ASYMPTOTIC RESULTS OF A HIGH DIMENSIONAL MANOVA TEST AND POWER COMPARISON WHEN THE DIMENSION IS LARGE COMPARED TO THE SAMPLE SIZE

Yasunori Fujikoshi**, Tetsuto Himeno* and Hirofumi Wakaki*

This paper is concerned with Dempster trace criterion for multivariate linear hypothesis which was proposed for high dimensional situation. First we derive asymptotic null and nonnull distributions of Dempster trace criterion when both the sample size and the dimension tend to infinity. Our approximations are examined through some numerical experiments. Next we compare the power of Dempster trace criterion with the ones of three classical criteria; likelihood ratio criterion, Lawley-Hotelling trace criterion, and Bartlett-Nanda-Pillai trace criterion when the dimension is large compared to the sample size.

Key words and phrases: Asymptotic null and nonnull distributions, Dempster trace criterion, high dimensional situation, MANOVA tests, power comparison.

1.   Introduction

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

$$(1.1) \qquad\qquad Y = A\Theta + \mathcal{E},$$

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

$$(1.2) \qquad\qquad H_0 : C\Theta = O \text{ vs } H_1 : C\Theta \neq O,$$

let S_h and S_e be the matrices of sums of squares and products due to the hypothesis and the error defined by

$$S_h = (C\hat{\Theta})' \left[C(A'A)^{-1}C' \right]^{-1} C\hat{\Theta},$$
$$S_e = (Y - A\hat{\Theta})'(Y - A\hat{\Theta}),$$

respectively, where C is a $q \times k$ known matrix with $\mathrm{rank}(C) = q$, and $\hat{\Theta} = (A'A)^{-1}A'Y$. Then S_h and S_e are independently distributed as a noncentral

Wishart distribution $W_p(q, \Sigma; MM')$ and a central Wishart distribution $W_p(n, \Sigma)$, where $n = N - k$, and M is a $p \times q$ matrix such that

$$(1.3) \qquad MM' = (C\Theta)' \left[C(A'A)^{-1}C' \right]^{-1} C\Theta.$$

Under the assumption that $n \geq p$, the following three well known statistics have been used:
  (i) Likelihood Ratio statistic: $-\log(\mid S_e \mid / \mid S_e + S_h \mid)$
  (ii) Lawley-Hotelling trace criterion: $\operatorname{tr} S_h S_e^{-1}$
  (iii) Bartlett-Nanda-Pillai trace criterion: $\operatorname{tr} S_h (S_e + S_h)^{-1}$.
When $n < p$, 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. For testing (1.2), we can write the corresponding statistic as
 (iv) Dempster trace criterion:

$$(1.4) \qquad T_D = (\operatorname{tr} S_h)/(\operatorname{tr} S_e),$$

which may be called Dempster trace criterion. For the null distribution of T_D, it has been proposed to use F-approximations by Dempster (1958, 1960), Takeda and Goto (1999), etc. It seems that the F-approximations are good in some situations with $\Sigma = \lambda I$. The test was applied by Dempster (1960) to an example which consist of 62 biological measurements and 12 subjects, in order to examine whether there is evidence that the 62 items could be used to distinguish between alcoholic or non-alchoholic. In general, it is getting important to develop multivariate theory for analyzing multivariate datasets with fewer observations than the dimension, or with large dimension in comparison of the number of samples, in various applied area.

Note that the criterion has been proposed for high dimensional case, and so, it is important to study its asymptotic behavior when the dimension p is large compared to the number n. In this paper we study its distribution under a high dimensional framework:

$$(1.5) \qquad \text{A1}: \quad q; \text{fix}, \quad n \to \infty, \quad p \to \infty, \quad \frac{p}{n} \to c \in (0, \infty).$$

Further, in addition to A1 we will assume

$$(1.6) \qquad \text{A2}: \quad \frac{1}{p} \operatorname{tr} \Sigma^k = O(1), \quad (k = 1, 2),$$

for the null case, and in addition to A1 and A2

$$(1.7) \qquad \text{A3}: \quad \frac{1}{p} \operatorname{tr} \Sigma^k \Omega = O(1), \quad (k = 1, 2),$$

for the nonnull case, where

$$\Omega = \Sigma^{-1/2} \Theta' C' (C(A'A)^{-1}C')^{-1} C\Theta \Sigma^{-1/2}.$$

The assumption (1.6) will be natural. The assumption (1.7) corresponds to local alternatives when $\Sigma = \lambda I$.

In Section 2 we derive asymptotic null distribution of T_D. Our approximations are numerically examined through some experiments. On the other hand, Wakaki *et al.* (2002) derived asymptotic distributions of the null and nonnull distributions of the three classical test statistics under the assumption (1.5) with $c < 1$ and the assumptions (1.6) and (1.7). In Section 3 we compare the power of Dempster trace criterion with the ones of three classical tests. Naturally it is seen that T_D will be more powerful than the classical tests as p becomes near to n.

2. Dempster's test

In this section we derive the limiting null and nonnull distributions of Dempster test statistic.

2.1. Limiting null distribution

Under the null hypothesis, S_h and S_e are independently distributed as the central Wishart distributions, $W_p(q, \Sigma)$ and $W_p(n, \Sigma)$, respectively. Let

$$\tilde{T}_D = \sqrt{p}\left\{ n\frac{\mathrm{tr}S_h}{\mathrm{tr}S_e} - q \right\}.$$

Let U and V be defined by

$$(2.1) \qquad U = \frac{\mathrm{tr}S_h - q\mathrm{tr}\Sigma}{\sqrt{2q\mathrm{tr}\Sigma^2}}, \quad V = \frac{\mathrm{tr}S_e - n\mathrm{tr}\Sigma}{\sqrt{2n\mathrm{tr}\Sigma^2}}.$$

For our asymptotics, we assume (1.5) and (1.6). Then, considering the characteristic functions of U and V, it is seen that U and V are asymptotically distributed as the nomal distribution $N(0,1)$. Using (2.1), we can expand $\tilde{T}_D$ as

$$\tilde{T}_D = \sqrt{p}\left\{ \frac{U\sqrt{2npq}\sqrt{(\mathrm{tr}\Sigma^2)/p} + pq\sqrt{n}(\mathrm{tr}\Sigma)/p}{V\sqrt{2p}\sqrt{(\mathrm{tr}\Sigma^2)/p} + p\sqrt{n}(\mathrm{tr}\Sigma)/p} - q \right\}$$
$$= \frac{\sqrt{2q(\mathrm{tr}\Sigma^2)/p}}{(\mathrm{tr}\Sigma)/p}U + o(1).$$

Therefore, we obtain the following theorem.

THEOREM 2.1. *Under the asymptotic framework* (1.5) *and the assumption* (1.6), *it holds that*

$$\frac{\tilde{T}_D}{\sigma_D} \overset{d}{\to} N(0,1),$$

where $\overset{d}{\to}$ denotes convergence in distribution, and

$$\sigma_D = \frac{\sqrt{2q(\mathrm{tr}\,\Sigma^2)/p}}{(\mathrm{tr}\,\Sigma)/p}.$$

For a practical situation, we need to use an estimator instead of σ_D since Σ is usually unknown. Srivastava (2003) has proposed a high dimensional consistent estimator, i.e., (n,p)-consistent estimator given by

$$\hat{\sigma}_D = \frac{\sqrt{2q\left\{(\mathrm{tr}S_e^2)/n^2 - (\mathrm{tr}S_e)^2/n^3\right\}/p}}{(\mathrm{tr}S_e)/(np)}.$$

Table 2.1. Upper 5 percent points.

q	n	p	Simu	σ_D	$\hat{\sigma}_D$
			$\Sigma = \mathrm{diag}(1,\dots,1)$
2	40	40	3.589	3.290	3.354
2	40	80	3.541	3.290	3.328
2	80	40	3.515	3.290	3.391
2	80	80	3.459	3.290	3.339
2	120	120	3.427	3.290	3.304
2	120	200	3.406	3.290	3.337
2	200	120	3.399	3.290	3.301
2	200	200	3.398	3.290	3.283
4	40	40	5.158	4.652	4.860
4	40	80	5.039	4.652	4.652
4	80	40	4.969	4.652	4.648
4	80	80	4.910	4.652	4.665
4	120	120	4.835	4.652	4.648
4	120	200	4.803	4.652	4.567
4	200	120	4.798	4.652	4.711
4	200	200	4.788	4.652	4.709

Table 2.2. Actual error probabilities of the first kind when the nominal level is 0.05.

q	n	p	σ_D	$\hat{\sigma}_D$
			$\Sigma = \mathrm{diag}(1,\dots,1)$
2	40	40	0.0640	0.0611
2	40	80	0.0617	0.0597
2	80	40	0.0604	0.0555
2	80	80	0.0586	0.0560
2	120	120	0.0564	0.0558
2	120	200	0.0557	0.0532
2	200	120	0.0551	0.0546
2	200	200	0.0551	0.0553
4	40	40	0.0671	0.0595
4	40	80	0.0632	0.0632
4	80	40	0.0606	0.0607
4	80	80	0.0595	0.0591
4	120	120	0.0563	0.0564
4	120	200	0.0557	0.0589
4	200	120	0.0548	0.0528
4	200	200	0.0550	0.0528

Table 2.3. Maximal value of $q = 2, 4, 6, 8, 10$.

p	20	40	60	80	100	120	140	160	180	200
					n					
20									8	8
40					6	4	6	10	8	8
60			2	2	6	8	10	10	10	10
80			4	4	8	10	8	10	10	10
100			4	6	6	8	8	10	10	10
120			2	6	8	10	10	10	10	10
140		2	4	4	8	10	10	10	10	10
160			4	4	8	10	10	10	10	10
180		2	4	4	10	10	10	10	10	10
200		2	2	6	8	10	10	10	10	10

We attempted to clear numerical accuracy of the limiting distribution in Theorem 2.1. Note that the null distribution of T_D depends on Σ through its characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_p$, and hence we may assume that $\Sigma = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$. We choosed the values of q, n, p, and $(\lambda_1, \ldots, \lambda_p)$ as follows:

$$q;\ 2, 4, 6, 8, 10,$$
$$n, p;\ 20, 40, 60, 80, 100, 120, 140, 160, 180, 200,$$
$$(\lambda_1, \ldots, \lambda_p) = (1, \ldots, 1).$$

The error probabilities of the first kind are almost monotone decreasing for n and p, monotone increasing for q, and lager than 0.05. If n or p is smaller than 50, the error probabilities of the first kind are almost lager than 0.06. If n and p are larger than 100, the error probabilities of the first kind are almost smaller than 0.06 and larger than 0.05. We show a part of these results on Table 2.1 and Table 2.2. Table 2.1 gives the estimated upper 5 percent points based on Monte Carlo simulation, the limiting distribution in Theorem 2.1 and the limiting distribution when σ_D is replaced by $\hat{\sigma}_D$, respectively. Table 2.2 gives the actual error probabilities of the first kind by using the approximated percent points. Table 2.3 gives the maximal values of q in the above range such that the error probability of the first kind is smaller than 0.06 and larger than 0.05. It should be noted that our approximation method underestimates the percent point. Further, the approximations are rather improving by using $\hat{\sigma}_D$ instead of σ_D. However, the approximations are not very accurate except the case when n is large in comparison with p. In order to get more accurate approximations it is expected to obtain asymptotic expansions up to $O(n^{-1/2})$ or $O(n^{-1})$ under the high dimensional framework (1.5) and the assumption (1.6) with $k = 1 \sim 3$ or $k = 1 \sim 4$.

2.2. Limiting nonnull distribution

Under alternative hypotheses, S_e and S_h are independently distributed as the central and noncentral Wishart distributions, $W_p(n, \Sigma)$ and $W_p(q, \Sigma, MM')$, respectively, where M is defined in (1.3). Let

$$T_D^* = \sqrt{p} \left\{ n \frac{\mathrm{tr} S_h}{\mathrm{tr} S_e} - q - \frac{\mathrm{tr} \Sigma \Omega}{\mathrm{tr} \Sigma} \right\}.$$

Let U and V be defined by

$$(2.2) \qquad U = \frac{1}{\sqrt{p}} (\mathrm{tr} S_h - q \mathrm{tr} \Sigma - \mathrm{tr} \Sigma \Omega), \quad V = \frac{1}{\sqrt{np}} (\mathrm{tr} S_e - n \mathrm{tr} \Sigma).$$

For our asymptotics, we assume (1.5), (1.6) and (1.7). Then U and V are asymptotically independent and normal. More precisely,

$$\frac{V}{\sqrt{2(\mathrm{tr}\, \Sigma^2)/p}} \xrightarrow{d} N(0, 1).$$

Further, using the characteristic function of the noncentral Wishart distribution (see, chapter 10 of Muirhead (1982)), the cumulant generating function of U can be expanded as

$$
\begin{aligned}
\log E[\exp(itU)] &= \log E\left[\exp\left(\frac{it}{\sqrt{p}}\mathrm{tr}S_h\right)\right] - \frac{it}{\sqrt{p}}(q\mathrm{tr}\Sigma + \mathrm{tr}\Sigma\Omega) \\
&= -\frac{q}{2}\log\det\left(I_p - \frac{2it}{\sqrt{p}}\Sigma\right) - \frac{1}{2}\mathrm{tr}\Omega \\
&\quad + \frac{1}{2}\mathrm{tr}\Omega\left(I_p - \frac{2it}{\sqrt{p}}\Sigma\right)^{-1} - \frac{it}{\sqrt{p}}(q\mathrm{tr}\Sigma + \mathrm{tr}\Sigma\Omega) \\
&= \frac{q}{2}\left\{\frac{2it}{\sqrt{p}}\mathrm{tr}\Sigma + \frac{2(it)^2}{p}\mathrm{tr}\Sigma^2\right\} - \frac{1}{2}\mathrm{tr}\Omega \\
&\quad + \frac{1}{2}\left\{\mathrm{tr}\Omega + \frac{2it}{\sqrt{p}}\mathrm{tr}\Sigma\Omega + \frac{4(it)^2}{p}\mathrm{tr}\Sigma^2\Omega\right\} \\
&\quad - \frac{it}{\sqrt{p}}(q\mathrm{tr}\Sigma + \mathrm{tr}\Sigma\Omega) + o(1) \\
&= (it)^2\left\{q\frac{\mathrm{tr}\Sigma^2}{p} + 2\frac{\mathrm{tr}\Sigma^2\Omega}{p}\right\} + o(1).
\end{aligned}
$$

Using (2.2) we obtain

$$
\begin{aligned}
T_D^* &= \sqrt{p}\left\{\frac{\sqrt{np}U + pq\sqrt{n}(\mathrm{tr}\Sigma)/p + p\sqrt{n}(\mathrm{tr}\Sigma\Omega)/p}{\sqrt{p}V + p\sqrt{n}(\mathrm{tr}\Sigma)/p} - q - \frac{\mathrm{tr}\Sigma\Omega}{\mathrm{tr}\Sigma}\right\} \\
&= \frac{1}{(\mathrm{tr}\Sigma)/p}U + o(1).
\end{aligned}
$$

Therefore we obtain the following theorem.

THEOREM 2.2. *Under the asymptotic framework* (1.5) *and the assumptions* (1.6) *and* (1.7), *it holds that*

$$
\frac{T_D^*}{\sigma_D^*} \xrightarrow{d} N(0,1),
$$

where

$$
\sigma_D^* = \frac{\sqrt{2q(\mathrm{tr}\Sigma^2)/p + 4(\mathrm{tr}\Sigma^2\Omega)/p}}{(\mathrm{tr}\Sigma)/p}.
$$

In a special case $\Omega = 0$, we get the limiting null distribution in Theorem 2.1.

3. Power comparison

In this section we compare the power of Dempster test with ones of likelihood ratio test, Lawley-Hotelling test, and Bartlett-Nanda-Pillai test. Let

$$
\delta_D = \tilde{T}_D - T_D^* = \sqrt{p}\frac{\mathrm{tr}\Sigma\Omega}{\mathrm{tr}\Sigma},
$$

then the power of $\tilde{T}_D$ with significance level α can be expressed as

$$P_D = Pr(\tilde{T}_D > \sigma_D z_\alpha) = Pr(T_D^* > \sigma_D z_\alpha - \delta_D),$$

where z_α is the upper 100α % point of the standard nomal distribution. Using Theorem 2.2, the asymptotic power is

$$\lim_{p \to \infty} P_D = \lim_{p \to \infty} \Phi\left(\frac{\delta_D - \sigma_D z_\alpha}{\sigma_D^*}\right).$$

If the order of $\text{tr}\Sigma^k\Omega$ $(k = 1, 2)$ is larger than $\sqrt{p}$, $\delta_D \to \infty$ so that the asymptotic power tends to one, while if the order of $\text{tr}\Sigma^k\Omega$ $(k = 1, 2)$ is smaller than $\sqrt{p}$, the asymptotic power is α since $\delta_D \to 0$ and $\sigma_D^* \to \sigma_D$. When $\text{tr}\Sigma^k\Omega = O(\sqrt{p})$ $(k = 1, 2)$, we obtain the asymptotic result easily, then

$$\lim_{p \to \infty} P_D = \Phi\left(\frac{\delta_D}{\sigma_D} - z_\alpha\right)$$

(3.1)
$$= \Phi\left(\frac{\text{tr}\Sigma\Omega}{\sqrt{2q\text{tr}\Sigma^2}} - z_\alpha\right).$$

On the other hand, let

$$T_{LR} = -\sqrt{p}\left(1 + \frac{m}{p}\right)\left\{\log\frac{|S_e|}{|S_e + S_h|} + q\log\left(1 + \frac{p}{m}\right)\right\},$$

$$T_H = \sqrt{p}\left\{\frac{m}{p}\text{tr}S_h S_e^{-1} - q\right\},$$

$$T_{BNP} = \sqrt{p}\left(1 + \frac{p}{m}\right)\left\{\left(1 + \frac{m}{p}\right)\text{tr}S_h(S_e + S_h)^{-1} - q\right\},$$

then under the assumption $\text{tr}\Omega = O(\sqrt{p})$, the power of T_G (G=LR, H, BNP) with significance level α can be expressed (see Wakaki *et al.* (2002)) as

$$\lim_{p \to \infty} P_G = \Phi\left(\frac{\delta_0}{\sigma} - z_\alpha\right)$$

(3.2)
$$= \Phi\left(\frac{\text{tr}\Omega/\sqrt{p}}{\sqrt{2q(1 + r)}} - z_\alpha\right),$$

where $r = p/m$ and $m = n - p + q$. Note that Ω in the case T_D is used for $\text{tr}\Omega$ in the case T_G. Comparing (3.1) with (3.2) and neglecting the terms of $o(1)$,

$$\frac{(\text{tr}\Omega)/\sqrt{p}}{\sqrt{1 + r}} > \frac{\text{tr}\Sigma\Omega}{\sqrt{\text{tr}\Sigma^2}} \Rightarrow P_G > P_D,$$

$$\frac{(\text{tr}\Omega)/\sqrt{p}}{\sqrt{1 + r}} < \frac{\text{tr}\Sigma\Omega}{\sqrt{\text{tr}\Sigma^2}} \Rightarrow P_G < P_D.$$

By the definition of Ω, i.e., $\Omega = \Sigma^{-1/2}MM'\Sigma^{-1/2}$ we have

$$
\begin{aligned}
\frac{\mathrm{tr}\Sigma\Omega}{(\mathrm{tr}\Omega/\sqrt{p})\sqrt{\mathrm{tr}\Sigma^2}} &= \frac{\mathrm{tr}MM'}{\mathrm{tr}\Sigma^{-1}MM'\sqrt{(\mathrm{tr}\Sigma^2)/p}} \\
&= \frac{(\mathrm{vec}(M))'\mathrm{vec}(M)}{(\mathrm{vec}(M))'(I_q \otimes \Sigma^{-1})\mathrm{vec}(M)\sqrt{\mathrm{tr}\Sigma^2/p}} \\
&\geq \frac{1}{\max_j \lambda_j^{-1}\sqrt{\mathrm{tr}\Sigma^2/p}} \\
&= \frac{\min_j \lambda_j}{\sqrt{\mathrm{tr}\Sigma^2/p}} \\
&\geq \frac{\min_j \lambda_j}{\max_j \lambda_j},
\end{aligned}
$$

where λ_i's are the characteristic roots of Σ. Thus, neglecting the terms of $o(1)$ we can see that

$$
R_\lambda = \frac{\min_j \lambda_j}{\max_j \lambda_j} > \frac{1}{\sqrt{1+r}} \Rightarrow P_G < P_D.
$$

In particular,

(1) If $\Sigma = aI_p$ (a: constant), then $P_G < P_D$.
(2) If R_λ is near to one, then $P_G < P_D$.
(3) If R_λ is small and p is small, then $P_G > P_D$.
(4) If c is close to one, r becomes large since $r = p/m \to c/(1-c)$, and hence $P_G < P_D$.

Acknowledgements

REFERENCES

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.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, New York.

Srivastava, M. (2003). Multivariate analysis with fewer observations than the dimension, submitted for publication.

Takeda, J. and Goto, M. (1999). Comparison of highly multivariate small samples — Estimation of effective dimensionality in the Dempster's approximation test, *Japanese J. Applied Statist.*, **28**, 163–177 (in Japanese).

Wakaki, H., Fujikoshi, Y. and Ulyanov, V. (2002). Asymptotic expansions of the distributions of MANOVA test statistics when the dimension is large, TR No. 02-10, Statistical Research Group, Hiroshima University.