\sqrt{\frac{2y^2}{(1-y)^3 p}}\,\xi_\alpha - \frac{y\|\delta\|^2}{\tau p(1-y)} + o(\frac{1}{\sqrt{n}})}{\sqrt{\sum_{i=1}^p 2d_i^2}} \Big| D_n\Big)\Big]$$

$$= \Phi\Big(-\xi_\alpha + \sqrt{\frac{n(1-y)}{2y}}\,\kappa(1-\kappa)\|\delta\|^2\Big) + o(1). \tag{2.10}$$

The proof of Theorem 2.1 is now complete.

3. Discussion on Dempster's Non-Exact Test

Dempster (1958, 1960) proposed a non-exact test for the hypothesis described in Section 2, with the dimension of data possibly greater than the sample degrees of freedom. First, let us briefly describe his test. Denote $N = N_1 + N_2$, $X' = (\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1N_1}, \mathbf{x}_{21}, \ldots, \mathbf{x}_{2N_2})$ and by $H' = (\frac{1}{\sqrt{N}}\mathbf{J}_N, (\sqrt{\frac{N_2}{N_1(N_1+N_2)}}\mathbf{J}'_{N_1}, -\sqrt{\frac{N_1}{N_2(N_1+N_2)}}\mathbf{J}'_{N_2})', \mathbf{h}_3, \ldots, \mathbf{h}_N)$ a suitably chosen orthogonal matrix, where $\mathbf{J}_d$ is a d dimensional column vector of 1's. Let $Y = HX = (\mathbf{y}_1, \ldots, \mathbf{y}_N)'$. Then, the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are independent normal random vectors with $E(\mathbf{y}_1) = (N_1\mu_1 + N_2\mu_2)/\sqrt{N}$, $E(\mathbf{y}_2) = \tau^{-1/2}(\mu_1 - \mu_2)$, $E(\mathbf{y}_j) = \mathbf{0}$, for $3 \leq j \leq N$, $\mathrm{Cov}(\mathbf{y}_j) = \Sigma$, $1 \leq j \leq N$. Then, Dempster proposed his non-exact significance test statistic $F = Q_2/(\sum_{i=3}^N Q_i/n)$, where $Q_i = \mathbf{y}_i'\mathbf{y}_i$, $n = N-2$. He used the so-called χ^2 approximation technique, assuming Q_i is approximately distributed as $m\chi_r^2$, where the parameters m and r may be solved by the method of moments. Then, the distribution of F is approximately $F_{r,nr}$. But generally the parameter r (its explicit form is given in (3.3) below) is unknown. He estimated r by either of the following two ways.

Approach 1: $\hat{r}$ is the solution of the equation

$$t = \Big(\frac{1}{\hat{r}_1} + \frac{1+\frac{1}{n}}{3\hat{r}_1^2}\Big)(n-1), \tag{3.1}$$

Approach 2: $\hat{r}$ is the solution of the equation

$$t + w = \Big(\frac{1}{\hat{r}_2} + \frac{1+\frac{1}{n}}{3\hat{r}_2^2}\Big)(n-1) + \Big(\frac{1}{\hat{r}_2} + \frac{3}{2\hat{r}_2^2}\Big)\binom{n}{2}, \tag{3.2}$$

where $t = n[\ln(\frac{1}{n}\sum_{i=3}^N Q_i)] - \sum_{i=3}^N \ln Q_i$, $w = -\sum_{3\leq i<j\leq N} \ln\sin^2\theta_{ij}$, and θ_{ij} is the angle between the vectors of $\mathbf{y}_i, \mathbf{y}_j$, $3 \leq i < j \leq N$. Dempster's test is then to reject H_0 if $F > F_\alpha(\hat{r}, n\hat{r})$.

By elementary calculus, we have

$$r = \frac{(\mathrm{tr}(\Sigma))^2}{\mathrm{tr}(\Sigma^2)} \quad \text{and} \quad m = \frac{\mathrm{tr}(\Sigma^2)}{\mathrm{tr}\Sigma}. \tag{3.3}$$

From (3.3) and the Cauchy-Schwarz inequality, it follows that $r \leq p$. On the other hand, under regular conditions, both $\mathrm{tr}(\Sigma)$ and $\mathrm{tr}(\Sigma^2)$ are of the order $O(n)$, and hence, r is of the same order. Under wider conditions (3.7) and (3.8) given in Theorem 3.1 below, it can be proved that $r \to \infty$. Further, we may prove that $t \sim (n/r)\mathcal{N}(1, n^{-1/2})$ and $w \sim \frac{n(n-1)}{2r}\mathcal{N}(1, \frac{4}{n(n-1)} + \frac{8}{nr})$. From these estimates, one may conclude that both $\hat{r}_1$ and $\hat{r}_2$ are ratio-consistent (in the sense that $\frac{\hat{r}}{r} \to 1$). Therefore, the solutions of equations (3.1) and (3.2) should satisfy

$$\hat{r}_1 = \frac{n}{t} + O(1) \tag{3.4}$$

and

$$\hat{r}_2 = \frac{1}{w}\binom{n}{2} + O(1) \tag{3.5}$$

respectively. Since the random effect may cause an error of order $O(1)$, one may simply choose the estimates of r as $\frac{n}{t}$ or $\frac{1}{w}\binom{n}{2}$.

In the remainder of this section, we derive an asymptotic power function of Dempster's non-exact test, under the conditions: $p/n \to y > 0$, $N_1/(N_1 + N_2) \to \kappa \in (0, 1)$ and the parameter r is known. The reader should note that the limiting ratio y is allowed to be greater than one in this case, which is different from that assumed in Section 2. When r is unknown, substituting r by the estimators $\hat{r}_1$ or $\hat{r}_2$ may cause an error of high order smallness in the approximation of the power function of Dempster's non-exact test, as will be seen in the proof of Theorem 3.1.

Similar to Lemma 2.1 one may show the following lemma.

Lemma 3.1. *When $n, r \to \infty$,*

$$F_\alpha(r, nr) = 1 + \sqrt{2/r}\,\xi_\alpha + o(1/\sqrt{r}). \tag{3.6}$$

Then we have the following approximation of the power function of Dempster's test.

Theorem 3.1. *If*

$$\mu'\Sigma\mu = o(\tau \mathrm{tr}\Sigma^2), \tag{3.7}$$

$$\lambda_{\max} = o(\sqrt{\mathrm{tr}\Sigma^2}), \tag{3.8}$$

and r is known, then

$$\beta_{\mathrm{D}}(\mu) - \Phi(-\xi_\alpha + \frac{n\kappa(1-\kappa)\|\mu\|^2}{\sqrt{2\mathrm{tr}\Sigma^2}}) \to 0, \tag{3.9}$$

where $\mu = \mu_1 - \mu_2$.

Remark 3.1. In usual cases when considering the asymptotic power of Dempster's test, the quantity $\|\mu\|^2$ is ordinarily assumed to have the same order as $1/\sqrt{n}$ and $\mathrm{tr}(\Sigma^2)$ to have order n. Thus, the quantities $n\|\mu\|^2/\sqrt{\mathrm{tr}\Sigma^2}$ and $\sqrt{n}\|\delta\|^2$ are both bounded away from zero and infinity. The expression of the asymptotic power of Hotelling test is involved with a factor $\sqrt{1-y}$ which disappears in the expression of the asymptotic power of Dempster's test. This reveals the reason why the power of the Hotelling test increases much slower than that of the Dempster test as the non-central parameter increases if y is close to one.

Proof. Let $\delta = (\delta_1, \ldots, \delta_p)' = \Sigma^{-\frac{1}{2}}\mu$. Then,

$$\beta_{\mathrm{D}}(\mu) = P\Big(\frac{\sum_{i=1}^p (y_i^2 + 2\tau^{-1/2}\delta_i y_i + \tau^{-1}\delta_i^2)\lambda_i}{\sum_{j=1}^n \sum_{i=1}^p z_{ij}^2 \lambda_i} > n^{-1} F_\alpha(r, nr)\Big), \tag{3.10}$$

where y_i, $z_{ij}, i = 1, \ldots, p, j = 1, \ldots, n$ are i.i.d. $\mathcal{N}(0,1)$ variables and $\lambda_1, \ldots, \lambda_p$ are eigenvalues of Σ. By the Central Limit Theorem, the laws of large numbers, (3.7) and (3.8), one may easily show that:

$$\frac{\sum_{i=1}^p (y_i^2 - 1 + 2\tau^{-1/2}\delta_i y_i)\lambda_i}{\sqrt{2\mathrm{tr}\Sigma^2}} \cong \frac{\sum_{i=1}^p (y_i^2 - 1 + 2\tau^{-1/2}\delta_i y_i)\lambda_i}{\sqrt{2\mathrm{tr}\Sigma^2 + 4\tau^{-1}\mu'\Sigma\mu}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.11}$$

and

$$\sum_{j=1}^n \sum_{i=1}^p z_{ij}^2 \lambda_i = n(\mathrm{tr}\Sigma)\Big(1 + \sqrt{2/nr}\,\mathcal{N}(0,1) + o_p(\sqrt{1/nr})\Big). \tag{3.12}$$

Noting that $\sum_{i=1}^p \delta_i^2 \lambda_i = \|\mu\|^2$ and $r = \frac{(\mathrm{tr}\Sigma)^2}{\mathrm{tr}\Sigma^2}$, the result (3.9) follows from (3.7) and Lemma 3.1, immediately. The proof of Theorem 3.1 is now complete.

4. A New Approach to Test H_0

In this section, we propose a new test for H_0. Instead of the normality of the underlying distributions, we assume:

(a) $\mathbf{x}_{ij} = \Gamma \mathbf{z}_{ij} + \mu_j$, $i = 1, \ldots, N_j$, $j = 1, 2$, where Γ is a $p \times m$ matrix ($m \le \infty$) with $\Gamma\Gamma' = \Sigma$ and $\mathbf{z}_{ij}$ are i.i.d. random m-vectors with independent components satisfying $E\mathbf{z}_{ij} = \mathbf{0}$, $\mathrm{Var}(\mathbf{z}_{ij}) = I_m$, $Ez_{ijk}^4 = 3 + \Delta < \infty$ and $E\prod_{k=1}^m z_{ijk}^{\nu_k} = 0$ (and 1) when there is at least one $\nu_k = 1$ (there are two ν_k's equal to 2, correspondingly), whenever $\nu_1 + \cdots + \nu_m = 4$;

(b) $p/n \to y > 0$ and $N_1/(N_1 + N_2) \to \kappa \in (0,1)$;

(c) (3.7) and (3.8) are true.

Here and later, it should be noted that all random variables and parameters depend on n. For simplicity we omit the subscript n from all random variables except those statistics defined later.

Now, we begin to construct our test. Consider the statistic

$$M_n = (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) - \tau \mathrm{tr} S_n, \tag{4.1}$$

where $S_n = \frac{1}{n}A$, $\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2$ and A are defined in Section 2. Under H_0, we have $EM_n = 0$. If the conditions (a) - (c) are true, it may be proved (see the Appendix) that under H_0,

$$Z_n = \frac{M_n}{\sqrt{\mathrm{Var} M_n}} \to \mathcal{N}(0,1), \text{ as } n \to \infty. \tag{4.2}$$

If the underlying distributions are normal as described in Section 2, then under H_0 we have

$$\sigma_M^2 := \mathrm{Var} M_n = 2\tau^2(1 + \frac{1}{n})\mathrm{tr}\Sigma^2. \tag{4.3}$$

If the underlying distributions are not normal but satisfy the conditions (a) - (c), one may show (see the Appendix) that

$$\mathrm{Var} M_n = \sigma_M^2(1 + o(1)). \tag{4.4}$$

Hence (4.2) is still true if the denominator of Z_n is replaced by σ_M. Therefore, to complete the construction of our test statistic, we need only find a ratio-consistent estimator of $\mathrm{tr}(\Sigma^2)$ and substitute it into the denominator of Z_n. It seems that a natural estimator of $\mathrm{tr}\Sigma^2$ should be $\mathrm{tr}S_n^2$. However, unlike the case where p is fixed, $\mathrm{tr}S_n^2$ is generally neither unbiased nor ratio-consistent even under the normal assumption. If $nS_n \sim W_p(n, \Sigma)$, it is routine to verify that

$$B_n^2 = \frac{n^2}{(n+2)(n-1)}\Big(\mathrm{tr}S_n^2 - \frac{1}{n}(\mathrm{tr}S_n)^2\Big),$$

is an unbiased and ratio-consistent estimator of $\mathrm{tr}\Sigma^2$. Here, it should be noted that $\mathrm{tr}S_n^2 - \frac{1}{n}(\mathrm{tr}S_n)^2 \geq 0$, by the Cauchy-Schwarz inequality. In the Appendix, we shall prove that B_n^2 is still a ratio-consistent estimator of $\mathrm{tr}\Sigma^2$ under the Conditions (a) - (c). Replacing $\mathrm{tr}\Sigma^2$ in (4.3) by the ratio-consistent estimator B_n^2, we obtain our test statistic

$$\begin{aligned} Z &= \frac{(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) - \tau \ \mathrm{tr}S_n}{\tau\sqrt{\frac{2(n+1)n}{(n+2)(n-1)}\Big(\mathrm{tr}S_n^2 - n^{-1}(\mathrm{tr}S_n)^2\Big)}} \\ &= \frac{\frac{N_1N_2}{N_1+N_2}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) - \mathrm{tr}S_n}{\sqrt{\frac{2(n+1)}{n}}B_n} \to \mathcal{N}(0,1). \end{aligned} \tag{4.5}$$

Due to (4.5) the test rejects H_0 if $Z > \xi_\alpha$. Regarding the asymptotic power of our new test, we have the following theorem.

Theorem 4.1. *Under the Conditions in* (a) - (c),

$$\beta_{\rm BS}(\mu) - \Phi\Big(- \xi_\alpha + \frac{n\kappa(1-\kappa)\|\mu\|^2}{\sqrt{2{\rm tr}\Sigma^2}}\Big) \to 0. \tag{4.6}$$

Proof. Let $\overline{\mathbf{z}}_j$ be the sample mean of $\mathbf{z}_{ij}$, $i = 1, \ldots, n_j$, $j = 1, 2$ and let

$$M_n^0 = (\overline{\mathbf{z}}_1 - \overline{\mathbf{z}}_2)'\Gamma'\Gamma(\overline{\mathbf{z}}_1 - \overline{\mathbf{z}}_2) - {\rm tr}(S_n).$$

Then, M_n^0 has the same distribution as M_n under H_0. Thus, ${\rm Var}(M_n^0) = \sigma_M^2(1+o(1))$ and $M_n^0/\sqrt{{\rm Var}(M_n^0)} \to \mathcal{N}(0,1)$. Note that $M_n = M_n^0 - 2\mu'(\overline{\mathbf{z}}_1 - \overline{\mathbf{z}}_2) + \|\mu\|^2$ and by (3.7)

$${\rm Var}(\mu'(\overline{\mathbf{z}}_1 - \overline{\mathbf{z}}_2)) = \tau\mu'\Sigma\mu = o(\tau^2{\rm tr}(\Sigma^2)).$$

Hence, ${\rm Var}(M_n^0)/\,{\rm Var}(M_n) \to 1$ and consequently $\frac{M_n - \|\mu\|^2}{\sqrt{{\rm Var}(M_n^0)}} \to \mathcal{N}(0,1)$.

Note that $\frac{2(n+1)}{n}B_n^2/\,{\rm Var}(M_n^0) \to 1$. Hence,

$$Z - \frac{n\kappa(1-\kappa)\|\mu\|^2}{\sqrt{2{\rm tr}(\Sigma^2)}} \to \mathcal{N}(0,1).$$

This implies that

$$\begin{aligned}\beta_{\rm BS}(\mu) &= P_{H_1}(Z > \xi_\alpha)\\ &= P\Big(\frac{M_n - \|\mu\|^2}{\sqrt{{\rm Var}\, M_n^0}} > \xi_\alpha - \frac{n\kappa(1-\kappa)\|\mu\|^2}{\sqrt{2{\rm tr}\Sigma^2}} + o(1)\Big)\\ &= \Phi\Big(- \xi_\alpha + \frac{n\kappa(1-\kappa)\|\mu\|^2}{\sqrt{2{\rm tr}\Sigma^2}}\Big) + o(1), \end{aligned} \tag{4.7}$$

which completes the proof of the theorem.

5. Discussions and Simulations

Comparing Theorems 2.1, 3.1 and 4.1, we find that from the point of view of large sample theory, Hotelling's test is less powerful than the other two tests, when y is close to one, and that the latter two tests have the same asymptotic power function. Our simulation results show that even for moderate sample and dimension sizes, Hotelling's test is still less powerful than the other two tests when the underlying covariance structure is reasonably regular (i.e., the structure of Σ does not cause a too large difference between $\mu'\Sigma^{-1}\mu$ and $\sqrt{n}\|\mu\|^2/\sqrt{{\rm tr}(\Sigma^2)}$), whereas the Type I error does not change much in the latter two tests.

It would not be hard to see that using the approach of this paper, one may easily derive similar results for the one-sample problem, namely, Hotelling's test is less powerful than a non-exact test which can be defined as in Section 4, when the dimension of data is high.

Now, we would like to explain why this phenomenon happens. The reason for the less powerfulness of Hotelling's test is the "inaccuracy" of the estimator of the covariance matrix. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. random p-vectors of mean 0 and variance-covariance matrix I_p. By the law of large numbers, the sample covariance matrix $S_n = n^{-1} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'$ should be "close" to the identity I_p with an error of the order $O_p(1/\sqrt{n})$ when p is fixed. However, when p is proportional to n (say $p/n \to y \in (0, 1)$), the ratio of the largest and the smallest eigenvalues of S_n tends to $(1+\sqrt{y})^2/(1-\sqrt{y})^2$ (see, e.g., references Bai, Silverstein and Yin (1988), Bai and Yin (1993), Geman (1980), Silverstein (1985) and Yin, Bai and Krishnaiah (1988)). More precisely, in the Theory of spectral analysis of large dimensional random matrices, it has been proven that the empirical distribution of the eigenvalues of S_n tends to an limiting distribution spreading over $[(1-\sqrt{y})^2,\ (1+\sqrt{y})^2]$ as $n \to \infty$ (see e.g., Jonsson (1982), Wachter (1978), Yin (1986) and Yin, Bai and Krishnaiah (1983)). These show that S_n is not close to I_p. Especially when y is "close to" one, then S_n has many small eigenvalues and hence S_n^{-1} has many huge eigenvalues. This will cause the deficiency of the T^2 test. We believe that in many other multivariate statistical inferences with an inverse of a sample covariance matrix involved, the same phenomenon should exist (as another example, see Saranadasa (1991, 1993)).

Here we would like to explain our quotation-marked " 'close to' one". Note that the limiting ratio of the largest to the smallest eigenvalues of S_n tends to $(1+\sqrt{y})^2/(1-\sqrt{y})^2$. For our simulation example, $y = 0.93$ and the ratio of the extreme eigenvalues is about 3039. That is very serious. Even for y as small as 0.1 or 0.01, the ratio can be as large as 3.705 and 1.494. These show that it is not necessary to require the dimension of data to be very close to the degrees of freedom to make the effect of high dimension visible. In fact, this has been shown by our simulation for $p = 4$.

Dempster's test statistic depends on the choice of vectors $h_3, h_4, \ldots, h_N$ because different choices of these vectors would produce different estimates of the parameter r. On the other hand, the estimation of r and the rounding of the estimates may cause an error (probably an error of second order smallness) in Dempster's test. Thus, we conjecture that our new test can be more powerful than Dempster's in their second terms of an Edgeworth type expansion of their power functions. This conjecture was strongly supported by our simulation results. Because our test statistic is mathematically simple, it is not difficult to get an Edgeworth expansion by using the results obtain in Babu and Bai (1993), Bai and Rao (1991) or Bhattacharya and Ghosh (1978). It seems difficult to get a similar expansion for Dempster's test due to his complicated estimation of r.

We conducted our simulation study to compare the power of the three tests for both normal and non-normal cases. Let $N_1 = 25$, $N_2 = 20$, and $p = 40$. For the non-normal case, observations were generated by the following moving average model: Let $\{U_{ijk}\}$ be a set of independent gamma variables with shape parameter 4 and scale parameter 1. Define

$$X_{ijk} = U_{ijk} + \rho U_{i,j+1,k} + \mu_{jk}, \quad (j = 1, \ldots, p,\ i = 1, \ldots, N_k,\ k = 1, 2),$$

where ρ and the μ's are constants. Under this model, $\Sigma = (\sigma_{ij})$ with $\sigma_{ii} = 4(1 + \rho^2)$, $\sigma_{i,i\pm 1} = 4\rho$ and $\sigma_{ij} = 0$ for $|i - j| > 1$. For the normal case, the covariance matrices were chosen to be $\Sigma = I_p$ and $\Sigma = (1 - \rho)I_p + \rho\mathbf{J}_p$, with $\rho = 0.5$, where $\mathbf{J}$ is a $p \times p$ matrix with all entries one. Simulation was also conducted for small p (chosen as $p = 4$). The tests were made for size $\alpha = 0.05$ with 1000 repetitions. The power is evaluated at standard parameter $\eta = \|\mu_1 - \mu_2\|^2/\sqrt{\mathrm{tr}\Sigma^2}$. The simulation for the non-normal case was conducted for $\rho = 0, .3, .6$ and $.9$ (Table 5.1 and Figure 5.1). All three tests have almost the same significance level. Under the alternative hypothesis, the power curves of Dempster's test and our test are rather close but that of our test is always higher than Dempster's test. Theoretically, the power function for Hotelling's test should increase very lowly when the noncentral parameter increases. This was also demonstrated by our simulation results. The reader should note that there are only 1000 repetations for each value of noncentral parameter in our simulation which may cause an error of $\sqrt{1/1000} = 0.0316$ by Central Limit Theorem, it is not surprising the simulated power function of the Hotelling's test, whose magnitude is only around 0.05, seems not increasing at some points of the noncentral parameter.

Similar tables are presented for the normal case (Table 5.2 and Figure 5.2). For higher dimension cases the power functions of Dempster's test and our test are almost the same and our method is not worse than Hotelling's test even for $p = 4$.

Acknowledgement

The research of the first author was partially supported by US NSF grant DMS 9408799 and partially by ROC NSC grant NSC84-2816-M110-009L.

Table 5.1. Simulated power functions of the three tests with multivariate Gamma distribution.

	$N = 45, p = 40, \alpha = .05$											
	$\rho = 0, \phi = 1$			$\rho = .3, \phi = 3.4$			$\rho = .6, \phi = 15.6$			$\rho = .9, \phi = 235.8$		
η	H	D	BS	H	D	BS	H	D	BS	H	D	BS
0.00	.044	.047	.052	.044	.047	.055	.052	.047	.055	.055	.046	.058
0.15	.059	.289	.316	.071	.293	.319	.092	.282	.315	.109	.277	.305
0.30	.084	.602	.633	.113	.643	.667	.147	.639	.677	.177	.609	.643
0.45	.118	.841	.870	.146	.875	.898	.204	.879	.897	.242	.848	.873
0.60	.149	.947	.960	.190	.966	.971	.251	.968	.976	.331	.958	.961
0.75	.181	.982	.988	.240	.996	.997	.318	.996	.999	.401	.994	.994
0.90	.212	.997	.999	.276	1.000	1.000	.380	1.000	1.000	.469	1.000	1.000
1.05	.251	1.000	1.000	.328	1.000	1.000	.444	1.000	1.000	.524	1.000	1.000
1.20	.293	1.000	1.000	.374	1.000	1.000	.504	1.000	1.000	.589	1.000	1.000
1.35	.327	1.000	1.000	.427	1.000	1.000	.537	1.000	1.000	.636	1.000	1.000

Table 5.2. Simulated power functions of the three tests with multivariate normal distribution.

	$N = 45, p = 40, \alpha = .05$						$N = 45, p = 40, \alpha = .05$					
	$\rho = 0, \phi = 1$			$\rho = .5, \phi = 41$			$\rho = 0, \phi = 1$			$\rho = .5, \phi = 5$		
η	H	D	BS	H	D	BS	H	D	BS	H	D	BS
0.00	.040	.050	.062	.050	.081	.074	.042	.050	.069	.060	.084	.085
0.15	.071	.284	.313	.059	.336	.318	.242	.267	.320	.177	.268	.297
0.30	.094	.634	.680	.061	.521	.510	.458	.498	.572	.261	.487	.506
0.45	.129	.871	.890	.055	.669	.661	.651	.697	.748	.356	.582	.608
0.60	.166	.965	.975	.071	.779	.771	.774	.824	.860	.459	.767	.804
0.75	.193	.988	.990	.061	.863	.864	.896	.933	.954	.574	.828	.848
0.90	.237	.994	.997	.083	.903	.899	.951	.967	.976	.837	.947	.953
1.05	.278	1.000	1.000	.085	.944	.945	.967	.979	.986	.855	.974	.974
1.20	.261	1.000	1.000	.086	.961	.958	.976	.989	.990	.895	.980	.982
1.35	.331	1.000	1.000	.101	.973	.973	.995	1.000	1.000	.937	.991	.992

H: Hotelling's F test, D: Dempster's non exact F test, BS: Proposed normal test, $\eta = \frac{\|\mu_1 - \mu_2\|^2}{\sqrt{\mathrm{tr}\Sigma^2}}$ and $\phi = \frac{\lambda_{\max}}{\lambda_{\min}}$.

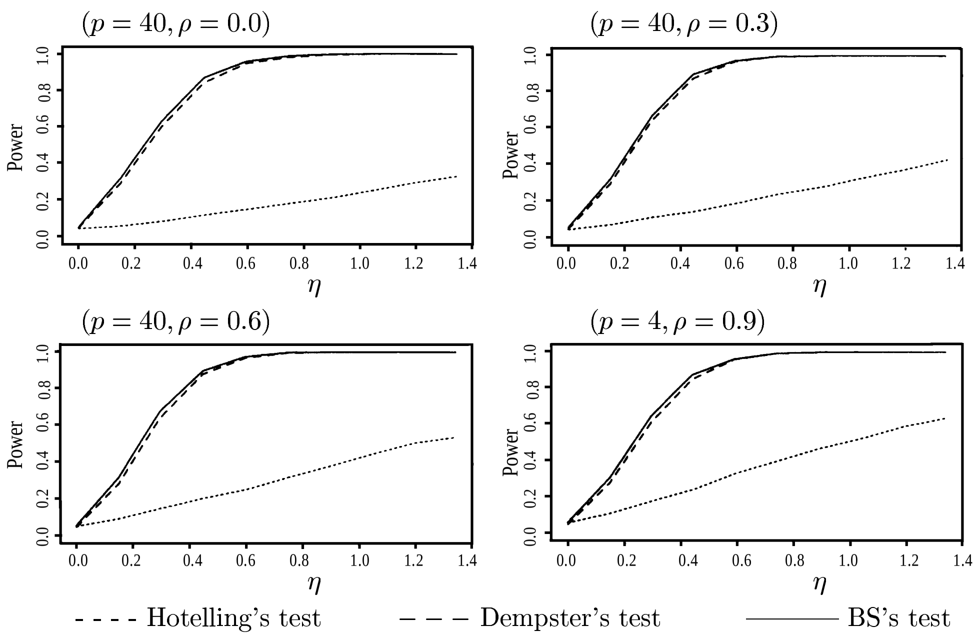

Figure 5.1. Simulated power functions of the three tests with multivariate Gamma distribution.

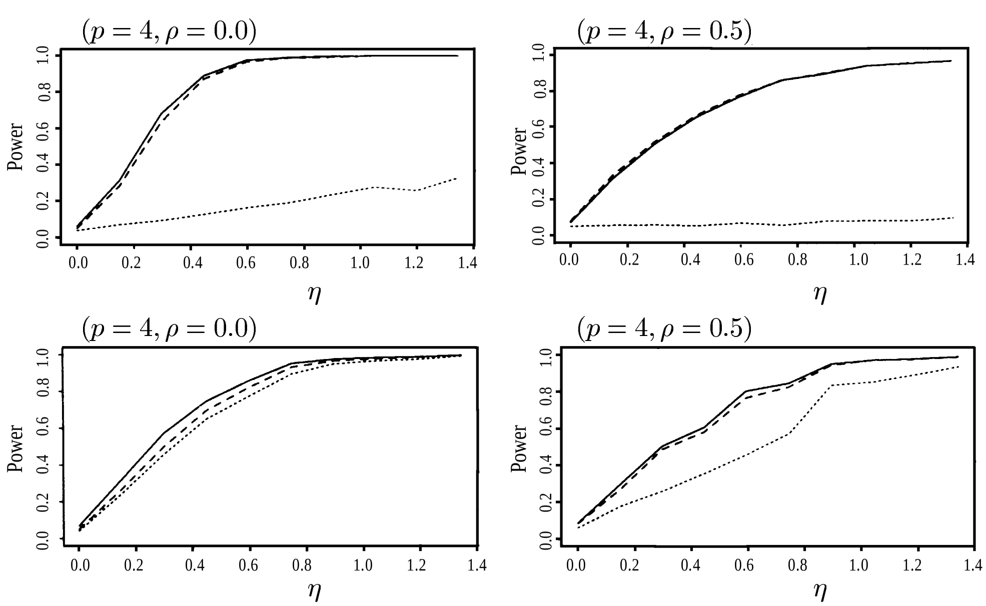

Figure 5.2. Simulated power functions of the three tests with multivariate normal distribution.

Appendix. Asymptotics Related to the Statistic M_n

A.1. The proof of (4.4):

By definition, we have

$$M_n = (1+\tau N_1 n^{-1})\|\overline{\mathbf{x}}_1\|^2 + (1+\tau N_2 n^{-1})\|\overline{\mathbf{x}}_2\|^2 - 2\overline{\mathbf{x}}_1'\overline{\mathbf{x}}_2 - \tau n^{-1}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\|\mathbf{x}_{ij}\|^2.$$

Under H_0, we may assume $\mu_1 = \mu_2 = 0$. Write $\Gamma = [\Gamma_1, \ldots, \Gamma_p]' = [\gamma_{k,\ell}]$ and $\Gamma'\Gamma = [\nu_{k\ell}]$. Then, by Conditions (a) - (c), we have

$$\mathrm{Var}(\tau n^{-1}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\|\mathbf{x}_{ij}\|^2) = \tau^2 n^{-2} E\Big(\sum_{j=1}^{2}\sum_{i=1}^{N_j}\sum_{k=1}^{p}[(\Gamma_k'\mathbf{z}_{ij})^2 - \|\Gamma_k\|^2]\Big)^2$$

$$= \tau^2 n^{-2} N\Big[2\mathrm{tr}(\Sigma^2) + \Delta\sum_{\ell=1}^{m}\nu_{\ell\ell}^2\Big] \le C\tau^2 n^{-1}[2\mathrm{tr}\Sigma^2 + \Delta\lambda_{\max}\mathrm{tr}\Sigma] = o(\sigma_M^2).$$

Similarly, we may show that

$$\mathrm{Var}(\overline{\mathbf{x}}_1'\overline{\mathbf{x}}_2) = N_1^{-2}N_2^{-2}E\Big(\sum_{i=1}^{N_1}\sum_{\ell=1}^{N_2}\mathbf{x}_{i1}'\mathbf{x}_{\ell 2}\Big)^2 = \frac{1}{N_1N_2}\mathrm{tr}(\Sigma^2),$$

$\mathrm{Var}(\|\overline{\mathbf{x}}_1\|^2) = \frac{2}{N_1^2}\mathrm{tr}(\Sigma^2) + \frac{\Delta}{N_1^3}\sum_{\ell=1}^{m}\nu_{\ell\ell}^2$, $\mathrm{Var}(\|\overline{\mathbf{x}}_2\|^2 = \frac{2}{N_2^2}\mathrm{tr}(\Sigma^2) + \frac{\Delta}{N_2^3}\sum_{\ell=1}^{m}\nu_{\ell\ell}^2$, $\mathrm{Cov}(\|\overline{\mathbf{x}}_1\|^2,\ \|\overline{\mathbf{x}}_2\|^2) = 0$ and $\mathrm{Cov}(\overline{\mathbf{x}}_1'\overline{\mathbf{x}}_2,\ \|\overline{\mathbf{x}}_j\|^2) = 0$ for $j = 1, 2$. Therefore, by the fact that $\sum_{\ell=1}^{m}\nu_{\ell\ell}^2 \le p\lambda_{\max}^2$, we have

$$\mathrm{Var}(M_n) = 2\tau^2\mathrm{tr}(\Sigma^2) + \Delta\Big(\frac{1}{N_1^3} + \frac{1}{N_2^3}\Big)\Big[\sum_{\ell=1}^{m}\nu_{\ell\ell}^2\Big] = \sigma_M^2(1+o(1)).$$

The proof of (4.4) is then complete.

A.2. The asymptotic normality of Z_n under H_0: From the proof of A.1, one can see that $\tau(\mathrm{tr}(S_n) - \mathrm{tr}(\Sigma))/\sigma_M \to 0$. Therefore, to show that $Z_n \to \mathcal{N}(0,1)$, we need only show that $[\|\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\|^2 - E(\|\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\|^2)]/\sigma_M \to \mathcal{N}(0,1)$.

We may rewrite

$$\|\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\|^2 - E(\|\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\|^2) = \sum_{k=1}^{p}[(\Gamma_k'(\overline{\mathbf{z}}_1 - \overline{\mathbf{z}}_2))^2 - \|\Gamma_k\|^2]$$

$$:= \sum_{\ell=1}^{m}\sum_{h=1}^{N}[U_{\ell,h} + V_{\ell,h} + \nu_{\ell\ell}(w_{\ell,h}^2 - E(w_{\ell,h}^2))],$$

where $\overline{\mathbf{z}}_j = N_j^{-1}\sum_{i=1}^{N_j}\mathbf{z}_{ij}$, $\overline{z}_{\cdot jk}$ denotes the kth component of $\overline{\mathbf{z}}_j$ and

$$U_{\ell,h} = 2\sigma_M^{-1}\Big[w_{\ell,h}\nu_{\ell\ell}\sum_{k_1=1}^{h-1} w_{\ell,k_1}\Big]$$
$$V_{\ell,h} = 2\sigma_M^{-1}w_{\ell,h}\sum_{\ell_1=1}^{\ell-1}\nu_{\ell_1\ell}(\overline{z}_{\cdot 1\ell_1} - \overline{z}_{\cdot 2\ell_1}),$$

with the convention that $\sum_{\ell_1=1}^{0} = 0$ and the notation

$$w_{\ell,h} = \begin{cases} \frac{1}{N_1}z_{h,1,\ell}, & \text{if } h = 1,2,\ldots,N_1, \\ \frac{1}{N_2}z_{h-N_1,2,\ell}, & \text{if } h = N_1+1,\ldots,N. \end{cases}$$

Since $\mathrm{Var}(\sum_{\ell=1}^{m}\sum_{h=1}^{N}(w_{\ell,h}^2 - E(w_{\ell,h}^2))) = (\Delta+2)(\frac{1}{N_1^3}+\frac{1}{N_1^3})\sum_{\ell=1}^{m}\nu_{\ell\ell}^2/\sigma_M^2 \to 0$, we need only show that $\sum_{\ell=1}^{m}\sum_{h=1}^{N}[U_{\ell,h}+V_{\ell,h}] \to \mathcal{N}(0,1)$.

Note that $\{U_{N(\ell-1)+k}^* = U_{\ell,k}+V_{\ell,k}\}$ forms a sequence of martingale differences with σ-fields $\mathcal{F}_{N(\ell-1)+h} = \mathcal{F}(z_{ijt}, j = 1,2; t < \ell; i = 1,\ldots,N_j$ and $w_{\ell,i}, i \le h)$. Then the asymptotic normality may be proved by employing Corollary 3.1 in Hall (1980) with routine verification of the following:

$$\sum_{\ell=1}^{m}\sum_{h=1}^{N} E(U_{\ell,h}^4 + V_{\ell,h}^4) \to 0$$

and

$$\mathrm{Var}\Big(\sum_{\ell=1}^{m}\sum_{h=1}^{N} E[(U_{\ell,h}^2+V_{\ell,h}^2)|\mathcal{F}_{N(\ell-1)+h}]\Big) \to 0.$$

The proof of (4.2) is now complete.

A.3. The ratio-consistency of B_n^2: We only need show that $\tilde{B}_n^2 = \mathrm{tr}S_n^2 - \frac{1}{n}(\mathrm{tr}S_n)^2$ is ratio-consistent for $\mathrm{tr}(\Sigma^2)$. Without loss of generality, we assume that $\mu_1 = \mu_2 = 0$. Note that

$$S_n = n^{-1}\Big[\sum_{j=1}^{2}\sum_{i=1}^{N_j}\mathbf{x}_{ij}\mathbf{x}_{ij}' - N_1\overline{\mathbf{x}}_1\overline{\mathbf{x}}_1' - N_2\overline{\mathbf{x}}_2\overline{\mathbf{x}}_2'\Big].$$

Since $E\overline{\mathbf{x}}_j'\overline{\mathbf{x}}_j = N_j^{-1}\mathrm{tr}(\Sigma) = o(\sqrt{\mathrm{tr}(\Sigma^2)})$, $j = 1,2$, it follows that, $\overline{\mathbf{x}}_j'\overline{\mathbf{x}}_j = o(\sqrt{\mathrm{tr}(\Sigma^2)})$. Therefore, we need only show that

$$\hat{B}_n^2 = \mathrm{tr}\Big(\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\mathbf{x}_{ij}\mathbf{x}_{ij}'\Big)^2 - \frac{1}{n}\Big(\mathrm{tr}(\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\mathbf{x}_{ij}\mathbf{x}_{ij}')\Big)^2$$

is a ratio-consistent estimator of $\mathrm{tr}(\Sigma^2)$.

By elementary calculation, we have $E(\mathrm{tr}(\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\mathbf{x}_{ij}\mathbf{x}'_{ij})) = \frac{N}{n}\mathrm{tr}(\Sigma)$ and $\mathrm{Var}(\mathrm{tr}(\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\mathbf{x}_{ij}\mathbf{x}'_{ij})) = O(\mathrm{tr}(\Sigma^2))$. These, together with $p^{-1/2}\mathrm{tr}(\Sigma) = o(\sqrt{\mathrm{tr}(\Sigma^2)})$, imply that

$$\frac{1}{n}\Big(\mathrm{tr}(\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\mathbf{x}_{ij}\mathbf{x}'_{ij})\Big)^2 = \frac{N}{n^2}(\mathrm{tr}(\Sigma))^2 + o_p(\mathrm{tr}(\Sigma^2)).$$

Rewrite

$$\begin{aligned}
&\mathrm{tr}\Big(\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\mathbf{x}_{ij}\mathbf{x}'_{ij}\Big)^2\\
&= \frac{N^2}{n^2}\mathrm{tr}(\Sigma^2) + \frac{2N}{n^2}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\mathrm{tr}((\Gamma'\Gamma)^2(\mathbf{z}_{ij}\mathbf{z}'_{ij} - I_m))\\
&\quad + \frac{1}{n^2}\sum_{j=1}^{2}\sum_{j'=1}^{2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_{j'}}(\mathrm{tr}((\Gamma'\Gamma)(\mathbf{z}_{ij}\mathbf{z}'_{ij} - I_m))(\Gamma'\Gamma)(\mathbf{z}_{i'j'}\mathbf{z}'_{i'j'} - I_m))\\
&:= \frac{N^2}{n^2}\mathrm{tr}(\Sigma^2) + H_1 + H_2.
\end{aligned}$$

We have $E(H_1) = 0$ and $\mathrm{Var}(H_1) = \frac{4N^3}{n^4}\Big[2\mathrm{tr}(\Sigma^4) + \sum_{i=1}^{m}([(\Gamma'\Gamma)^2]_{ii}^2\Big] = o(\mathrm{tr}^2(\Sigma^2))$. Thus,

$$H_1 = o_p(\mathrm{tr}(\Sigma^2)).$$

Write $H_2 = H_{21} + H_{22} + H_{23} + H_{24} + H_{25}$, where

$$\begin{aligned}
H_{21} &= \frac{1}{n^2}\sum_{(ij)\neq(i'j')}(\mathrm{tr}((\Gamma'\Gamma)(\mathbf{z}_{ij}\mathbf{z}'_{ij} - I_m))(\Gamma'\Gamma)(\mathbf{z}_{i'j'}\mathbf{z}'_{i'j'} - I_m)),\\
H_{22} &= \frac{1}{n^2}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\sum_{(k'\neq\ell;\ \ell'\neq k)}\nu_{k,\ell}\nu_{k',\ell'}(z_{ij\ell}z_{ijk'}z_{ij\ell'}z_{ijk})
\end{aligned}$$

and

$$\begin{aligned}
H_{23} &= \frac{2}{n^2}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\sum_{\ell\neq k\neq\ell'}\nu_{k,\ell}\nu_{\ell,\ell'}((z_{ij\ell}^2 - 1)(z_{ij\ell'}z_{ijk}))\\
H_{24} &= \frac{2}{n^2}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\sum_{\ell\neq k}\nu_{k,\ell}\nu_{\ell,\ell}((z_{ij\ell}^2 - 1)(z_{ij\ell}z_{ijk}))\\
H_{25} &= \frac{1}{n^2}\sum_{j=1}^{2}\sum_{i=1}^{N_j}\sum_{k,\ell=1}^{m}\nu_{k,\ell}^2(z_{ij\ell}^2 - 1)(z_{ijk}^2 - 1).
\end{aligned}$$

We have $E(H_{21}) = 0$, $E(H_{22}) = \frac{N}{n^2}[\mathrm{tr}(\Sigma^2) + \mathrm{tr}^2(\Sigma) - 2\sum_{k=1}^m \nu_{kk}^2 = \frac{N}{n^2}\mathrm{tr}^2(\Sigma) + o(\mathrm{tr}(\Sigma^2))$ and

$$\mathrm{Var}(H_{21}) = \frac{2N(N-1)}{n^4}\Big[4\mathrm{tr}^2(\Sigma^2) + 4\Delta \sum_{i,j,t=1}^m \nu_{ij}^2\nu_{it}^2 + \Delta^2 \sum_{i,j=1}^m \nu_{ij}^4\Big] = o(\mathrm{tr}^2(\Sigma^2)).$$

Similarly, we may show that $\mathrm{Var}(H_{22})$ and $\mathrm{Var}(H_{23})$ have the same order. Finally, one may show that

$$\begin{aligned} E|H_{24}| &\le \frac{CN}{n^2}\sum_{\ell=1}^m \nu_{\ell,\ell} E\left|\sum_{k=1}^m \nu_{k,\ell} z_{11k}\right| \\ &\le \frac{CN}{n^2}\sum_{\ell=1}^m \nu_{\ell,\ell}\sqrt{\sum_{k=1}^m \nu_{k,\ell}^2} \\ &\le \frac{CN}{n^2}\lambda_{\max}\mathrm{tr}(\Sigma) = o(\mathrm{tr}(\Sigma)). \end{aligned}$$

and

$$E|H_{25}| \le \frac{CN}{n^2}\sum_{k,\ell=1}^m \nu_{k,\ell}^2 = o(\mathrm{tr}(\Sigma)).$$

Combining the above, we obtain $H_2 = \frac{N}{n^2}\mathrm{tr}^2(\Sigma) + o_p(\mathrm{tr}(\Sigma^2))$. Thus, $\hat{B}_n^2 = \mathrm{tr}(\Sigma^2)[1 + o_p(1)]$ and consequently, the ratio-consistency of $\hat{B}_n^2$ follows.

References

Babu, G. J. and Bai, Z. D. (1993). Edgeworth expansions of a function of sample means under minimal moment conditions and partial Cramer's conditions. Sankhyā Ser.A **55**, 244-258.

Bai, Z. D., Krishnaiah, P. R. and Zhao, L. (1989). On rates of convergence of efficient detection criteria in signal processing with white noise IEEE Information **35**, 380-388.

Bai, Z. D. and Rao, C. R. (1991). Edgeworth expansion of a function of sample means. Ann. Statist. **19**, 1295-1315.

Bai, Z. D. Silverstein, J. W. and Yin, Y. Q. (1988). A note on the largest eigenvalue of a large dimensional sample covariance matrix. J. Multivariate Anal. **26**, 166-168.

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of large dimensional sample covariance matrix. Ann. Probab. **21**, 1275-1294.

Bhattacharya, R. N. and Ghosh, J. K. (1988). On moment conditions for valid formal Edgeworth expansions. J. Multivariate Anal. **27**, 68-79.

Chung, J. H. and Fraser, D. A. S. (1958). Randomization tests for a multivariate two-sample problem. J. Amer. Statist. Assoc. **53**, 729-735.

Dempster, A. P. (1958). A high dimensional two sample significance test. Ann. Math. Statist. **29**, 995-1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. Biometrics **16**, 41-50.

Geman, S. (1980). A limit theorem for the norm of random matrices. Ann. Probab. **8**, 252-261.

Hall, P. G. and Heyde, C. C. (1980). Martingale Limit Theory and Its Applications. Academic Press, New York.

Huber, Peter J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo Ann. Statist. **1**, 799-821.

Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. J. Multivariate Anal. **12**, 1-38.

Loève, M. (1977). Probability Theory, 4th Ed. Springer-Verlag, New York.

Narayanaswamy, C. R. and Raghavarao, D. (1991). Principal component analysis of large dispersion matrices. Appl. Statist. **40**, 309-316.

Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency Ann. Statist. **12**, 1298-1309.

Portnoy, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large: II. Normal approximation (Corr: 91V19 p2282) Ann. Statist. **13**, 1403-1417.

Saranadasa, H. (1991). Discriminant analysis based on experimental design concepts, Ph.D. Thesis, Department of Statistics, Temple University.

Saranadasa, H. (1993). Asymptotic expansion of the misclassification probabilities of D- and A-criteria for discrimination from two high dimensional populations using the theory of large dimensional random matrices. J. Multivariate Anal. **46**, 154-174.

Silverstein, J. W. (1985). The smallest eigenvalue of a large dimensional Wishart matrix. Ann. Probab. **13**, 1364-1368.

Wachter, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. Ann. Probab. **6**, 1-18.

Yin, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. J. Multivariate Anal. **20**, 50-68.

Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1983). Limiting behavior of the eigenvalues of a multivariate F matrix. J. Multivariate Anal. **13**, 508-516.

Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1988). On the limit of the Largest eigenvalue of the large dimensional sample covariance matrix. Probab. Theory Related Fields **78**, 509-521.

Zhao, L. C., Krishnaiah, P. R. and Bai, Z. D. (1986a). On detection of the number of signals in presence of white noise J. Multivariate Anal. **20**, 1-25.

Zhao, L. C., Krishnaiah, P. R. and Bai, Z. D. (1986b). On detection of the number of signals when the noise covariance matrix is arbitrary J. Multivariate Anal. **20**, 26-49.

Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung 80424, Taiwan.

The R. W. Johnson Pharmaceutical Research Institute, Preclinical Biostatistics, Welsh and Mckean Road, Spring House, PA 19477-0776, U.S.A.

(Received July 1993; accepted April 1995)