Journal of the Korean Statistical Society (2004), **33**: *1, pp 1–24*

INVITED PAPER

MULTIVARIATE ANALYSIS FOR THE CASE WHEN THE DIMENSION IS LARGE COMPARED TO THE SAMPLE SIZE†

YASUNORI FUJIKOSHI[1]

ABSTRACT

This paper is concerned with statistical methods for multivariate data when the number p of variables is large compared to the sample size n. Such data appear typically in analysis of DNA microarrays, curve data, financial data, *etc.* However, there is little statistical theory for high dimensional data. On the other hand, there are some asymptotic results under the assumption that both n and p tend to ∞, in some ratio $p/n \to c$. The results suggest that the new asymptotic results are more useful and insightful than the classical large sample asymptotics. The main purpose of this paper is to review some asymptotic results for high dimensional statistics as well as classical statistics under a high dimensional asymptotic framework.

AMS 2000 subject classifications. Primary 62H10; Secondary 62E20.
Keywords. Asymptotic distribution, discriminant functions, eigenvalues, high dimensional inference, MANOVA tests, tests for covariance matrices.

1. INTRODUCTION

In many important statistical applications we have encountered multivariate data such that the number p of variables is larger than the number n of samples. In general, the terminology of high dimensional data is used for a set of data where the number p of dependent or independent variables is large compared to the number n of samples. Such data appear typically in analysis of DNA

†Portions of this work were presented in the Ilsong Invited Lecture given at the 2003 Autumn Conference of the Korean Statistical Society in Seoul, Korea.

[1]Department of Statistics, Graduate School of Science, Hiroshima University, Kagamiyama Higashi-Hiroshima 739-8526, Japan

microarrays, curve data, financial data, *etc.* However, it may be noted that there is little statistical theory of high dimensional multivariate data.

One of the approaches for high dimensional multivariate data is to extend classical methods to the corresponding modified methods. The other approach is to consider methods for reducing the variables used by deleting a set of redundant variables or constructing a set of useful combined variables. It may be noted that even in the latter approach, we might have a set of multivariate data where the number of variables used is relatively large compared to the sample size.

On the other hand, there are some asymptotic results under a high dimensional framework such that both n and p tend to ∞, in some ratio $p/n \to c$. We denote such asymptotic as (n, p)-asymptotic, and denote the large sample asymptotic as (n)-asymptotic. Through some (n, p)-asymptotic results, it may be noted that such new asymptotic results are more useful and insightful than the classical large sample asymptotic results.

The main purpose of this paper is to review asymptotic results of some statistics for high dimensional statistical procedures as well as classical statistical procedures under a high dimensional framework. In Section 2 we consider MANOVA tests. In Section 3 we consider some discriminant functions. In Section 4 we consider test statistics for covariance matrices. The distributions of eigenvalues for sample covariance matrix and MANOVA matrix are considered in Section 5.

2. MANOVA Tests

2.1. Test statistics

Let $\mathbf{Y}$ be an $N \times p$ observation matrix which is obtained by independently observing a p dimensional variate $\mathbf{y} = (y_1, \ldots, y_p)'$ for N subjects. A multivariate linear model for $\mathbf{Y}$ is expressed as

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\Theta} + \boldsymbol{\mathcal{E}}, \tag{2.1}$$

where $\mathbf{A}$ is a known $N \times k$ design matrix with $\text{rank}(\mathbf{A}) = k$, $\boldsymbol{\Theta}$ is a $k \times p$ unknown parameter matrix, and $\boldsymbol{\mathcal{E}}$ is an $N \times p$ error matrix. It is assumed that the rows of $\boldsymbol{\mathcal{E}}$ are independently distributed as $N_p(\mathbf{0}, \boldsymbol{\Sigma})$. For testing

$$H_0 : \mathbf{C}\boldsymbol{\Theta} = \mathbf{0} \ \ vs. \ H_1 : \mathbf{C}\boldsymbol{\Theta} \neq \mathbf{0},$$

let $\mathbf{S}_h$ and $\mathbf{S}_e$ be the matrices of sums of squares and products due to the hypothesis and the error defined by

$$\mathbf{S}_h = (\mathbf{C}\widehat{\boldsymbol{\Theta}})' \left\{\mathbf{C}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}'\right\}^{-1} \mathbf{C}\widehat{\boldsymbol{\Theta}},$$
$$\mathbf{S}_e = (\mathbf{Y} - \mathbf{A}\widehat{\boldsymbol{\Theta}})'(\mathbf{Y} - \mathbf{A}\widehat{\boldsymbol{\Theta}}),$$

respectively, where $\mathbf{C}$ is a $q \times k$ known matrix with $\mathrm{rank}(\mathbf{C}) = q$, and $\widehat{\boldsymbol{\Theta}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}$. Then $\mathbf{S}_h$ and $\mathbf{S}_e$ are independently distributed as a noncentral Wishart distribution $W_p(q, \boldsymbol{\Sigma}; \mathbf{M}\mathbf{M}')$ and a central Wishart distribution $W_p(n, \boldsymbol{\Sigma})$, where $n = N - k$, and $\mathbf{M}$ is a $p \times q$ matrix such that

$$\mathbf{M}\mathbf{M}' = (\mathbf{C}\boldsymbol{\Theta})' \left\{\mathbf{C}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}'\right\}^{-1} \mathbf{C}\boldsymbol{\Theta}.$$

Under the assumption that $n \geq p$, the following three well known statistics have been used:

(i) Likelihood ratio statistic :
$$-\log\left(\frac{|\,\mathbf{S}_e\,|}{|\,\mathbf{S}_e + \mathbf{S}_h\,|}\right) = \log \prod_{j=1}^{t} (1 + \ell_j)\,;$$
(ii) Lawley-Hotelling trace criterion :
$$\mathrm{tr}\left(\mathbf{S}_h\mathbf{S}_e^{-1}\right) = \sum_{j=1}^{t} \ell_j\,;$$
(iii) Bartlett-Nanda-Pillai trace criterion :
$$\mathrm{tr}\left\{\mathbf{S}_h(\mathbf{S}_e + \mathbf{S}_h)^{-1}\right\} = \sum_{j=1}^{t} \frac{\ell_j}{1 + \ell_j}\,,$$

where $t = \min(p, q)$ and $\ell_1 > \cdots > \ell_t > 0$ are the non-zero eigenvalues of $\mathbf{S}_h\mathbf{S}_e^{-1}$.

When we consider the distributions of the above invariant tests, without loss of generality we may assume that

$$\mathbf{S}_h = \mathbf{Z}\mathbf{Z}', \quad \mathbf{Z} \sim N_{p\times q}(\boldsymbol{\Sigma}^{-1/2}\mathbf{M}, \mathbf{I}_q \otimes \mathbf{I}_q), \quad \mathbf{S}_e \sim W_p(n, \mathbf{I}_p). \tag{2.2}$$

The following lemma by Wakaki *et al.* (2003) is fundamental for deriving asymptotic distributions of the three test statistics under a high dimensional framework.

LEMMA 2.1. *Let*

$$\mathbf{B} = \mathbf{Z}\mathbf{Z}' \textit{ and } \mathbf{W} = \mathbf{B}^{1/2}(\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}')^{-1}\mathbf{B}^{1/2}.$$

Then $\mathbf{B}$ *and* $\mathbf{W}$ *are independently distributed as a noncentral Wishart distribution* $W_q(p, \mathbf{I}_q, \mathbf{\Omega})$ *and a central Wishart distribution* $W_q(m, \mathbf{I}_q)$, *respectively, where* $m = n - p + q$ *and the noncentrality matrix* $\mathbf{\Omega}$ *is given by*

$$\mathbf{\Omega} = \mathbf{M}'\mathbf{\Sigma}^{-1}\mathbf{M}. \tag{2.3}$$

Note that the non-zero eigenvalues of $\mathbf{S}_h\mathbf{S}_e^{-1}$ are equal to the ones of $\mathbf{BW}^{-1}$. Therefore, the three statistics can be represented as

(i) $\dfrac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_h|} = \dfrac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|}$;

(ii) $\mathrm{tr}\left(\mathbf{S}_h\mathbf{S}_e^{-1}\right) = \mathrm{tr}\left(\mathbf{BW}^{-1}\right)$;

(iii) $\mathrm{tr}\left\{\mathbf{S}_h(\mathbf{S}_e + \mathbf{S}_h)^{-1}\right\} = \mathrm{tr}\left\{\mathbf{B}(\mathbf{W} + \mathbf{B})^{-1}\right\}$.

Here the statistics can be represented in terms of independent Wishart matrices $\mathbf{B}$ and $\mathbf{W}$ of size $q \times q$.

2.2. Null distributions

Our interest is to consider asymptotic expansions of the null distributions of the above three statistics under (A0):

$$\text{(A0)}\ q : \text{fixed},\ p/n \to c \in (0, 1). \tag{2.4}$$

This assumption implies that $m = n - p + q$ also tends to infinity. For the LR statistic, there are some works. It is known (see, *e.g.*, Anderson, 1984) that $|\mathbf{S}_e|/|\mathbf{S}_e + \mathbf{S}_h|$ is distributed as a Lambda distribution $\Lambda_{p,q,n}$. Mudholkar and Trivedi (1980) have proposed a normal approximation. Their idea is to improve an approximation by a power transformation such that a leading term in an expansion for skewness of transformed statistic vanishes. The saddlepoint approximations have been proposed by Srivastava and Yau (1989). Tonda and Fujikoshi (2003) obtained an asymptotic expansion of the null distribution of the LR statistic by using the property of $\Lambda_{p,q,n} = \Lambda_{q,p,n-p+q}$.

Let

$$\begin{aligned}
T_{LR} &= -\sqrt{p}\left(1 + \frac{m}{p}\right)\left\{\log\frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_h|} + q\log\left(1 + \frac{p}{m}\right)\right\},\\
T_{LH} &= \sqrt{p}\left\{\frac{m}{p}\mathrm{tr}\left(\mathbf{S}_h\mathbf{S}_e^{-1}\right) - q\right\},\\
T_{BNP} &= \sqrt{p}\left(1 + \frac{p}{m}\right)\left[\left(1 + \frac{m}{p}\right)\mathrm{tr}\left\{\mathbf{S}_h(\mathbf{S}_e + \mathbf{S}_h)^{-1}\right\} - q\right].
\end{aligned} \tag{2.5}$$

Then we can treat the three statistics as T_G ($G = LR, LH$ and BNP) in a unified way. Expanding T_G in terms of

$$\mathbf{U} = \sqrt{p}\left(\frac{1}{p}\mathbf{B} - \mathbf{I}_q\right) \text{ and } \mathbf{V} = \sqrt{m}\left(\frac{1}{m}\mathbf{W} - \mathbf{I}_q\right),$$

we can write the characteristic function of T_G as

$$C_{T_G}(t) = \exp\left\{q(1+r)(it)^2\right\}\left\{1 + \frac{1}{\sqrt{p}}\left(itb_1 + (it)^3 b_3\right)\right\} + O(p^{-1}), \qquad (2.6)$$

where $r = p/m$, $b_1 = q(q+1)\{c(1+r)+r\}$ and $b_3 = 4q(1+r)\{c(1+r)+r\} + 4q(1-r^2)/3$, with $c = -p(m+p)^{-1}/2,\ 0,\ -p(m+p)^{-1}$ for $G = LR, LH$ and BNP, respectively.

By inverting (2.6) we obtain an asymptotic expansion of the distribution function of T_G as in the following theorem (Wakaki *et al.*, 2003).

THEOREM 2.1. *Let T_G be the transformed test statistics for $G = LR, LH$ and BNP defined by (2.5). Then*

$$\Pr\left(\frac{1}{\sigma}T_G \le z\right) = \Phi(z) - \phi(z)\frac{1}{\sqrt{p}}\left\{\frac{1}{\sigma}b_1 + \frac{1}{\sigma^3}b_3(z^2-1)\right\} + O(p^{-1}),$$

where Φ and ϕ are the distribution and the density functions of the standard normal distribution, $\sigma = \sqrt{2q(1+r)}$, and b_j's are given in (2.6).

From the above theorem we obtain Cornish-Fisher expansion of the upper percent point of the distribution as in the following corollary.

COROLLARY 2.1. *Let z_α be the upper $100(1-\alpha)\%$ point of the standard normal distribution, and let*

$$z_G(\alpha) = z_\alpha + \frac{1}{\sqrt{p}}\left\{\frac{b_1}{\sigma} + (z_\alpha^2 - 1)\frac{b_3}{\sigma^3}\right\}.$$

Then

$$\Pr\left(\frac{1}{\sigma}T_G \le z_G(\alpha)\right) = 1 - \alpha + O(p^{-1}).$$

The terms of $O(p^{-1})$ in Theorem 2.1 and Corollary 2.1 have been also obtained. For numerical accuracy of the expansions, see Wakaki *et al.* (2003), Tonda and Fujikoshi (2003), *etc.*

Now we shall examine a relationship between (n)- and (n,p)- asymptotics. For example, we have that for Lawley-Hotelling trace criterion $\mathrm{tr}\left(\mathbf{S}_h\mathbf{S}_e^{-1}\right)$,

$$\frac{1}{\sqrt{p}}\left\{m\mathrm{tr}\left(\mathbf{S}_h\mathbf{S}_e^{-1}\right)-pq\right\} \xrightarrow{D} N(1,2q(1+r)) : (n,p)\text{-asymptotics},$$

where $\xrightarrow{D}$ denotes convergence in distribution. On the other hand, first consider asymptotic distribution under the large sample framework such that n tends to infinity and p and q are fixed, and then consider asymptotic distribution of the resultant result when p tends to infinity. Then we have

$$\frac{1}{\sqrt{p}}\left\{m\mathrm{tr}(\mathbf{S}_h\mathbf{S}_e^{-1})-pq\right\} \xrightarrow{D} \frac{1}{\sqrt{p}}\left(\chi^2_{pq}-pq\right) : (n)\text{-asymptotics},$$
$$\frac{1}{\sqrt{p}}\left(\chi^2_{pq}-pq\right) \xrightarrow{D} N(1,2q) : (p)\text{-asymptotics}.$$

The final results in the two approaches are different in variances $2q(1+r)$ and $2q$ in their asymptotic normality. So, if $r=p/(n-p+q)$ is large, the difference becomes large. In other words, the existing (n)-asymptotics break down when p goes to infinity with n.

2.3. Non-null distributions

Note that the noncentrality matrix $\mathbf{\Omega}$ depends on n and p. It is natural to assume that

$$\text{(B0)}\ \mathbf{\Omega}=O(p), \tag{2.7}$$

which is equivalent to $\mathbf{\Omega}=O(n)$ under our high dimensional framework. Let

$$\begin{aligned}
T^*_{LR} &= -\sqrt{p}\left(1+\frac{m}{p}\right)\left\{\log\frac{|\mathbf{S}_e|}{|\mathbf{S}_e+\mathbf{S}_h|}+\log\left|\left(1+\frac{p}{m}\right)\mathbf{I}_q+\frac{1}{m}\mathbf{\Omega}\right|\right\},\\
T^*_{LH} &= \sqrt{p}\left\{\frac{m}{p}\mathrm{tr}\left(\mathbf{S}_h\mathbf{S}_e^{-1}\right)-\mathrm{tr}\left(\mathbf{I}_q+\frac{1}{p}\mathbf{\Omega}\right)\right\},\\
T^*_{BNP} &= \sqrt{p}\left(1+\frac{p}{m}\right)\left[\left(1+\frac{m}{p}\right)\mathrm{tr}\left\{\mathbf{S}_h(\mathbf{S}_e+\mathbf{S}_h)^{-1}\right\}\right.\\
&\qquad \left.-\,\mathrm{tr}\left\{\left(\mathbf{I}_q+\frac{1}{m+p}\mathbf{\Omega}\right)^{-1}\left(\mathbf{I}_q+\frac{1}{p}\mathbf{\Omega}\right)\right\}\right].
\end{aligned} \tag{2.8}$$

Then, we have the following result (Wakaki *et al.*, 2003).

THEOREM 2.2. *Let T_G^* be the transformed test statistics for $G = LR, LH$ and BNP defined by (2.8). Then, under the assumptions (A0) in (2.4) and (B0) in (2.7),*

$$T_G^* \xrightarrow{D} N(0, {\sigma_G^*}^2),$$

where $G = LR, LH, BNP$, and

$${\sigma_G^*}^2 = \mathrm{tr}\left\{\mathbf{I}_q + \frac{1}{(1+r)p}\mathbf{\Omega}\right\}^{-2(1+d_G)}\left\{\mathbf{I}_q + \frac{2}{p}\mathbf{\Omega} + r(\mathbf{I}_q + \frac{1}{p}\mathbf{\Omega})^2\right\},$$

and $d_G = 0, -1, 1$ for $G = LR, LH, BNP$, respectively.

From this theorem we can obtain asymptotic results of the powers of the three tests.

COROLLARY 2.2. *Let β_G be the power of test T_G with significance level α. If $\mathbf{\Omega} = O(\sqrt{p})$,*

$$\lim_{n,p\to\infty}\beta_G = \Phi\left(\frac{1}{\sqrt{2q(1+r)}}\mathrm{tr}\left(\frac{1}{\sqrt{p}}\mathbf{\Omega}\right) - z_\alpha\right).$$

Further, if the order of $\mathbf{\Omega}$ is larger than $\sqrt{p}$, the asymptotic power is one, while if the order of $\mathbf{\Omega}$ is smaller than $\sqrt{p}$, the asymptotic power is α.

2.4. Test for high dimensional data

When $n < p$, $\mathbf{S}_e$ becomes singular, and it will be impossible to use the classical statistics. For such cases, a non-exact test was first proposed by Dempster (1958, 1960) for one and two sample cases. The statistic may be generalized as $\mathrm{tr}(\mathbf{S}_h)\{\mathrm{tr}(\mathbf{S}_e)\}^{-1}$. For high dimensional asymptotics of $\mathrm{tr}(\mathbf{S}_h)\{\mathrm{tr}(\mathbf{S}_e)\}^{-1}$ we assume that

$$\text{(A1)}\ \frac{1}{p}\mathrm{tr}(\mathbf{\Sigma}^k) = O(1),\ \ k = 1, 2, \tag{2.9}$$

$$\text{(B1)}\ \frac{1}{p}\mathrm{tr}(\mathbf{\Sigma}^k\tilde{\mathbf{\Omega}}) = O(1),\ \ k = 1, 2, \tag{2.10}$$

where $\tilde{\mathbf{\Omega}} = \mathbf{\Sigma}^{-1/2}\mathbf{M}\mathbf{M}'\mathbf{\Sigma}^{-1/2}$. Consider a transformed statistic defined by

$$T_D^* = \sqrt{p}\left\{n\frac{\mathrm{tr}(\mathbf{S}_h)}{\mathrm{tr}(\mathbf{S}_e)} - q - \frac{\mathrm{tr}(\mathbf{\Sigma}\tilde{\mathbf{\Omega}})}{\mathrm{tr}(\mathbf{\Sigma})}\right\}. \tag{2.11}$$

Let

$$u = \frac{1}{\sqrt{p}}\Big\{\mathrm{tr}(\mathbf{S}_h) - q\mathrm{tr}(\mathbf{\Sigma}) - \mathrm{tr}(\tilde{\mathbf{\Omega}}\mathbf{\Sigma})\Big\} \text{ and } v = \frac{1}{\sqrt{np}}\Big\{\mathrm{tr}(\mathbf{S}_e) - n\mathrm{tr}(\mathbf{\Sigma})\Big\}.$$

Then, it is seen that u and v are asymptotically independently distributed as $N(0, \sigma_u^2)$ and $N(0, \sigma_v^2)$, respectively, where $\sigma_u^2 = 2qp^{-1}\mathrm{tr}(\mathbf{\Sigma}^2) + 4p^{-1}\mathrm{tr}(\mathbf{\Sigma}^2\tilde{\mathbf{\Omega}})$ and $\sigma_v^2 = 2p^{-1}\mathrm{tr}(\mathbf{\Sigma}^2)$. Expressing T_D^* in terms of u and v, Fujikoshi *et al.* (2003b) obtained the following result.

THEOREM 2.3. *Let T_D^* be the transformed test statistic defined by (2.11). Then, under the assumptions (A0) with $c > 0$, (A1) and (B1),*

$$T_D^* \xrightarrow{D} N(0, {\sigma_D^*}^2),$$

where

$${\sigma_D^*}^2 = \left\{\frac{1}{p}\mathrm{tr}(\mathbf{\Sigma})\right\}^{-2}\left[2q\left\{\frac{1}{p}\mathrm{tr}(\mathbf{\Sigma}^2)\right\} + 4\left\{\frac{1}{p}\mathrm{tr}(\mathbf{\Sigma}^2\tilde{\mathbf{\Omega}})\right\}\right].$$

In particular, under the hypothesis

$$T_D = \sqrt{p}\left\{n\frac{\mathrm{tr}(\mathbf{S}_h)}{\mathrm{tr}(\mathbf{S}_e)} - q\right\} \xrightarrow{D} N(0, {\sigma_D}^2),$$

where

$${\sigma_D}^2 = 2q\left\{\frac{1}{p}\mathrm{tr}(\mathbf{\Sigma})\right\}^{-2}\left\{\frac{1}{p}\mathrm{tr}(\mathbf{\Sigma}^2\mathbf{\Omega})\right\}.$$

COROLLARY 2.3. *Let β_D be the power of test T_D with significance level α. If $\mathbf{\Omega} = O(\sqrt{p})$,*

$$\lim_{n,p\to\infty} \beta_D = \Phi\left(\frac{1}{\sqrt{2q}}\frac{\mathrm{tr}(\mathbf{\Sigma}\tilde{\mathbf{\Omega}})}{\sqrt{\mathrm{tr}(\mathbf{\Sigma}^2)}} - z_\alpha\right). \tag{2.12}$$

Further, if the order of $\mathbf{\Omega}$ is larger than $\sqrt{p}$, the asymptotic power is one, while if the order of $\mathbf{\Omega}$ is smaller than $\sqrt{p}$, the asymptotic power is α.

Note that $\mathrm{tr}(\mathbf{\Omega}) = \mathrm{tr}(\tilde{\mathbf{\Omega}})$. Therefore, from Corollaries 2.2 and 2.3 we have

$$\frac{\mathrm{tr}\left(\tilde{\mathbf{\Omega}}/\sqrt{p}\right)}{\sqrt{1+r}} > \frac{\mathrm{tr}(\mathbf{\Sigma}\tilde{\mathbf{\Omega}})}{\sqrt{\mathrm{tr}(\mathbf{\Sigma}^2)}} \Rightarrow \lim_{n,p\to\infty} \beta_G > \lim_{n,p\to\infty} \beta_D,$$

$$\frac{\mathrm{tr}\left(\tilde{\mathbf{\Omega}}/\sqrt{p}\right)}{\sqrt{1+r}} < \frac{\mathrm{tr}(\mathbf{\Sigma}\tilde{\mathbf{\Omega}})}{\sqrt{\mathrm{tr}(\mathbf{\Sigma}^2)}} \Rightarrow \lim_{n,p\to\infty} \beta_G < \lim_{n,p\to\infty} \beta_D.$$

For example, if $\mathbf{\Sigma} = a\mathbf{I}_p$ (a: constant), $\lim_{n,p\to\infty}\beta_G < \lim_{n,p\to\infty}\beta_D$. Further, if p is near to n, T_D is more powerful than T_G.

It may be noted that Theorem 2.3 and Corollary 2.3 hold if the assumption $p/n \to c \in (0,+\infty)$ is replaced by $n = O(p^\delta)$ with $0 < \delta \le 1$. Further, for a practical use of the asymptotic result of T_D we need to estimate σ_D^2. It is suggested to use

$$\hat{\sigma}_D^2 = 2pq\frac{\mathrm{tr}(\mathbf{S}_e^2) - n^{-1}\{\mathrm{tr}(\mathbf{S}_e)\}^2}{\{\mathrm{tr}(\mathbf{S}_e)\}^2}$$

which is a (n,p)-consistent estimator (Srivastava, 2003).

2.5. Some other results

When $n \ge p$, the LR criterion for testing a linear hypothesis is based on

$$\Lambda = \frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_h|} = \prod_{j=1}^{t}(1+\ell_j)^{-1},$$

where $t = \min(p,q)$, and $\ell_1 > \cdots > \ell_t > 0$ are nonzero roots of $\mathbf{S}_h\mathbf{S}_e^{-1}$. When $n < p$, $\mathbf{S}_e$ becomes singular. Srivastava (2003) proposed to use

$$\Lambda^+ = \prod_{j=1}^{\tilde{t}}(1+d_j)^{-1},$$

where $\tilde{t} = \min(p,q)$, $d_1 > \cdots > d_{\tilde{t}} > 0$ are nonzero roots of $\mathbf{S}_h\mathbf{S}_e^+$, and $\mathbf{S}_e^+$ is the Moore-Penrose inverse matrix of $\mathbf{S}_e$. Consider a spectral decomposition

$$\mathbf{S}_e = \mathbf{HMH}',\ \mathbf{H}'\mathbf{H} = \mathbf{I}_k,\ \mathbf{M} = \mathrm{diag}(m_1,\dots,m_k),\ m_1 \ge \cdots \ge m_k > 0.$$

Then, the Moore-Penrose inverse matrix is defined by $\mathbf{S}_e^+ = \mathbf{HM}^{-1}\mathbf{H}'$. Some distributional results have been studied by Srivastava (2003).

In canonical discriminant analysis, it is important to decide the number of useful canonical discriminant functions which is defined by $\mathrm{rank}(\mathbf{\Omega})$. For testing the hypothesis

$$H_0 : \mathrm{rank}(\mathbf{\Omega}) = k$$

the following three statistics have been proposed:

$$\text{(i)}\ \log\prod_{j=k+1}^{q}(1+l_j);\quad \text{(ii)}\ \sum_{j=k+1}^{q} l_j;\quad \text{(iii)}\ \sum_{j=k+1}^{q}\frac{l_j}{1+l_j}.$$

Fujikoshi *et al.* (2003a) derived high dimensional asymptotic results as in Sections 2.2 and 2.3 for the above dimensionality tests. For the null distributions, see also Section 5.3.

3. DISCRIMINANT FUNCTIONS

We consider a discriminant problem of classifying a $p \times 1$ observation vector $\mathbf{x}$ as coming one of two normal populations $\Pi_1 : N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\Pi_2 : N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$. Here all the parameters $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ are unknown. Suppose that random samples of sizes N_1 and N_2 are available from Π_1 and Π_2, respectively. Let $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$ and $\mathbf{S}$ be the sample mean vectors and the pooled sample covariance matrix. Further, let Δ and D be the population and sample Mahalanobis distances. Let

$$N = N_1 + N_2, \quad n = N - 2, \quad m = n - p.$$

3.1. The case $n > p$

When $n > p$, there are two typical classification statistics W and Z defined by

$$W = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\left\{\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)\right\} \text{ and}$$
$$Z = a_2(\mathbf{x} - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_2) - a_1(\mathbf{x} - \bar{\mathbf{x}}_1)'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_1),$$

respectively, where $a_i = N_i/(N_i + 1)$, $i = 1, 2$. The rule is usually to classify $\mathbf{x}$ as coming from Π_1 if $W > 0$ (or $Z > 0$) and from Π_2 if $W \leq 0$ (or $Z \leq 0$). One of the important problems on the classification procedures is to evaluate the expected or unconditional probabilities of misclassification (EPMC), *i.e.*,

$$e_T(2|\,1) = \Pr(T \leq 0\,|\,\mathbf{x} \in \Pi_1),$$
$$e_T(1|\,2) = \Pr(T > 0\,|\,\mathbf{x} \in \Pi_2),$$

where $T = W$ or Z. The two discriminant functions W and Z can be treated in a unified way, by considering the discriminant function defined by

$$T = \frac{1}{2}\left\{(\mathbf{x} - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_2) - b(\mathbf{x} - \bar{\mathbf{x}}_1)'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_1)\right\},$$

where b is a constant. The rule is to classify $\mathbf{x}$ as coming from Π_1 if $T > c$ and from Π_2 if $T \leq c$, where c is a constant, particularly $c = 0$. The important special cases are given by putting b as follows.

$$\text{(i) if } b = 1,\ T = W; \quad \text{(ii) if } b = a_1 a_2^{-1},\ T = \frac{1}{2}a_2^{-1}Z.$$

We assume the following asymptotic framework:

$$\lim_{p\to\infty} \frac{p}{N_i} = \lambda_i\ (< 1),\ i = 1,\ 2, \qquad \lim_{p\to\infty} \Delta^2 = d_0^2. \tag{3.1}$$

Using that T can be expressed as a function of several independent normal and χ^2 variables, Fujikoshi and Seo (1998) obtained asymptotic distribution of T under the assumption (3.1) whose special case can be stated as in the following theorem.

THEOREM 3.1. *Suppose that* $\mathbf{x}$ *comes from* Π_i: $N_p(\boldsymbol{\mu}_i, \Sigma)$, $i = 1, 2$. *Then under (3.1),* W *and* $Z/2$ *are asymptotically distributed as* $N(\zeta_W^{(i)}, \sigma^2)$ *and* $N(\zeta_Z^{(i)}, \sigma^2)$, *respectively, where*

$$\begin{aligned}
\zeta_W^{(i)} &= \frac{1}{2}\left(\frac{N}{N-p}\right)\left\{\Delta^2 + \frac{p}{N_1 N_2}(-1)^{i+1}(N_1 - N_2)\right\}, \ i = 1, 2, \\
\zeta_Z^{(1)} &= \zeta_Z^{(2)} = \frac{1}{2}\left(\frac{N}{N-p}\right)\Delta^2, \\
\sigma^2 &= \left(\frac{N}{N-p}\right)^3\left(\Delta^2 + \frac{pN}{N_1 N_2}\right).
\end{aligned}$$

COROLLARY 3.1. *Under (3.1) the EPMC's for* W *and* Z *are asymptotically evaluated as follows:*

$$\begin{aligned}
\lim e_W(2|\,1) &= \Phi(\gamma_W^{(1)}), \quad \lim e_W(1|\,2) = \Phi(\gamma_W^{(2)}), \\
\lim e_Z(2|\,1) &= \Phi(\gamma_Z^{(1)}), \quad \lim e_Z(1|\,2) = \Phi(\gamma_Z^{(2)}),
\end{aligned}$$

where

$$\begin{aligned}
\gamma_W^{(i)} &= -\frac{1}{2}\left(\frac{N-p}{N}\right)^{1/2}\left\{\Delta^2 + \frac{p}{N_1 N_2}(-1)^{i+1}(N_1 - N_2)\right\} \\
&\quad \times \left(\Delta^2 + \frac{pN}{N_1 N_2}\right)^{-1/2}, \ i = 1, 2, \\
\gamma_Z^{(1)} &= \gamma_Z^{(2)} = -\frac{1}{2}\left(\frac{N-p}{N}\right)^{1/2}\Delta^2\left(\Delta^2 + \frac{pN}{N_1 N_2}\right)^{-1/2}.
\end{aligned}$$

In a special case $N_1 = N_2$ we obtain

$$\gamma_W^{(1)} = \gamma_W^{(2)} = -\frac{1}{2}\left(\frac{N-p}{N}\right)^{1/2}\Delta^2\left(\Delta^2 + \frac{pN}{N_1 N_2}\right)^{-1/2}.$$

which was proposed by Raudys (1972).

For high dimensional asymptotic approximations of $e(2|\,1)$, it may be noted that there are some earlier works by Deev (1970), Raudys (1972), *etc.* Thier asymptotic approximation proposed for $N_1 = N_2$ has been compared with the classical asymptotic approximations by Wyman *et al.* (1990). For an extension to the case $N_1 \neq N_2$, see Fujikoshi and Seo (1998). The comparison shows that the high dimensional approximations are also extremely accurate. Saranadasa (1993) proposed an approximation for $e_Z(2|1)$ which is different from the one as in Corollary 3.1.

For a practical use of Corollary 3.1, we need to estimate Δ^2. It is recommended to use

$$D_u^2 = \frac{n-p-1}{n}D^2 - \frac{pN}{N_1N_2}$$

which is unbiased and consistent under our framework.

3.2. Error bound for asymptotic approximations

Lachenbruch (1968) has proposed the approximations for $e_W(2|1)$ and $e_W(1|2)$, which are expressed as

$$e_W(2|1) \simeq \Phi(\gamma_L^{(1)}), \quad e_W(1|2) \simeq \Phi(\gamma_L^{(2)}), \tag{3.2}$$

where for $i = 1, 2$,

$$\begin{aligned}\gamma_L^{(i)} &= -\frac{1}{2}\left\{\frac{m(m-3)}{(n-1)(m-1)}\right\}^{1/2} \\ &\quad \times \left\{\Delta^2 + \frac{p}{N_1N_2}(-1)^{i+1}(N_1 - N_2)\right\}\left(\Delta^2 + \frac{pN}{N_1N_2}\right)^{-1/2}.\end{aligned}$$

The approximations were proposed, without considering their asymptotic properties. In fact, he proposed them by substituting their expectations to the mean and variance in the conditional normality. From Corollary 3.1, the above approximations can be justified as the asymptotic results under (3.1).

In this section we are interested in deriving their error bounds whose orders are the first order with respect to $(N_1^{-1}, N_2^{-1}, p^{-1})$. Suppose that $\mathbf{x}$ belongs to Π_1. Then we can express W as

$$W = V^{-1/2}\tilde{Z} - U, \tag{3.3}$$

where

$$\begin{aligned}V &= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \\ \tilde{Z} &= V^{-1/2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1), \\ U &= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_1) - \frac{1}{2}D^2.\end{aligned}$$

Here $\tilde{Z}$ is distributed as $N(0,1)$, and $\tilde{Z}$ and (U,V) are independent. Therefore we can write

$$\begin{aligned} e_W(2|1) &= \mathrm{P}(W \le 0 \mid \mathbf{x} \in \Pi_1) \\ &= E_{(U,V)}\left\{\Phi(V^{-1/2}U)\right\}. \end{aligned} \tag{3.4}$$

Note that

$$\begin{aligned} E(U) &= -\frac{n}{2(m-1)}\left\{\Delta^2 + \frac{(N_1-N_2)p}{N_1N_2}\right\} = u_0 \quad (m>1), \\ E(V) &= \frac{n^2(n-1)}{m(m-1)(m-3)}\left\{\Delta^2 + \frac{Np}{N_1N_2}\right\} = v_0 \ (m>3), \end{aligned}$$

and hence $\gamma_L^{(1)} = E(V)^{-1/2}E(U)$. Therefore, the first approximation in (3.2) can be regarded as the one obtained from (3.4) by substituting $(E(U), E(V))$ to (U,V), *i.e.*

$$\Phi\left(\gamma_L^{(1)}\right) = \Phi\left(\{E(V)\}^{-1/2}E(U)\right). \tag{3.5}$$

More precisely, first we consider approximating $\Phi(v^{-1/2}u)$ by its Taylor type approximation at $(u,v) = (u_0, v_0)$ up to the first order and to evaluate its remainder terms which are the second order with respect to $(u-u_0, v-v_0)$. Then, we consider taking the mean of resultant inequality with respect to (U,V).

The variances of U and V are given as follows:

$$\begin{aligned} \mathrm{Var}(U) &= \frac{n^2}{2m(m-1)(m-3)}\left[\frac{1}{m-1}\Delta^4 + \frac{2(n-1)}{mN_2}\Delta^2\left\{1+\frac{N_1-N_2}{(m-1)N_1}\right\}\right. \\ &\quad \left. + \frac{2(n-1)p}{N_1N_2}\left\{\frac{1}{m} + \frac{(N_1-N_2)^2}{2(m-1)N_1N_2}\right\}\right], \quad m-3>0, \\ \mathrm{Var}(V) &= \frac{2(n-1)n^4}{m(m-1)^2(m-3)^2}\left[\frac{1}{m}\left\{1+\frac{8(m-4)}{(m-5)(m-7)}\right\}\right. \\ &\quad \times\left\{\frac{p-1}{m}\left(\Delta^2+\frac{pN}{N_1N_2}\right)^2 + \frac{N(n-3)}{N_1N_2}\left(2\Delta^2+\frac{pN}{N_1N_2}\right)\right\} \\ &\quad \left. + \frac{4(n-1)(m-4)}{m(m-5)(m-7)}\left(\Delta^2+\frac{pN}{N_1N2}\right)^2\right], \quad m-7>0. \end{aligned}$$

THEOREM 3.2. *The error when we approximate $e_W(2|1)$ by $\Phi(\gamma_L^{(1)})$ is bounded by B, i.e.,*

$$\left|e_W(2|1) - \Phi(\gamma_L^{(1)})\right| \le B. \tag{3.6}$$

where

$$B = \beta_{2,0}v_0^{-1}\mathrm{Var}(U) + \beta_{2,2}v_0^{-2}\mathrm{Var}(V) + \beta_{2,1}v_0^{-3/2}\{\mathrm{Var}(U)\cdot\mathrm{Var}(V)\}^{1/2},$$

and the constants $\beta_{2,j}$ *are given by*

$$\beta_{2,0} = \frac{1}{2}h_1, \quad \beta_{2,1} = \frac{1}{2}h_2, \quad \beta_{2,2} = \frac{1}{2}\left(\sqrt{1+h_1} + \frac{1}{2}\sqrt{h_3}\right)^2.$$

Here $h_j = \sup|h_j(x)\phi(x)|$ *and* $h_j(x)$ *is the* j^{th} *Hermite polynomial.*

Note that the order of B is O_1^*, where O_j^* denotes the j^{th} order with respect to $(N_1^{-1}, N_2^{-1}, p^{-1})$. It has been also proved (Fujikoshi, 2000) that the constants $\beta_{i,j}$ may be replaced by their improved constants,

$$\tilde{\beta}_{2,0} = 0.121, \quad \tilde{\beta}_{2,1} = 0.2, \quad \tilde{\beta}_{2,2} = 0.5.$$

3.3. The case $p > n$

When $p > n$, the classical discrimination rule becomes inapplicable since $\mathbf{S}$ is singular. As one of the approaches for the case, Saranadasa (1993) proposed a method to look at the classification problem in terms of one way MANOVA problem. The method is to classify a new observation by minimizing a suitable norm of the matrix of sums of squares and products due to within groups. Let $\mathbf{E}_i$ be the new matrix of sums of squares and products due to within groups when $\mathbf{x}$ is placed in Π_i. Then

$$\mathbf{E}_i = n\mathbf{S} + a_i(\mathbf{x} - \bar{\mathbf{x}}_1)(\mathbf{x} - \bar{\mathbf{x}}_1)'.$$

Using trace criterion, we may classify $\mathbf{x}$ into Π_1 if

$$\mathrm{tr}(\mathbf{E}_1) < \mathrm{tr}(\mathbf{E}_2).$$

This is equivalent to the classification procedure based on the classification statistic defined by

$$T = a_1(\mathbf{x} - \bar{\mathbf{x}}_1)'(\mathbf{x} - \bar{\mathbf{x}}_1) - a_2(\mathbf{x} - \bar{\mathbf{x}}_2)'(\mathbf{x} - \bar{\mathbf{x}}_2). \tag{3.7}$$

Saranadasa (1993) obtained asymptotic expansions for $e_T(2|1)$ and $e_T(1|2)$ by expressing T as a sum of p independent but not necessarily identically distributed random variables. The leading terms are given as in the following theorem.

THEOREM 3.3. *Let T be the classification statistic defined by (3.7). Then, for large p,*

$$e_T(2|\,1) \simeq \Phi(\gamma_T^{(1)}), \quad e_T(1|\,2) \simeq \Phi(\gamma_T^{(2)}),$$

where

$$\gamma_T^{(1)} = \frac{a_2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{2\sqrt{(1 - a_1 a_2)\mathrm{tr}(\boldsymbol{\Sigma}^2) + a_2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}},$$

and $\gamma_T^{(2)}$ is defined from $\gamma_T^{(1)}$ by substituting a_2 for a_1.

Along the idea of regularized discriminant analysis due to Friedman (1989), Loh (1997) studied an adapting discriminant function defined by

$$W_\lambda = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'(\mathbf{S} + \lambda \mathbf{I})^{-1}\left\{\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)\right\}, \tag{3.8}$$

where $\lambda = \lambda(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{S})$ and λ is a smooth function. Some large sample properties have been obtained, but its high dimensional properties have not been studied.

4. TESTS FOR COVARIANCE MATRIX

4.1. Preliminaries

Let $\mathbf{S}$ be the sample covariance matrix based on a sample of size $N = n + 1$ from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We consider two common testing problems for the covariance matrix $\boldsymbol{\Sigma}$: (i) the covariance matrix $\boldsymbol{\Sigma}$ is equal to the identity matrix $\mathbf{I}$; (ii) the covariance matrix $\boldsymbol{\Sigma}$ is proportional to the identity matrix (sphericity). The identity in (i) can be replaced with any other given matrix $\boldsymbol{\Sigma}_0$. For each testing problems, we have the likelihood ratio test statistics and the other tests based on quadratic forms of the eigenvalues of $\mathbf{S}$.

Our interest is to consider asymptotic properties of these tests for the situations where (1) $n > p$, but p is relatively large, and (2) $p > n$ or $p \gg n$. It is assumed that

$$\text{(C0)}\ \frac{p}{n} \to c \in (0, +\infty). \tag{4.1}$$

If $c > 1$, $\mathbf{S}$ becomes singular. So, we cannot use the likelihood ratio tests since they depend on $|\mathbf{S}|$. In addition to the assumption (C0), we use the following assumptions:

(C1) $\dfrac{1}{p}\mathrm{tr}(\Sigma^k) = O(1),\ k = 1, 2.$

(C2) $\dfrac{1}{p}\mathrm{tr}(\Sigma^k) = O(1),\ k = 3, 4.$

Let $\lambda_1 \geq \cdots \geq \lambda_p$ be the eigenvalues of $\mathbf{\Sigma}$, and let

$$\frac{1}{p}\sum_{i=1}^{p}\lambda_i = \alpha, \quad \frac{1}{p}\sum_{i=1}^{p}(\lambda_i - \alpha)^2 = \delta^2.$$

Then $p^{-1}\mathrm{tr}\{(\mathbf{\Sigma} - \mathbf{I})^2\} = (\alpha - 1)^2 + \delta^2$.

Note that $n\mathbf{S}$ is distributed as a central Wishart distribution $W_p(n, \mathbf{\Sigma})$. Then the following two lemmas were given in Ledoit and Wolf (2002).

LEMMA 4.1 (Law of large numbers). *Under Assumptions (C0)–(C2),*

(1) $\frac{1}{p}\mathrm{tr}(\mathbf{S}) \xrightarrow{P} \alpha$,

(2) $\frac{1}{p}\mathrm{tr}(\mathbf{S}^2) \xrightarrow{P} (1+c)\alpha^2 + \delta^2$,

where $\xrightarrow{P}$ *denotes convergence in probability.*

LEMMA 4.2 (Central limit theorem). *Under Assumptions (C0), (C1), if* $\delta^2 = 0$, *then*

$$n \times \begin{bmatrix} \frac{1}{p}\mathrm{tr}(\mathbf{S}) \\ \frac{1}{p}\mathrm{tr}(\mathbf{S}^2) - \frac{n+p+1}{n}\alpha^2 \end{bmatrix} \xrightarrow{D} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}\right)$$

where $\xrightarrow{D}$ *denotes convergence in distribution, and*

$$\gamma_{11} = 2c^{-1}\alpha^2, \quad \gamma_{12} = \gamma_{21} = 4(1+c^{-1})\alpha^2, \quad \gamma_{22} = 4(2c^{-1} + 5 + 2c)\alpha^4.$$

4.2. Testing problem (i)

For testing

$$H_1 : \mathbf{\Sigma} = \mathbf{I}_p \ vs.\ K_1 : \mathbf{\Sigma} \neq \mathbf{I}_p,$$

we have two typical test statistics given by

$$\begin{aligned} -2\log LR_1 &= n\mathrm{tr}(\mathbf{S}) - n\log|\mathbf{S}| - pn, \\ T_1 &= \frac{1}{p}\mathrm{tr}\left\{(\mathbf{S} - \mathbf{I}_p)^2\right\}, \end{aligned}$$

where LR_1 is a modified LR criterion. The first statistic $-2\log LR_1$ can be used only for $n > p$, while the statistic T_1 can be used for any p and n. It is easy to see that under the null hypothesis

$$\begin{aligned} nT_1 - p &\xrightarrow{D} \frac{2}{p}\chi^2_{p(p+1)/2} - p \ : (n)\text{-asymptotics}, \\ \frac{2}{p}\chi^2_{p(p+1)/2} - p &\xrightarrow{D} N(1,4) \ : (p)\text{-asymptotics}. \end{aligned}$$

Ledoit and Wolf (2002) studied asymptotic behaviors of T_1 under the assumptions (C0)–(C2): They showed

$$\begin{aligned} T_1 &\xrightarrow{P} c\alpha^2 + (\alpha-1)^2 + \delta^2 \ : (n,p)\text{-asymptotics}, \\ nT_1 - p &\xrightarrow{D} N(1, 4+8c) \ : (n,p)\text{ -asymptotics}. \end{aligned}$$

These imply that (1) the existing n-consistency of T_1 does not extend to (n,p)-consistency, and (2) the n-asymptotic distribution of T_1 breaks down when p goes to infinity with n. As one of the test statistics overcoming these weak points, they proposed a modified statistic defined by

$$\tilde{T}_1 = \frac{1}{p}\mathrm{tr}\left\{(\mathbf{S}-\mathbf{I}_p)^2\right\} - \frac{1}{n}\left\{\frac{1}{p}\mathrm{tr}(\mathbf{S})\right\}^2 + \frac{p}{n}.$$

Wakaki (2003) has obtained asymptotic expansions of the null and nonnull distributions of the modified likelihood ratio criterion $\lambda = LR_1$ under a high dimensional framework (C0) with $0 < c < 1$. It is well known (see, *e.g.*, Anderson, 1984) that the characteristic function of $\log\lambda$ is expressed as

$$\begin{aligned} E\{\exp(it\log\lambda)\} &= E(\lambda^{it}) \\ &= 2^{pnit/2}\Gamma_p(n(1+it)/2)\left\{\Gamma_p(n/2))\right\}^{-1} \\ &\quad \times|\mathbf{I}_p + it\mathbf{\Sigma}|^{-n(1+it)/2}|\mathbf{\Sigma}|^{nit/2}, \end{aligned}$$

where $\Gamma_p(a) = \pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(a-(i-1)/2)$. Therefore, the cumulant generating function of $2n^{-1}\log\lambda$ is expanded as

$$\begin{aligned} K(t) &= \log E\{\exp\left(2n^{-1}it\log\lambda\right)\} \\ &= it\mu_n + \frac{1}{2}(it)^2\sigma_n^2 + \frac{1}{6}(it)^3\gamma_{3,n} + \cdots, \end{aligned}$$

where

$$\begin{aligned}\mu_n &= p\log 2 + \log|\mathbf{\Sigma}| - \mathrm{tr}(\mathbf{\Sigma}) + \psi_p\left(\frac{n}{2}\right),\\ \sigma_n^2 &= \psi'\left(\frac{n}{2}\right) - \frac{2}{n}p + \frac{2}{n}\mathrm{tr}\left\{(\mathbf{\Sigma} - \mathbf{I}_p)^2\right\},\\ \gamma_{k,n} &= \psi_p^{(k-1)}\left(\frac{n}{2}\right) - \left(-\frac{2}{n}\right)^{(k-1)}(k-1)!\\ &\quad \times\mathrm{tr}\left[\mathbf{\Sigma}^{k-1}\left\{\Sigma - k(k-1)^{-1}\mathbf{I}_p\right\}\right], \quad k = 3, 4, \ldots\end{aligned}$$

Here ψ is the digamma function defined by

$$\psi(a) = \frac{d}{da}\log\Gamma(a) = -C + \sum_{k=0}^{\infty}\left(\frac{1}{1+k} - \frac{1}{a+k}\right) = O(\log a).$$

Note that

$$\psi^{(s)}(a) = \sum_{k=0}^{\infty}\frac{-s(-1)^s}{(a+k)^{s+1}} = O(a^{-s}), \quad \psi_p(a) = \sum_{j=1}^{p}\psi\left(a - \frac{1}{2}(j-1)\right).$$

THEOREM 4.1. *Under Assumptions (C0)-(C2), the distribution function of* LR_1 *can be expanded as*

$$\Pr\left(\frac{2n^{-1}\log LR_1 - \mu_n}{\sigma_n} \le x\right) = \Phi(x) - \frac{1}{6}\gamma_{n,3}\phi(x)(x^2 - 1) + O(n^{-2}).$$

Wakaki (2003) noted that the new approximation is more accurate than the classical χ^2-type asymptotic expansions.

4.3. Testing problem (ii)

For testing

$$H_2 : \mathbf{\Sigma} = \sigma^2\mathbf{I}_p \ vs.\ K_2 : \mathbf{\Sigma} \neq \sigma^2\mathbf{I}_p,$$

we have two typical test statistics

$$\begin{aligned}-2\log LR_2 &= n\left[\log\left\{\frac{1}{p}\mathrm{tr}(\mathbf{S})\right\}^p - \log|\mathbf{S}|\right],\\ T_2 &= \frac{1}{p}\mathrm{tr}\left[\left\{\frac{\mathbf{S}}{(1/p)\mathrm{tr}(\mathbf{S})} - \mathbf{I}_p\right\}^2\right].\end{aligned}$$

It has been shown (Ledoit and Wolf, 2002) that T_2 is (n,p)-consistent, and its (n)-asymptotic distribution remains valid if p goes to infinity with n, even for the case $p > n$.

Wakaki (2003) has studied asymptotic distribution of the null distribution of $-2 \log LR_2$ under the assumptions (C0)–(C2) with $0 < c < 1$.

5. DISTRIBUTIONS OF EIGENVALUES

5.1. Spectral distribution

Let $\ell_1 \geq \cdots \geq \ell_p$ be the eigenvalues of a $p \times p$ symmetric matrix $\mathbf{S}$. Then the spectral distribution of the matrix $\mathbf{S}$ is defined by

$$F_n(x) = \frac{1}{p} \sharp\{\ell_i : \ell_i \leq x\}, \tag{5.1}$$

where $\sharp\{\cdot\}$ denotes the number of elements of the set $\{\ \}$. If the empirical distribution function $F_n(x)$ converges to F, then F is called the limiting spectral distribution (LSD) function of $\mathbf{S}$.

Let $\mathbf{S}$ be the matrix defined by

$$\mathbf{S} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j{}', \tag{5.2}$$

where $\mathbf{x}_j = (x_{1j}, \ldots, x_{pj})'$ and x_{ij} are *iid* random variables with mean 0 and variance σ^2.

In spectral analysis of high dimensional random matrix the sample covariance matrix is defined by (5.2), not by $\mathbf{S} = n^{-1} \sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$. This is not essential under the assumption of normality, but is essential under the assumption of non-normality.

The LSD of $\mathbf{S}$ was first obtained by Marčenko and Pastur (1967). Subsequent work was done by Jonsson (1982), Yin (1986), Bai *et al.* (1987), *etc.* For a review of LSD, see Bai (1999).

THEOREM 5.1. *Let F_n be the spectral distribution of the sample covariance matrix defined by (5.2). Under the assumption that $p/n \to c \in (0, 1)$ as $p \to \infty$. Then*

$$F_n \xrightarrow{a.s.} F,$$

where $\xrightarrow{a.s.}$ denotes almost surely convergence,

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} \dfrac{1}{2\pi c x \sigma^2} \sqrt{(x-a)(b-x)}, & a < x < b, \\ 0, & \textit{otherwise}, \end{cases}$$

and the constants a and b are given by $a = (1 - \sqrt{c})^2\sigma^2$ and $b = (1 + \sqrt{c})^2\sigma^2$, respectively.

The LSD is useful in obtaining asymptotic behavior of a symmetric function of the eigenvalues of $\mathbf{S}$. In fact, we have

$$\begin{aligned} T_n &\equiv \frac{1}{p}\{\phi(\ell_1) + \cdots + \phi(\ell_p)\} \\ &= \int_0^\infty \phi(x)dF_n(x). \end{aligned} \tag{5.3}$$

Then, under appropriate regularity conditions we can see that T_n converges to

$$\int_0^\infty \phi(x)dF(x) \tag{5.4}$$

under a high dimensional framework.

Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be independently distributed as Wishart distributions $W_p(m, \mathbf{I}_p)$ and $W_p(n, \mathbf{I}_p)$, respectively. Assume that

$$\frac{p}{m} \to c_1 > 0 \quad \text{and} \quad \frac{p}{n} \to c_2 \in \left(0, \frac{1}{2}\right). \tag{5.5}$$

Then, the LSD of $\mathbf{F} = \mathbf{S}_1\mathbf{S}_2^{-1}$ was also studied by many authors. Bai (1999) extended the result to a non-normal case.

5.2. The maximum eigenvalue

In this section we consider asymptotic distribution of the maximum eigenvalue ℓ_1 of $\mathbf{S}$. Under certain condition on moments, Geman (1980) showed the almost sure ($a.s.$) convergence given by

$$n^{-1}\ell_1 \xrightarrow{a.s.} (1 + \sqrt{c})^2,$$

that is, $\ell_1 \sim (\sqrt{n} + \sqrt{p})^2$. The result was generalized by Yin *et al.* (1988) under the assumption that the entries of $\mathbf{X}$ have finite fourth moment.

Suppose that the entries of $\mathbf{X}$ are independently distributed as $N(0,1)$, and hence $n\mathbf{S}$ is distributed as $W_p(n, \mathbf{I}_p)$. Then, the limiting distribution of ℓ_1 was derived by Johnstone (2001). The center and scaling constants are defined by

$$\mu_{np} = (\sqrt{n-1} + \sqrt{p})^2, \tag{5.6}$$

$$\sigma_{np} = (\sqrt{n-1} + \sqrt{p})\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3}. \tag{5.7}$$

The limiting distribution function is defined by

$$F_1(s) = \exp\left[-\frac{1}{2}\int_s^\infty \{q(x) + (x-s)q^2(x)\}\,dx\right], \quad s \in R,$$

where $q(x)$ solves the Painlevé II differential equation

$$\begin{aligned} q''(x) &= xq(x) + 2q^3(x), \\ q(x) &\sim \mathrm{Ai}(x) \text{ as } x \to +\infty \end{aligned}$$

and $\mathrm{Ai}(x)$ denotes the Airy function. The final result is given as in the following Theorem 5.2.

THEOREM 5.2. *Suppose that* $n\mathbf{S}$ *is distributed as* $W_p(n, \mathbf{I}_p)$, *and let* ℓ_1 *be the maximum eigenvalue of* $\mathbf{S}$. *Then, under that* $p/n \to c \in (0, 1]$,

$$w_1 \equiv \frac{n\ell_1 - \mu_{np}}{\sigma_{np}} \xrightarrow{D} F_1.$$

The limiting distribution function is complicated, but the function has been numerically evaluated.

5.3. The eigenvalues in MANOVA

Let $\mathbf{S}_h$ and $\mathbf{S}_e$ be the matrices of sums of squares and products due to the hypothesis and the error in a MANOVA model. Further, we assume that $\mathbf{S}_h$ and $\mathbf{S}_e$ are independently distributed as a noncentral Wishart distribution $W_p(q, \boldsymbol{\Sigma}; \mathbf{MM}')$ and a central Wishart distribution $W_p(n, \boldsymbol{\Sigma})$, respectively. Let $\ell_1 > \cdots > \ell_t > 0$ be the non-zero eigenvalues of $\mathbf{S}_h\mathbf{S}_e^{-1}$, where $t = \min(p, q)$. Let $\boldsymbol{\Omega} = \mathbf{M}'\boldsymbol{\Sigma}^{-1}\mathbf{M}$ be the noncentrality matrix. Assume that $\mathrm{rank}(\boldsymbol{\Omega}) = k$, and let $\omega_1 \geq \cdots \geq \omega_k > 0$ be the non-zero eigenvalues of $\boldsymbol{\Omega}$. The following result was obtained by Fujikoshi *et al.* (2003), by using perturbation expansion of the eigenvalues.

THEOREM 5.3. *Assume that* $\boldsymbol{\Omega} = O(p)$ *and*

$$\omega_1 > \cdots > \omega_k > \omega_{k+1} = \cdots = \omega_q = 0.$$

Then, under the assumption that $p/n \to c \in (0, 1)$, *we have the following results.*

(1) $\ell_1, \ldots, \ell_k, \{\ell_{k+1}, \ldots, \ell_q\}$ *are asymptotically independent.*

(2) *For* $\alpha = 1, \dots, k$,

$$\sqrt{m}(\ell_\alpha - \mu_\alpha) \xrightarrow{D} N(0, \sigma_\alpha^2),$$

where $\mu_\alpha = r(1+\omega_\alpha p^{-1})$, $m = n-p+q$, $r = p/m$, $\sigma_\alpha^2 = 2(-r+2\mu_\alpha+\mu_\alpha^2)$.

(3) *Let* $y_j = \sqrt{m}(\ell_{k+j} - r)/\sqrt{r(1+r)}$, $j = 1, \dots, a = q-k$. *Then, the limiting density function of* $(y_1, \dots, y_k)$ *is given by*

$$\frac{\pi^{a(a-1)/4}}{\Gamma_a(a/2)2^{a(a+3)/4}} \exp\left(-\sum_{j=1}^{a} y_j^2\right) \prod_{i<j}^{a} (y_i - y_j),$$

where $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma(a - (i-1)/2)$.

Let

$$T_{LR}^{(k)} = -\sqrt{p}\left(1 + \frac{m}{p}\right)\left\{\log \prod_{j=k+1}^{q} (1+l_j)^{-1} + (q-k)\log\left(1+\frac{p}{m}\right)\right\},$$

$$T_{LH}^{(k)} = \sqrt{p}\left(\frac{m}{p}\sum_{j=k+1}^{q} l_j - (q-k)\right),$$

$$T_{BNP}^{(k)} = \sqrt{p}\left(1+\frac{p}{m}\right)\left\{\left(1+\frac{m}{p}\right)\sum_{j=k+1}^{q}\frac{l_j}{1+l_j} - (q-k)\right\}.$$

Then, from Theorem 5.3 (3) we have

$$T_G^{(k)} \xrightarrow{D} N\big(0, 2(q-k)(1+r)\big),$$

for the null distribution of $T_G^{(k)}$ for $G = LR, LH, BNP$. Fujikoshi *et al.* (2003a) have noted that the high dimensional approximations are numerically more accurate than the classical large sample approximations.

ACKNOWLEDGEMENTS

The author would like to thank two associate editors for their careful readings and useful comments.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley & Sons, New York.

Bai, Z. D. (1999). "Methodologies in spectral analysis of large-dimensional random matrices, a review", *Statistica Sinica*, **9**, 611–677.

Bai, Z. D., Yin, Y. Q. and Krishnaiah, P. R. (1987). "On the limiting empirical distribution function of the eigenvalues of a multivariate F matrix", *Theory of Probability and Its Applications*, **32**, 490–500.

Deev, A. D. (1970). "Representation of statistics of discriminant analysis and asymptotic expansion when space dimensions are comparable with sample size", *Soviet Mathematics Doklady*, **11**, 1547–1550.

Dempster, A. P. (1958). "A high dimensional two sample significance test", *The Annals of Mathematical Statistics*, **29**, 995–1010.

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

Friedman, J. H. (1989). "Regularized discriminant analysis", *Journal of the American Statistical Association*, **84**, 165–175.

Fujikoshi, Y. (2000). "Error bounds for asymptotic approximations of the linear discriminant function when the sample sizes and dimensionality are large", *Journal of Multivariate Analysis*, **73**, 1–17.

Fujikoshi, Y., Himeno, T. and Wakaki, H. (2003a). "Asymptotic results in canonical discriminant analysis when the dimension is large compared to the sample size", Technical report No. 03–17, Statistical Research Group, Hiroshima University.

Fujikoshi, Y., Himeno, T. and Wakaki, H. (2003b). "Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size", Technical report No. 03–18, Statistical Research Group, Hiroshima University.

Fujikoshi, Y. and Seo, T. (1998). "Asymptotic approximations for EPMC's of the linear and the quadratic discriminant functions when the samples sizes and the dimension are large", *Random Operators and Stochastic Equations*, **6**, 269–280.

Geman, S. (1980). "A limit theorem for the norm of random matrices", *The Annals of Probability*, **8**, 252–261.

Johnstone, I. M. (2001). "On the distribution of the largest eigenvalue in principal component analysis", *The Annals of Statistics*, **29**, 295–327.

Jonsson, D. (1982). "Some limit theorems for eigenvalues of a sample covariance matrix", *Journal of Multivariate Analysis*, **12**, 1-38.

Lachenbruch, P. A. (1968). "On expected probabilities of misclassification in discriminant analysis, necessary sample size, and a relation with the multiple correlation coefficient", *Biometrics*, **24**, 823–834.

Ledoit, O. and Wolf, M. (2002). "Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size", *The Annals of Statistics*, **30**, 1081–1102.

Loh, W-L. (1997). "Linear discrimination with adaptive ridge classification rules", *Journal of Multivariate Analysis*, **62**, 169–180.

Marčenko, V. A. and Pastur, L. A. (1967). "Distribution of eigenvalues of some sets of random matrices", *Mathematics of the USSR-Sbornik*, **1**, 457–483.

Mudholkar, G. S. and Trivedi, M. C. (1980). "A normal approximation for the distribution of the likelihood ratio statistic in multivariate analysis of variance", *Biometrika*, **67**, 485–488.

RAUDYS, S. (1972). "On the amount of a priori information in designing the classification algorithm", *Proceedings of Academy of Sciences of the USSR, Tech. Cyber.*, **4**, 168–174.

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", *Journal of Multivariate Analysis*, **46**, 154–174.

SRIVASTAVA, M. S. (2003). "Multivariate analysis with fewer observations than the dimension", submitted.

SRIVASTAVA, M. S. AND YAU, W. K. (1989). "Saddlepoint method for obtaining tail probability of Wilks' likelihood ratio test", *Journal of Multivariate Analysis*, **31**, 117–126.

TONDA, T. AND FUJIKOSHI, Y. (2003). "Asymptotic expansion of the null distribution of LR statistic for multivariate linear hypothesis when the dimension is large", to appear in *Communications in Statistics-Theory and Methods.*

WAKAKI, H. (2003). "Asymptotic expansions for likelihood ratio test statistics on covariance matrices when the sample size and the dimension are large", reported in the Meeting of Japan Statistical Society, September.

WAKAKI, H., FUJIKOSHI, Y. AND ULYANOV, V. (2003). "Asymptotic expansions of the distributions of MANOVA test statistics when the dimension is large", submitted.

WYMAN, F. J., YOUNG, D. M. AND TURNER, D. W. (1990). "A comparison of asymptotic error rate expansions for the sample linear discriminant function", *Pattern Recognition*, **23**, 775–783.

YIN, Y. Q. (1986). "Limiting spectral distribution for a class of random matrices", *Journal of Multivariate Analysis*, **20**, 50–68.

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", *Probability Theory and Related Fields*, **78**, 509–521.