Interval estimation in two-group discriminant analysis under heteroscedasticity for large dimension

Takayuki Yamada

Education Center, Kagoshima University, 1-21-30 Korimoto, Kagoshima 890-0065, Japan
E-mail address: yamada@gm.kagoshima-u.ac.jp

Abstract

This paper is concerned with the problem of discriminating between two populations with heteroscedastic multivariate normal distributions based on an observation vector $\boldsymbol{x}$. We give the limiting distribution of the unbiased estimator for the log odds ratio of the posterior probabilities as the sample sizes N_i $(i = 1, 2)$ and the dimension p go to infinity together with the ratio $p/(N_i-1)$ converging a finite non-zero constant $c_i \in (0, 1)$ for the case in which the prior probabilities are equal. Approximated interval estimation for the log odds ratio is derived. Simulation results indicate that our estimation has good accuracy compared with the classical results.

1 Introduction

This paper is concerned with the problem of discriminating between two populations (Π_1, Π_2) based on an informative p variate observation vector $\boldsymbol{x}$. For this problem, Bayes's rule gives a role to the odds ratio to obtain a criteria for discrimination. The logarithm of this odds ratio is expressed as $\pi + \xi(\boldsymbol{x})$, where $\pi = \log(\pi_1/\pi_2)$, with π_i being the prior probability of $\boldsymbol{x}$ from Π_i, and where $\xi(\boldsymbol{x}) = \log\{f(\boldsymbol{x}|1)/f(\boldsymbol{x}|2)\}$, with $f(\boldsymbol{x}|i)$ being the conditional probability density function under the condition that $\boldsymbol{x}$ is belonging to Π_i.

Assume that the prior probabilities are equal. We assume further that the underlying probability distribution for $\boldsymbol{x} \in \Pi_i$ is p variate normal $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. In order to estimate the parameter of interest $\xi(\boldsymbol{x})$, we use training data of random samples $\{\boldsymbol{x}_{1,1}, \ldots, \boldsymbol{x}_{1,N_1}\}$ and $\{\boldsymbol{x}_{2,1}, \ldots, \boldsymbol{x}_{2,N_2}\}$.

For the case in which $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, confidence interval of $\xi_{\rm E}(\boldsymbol{x}) = \xi(\boldsymbol{x}) = \xi(\boldsymbol{x}; \Theta_{\rm E})$ has been proposed by Critchley and Ford [1], Rigby [8], Davis [4], and Critchley et al. [3], where $\Theta_{\rm E} = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}\}$.

Consider the case in which $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. It can be expressed that

$$\xi_{\rm U}(\boldsymbol{x}) = \xi(\boldsymbol{x}) = \xi(\boldsymbol{x}; \Theta_{\rm U}) = -\frac{1}{2}\{\delta_{1\rm U}(\boldsymbol{x}) - \delta_{2\rm U}(\boldsymbol{x})\}$$

with

$$\delta_{i\rm U}(\boldsymbol{x}) = \delta_i(\boldsymbol{x}) + \log|\boldsymbol{\Sigma}_i| \quad (i = 1, 2),$$

where $\Theta_{\rm U} = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\}$,

$$\delta_i(\boldsymbol{x}) = \delta(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = (\boldsymbol{x} - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i).$$

The uniform minimum variance unbiased estimator of $\xi_{\rm U}(\boldsymbol{x})$ is given by

$$\hat{\xi}_{\rm U}(\boldsymbol{x}) = \xi(\boldsymbol{x}; \widehat{\Theta_{\rm U}}) = -\frac{1}{2}\{\widehat{\delta_{1\rm U}(\boldsymbol{x})} - \widehat{\delta_{2\rm U}(\boldsymbol{x})}\}$$

with

$$\widehat{\delta_{i\rm U}(\boldsymbol{x})} = \frac{\delta(\boldsymbol{x}; \bar{\boldsymbol{x}}_i, \boldsymbol{S}_i)}{c_1(n_i)} - \frac{p}{N_i} + \log|\boldsymbol{S}_i| + p\log n_i - c_2(n_i) \quad (i = 1, 2),$$

where $\widehat{\Theta_{\rm U}} = \{\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, \boldsymbol{S}_1, \boldsymbol{S}_2\}$, $c_1(n_i) = n_i/(n_i - p - 1)$, $c_2(n_i) = p\log 2 + \sum_{j=1}^{p}\psi((n_i - p + j)/2)$, $\psi = (d/dy)\log\Gamma(y)$,

$$\bar{\boldsymbol{x}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\boldsymbol{x}_{ij}, \quad \boldsymbol{S}_i = \frac{1}{n_i}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)', \quad n_i = N_i - 1.$$

Critchley et al. [2] proposed asymptotic approximation of the confidence interval for $\xi_{\rm U}(\boldsymbol{x})$, which the one with the confidence level $1-\alpha$ is given as follows:

$$\mathrm{CI}_{\mathrm{CFR}} : \hat{\xi}_{\rm U}(\boldsymbol{x}) - \sqrt{\nu_{\rm U}\{\hat{\delta}_1(\boldsymbol{x}), \hat{\delta}_2(\boldsymbol{x})\}} z_{1-\alpha/2} \leqq \xi_{\rm U}(\boldsymbol{x}) \leqq \hat{\xi}_{\rm U}(\boldsymbol{x}) + \sqrt{\nu_{\rm U}\{\hat{\delta}_1(\boldsymbol{x}), \hat{\delta}_2(\boldsymbol{x})\}} z_{1-\alpha/2},$$

where z_α satisfies $\Phi(z_\alpha) = \alpha$, with $\Phi(.)$ being the cumulative distribution function of the standard normal distribution $N(0,1)$; $\hat{\delta}_i(\boldsymbol{x}) = \delta(\boldsymbol{x}; \bar{\boldsymbol{x}}_i, \boldsymbol{S}_i)$ for $i=1,2$;

$$\nu_{\rm U}\{\delta_1(\boldsymbol{x}), \delta_2(\boldsymbol{x})\} = \sum_{i=1}^{2} \left[\frac{\{\delta_i(\boldsymbol{x})\}^2}{2(m_i-4)} + \left\{ \frac{1}{n_i} - \frac{n_i-2}{(m_i-1)(m_i-4)} \right\} \delta_i(\boldsymbol{x}) + \frac{p(n_i-2)}{2(m_i-1)(m_i-4)} \right],$$

where $m_i = n_i - p$.

In the age of Big Data, we always encounter the case that both the dimension and the sample size are very large. For example, financial data, consumer data, network data and medical data have this feature. Generally, classical statistical methods established on the case that the dimension is smaller than sample size become poor performance when the dimension becomes large. So it is meaningful to find new method to resolve the problem of multivariate analysis for large dimensional case.

Yamada et al. [10] showed the asymptotic normality of the uniform minimum variance unbiased estimator for $\xi_{\rm E}(\boldsymbol{x})$ as the dimension and sample sizes approach infinity, and proposed confidence interval based on the asymptotic distribution. They investigated the actual confidence of the confidence interval by simulation, and presented the usefulness compared with the classical confidence intervals (Critchley and Ford [1], Rigby [8], Davis [4], and Critchley et al. [3]) for the case in which the dimension is relatively large but is less than the total sample sizes.

The usefulness for asymptotic distribution as the dimension and sample size approach infinity is written in literature, cf. Srivastava [9] and Fujikoshi et al. [5].

In this article, we will show the asymptotic normality for $\hat{\xi}_{\rm U}(\boldsymbol{x})$ for the case in which $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$ under the high-dimensional asymptotic framework A:

$$\mathrm{A} : n_i \to \infty, p \to \infty, p/n_i \to c_i \in (0,1) \quad (i=1,2).$$

Hall et al. [6] showed that $\sqrt{\boldsymbol{z}'\boldsymbol{z}} = \sqrt{p} + O_p(1)$ as $p \to \infty$ for $\boldsymbol{z} \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$. From this property, $\delta_i(\boldsymbol{x})/p \to 1$ as $p \to \infty$ if $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. Indeed, we suppose the case that $\boldsymbol{\Sigma}_1$ is little different from $\boldsymbol{\Sigma}_2$. Motivating them, we assume

$$\mathrm{C} : \frac{\delta_i(\boldsymbol{x})}{p} \to c_i \in (0,\infty) \text{ as } p \to \infty.$$

At first, we mention two-fundamental results, which are as follows.

Theorem 1. *Assume that the condition* C *holds. For fixed* $\boldsymbol{x}$*, under the high-dimensional asymptotic framework* A,

$$\sqrt{\frac{2m}{\delta^*(\boldsymbol{x};\Theta_{\rm U})}} \left\{ \hat{\xi}_{\rm U}(\boldsymbol{x}) - \xi_{\rm U}(\boldsymbol{x}) \right\} \xrightarrow{\mathcal{D}} N(0,1),$$

where "$\xrightarrow{\mathcal{D}}$*" denotes convergence in distribution,* $m = m_1 + m_2$*, and*

$$\delta^*(\boldsymbol{x};\Theta_{\rm U}) = \sum_{i=1}^{2} \frac{m}{m_i} \{\delta_i(\boldsymbol{x})\}^2 .$$

Theorem 2. *Assume that the condition* C *holds. For fixed* $\boldsymbol{x}$*, under the high-dimensional asymptotic framework* A,

$$\frac{\widehat{\delta^*(\boldsymbol{x};\Theta_{\rm U})}}{\delta^*(\boldsymbol{x};\Theta_{\rm U})} \xrightarrow{p} 1,$$

where "$\xrightarrow{p}$" denotes convergence in probability, where

$$\delta^*\widehat{(\boldsymbol{x};\Theta_{\mathrm{U}})} = \sum_{i=1}^{2} \frac{m}{m_i} \left\{ \frac{1}{c_1(n_i)} \hat{\delta}_i(\boldsymbol{x}) - \frac{p}{N_i} \right\}^2.$$

From the above theorems and Slutsky's theorem, we find that

$$\sqrt{\frac{2m}{\delta^*\widehat{(\boldsymbol{x};\Theta_{\mathrm{U}})}}} \left\{ \hat{\xi}_{\mathrm{U}}(\boldsymbol{x}) - \xi_{\mathrm{U}}(\boldsymbol{x}) \right\} \xrightarrow{\mathcal{D}} N(0,1).$$

Based on this asymptotic normality, we provide an approximation of the confidence interval for $\xi_{\mathrm{U}}(\boldsymbol{x})$, which the one with the confidence level $1-\alpha$ is given as follows.

$$\mathrm{CI_P} : \hat{\xi}_{\mathrm{U}}(\boldsymbol{x}) - \sqrt{\frac{\delta^*\widehat{(\boldsymbol{x};\Theta_{\mathrm{U}})}}{2m}} z_{1-\alpha/2} \leqq \xi_{\mathrm{U}}(\boldsymbol{x}) \leqq \hat{\xi}_{\mathrm{U}}(\boldsymbol{x}) + \sqrt{\frac{\delta^*\widehat{(\boldsymbol{x};\Theta_{\mathrm{U}})}}{2m}} z_{1-\alpha/2}.$$

The rest of this paper is organized as follows. In Section 2, we compare the actual confidence of our proposed confidence interval with the classical ones mentioned above through simulation. The proof of Theorem 1 and Theorem 2 are given in Section 3.

2 Simulation

In order to see the performance of the proposed confidence intervals, we conduct some simulations for the actual confidence level based on 10,000 repetitions. We then compare the confidence intervals proposed here with the ones treated in Hirst et al. [7].

Method U1. The confidence interval $\mathrm{CI_{CFR}}$ is used.

Method U2. The confidence interval given by Critchley et al. [3] under heteroscedasticity is used.

Method U3. The method of Rigby [8] is used.

Method U4. The confidence interval $\mathrm{CI_p}$ is used.

We use the same notation as Critchley and Ford [1]. Let $\boldsymbol{P}$ be a matrix such that $\boldsymbol{\Sigma}_1^{-1} = \boldsymbol{P}'\boldsymbol{P}$. Let $\boldsymbol{Q}'$ be an orthogonal matrix whose first column is proportional to $\boldsymbol{P}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$.
Transforming $\boldsymbol{t} \to \boldsymbol{t}^* = \boldsymbol{A}\boldsymbol{t} - \boldsymbol{c}$, where $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{P}$, $\boldsymbol{c} = \boldsymbol{A}\boldsymbol{\mu}_1 + (\Delta_1/2)\boldsymbol{e}_1$, $\Delta_1 = \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}$, $\boldsymbol{e}_1 = (1,0,\ldots,0)'$, we have

$$\begin{aligned}
&\bar{\boldsymbol{x}}_1 \to \bar{\boldsymbol{x}}_1^* \sim N_p(-(\Delta_1/2)\boldsymbol{e}_1, N_1^{-1}\boldsymbol{I}_p), \quad \bar{\boldsymbol{x}}_2 \to \bar{\boldsymbol{x}}_2^* \sim N_p((\Delta_1/2)\boldsymbol{e}_1, N_2^{-1}\boldsymbol{\Sigma}_2^*),\\
&\boldsymbol{S}_1 \to \boldsymbol{S}_1^* \sim W_p(\boldsymbol{n}_1, \boldsymbol{I}_p),\\
&\boldsymbol{S}_2 \to \boldsymbol{S}_2^* \sim W_p(\boldsymbol{n}_2, \boldsymbol{\Sigma}_2^*),\\
&\boldsymbol{x} \to \boldsymbol{x}^*,
\end{aligned}$$

where $\boldsymbol{\Sigma}_2^* = \boldsymbol{Q}\boldsymbol{P}\boldsymbol{\Sigma}_2\boldsymbol{P}'\boldsymbol{Q}'$. So, we generate random samples from $\Pi_1 : N_p(-(\Delta_1/2)\boldsymbol{e}_1, \boldsymbol{I}_p)$ and $\Pi_2 : N_p((\Delta_1/2)\boldsymbol{e}_1, \boldsymbol{\Sigma}_2^*)$. As the covariance matrix $\boldsymbol{\Sigma}_2^*$, we set the following three cases.

M1. $\boldsymbol{\Sigma}_2^* = 1.2\boldsymbol{I}_p$.

M2. $\boldsymbol{\Sigma}_2^* = \boldsymbol{D} = \mathrm{diag}(d_1,\ldots,d_p)$, where $d_i = 5 + (-1)^{i-1}\{1 - (i-1)/p\}$.

M3. $\boldsymbol{\Sigma}_2^* = \boldsymbol{D}^{1/2}(0.1^{|i-j|})\boldsymbol{D}^{1/2}$.

In our simulations, we considered the case in which $N_1 = N_2 = 100$, $p = 10, 30, 50, 70$, $\Delta_1 = 1.6832$, $\alpha = 0.05$, and

$$\boldsymbol{x}^* = \begin{pmatrix} x_1^* \\ y_1^* \boldsymbol{h} \end{pmatrix}, \quad \boldsymbol{h} = \frac{1}{\sqrt{\boldsymbol{z}'\boldsymbol{z}}}\boldsymbol{z}, \quad \boldsymbol{z} \sim N_{p-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1}).$$

For the setting of (x_1^*, y_1^*), Hirst et al. [7] treated the following 6 cases:

$$\begin{aligned} &\mathrm{A} = (2.4866, 0), \quad \mathrm{B} = (0.8416, 0), \quad \mathrm{C} = (0.8416, 1.645), \\ &\mathrm{D} = (0.4208, 0.5265), \quad \mathrm{E} = (0, 0), \quad \mathrm{F} = (0, 1.4134). \end{aligned}$$

In order to satisfy the condition C, we multiplied these cases by $\sqrt{p}$, i.e., we studied the following cases.

$$\begin{aligned} &\mathrm{A}_h = (2.4866\sqrt{p}, 0), \quad \mathrm{B}_h = (0.8416\sqrt{p}, 0), \quad \mathrm{C}_h = (0.8416\sqrt{p}, 1.645\sqrt{p}), \\ &\mathrm{D}_h = (0.4208\sqrt{p}, 0.5265\sqrt{p}), \quad \mathrm{F}_h = (0, 1.4134\sqrt{p}). \end{aligned}$$

We omitted the case $\mathrm{E}_h = (0, 0)$ because the condition C is not satisfied. We will show that the model M3 satisfies the condition C. The matrix $(\rho^{|i-j|})$ with $0 < \rho < 1$ is a Hermite Toeplitz matrix. There exist real-valued eigenvalues $\lambda_1 \geqq \cdots \geqq \lambda_p$ of $(\rho^{|i-j|})$ such that

$$\mu_{\min} = \min_{\theta \in [-\pi, \pi]} f(\theta) \leqq \lambda_p \leqq \cdots \leqq \lambda_1 \leqq \mu_{\max} = \max_{\theta \in [-\pi, \pi]} f(\theta),$$

where

$$f(\theta) = \sum_{k=-\infty}^{\infty} \rho^{|k|} e^{ik\theta} = 1 + \frac{\rho e^{i\theta}}{1 - \rho e^{i\theta}} + \frac{\rho e^{-i\theta}}{1 - \rho e^{-i\theta}}.$$

It is observed that

$$\mu_{\min} = -1 + \frac{2}{1 + \rho^2}.$$

So, for the model M3, we find that

$$\frac{\delta_2(\boldsymbol{x}^*)}{p} \leqq \frac{1}{p\mu_{\min}} (\boldsymbol{x}^* - (\Delta_1/2)\boldsymbol{e}_1)' \boldsymbol{D}^{-1} (\boldsymbol{x}^* - (\Delta_1/2)\boldsymbol{e}_1),$$

and so the condition C holds.

We can see from Table 1 that the actual confidence levels for U4 vary between 0.94 to 0.97. So, it can be observed that our proposed confidence interval always keep around the given confidence levels. By comparing the results of M1-M3, we can see the tendency that the actual confidence level decrease as the model for $\boldsymbol{\Sigma}_2^*$ becomes complicated. Our extra simulation results which does not written in this paper reveals that the actual confidence level gets away from the given confidence level as the dimension becomes close to sample size. It is also observed that many of the actual confidence levels for U2 and U3 are not in line with the predefined level $(= 1 - \alpha = 0.95)$. We can also check that the confidence interval based on U1 is conservative, in which case the interval contains too many true values, and so it is likely that the width is too large.

To sum up, our simulation demonstrate that the proposed confidence interval outperforms the classical methods for the case that each sample size and the dimension are large, but the dimension is less than and equal to a half of the smallest sample sizes.

3 Proof of theorem

In this section, we provide the proofs of Theorem 1 and Theorem 2.

Table 1: Actual confidence level based on 10,000 repetitions when $N_1 = N_2 = 100$ and $\alpha = 0.05$

			Feature vector				
p	Method	$\mathbf{\Sigma}_2^*$	A_h	B_h	C_h	D_h	F_h
10	U1	M1	0.98	0.96	0.98	0.98	0.98
		M2	0.97	0.97	0.97	0.97	0.97
		M3	0.97	0.97	0.97	0.97	0.97
	U2	M1	0.94	0.93	0.94	0.94	0.94
		M2	0.93	0.92	0.93	0.92	0.93
		M3	0.93	0.92	0.93	0.92	0.94
	U3	M1	0.93	0.91	0.93	0.92	0.94
		M2	0.89	0.88	0.89	0.89	0.90
		M3	0.89	0.89	0.90	0.90	0.91
	U4	M1	0.95	0.94	0.96	0.96	0.96
		M2	0.95	0.95	0.95	0.95	0.95
		M3	0.95	0.95	0.94	0.96	0.95
30	U1	M1	0.99	1.00	1.00	1.00	1.00
		M2	0.99	0.99	0.99	0.99	0.99
		M3	0.99	0.99	0.99	0.99	0.99
	U2	M1	0.88	0.78	0.89	0.83	0.89
		M2	0.55	0.52	0.60	0.58	0.63
		M3	0.55	0.53	0.61	0.58	0.65
	U3	M1	0.86	0.73	0.87	0.79	0.88
		M2	0.47	0.43	0.52	0.50	0.56
		M3	0.47	0.44	0.53	0.50	0.57
	U4	M1	0.96	0.95	0.96	0.96	0.96
		M2	0.95	0.95	0.95	0.96	0.95
		M3	0.94	0.95	0.95	0.96	0.95
50	U1	M1	1.00	1.00	1.00	1.00	1.00
		M2	1.00	1.00	1.00	1.00	1.00
		M3	1.00	1.00	1.00	1.00	1.00
	U2	M1	0.77	0.59	0.79	0.68	0.82
		M2	0.12	0.10	0.16	0.13	0.22
		M3	0.13	0.09	0.18	0.13	0.22
	U3	M1	0.75	0.55	0.77	0.64	0.80
		M2	0.09	0.07	0.13	0.10	0.17
		M3	0.09	0.06	0.14	0.10	0.18
	U4	M1	0.96	0.96	0.97	0.97	0.96
		M2	0.94	0.94	0.94	0.96	0.94
		M3	0.94	0.95	0.95	0.96	0.94
70	U1	M1	1.00	1.00	1.00	1.00	1.00
		M2	1.00	1.00	1.00	1.00	1.00
		M3	1.00	1.00	1.00	1.00	1.00
	U2	M1	0.63	0.45	0.67	0.53	0.70
		M2	0.01	0.01	0.03	0.02	0.06
		M3	0.00	0.00	0.03	0.02	0.05
	U3	M1	0.62	0.43	0.65	0.51	0.69
		M2	0.01	0.00	0.03	0.01	0.05
		M3	0.00	0.00	0.03	0.01	0.04
	U4	M1	0.97	0.96	0.97	0.98	0.98
		M2	0.94	0.95	0.94	0.97	0.94
		M3	0.94	0.95	0.94	0.97	0.94

3.1 Proof of Theorem 1

In order to prove Theorem 1, we will use the following lemmas.

Lemma 1. *Let* $y \sim \chi^2(m+1)$, $\boldsymbol{z} \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, *and* y *is independent to* $\boldsymbol{z}$, *where* $m = n - p$, *with* $n = N - 1$. *Assume that* $\boldsymbol{b}'\boldsymbol{b}$ *converges a positive constant as* $p \to \infty$. *Then,*

$$\frac{1}{\sqrt{2}\boldsymbol{b}'\boldsymbol{b}}\frac{\sqrt{m}}{p}\left[\frac{\left(\frac{1}{\sqrt{N}}\boldsymbol{z} - \sqrt{p}\boldsymbol{b}\right)'\left(\frac{1}{\sqrt{N}}\boldsymbol{z} - \sqrt{p}\boldsymbol{b}\right)}{y/(m-1)} - \frac{p}{N} - p\boldsymbol{b}'\boldsymbol{b}\right] \xrightarrow{\mathcal{D}} N(0,1)$$

as $N, p \to \infty$ *under the condition that* $p/n \to c \in (0,1)$.

Lemma 2. *Let* $\boldsymbol{W} \sim W_p(n, \boldsymbol{I}_p)$. *Then*

$$p^{-1/2}(\log|\boldsymbol{W}| - E[\log|\boldsymbol{W}|]) \xrightarrow{p} 0$$

as $n, p \to \infty$ *under the condition that* $p/n \to c \in (0,1)$. *Here,* $E[\log|\boldsymbol{W}|] = c_2(n)$.

Proofs of Lemma 1 and Lemma 2 are given in Appendix.

Proof of Theorem 1. It holds that

$$\frac{\hat{\delta}_i(\boldsymbol{x})}{c_1(n_i)} \overset{\mathcal{D}}{=} (m_i - 1)\left(\sqrt{p}\boldsymbol{b}_i - \frac{1}{\sqrt{N_i}}\boldsymbol{z}_i\right)'\boldsymbol{W}_i^{-1}\left(\sqrt{p}\boldsymbol{b}_i - \frac{1}{\sqrt{N_i}}\boldsymbol{z}_i\right),$$

where $\boldsymbol{z}_i \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, $\boldsymbol{W}_i \sim W_p(n_i, \boldsymbol{I}_p)$, $\boldsymbol{b}_i = p^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}_i)$. Let $\boldsymbol{\Gamma}$ be an orthogonal matrix whose first column is proportional to $\sqrt{p}\boldsymbol{b}_i - (1/\sqrt{N_i})\boldsymbol{z}_i$. From the invariance property of Wishart distribution $W_p(n_i, \boldsymbol{I}_p)$ for the transformation: $\boldsymbol{W}_i \to \boldsymbol{\Gamma}'\boldsymbol{W}_i\boldsymbol{\Gamma}$, under the condition that $\boldsymbol{z}_i$ is given,

$$\left(\sqrt{p}\boldsymbol{b}_i - \frac{1}{\sqrt{N_i}}\boldsymbol{z}_i\right)'\boldsymbol{W}_i^{-1}\left(\sqrt{p}\boldsymbol{b}_i - \frac{1}{\sqrt{N_i}}\boldsymbol{z}_i\right) \overset{\mathcal{D}}{=} \frac{1}{w_{i,11\cdot 2}}\left(\sqrt{p}\boldsymbol{b}_i - \frac{1}{\sqrt{N_i}}\boldsymbol{z}_i\right)'\left(\sqrt{p}\boldsymbol{b}_i - \frac{1}{\sqrt{N_i}}\boldsymbol{z}_i\right), \quad (1)$$

where $w_{i,11\cdot 2} = w_{i,11} - \boldsymbol{w}_{i,21}'\boldsymbol{W}_{i,22}\boldsymbol{w}_{i,21}$ for the partition

$$\boldsymbol{W}_i = \begin{pmatrix} w_{i,11} & \boldsymbol{w}_{i,21}' \\ \boldsymbol{w}_{i,21} & \boldsymbol{W}_{i,22} \end{pmatrix}.$$

From Lemma 1,

$$\frac{1}{\sqrt{2}\boldsymbol{b}_i'\boldsymbol{b}_i}\frac{\sqrt{m_i}}{p}\left[\frac{\hat{\delta}_i(\boldsymbol{x})}{c_1(n_i)} - \frac{p}{N_i} - p\boldsymbol{b}_i'\boldsymbol{b}_i\right] \xrightarrow{\mathcal{D}} N(0,1).$$

under A and C. On the other hand, from Lemma 2,

$$\frac{\sqrt{m_i}}{p}(\log|\boldsymbol{S}_i| - \log|\boldsymbol{\Sigma}_i| + p\log n_i - c_2(n_i)) \xrightarrow{p} 0$$

under A. Using Slutsky's theorem, we find that

$$\sqrt{\frac{m_i}{2\delta_i(\boldsymbol{x})}}\left\{\widehat{\delta_{i\mathrm{U}}(\boldsymbol{x})} - \delta_{i\mathrm{U}}(\boldsymbol{x})\right\} \xrightarrow{\mathcal{D}} N(0,1)$$

under A and C, which implies the conclusion of Theorem 1. □

3.2 Proof of Theorem 2

We give a proof of Theorem 2.

Proof of Theorem 2. From the expression (1),

$$
\begin{aligned}
&E\left[\left\{\frac{\hat{\delta}_i(\boldsymbol{x})}{c_1(n_i)\delta_i(\boldsymbol{x})}-1-\frac{p}{N_i\delta_i(\boldsymbol{x})}\right\}^2\right]\\
&=E\left[\left(\frac{m_i-1}{w_{i,11.2}}-1\right)^2+\frac{4p}{N_i}\frac{(m_i-1)^2(\boldsymbol{b}_i'\boldsymbol{z}_i)^2}{\{\delta_i(\boldsymbol{x})\}^2w_{i,11.2}^2}+\frac{1}{N_i^2}\frac{(m_i-1)^2(\boldsymbol{z}_i'\boldsymbol{z}_i)^2}{\{\delta_i(\boldsymbol{x})\}^2w_{i,11.2}^2}\right.\\
&\quad\left.-\frac{2p}{N_i\delta_i(\boldsymbol{x})}\left\{\frac{m_i-1}{w_{i,11.2}\delta_i(\boldsymbol{x})}\left(\sqrt{p}\boldsymbol{b}_i-\frac{1}{\sqrt{N_i}}\boldsymbol{z}_i\right)'\left(\sqrt{p}\boldsymbol{b}_i-\frac{1}{\sqrt{N_i}}\boldsymbol{z}_i\right)-1\right\}+\frac{p^2}{N_i^2\{\delta_i(\boldsymbol{x})\}^2}\right]\\
&=\frac{2}{m_i-3}+\frac{m_i-1}{m_i-3}\left[\frac{1}{4N_i\delta_i(\boldsymbol{x})}+\frac{p(p+2)}{N_i^2\{\delta_i(\boldsymbol{x})\}^2}\right]-\frac{p^2}{N_i^2\{\delta_i(\boldsymbol{x})\}^2}\\
&\to 0
\end{aligned}
$$

under A and C. It follows from Chebyshev inequality that

$$
\frac{\hat{\delta}_i(\boldsymbol{x})}{c_1(n_i)\delta_i(\boldsymbol{x})}-\frac{p}{N_i\delta_i(\boldsymbol{x})}\xrightarrow{p}1.
$$

The conclusion of Theorem 2 follows from this and continuous mapping theorem. □

A Proof of Lemma

Firstly, we prove Lemma 1.

Proof of Lemma 1. It holds that

$$
\frac{m-1}{y}=1-\left(\frac{y}{m+1}-1\right)+\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right).
$$

Consider the probability convergence of

$$
\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)'\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)\cdot\left(\frac{y}{m+1}-1\right)=\left(\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}-2\sqrt{\frac{p}{N}}\boldsymbol{b}'\boldsymbol{z}+p\boldsymbol{b}'\boldsymbol{b}\right)\cdot\left(\frac{y}{m+1}-1\right).
$$

It can be expressed that

$$
\begin{aligned}
E\left[\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}\left(\frac{y}{m+1}-1\right)\right]=0,\quad \mathrm{Var}\left[\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}\left(\frac{y}{m+1}-1\right)\right]=\frac{p(p+2)}{N^2}\frac{2}{m+1},\\
E\left[\boldsymbol{b}'\boldsymbol{z}\left(\frac{y}{m+1}-1\right)\right]=0,\quad \mathrm{Var}\left[\boldsymbol{b}'\boldsymbol{z}\left(\frac{y}{m+1}-1\right)\right]=\frac{2}{m+1}\boldsymbol{b}'\boldsymbol{b}.
\end{aligned}
$$

Hence, from Chebyshev inequality,

$$
\begin{aligned}
\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}\left(\frac{y}{m+1}-1\right)&\xrightarrow{p}0,\\
\boldsymbol{b}'\boldsymbol{z}\left(\frac{y}{m+1}-1\right)&\xrightarrow{p}0
\end{aligned}
$$

as $N,p\to\infty$ under the condition that $p/n\to c\in(0,1)$. By virtue of these probability convergences, we find that

$$
\left(\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}-2\sqrt{\frac{p}{N}}\boldsymbol{b}'\boldsymbol{z}+p\boldsymbol{b}'\boldsymbol{b}\right)\cdot\left(\frac{y}{m+1}-1\right)-p\boldsymbol{b}'\boldsymbol{b}\left(\frac{y}{m+1}-1\right)\xrightarrow{p}0. \tag{2}
$$

Next, we focus on the probability convergence of

$$\begin{aligned}&\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)'\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)\cdot\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)\\&=\left(\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}-2\sqrt{\frac{p}{N}}\boldsymbol{b}'\boldsymbol{z}+p\boldsymbol{b}'\boldsymbol{b}\right)\cdot\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right).\end{aligned}$$

It holds that

$$\begin{aligned}&E\left[\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)\right]=0,\,\mathrm{Var}\left[\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)\right]=\frac{8p(p+2)}{N^2(m+1)(m-3)},\\&E\left[\boldsymbol{b}'\boldsymbol{z}\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)\right]=0,\,\mathrm{Var}\left[\boldsymbol{b}'\boldsymbol{z}\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)\right]=\frac{8}{(m+1)(m-3)}\boldsymbol{b}'\boldsymbol{b}.\end{aligned}$$

Hence, from Chebyshev inequality,

$$\begin{aligned}&\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)\xrightarrow{p}0,\\&\boldsymbol{b}'\boldsymbol{z}\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)\xrightarrow{p}0\end{aligned}$$

as $N,p\to\infty$ under the condition that $p/n\to c\in(0,1)$. From these probability convergences, we find that

$$\left(\frac{1}{N}\boldsymbol{z}'\boldsymbol{z}-2\sqrt{\frac{p}{N}}\boldsymbol{b}'\boldsymbol{z}+p\boldsymbol{b}'\boldsymbol{b}\right)\cdot\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)-p\boldsymbol{b}'\boldsymbol{b}\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)\xrightarrow{p}0. \quad (3)$$

Combining probability convergences given in (2) and (3), and taking consideration of the equality:

$$-\left(\frac{y}{m+1}-1\right)+\left(\frac{y}{m+1}+\frac{m-1}{y}-2\right)=\frac{m-1}{y}-1,$$

we have that

$$\frac{\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)'\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)}{y/(m-1)}-\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)'\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)-\left(\frac{m-1}{y}-1\right)\cdot p\boldsymbol{b}'\boldsymbol{b}\xrightarrow{p}0.$$

Furthermore, it holds that $E[|\boldsymbol{z}'\boldsymbol{z}/(\sqrt{p}N)]=\sqrt{p}/N$, which converges to 0 as $N,p\to\infty$ under the condition that $p/n\to c\in(0,1)$. From Markov inequality, $\{1/(\sqrt{p}N)\}\boldsymbol{z}'\boldsymbol{z}\xrightarrow{p}0$. In addition, since it is expressed that $E[p^{-1/2}\boldsymbol{b}'\boldsymbol{z}]=0$ and $\mathrm{Var}(p^{-1/2}\boldsymbol{b}'\boldsymbol{z})=p^{-1}\boldsymbol{b}'\boldsymbol{b}$, we find from Chebyshev inequality that $(1/\sqrt{p})\boldsymbol{b}'\boldsymbol{z}\xrightarrow{p}0$ as $p\to\infty$. Hence,

$$\frac{1}{\sqrt{p}}\left[\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)'\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)-\frac{p}{N}-p\boldsymbol{b}'\boldsymbol{b}\right]\xrightarrow{p}0.$$

On the other hand, since it holds that $E[|\sqrt{m}/y|]=\sqrt{m}/(m-1)\to 0$ as $N,p\to\infty$ under the condition that $p/n\to c\in(0,1)$, we find from Markov inequality that $\sqrt{m}/y\xrightarrow{p}0$. In addition, $\sqrt{m+1}\{y/(m+1)-1\}\xrightarrow{\mathcal{D}}N(0,2)$. From delta method and Slutsly theorem, $\sqrt{m}\{(m-1)/y-1\}\xrightarrow{\mathcal{D}}N(0,2)$. Combining these results,

$$\frac{1}{\sqrt{2}\boldsymbol{b}'\boldsymbol{b}}\frac{\sqrt{m}}{p}\left[\frac{\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)'\left(\frac{1}{\sqrt{N}}\boldsymbol{z}-\sqrt{p}\boldsymbol{b}\right)}{y/(m-1)}-\frac{p}{N}-p\boldsymbol{b}'\boldsymbol{b}\right]\xrightarrow{\mathcal{D}}N(0,1).$$

□

Before proving Lemma 2, we give a result concerning the boundaries of series.

Lemma 3. *Let $f(x)$ be a non-negative decreasing function. Then, for positive constant $a > 1$,*

$$\lim_{n\to\infty}\int_1^{n+1} f(x+a)dx < \sum_{k=1}^{\infty} f(k+a) < \lim_{n\to\infty}\int_1^{n+1} f(x+a-1)dx$$

Proof. Since $f(x)$ is a decreasing function, for positive integer ℓ, it holds that

$$\int_\ell^{\ell+1} f(x)dx > \{(\ell+1)-\ell\}\cdot f(\ell+1) = f(\ell+1),$$
$$\int_\ell^{\ell+1} f(x)dx < \{(\ell+1)-\ell\}\cdot f(\ell) = f(\ell),$$

and so

$$\sum_{k=1}^{n} f(k+a) < \int_1^{n+1} f(x+a-1)dx < \sum_{k=1}^{n} f(k+a-1).$$

□

Proof of Lemma 2. Characteristic function of $V = \log|\boldsymbol{W}|$ is given as

$$C(t) = \frac{\Gamma_p(\frac{n}{2}+it)}{\Gamma_p(\frac{n_1}{2})},$$

where $\Gamma_p(a/2) = \pi^{p(p-1)/4}\prod_{i=1}^p \Gamma((a-i+1)/2)$. The cumulant generating function $K(t) = \log C(t)$ can be expressed as

$$K(t) = \sum_{j=1}^{p}\left\{\log\Gamma\left(\frac{n-p+j}{2}+it\right) - \log\Gamma\left(\frac{n-p+j}{2}\right)\right\}.$$

Using Taylor expansion of $\log\Gamma((n-p+j)/2+x)$ at $x = 0$, we obtain an expansion of the cumulant generating function, formally, which is as follows.

$$K(t) = \sum_{s=1}^{\infty}\frac{\kappa^{(s)}}{s!}$$

with s-th cumulant $\kappa^{(s)}$, which is given as follows.

$$\kappa^{(s)} = \sum_{j=1}^{p}\psi^{(s-1)}\left(\frac{n-p+j}{2}\right),$$

where

$$\psi^{(s)}(a) = \begin{cases} -C + \displaystyle\sum_{k=0}^{\infty}\left(\frac{1}{1+k} - \frac{1}{k+a}\right) & (s=0), \\ \displaystyle\sum_{k=0}^{\infty}\frac{(-1)^{s+1}s!}{(k+a)^{s+1}} & (s \geq 1). \end{cases}$$

Here, C denotes Euler's constant. For $\mathrm{Var}(V) = \kappa^{(2)}$, it is found from Lemma 3 that

$$\sum_{j=1}^{p}\lim_{k\to\infty}\int_1^{k+1}\frac{1}{(x+a_j-1)^2}dx < \mathrm{Var}(V) < \sum_{j=1}^{p}\lim_{k\to\infty}\int_1^{k+1}\frac{1}{(x+a_j-2)^2}dx,$$

where $a_j = (n-p+j)/2$, and so

$$\sum_{j=1}^{p} \frac{2}{n-p+j} < \mathrm{Var}(V) < \sum_{j=1}^{p} \frac{2}{n-p+j-2}. \tag{4}$$

For the left-hand side of the inequality, using Lemma 3 again,

$$\sum_{j=1}^{p} \frac{2}{n-p+j} > 2\int_{1}^{p+1} \frac{1}{n-p+x}dx = 2\log\left(1+\frac{p}{n-p+1}\right).$$

On the other hand, the right-hand side of the inequality (4) is bounded by

$$\sum_{j=1}^{p} \frac{2}{n-p+1-2} = \frac{2p}{n-p-1}.$$

Thus $\mathrm{Var}(V)$ converges a positive constant as $n, p \to \infty$, $p/n \to c \in (0,1)$. The conclusion of Lemma 2 follows from Chebyshev inequality. □

References

[1] Critchley, F. and Ford, I. (1985). Interval estimation in discrimination: the multivariate normal equal covariance case. *Biometrika*, **72**, 109–116.

[2] Critchley, F., Ford, I. and Rijal, O. (1987). Uncertainly in discrimination. *Proc. Conf. DIANA II.* Prague: Math. Inst. of the Czechoslovak Academy of Sciences, pp. 83–106.

[3] Critchley, F., Ford, I. and Rijal, O. (1988). Interval estimation based on the profile likelihood: strong Lagrangian theory with applications to discrimination *Biometrika*, **75**, 21–28.

[4] Davis, A.W. (1987). Moments of linear discriminant functions and an asymptotic confidence interval for the log odds ratio. *Biometrika*, **74**, 829–840.

[5] Fujikoshi, Y., Ulyanov, V.V. and Shimizu, R. (2010). *Multivariate Statistics High-Dimensional and Large-Sample Approximations*, Wiley, Hoboken, NJ.

[6] Hall, P., Marron, J.S. and Neeman, A. (2005). Geometric representation of high dimension, row sample size data. *J. R. Statist. Soc. B*, **67**, 427–444.

[7] Hirst, D.J., Ford, I. and Critchley, F. (1990). An empirical investigation of methods for interval estimation of the log odds ratio in discriminant analysis. *Biometrika*, **77**, 609–615.

[8] Rigby, R.A. (1982). A credibility interval for the probability that a new observation belongs to one of two multivariate normal populations. *J.R. Statist. Soc. Ser. B*, **44**, 212–220.

[9] Srivastava, M.S. (2007). Multivariate theory for analyzing high dimensional data. *J. Japan Statist. Soc.*, **37**, 53–86.

[10] Yamada, T., Himeno, T. and Sakurai, T. (2017). Interval estimation in discriminant analysis for large dimension. Accepted in *Commun. Statist. Theory and Methods.*