High-dimensional Approximations of the Coefficients on Linear Discriminant Functions and Canonical Correlation Variates

Fujikoshi, Yasunori
Hiroshima University
1-3-1, Kagamiyama
HigashiHiroshima-shi, Hiroshima, 739-8526, Japan
E-mail: fujikoshi_y@yahoo.co.jp

1. Introduction

For discriminant analysis with $(q+1)$ p-dimensional populations, most inferential problems are based on two matrices S_b and S_w of sums of squares and products due to between-groups and within-groups, respectively. In this paper we are interesting in asymptotic behaviors of the coefficients on the linear discriminant functions, which are the characteristic vectors of $S_w^{-1}S_b$. Under normality S_b and S_w are independently distributed as Wishart distributions $W_p(q, \Sigma; \Omega)$ and $W_p(n-q-1, \Sigma)$, where n is the total sample size and Ω is the noncentrality matrix whose order is usually assumed to be $O(n)$. Since the large-sample asymptotic distributions of the characteristic roots were derived by Hsu (1941), the results have been extended (Fujikoshi (1977), Fujikoshi et al. (2010), Glynn (1980), Muirhead (1978, 1982), Sugiura (1976), etc.) by obtaining their asymptotic expansions and treating the characteristic vectors. However, these approximations become increasingly inaccurate as the value of p increases for a fixed value of n. On the other hand, we encounter to analyze high-dimensional data such that p is large compared to n.

Recently the distributions of the characteristic roots were studied in a high-dimensional situation where $p/n \to c \in (0, 1)$ by Fujikoshi et al. (2008). In this paper we extend the high-dimensional asymptotic results to the case that Ω has any multiplicity, and study asymptotic behavior of the transformed characteristic vecors, which may be applied for statistical inference of the coefficients on the linear discriminant functions. We also discuss with similar problems in canonical correlation analysis. Some relationships between high-dimensional and large-sample approximations are pointed. Further, it is noted that the consistency properties of the sample roots and the vectors in large-sample case do not hold in high-dimensional case.

2. Canonical Discriminant Analysis and Roots

Canonical discriminant analysis is a statistical procedure designed to discriminate between several different groups in terms of few discriminant functions, based on given multivariate observations on individuals in each population. Suppose that we have $(q+1)$ p-variate populations with mean vectors $\boldsymbol{\mu}_i$ and the same covariance marix Σ $(i = 1, \ldots, q+1)$, and n_i observations, $\boldsymbol{x}_{ij}(j = 1, \ldots, n_i)$, are available from the ith population. Let S_b and S_w be the matrices of sums of squares and products due to between groups and within groups, respectively. Then they are defined as

$$S_b = \sum_{i=1}^{q+1} n_i(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})', \quad S_w = \sum_{i=1}^{q+1}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)',$$

where $\bar{\boldsymbol{x}}_i$ is the sample mean vector of the ith population, and $\bar{\boldsymbol{x}}$ is the total mean vector.

We assume that the p-variate populations are normal. Then S_h and S_e are independently distributed as a noncentral Wishart distribution $W_p(q, \Sigma; \Sigma^{1/2}\Omega\Sigma^{1/2})$ and a central Wishart distribution

$W_p(n-q-1, \Sigma)$, respectively, where $n = \sum_{i=1}^{q+1} n_i$,

$$\Sigma^{1/2}\Omega\Sigma^{1/2} = \sum_{i=1}^{q+1} n_i(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})',$$

and $\bar{\boldsymbol{\mu}} = (1/n)(n_1\boldsymbol{\mu}_1 + \ldots + n_{q+1}\boldsymbol{\mu}_{q+1})$.

In canonical discriminant analysis, it is fundamental to study the distribution of the characteristic roots and vectors of $S_w^{-1}S_b$. Let $\ell_1 > \cdots > \ell_t > \ell_{t+1} = \cdots = \ell_p = 0$ be the characteristic roots, where $t = \min(p,q)$. In the following we consider the case $q \leq p$, so, $t = q$, since we are interesting in the case that p is large. Further, we assume that S_w is nonsingular, which is equivalent to assuming $n - q - 1 \geq p$. The matrix Ω is a population matrix corresponding to $S_w^{-1/2}S_bS_w^{-1/2}$. Let $\omega_1 \geq \ldots \geq \omega_p \geq 0$ be the characteristic roots of Ω. For their multiplicity, we assume that

$$\begin{aligned}
&\omega_1 = \cdots = \omega_{q_1} = n\lambda_1,\\
&\omega_{q_1+1} = \cdots = \omega_{q_1+q_2} = n\lambda_2,\\
&\qquad\vdots\\
&\omega_{q-q_r+1} = \cdots = \omega_q = n\lambda_r = 0,
\end{aligned}$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_r = 0$, $\sum_{\alpha=1}^{r} q_\alpha = q$. Suppose that

(1) $\quad$ A1 : q is fixed, and $\; c = p/n \to c_0 \in [0,1)$.

For $\alpha = 1, \ldots, r$, put

(2) $$Z_i = \frac{\sqrt{m}}{\sigma_\alpha}(\ell_i - \mu_\alpha), \quad i \in J_\alpha = \{q_1 + \cdots + q_{\alpha-1} + 1, \ldots, q_1 + \cdots + q_\alpha\},$$

where $m = n - p + q$, and

$$\mu_\alpha = c + \frac{n}{m}\lambda_\alpha, \quad \sigma_\alpha^2 = 2c(1+c) + 4(1+c)\frac{n}{m}\lambda_\alpha + 2\left(\frac{n}{m}\lambda_\alpha\right)^2.$$

Then, it is shown that the normalized roots $(Z_1, Z_2, \ldots, Z_q)$ are asymptotically distributed as

(3) $$\prod_{\alpha=1}^{r} f_1(\{z_i,\ i \in J_\alpha\}; q_\alpha),$$

where

$$f_1(\{z_i;\ i \in J_\alpha\}; q_\alpha) = \frac{\pi^{q_\alpha(q_\alpha-1)/4}}{2^{q_\alpha/2}\Gamma_{q_\alpha}(\frac{1}{2}q_\alpha)}\exp\left(-\frac{1}{2}\sum_{i\in J_\alpha} z_i^2\right)\prod_{i<j;i,j\in J_\alpha}(z_i - z_j).$$

The result when all the population roots are simple was derived by Fujikoshi, Himeno and Wakaki (2008). Further, putting $c = p/n = 0$ and $q/n = 0$ in (2) and (3), we can see that the result is reduced to the large-sample result due to Hsu (1941a). On the other side, Johnstone (2008) derived the asymptotic distribution of the largest root ℓ_1 when $p/n \to c_0 \in (0,1)$ and $q/n \to c_1 \in (0,1)$.

3. Characteristic Vectors in Canonical Discriminant Analysis

Let $\boldsymbol{h}_i$ be the characteristic vector corresponding to the i-th characteristic root ℓ_i of $S_w^{-1}S_b$ with $\boldsymbol{h}_i'S_w\boldsymbol{h}_i = n$, so that they are the solutions of

(4) $$S_b\boldsymbol{h}_i = \ell_iS_w\boldsymbol{h}_i, \quad \boldsymbol{h}_i'S_w\boldsymbol{h}_j = n\delta_{ij}, \quad i,j = 1, \ldots, p,$$

where $\ell_1 > \cdots > \ell_t > \ell_{t+1} = \cdots = \ell_p = 0$ and $t = \min(p,q)$. The functions $\boldsymbol{h}_i'\boldsymbol{x}, i = 1, \ldots, t$ are called canonical discriminant functions. We are interesting in asymptotic behavior under (1).

When $q = 1$, the non-zero characteristic root is only one. The characteristic root and vector are explicitly expressed as

$$\ell = \frac{n_1 n_2}{n(n-2)} D^2, \quad \boldsymbol{h} = \frac{\sqrt{n}}{D} S_w^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \tag{5}$$

where $D^2 = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$ and $(n-2)S = S_w$. For high-dimensional approximations, the following expressions are useful. Let $\boldsymbol{a}$ be any fixed p-dimensional vector, we have:

$$D^2 = \frac{y_1 + (z_1 + \tau)^2}{y_2}, \quad \tau = \sqrt{n_1 n_2 / n}\Delta, \quad \Delta = \left\{ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \right\}^{1/2},$$

$$\boldsymbol{a}' S_w^{-1} \sqrt{\frac{n_1 n_2}{n}} \{ \bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_1 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \} = \frac{(\boldsymbol{a}' \Sigma^{-1} \boldsymbol{a})^{1/2}}{y_2} \left\{ z_1 + z_2 \left(\frac{y_1 y_3}{y_4 (y_5 + z_2^2)} \right)^{1/2} \right\},$$

where z_1, z_2 are standard normal variates, $y_i, i = 1, \ldots, 5$ are chi-squared variates with f_i degrees of freedom, they are all independent, and

$$f_1 = f_3 = p - 1, \quad f_2 = n - p - 1, \quad f_4 = n - p, \quad f_5 = p - 2.$$

For a general $q (\leq p)$, consider the transformation from S_b and S_w to $\tilde{S}_b = \Gamma' \Sigma^{-1/2} S_b \Sigma^{-1/2} \Gamma$ and $\tilde{S}_w = \Gamma' \Sigma^{-1/2} S_w \Sigma^{-1/2} \Gamma$, where Γ is an orthogonal matrix such that $\Gamma' \Omega \Gamma = \mathrm{diag}(\omega_1, \ldots, \omega_p)$. We can write $\tilde{S}_b = Z Z'$, where Z is the random matrix whose elements are independent normal variates with the same variance 1 and

$$\mathrm{E}(Z') = M' = (D_{\sqrt{\omega}}, O)', \quad D_{\sqrt{\omega}} = \mathrm{diag}(\sqrt{\omega_1}, \ldots, \sqrt{\omega_q}).$$

Let $B = Z'Z$ and $W = B^{1/2} (Z' \tilde{S}_w Z)^{-1} B^{1/2}$. Then

$$B \sim W_q(n, I_q; D_\omega), \quad W \sim W_q(m, I_q)$$

and B and W are independent, where $D_\omega = \mathrm{diag}(\omega_1, \ldots, \omega_q)$. Note that the nonzero characteristic roots of $S_w^{-1} S_b$ are the same as the characteristic roots of $W^{-1} B$. Now we consider the characteristc roots and vectors of $W^{-1} B$ which are the solutions of

$$B \boldsymbol{a}_i = \ell_i W \boldsymbol{a}_i, \quad \boldsymbol{a}_i' W \boldsymbol{a}_j = n \delta_{ij}, \quad i, j = 1, \ldots, q. \tag{6}$$

The it holds that

$$\boldsymbol{a}_i = B^{-1/2} Z' \Gamma' \Sigma^{1/2} \boldsymbol{h}_i, \quad i = 1, \ldots, q. \tag{7}$$

Put

$$U = \frac{1}{\sqrt{p}} (B - p I_q - D_\omega), \text{ and } V = \frac{1}{\sqrt{m}} (W - m I_q)$$

which imply

$$B = m \left(D_\delta + \frac{1}{\sqrt{m}} \sqrt{r} U \right), \text{ and } W = m \left(I_q + \frac{1}{\sqrt{m}} V \right), \tag{8}$$

where $r = p/m$. Substituting (8) to (6), we obtain a perturbation expansion for $\boldsymbol{a}_i$ when ω_i is simple. Using the perturbation expansion we can derive high-dimensional asymptotic distribution of $\boldsymbol{a}_i$.

4. Canonical Correlation Analysis

Let $\boldsymbol{x}_1 = (x_1, \ldots, x_q)'$ and $\boldsymbol{x}_2 = (x_{q+1}, \ldots, x_{q+p})'$ be two random vectors having a joint $(q+p)$-variate normal distribution with mean vector and covariance matrix given by

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

respectively, where Σ_{12} is a $q \times p$ matrix. Without loss of generality we may assume $q \leq p$. Let S be the sample covariance matrix formed from a sample of size $N = n+1$ of $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$. Corresponding to a partition of $\boldsymbol{x}$, we partition S as

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}.$$

Let $\rho_1 \geq \cdots \geq \rho_q \geq 0$ and $r_1 > \cdots > r_q > 0$ be the population and the sample canonical correlations between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Note that $\rho_1^2 \geq \cdots \geq \rho_q^2 \geq 0$ and $r_1^2 > \cdots > r_q^2 > 0$ are the characteristic roots of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}$ and $S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}$, respectively. The sample canonical correlation vectors $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are the solutions of

$$\begin{aligned} S_{12}S_{22}^{-1}S_{21}\boldsymbol{a}_i &= r_i^2 S_{11}\boldsymbol{a}_i, \quad \boldsymbol{a}_i S_{11}\boldsymbol{a}_j = n\delta_{ij}, \quad i,j = 1,\ldots,p, \\ S_{21}S_{11}^{-1}S_{12}\boldsymbol{b}_j &= r_i^2 S_{22}\boldsymbol{b}_j, \quad \boldsymbol{a}_i S_{22}\boldsymbol{a}_j = n\delta_{ij}, \quad i,j = 1,\ldots,q. \end{aligned}$$

There are many results on asymptotic distributions of the canonical correlations and the canonical correlation vectors under a large sample framework; p and q are fixed, $n \to \infty$, since the first result was given by Hsu (1941b). However, it may be noted that these results will not work well as the dimension q or p becomes large.

In order to make up the weak point, it is important to study asymptotic behaviors of the canonical correlations and the canonical vectors when p and q are large. In this paper we consider them under a high-dimensional framework such that

$$\text{(9)} \qquad \mathrm{B1}: \ q; \text{fixed}, \quad p \to \infty, \quad n \to \infty, \quad m = n - p \to \infty, \quad c = p/n \to c_0 \in [0, 1).$$

When the population canonical correlation is simple, Fujikoshi and Sakurai (2009) obtained the following result:

$$\text{(10)} \qquad \sqrt{n}(r_\alpha^2 - \tilde{\rho}_\alpha^2) \overset{d}{\to} N(0, \sigma_\alpha^2), \quad \sqrt{n}(r_\alpha - \tilde{\rho}_\alpha) \overset{d}{\to} N(0, \frac{1}{4}\sigma_\alpha^2\tilde{\rho}_\alpha^{-2}),$$

where

$$\tilde{\rho}_\alpha = \{\rho_\alpha^2 + c(1-\rho_\alpha^2)\}^{1/2}, \quad \sigma_\alpha^2 = 2(1-c)(1-\rho_\alpha^2)^2\{2\rho_\alpha^2 + c(1-2\rho_\alpha^2)\}.$$

In particular, putting $c = 0$ in (10) we have the well known large sample results:

$$\text{(11)} \qquad \sqrt{n}(r_\alpha^2 - \rho_\alpha^2) \overset{d}{\to} N(0, 4\rho_\alpha^2(1-\rho_\alpha^2)^2), \quad \sqrt{n}(r_\alpha - \rho_\alpha) \overset{d}{\to} N(0, (1-\rho_\alpha^2)^2).$$

Here we note that the high-dimensional asymptotic result (10) depends on p through $c = p/n$, but the large-sample result (11) does not depend on p and are the same for all p.

In this paper we note that the high-dimensional results can be extended to the case when the population canonical correlations have multipilicity. We also give high-dimensional results on the canonical correlation vectors which are similar to the results in discriminant functions.

REFERENCES

Fujikoshi, Y. (1977). Asymptotic expansion for the distributions of the latent roots in MANOVA and the canonical correlations. *J. Multivariate Anal.*, **7**, 386–396.

Fujikoshi, Y., Himeno, T. and Wakaki, H. (2008). Asymptotic results in canonical discriminant analysis when the dimension is large compared to the sample. *J. Statist. Plann. Inference*, **138**, 3457–3466.

Fujikoshi, Y. and Sakurai, T. (2009). High-dimensional asymptotic expansions for the distributions of canonical correlations. *J. Multivariate Anal.*, **100**, 231–241.

Fujikoshi, Y., Ulyanov, V. and Shimizu, R. (2010). *Multivariate Analysis: High-Dimensional and Large-Sample Approximations.* Wiley, Hoboken, New Jersey.

Hsu, P. L. (1941a). On the limiting distribution of roots of a determinantal equation. *J. London Math. Soc.*, **16**, 183–194.

Hsu, P.L. (1941b). On the limiting distribution of roots of the canonical correlations. *Biometrika*, **32**, 38–45.

Glynn, W. J. (1980). Asymptotic representations of the densities of canonical correlations and latent roots in MANOVA when the populatio parameters have arbitrary multiplicities. *Ann. Statist.*, **8**, 468–477.

Muirhead, R. J. (1978), Latent roots and matrix variates. *Ann. Statist.*, **6**, 5–33.

Muirhead, R. J. (1982), *Aspect of Multivariate Statistical Theory*, Wiley, New York.

Johnstone, I. M.(2008). Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rate of convergence. *Ann. Statist.*, **36**, 2638–2716.

Sugiura, N. (1976). Asymptotic expansions of the distributions of the latent roots and latent vectors of the Wishart and multivariate F matrices. *J. Multivariate Anal.*, **6**, 500–525.

Sugiura, N. and Fujikoshi, Y. (1969). Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence. *Ann. Math. Statist.*, **40**, 942–952.