A BLOCK DIAGONAL COVARIANCE MATRIX TEST AND DISCRIMINANT ANALYSIS OF HIGH-DIMENSIONAL DATA

Poompong Kaewumpai

A Dissertation Submitted in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy (Statistics)
School of Applied Statistics
National Institute of Development Administration
2017

A BLOCK DIAGONAL COVARIANCE MATRIX TEST
AND DISCRIMINANT ANALYSIS OF
HIGH-DIMENSIONAL DATA

Poompong Kaewumpai
School of Applied Statistics

Professor .... Major Advisor
(Samruam Chongcharoen, Ph.D.)

The Examining Committee Approved This Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Philosophy (Statistics).

Assistant Professor .... Committee Chairperson
(Winai Bodhisuwan, Ph.D.)

Professor .... Committee
(Jirawan Jitthavech, Ph.D.)

Associate Professor .... Committee
(Vichit Lorchirachoonkul, Ph.D.)

Professor .... Committee
(Samruam Chongcharoen, Ph.D.)

Assistant Professor .... Acting Dean
(Jugkarin Sukmok, Ph.D.)

December 2017

ABSTRACT

Title of Dissertation	A Block Diagonal Covariance Matrix Test and Discriminant Analysis of High-Dimensional Data
Author	Mr. Poompong Kaewumpai
Degree	Doctor of Philosophy (Statistics)
Year	2017

In this dissertation, a new test statistic for testing for a block diagonal covariance matrix structure with a multivariate normal population where the number of variables p exceeds the number of observations n is proposed. Whereas classical approaches such as the likelihood ratio test cannot be applied when $p > n$, the proposed test statistic is based on the ratio of the estimators of $tr\Sigma^2$ and trD_{Σ}^2, where Σ is the population covariance matrix and D_{Σ} is the population covariance matrix under the null hypothesis. Furthermore, the asymptotic distribution of the proposed test statistic under the null hypothesis is standard normal. The performance of proposed test statistic was assessed using a simulation study, in which empirical type I error values and the empirical power were used to show its performance. The empirical type I error values were close to the significance level and the empirical power values were closed to 1 in all cases. Moreover, the performance of the proposed test was compared with another previously reported test statistic, and the empirical power values of the proposed test statistic were shown to be higher than those of the comparative test statistic in some cases.

Two new discriminant methods for high-dimensional data under the multivariate normal population with a block diagonal covariance matrix structure are also proposed. For the first method, the sample covariance matrix is approximated as a singular matrix based on the idea of reducing the dimensionality of the observations and using a well-conditioned covariance matrix. For the second method, a sample block diagonal covariance matrix is used instead. The performance of these two methods were compared with some of the previously reported methods via a simulation study, the results of which show that both proposed methods outperformed the other comparative methods under various conditions. In addition, the proposed test for testing block diagonal covariance matrix structure and the two new proposed methods for discriminant analysis were applied to a real-life dataset.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to everyone for helping in completing this dissertation. To be precise, I am grateful to my advisor, Professor Dr. Samruam Chongcharoen, for always giving invaluable suggestions, motivation, and encouragement; he was always available for consultation whenever I needed him. I also greatfully acknowledge my committee members: Professor Dr. Jirawan Jitthavech, Associate Professor Dr. Vichit Lorchirachoonkul and Assistant Winai Bodhisuwan for their time and helpful comments and suggestions.

I would like to express my deepest gratitude to Professor Dr. Prachoom Suwattee, Associate Professor Dr. Virool Boonyasombat, and Associate Professor Dr. Pachitjanut Siripanich for their help via knowledge transfer and dissemination. I am appreciative of my friends at the school of Applied Statistics, National Institute of Development Administration (NIDA), for their help, reinforcement, and sharing their knowledge and experience.

Finally, I am endlessly indepted to my parents and my older sister for their endless love, encouragement, and support throughout my graduate study.

Poompong Kaewumpai

December 2017

TABLE OF CONTENTS

LIST OF TABLES

Tables **Page**

CHAPTER 1

INTRODUCTION

1.1 Background

At present, data where the number of variables (denoted by p) exceeds the number of observations (denoted by n) is called the high-dimensional data, and it occurs in many scientific domains, such as genetics research, financial analysis, and computer vision.

For an example of genetics research, data from microarrays where the stress response of the microorganism *Escherichia coli* during the expression of a recombinant protein was collected by Schmidt-Heck, Guthke, Toepfer, Reischer, Duerrschmid and Bayer (2004) at the Institute of Applied Microbiology, University of Agricultural Sciences in Vienna. The data captured all 4,289 proteins encoded by 102 genes at 8, 15, 22, 45, 68, 90, 150, and 180 minutes after induction of the recombinant protein before comparison with pooled samples. Thus, this data contained 102 variables with a sample size of 8, and so high-dimensional data has occurred in this case.

In financial analysis, a company in the S&P 500 identified daily returns from 258 stocks. The relevant data was the closing prices or bid/ask average of these stocks for the trading days between 1 October 2013 and 31 December 2013 (a total of 64 days). This dataset was derived from the Center of Research in Security Prices Daily Stock in the Wharton Research Data Services. In this dataset, the sample size was 64 and the number of variables was 258, again displaying high dimensionality. Bao, Hu, Pan, and Zhou (2014) used this dataset in their simulation study.

In statistical analysis, some classical statistical methods fail to analyze high-dimensional data well because in the estimation of the covariance matrix Σ, the sample covariance matrix S is singular, which makes methods such as the likelihood ratio test is inapplicable. Hypothesis testing of whether Σ has a specific structure for a multivariate normal population, such as testing for sphericity $(H_0^1 : \Sigma = \sigma^2 I)$, complete independence $(H_0^2 : \Sigma = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_p^2))$, or independence between two subvectors $(H_0^3 : \Sigma = diag(\Sigma_{11}, \Sigma_{22}))$, is the important problem for high-dimensional data analysis. Classical methods that can be found in most multivariate statistical textbooks, such as Muirhead's (1982) and Johnson and Wichern's (2002), cannot be directly applied to derive test statistics for high-dimensional data. However, the problem for testing H_0^1 has been considered by Ledoit and Wolf (2002), Schott (2005), and Srivastava (2005). Ledoit and Wolf (2002) studied the limit distribution of some previously reported tests under $(n, p) \to \infty$ and $p/n \to c \neq 0$. They also introduced a new test statistic to test whether the covariance matrix is the identity matrix. Schott (2005) proposed a new test statistic based on the sample correlation matrix which can be used to test both H_0^1 and H_0^2. Srivastava (2005) developed test statistic for testing H_0^1 in high-dimensional data under the condition that $(\mathrm{tr}\,\Sigma^i / p) > 0$ exists, and also developed tests for testing H_0^2. Furthermore, Srivastava and Reid (2012) and Jiang, Bai, and Zheng (2013) developed the hypothesis testing of H_0^3. The asymptotic distribution of both tests were shown to be standard normal.

In this dissertation, testing for independence between m random vectors or, equivalently, testing for a block diagonal covariance matrix structure is of interest and may be considered as an extension of testing H_0^2 and H_0^3. The new test statistic is proposed which is based on the ratio of an unbiased and consistent estimator proposed by Srivastava (2005). The distribution of the proposed test statistic under H_0 is also derived. To evaluate the performance of this test statistic, a simulation study to calculate the empirical type I error rate and empirical power of the test was performed. Moreover, these measures were used to compare the performance of the proposed test statistic with some previously reported ones using a simulation study.

In addition, one of the multivariate tasks, discriminant analysis, was also studied under the population covariance matrix with a block diagonal structure, which can be tested for by the proposed test. The main purpose of discriminant analysis is to enable classification of new observations into one of g classes or populations. In high-dimensional data, classical discriminant analysis cannot be applied directly because the sample covariance matrix is singular, i.e. the inverse of the sample covariance matrix does not exist. Di Pillo (1976) stated that the performance of discriminant analysis in high-dimensional data is far from optimal, and the generalized inverse of the sample covariance matrix is usually used when it is singular. Despite its simplicity, this method might have poor performance since the generalized inverse will be very unstable because of the lack of some of the observations (Guo, Hastie, & Tibshirani, 2007).

In this situation, the challenging problem of discriminant analysis is the singularity of the sample covariance matrix. There are often two ways to address this problem, the first of which is a subspace approach (dimensionality reduction). For example, among these are the well-known Fisherfaces method (Belhumeur, Hespanha, & Kriegman, 1997) and Chen, Liao, Ko, Lin and Yu's (2000) direct linear discriminant analysis (D-LDA). Lu, Plataniotis and Venetsanopoulos (2003) proposed a new discriminant analysis method called regularized direct quadratic discriminant analysis (RD-QDA) by combining the D-LDA method with the regularized discriminant analysis method previously proposed by Friedman (1989). The second method is to apply linear algebra to solve the singularity problem. For example, Tian, Barbero, Gu, and Lee (1986) utilized the pseudo inverse to estimate the sample covariance matrix. Friedman (1989) used the regularization technique of discriminant analysis to shrink the sample covariance matrix. Additionally, Srivastava and Kubokawa (2007) used an empirical Bayes estimator of covariance matrix instead of the sample covariance matrix.

In this dissertation, two new discriminant methods are proposed to construct a method for dealing with high-dimensional data. Firstly, the dimensionality of the observations is reduced by taking the linear combinations of $\underset{\sim}{x}_{kh}$ to create $\underset{\sim}{y}_{kh} = H^T \underset{\sim}{x}_{kh}$, where H is the matrix obtained from the RD-QDA method (Lu et al., 2003), and then find a well-conditioned estimator for a large dimensional covariance matrix using the expression given by Ledoit and Wolf (2003, 2004). Secondly, the block diagonal of sample covariance matrix $S_{block} = diag(S_{11}, S_{22}, ..., S_{mm})$, where

$S_{ii}, i = 1, 2, .., m$ are submatrices on the diagonal of the pooled sample covariance matrix, is used. The two new discriminant methods were evaluated by performing a simulation study to calculate the misclassification rate, sensitivity, and specificity, and comparing them with some previously reported methods.

1.2 Objectives of the Study

In this dissertation, new statistical techniques in high-dimensional data are proposed and studied with the following objectives:

1) To propose a new test for testing a block diagonal covariance matrix structure in high-dimensional data with a multivariate normal population

2) To propose two new methods for discriminant analysis in high-dimensional data with a multivariate normal population.

3) To assess the performance of the proposed test statistic by considering its empirical type 1 errors and empirical power, and comparing them with some of the previously reported test statistics through a simulation study.

4) To assess the performance of the proposed methods by considering the misclassification rate, sensitivity, and specificity and comparing them with some of the previously reported methods through a simulation study.

1.3 Scope of the Study

In this study, the test statistic for testing a block diagonal covariance matrix structure and two methods for sample classification in high-dimensional data are proposed under the following conditions.

1.3.1 Testing for a Block Diagonal Covariance Matrix

1) The data are assumed to be from a multivariate normal distribution with a positive definite covariance matrix Σ with $p \times p$ dimensions, and mean vectors $\underset{\sim}{\mu}$, all of which are assumed to be unknown.

2) High-dimensional data means that the sample size is less than the number of variables $(n < p)$.

1.3.2 Discriminant Analysis

1) There are two different populations, each assumed to have a multivariate normal distribution with a common positive definite covariance matrix $\Sigma_1 = \Sigma_2 = \Sigma$ with $p \times p$ dimensions and mean vectors $\underset{\sim}{\mu}_h$, $h = 1, 2,$ all of which are assumed to be unknown.

2) The sample size from both populations are equal $(n = n_1 = n_2)$.

3) The probability that each observation comes from either population is equal.

4) A random sample of n observations from these populations with their true group labels is unknown.

5) High-dimensional data means that the degrees of freedom of the pooled sample covariance matrix is less than the number of variables $(\nu < p)$.

1.4 Usefulness of the Study

The new proposed test statistic for testing block diagonal covariance matrix structure and the two methods for sample classification may be beneficial for analyzing high-dimensional data in genetics and computer vision, or any other field involving high-dimensional data.

CHAPTER 2

LITERATURE REVIEW

This chapter contains a review of the literature on testing for a block diagonal covariance matrix in Section 2.1, followed by the classical approach and the high-dimensional approach. A review of discriminant analysis, both the classical approach and the high-dimensional approach, are provided in Section 2.2.

2.1 Testing Block Diagonal Covariance Matrix

In this section, the necessary notations that used in testing for a block diagonal covariance matrix are defined. These notations are used in both the classical approach and the high-dimensional approach.

Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n$ be i.i.d. as p dimensional random vectors which have a multivariate normal distribution with mean $\underset{\sim}{\mu}$ and positive definite covariance matrix Σ; $\underset{\sim}{\mu}$ and Σ are unknown parameters denoted by $\underset{\sim}{X}_k \sim N_p(\underset{\sim}{\mu}, \Sigma)$; and k represents the number of observation from a random sample, $k = 1, 2, ..., n$. The set of all random vectors can be used to construct an observations matrix X, such that

$$X = (\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n)^T = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{p1} \\ x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{pn} \end{pmatrix}.$$

The probability density function (pdf) for random vector $\underset{\sim}{X}_k$, $k = 1, 2, ..., n$ from a multivariate normal population is defined as

$$f(\underset{\sim}{X}_k) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\left(\underset{\sim}{X}_k - \underset{\sim}{\mu}\right)^T \Sigma^{-1} \left(\underset{\sim}{X}_k - \underset{\sim}{\mu}\right)/2},$$

where $E(\underset{\sim}{X}_k)=\underset{\sim}{\mu}=\begin{pmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_p\end{pmatrix}$ and $E(\underset{\sim}{X}_k-\underset{\sim}{\mu})(\underset{\sim}{X}_k-\underset{\sim}{\mu})^T=\Sigma=\begin{pmatrix}\sigma_{11}&\sigma_{12}&\cdots&\sigma_{1p}\\ \sigma_{12}&\sigma_{22}&\cdots&\sigma_{2p}\\ \vdots&\vdots&\ddots&\vdots\\ \sigma_{1p}&\sigma_{2p}&\cdots&\sigma_{pp}\end{pmatrix}$,

$k=1,2,...,n$.

Now consider for each vector $\underset{\sim}{X}_k$, $p\times 1$, we partition it into m components with each group of sizes $p_i\times 1,\ i=1,2,...,m$. We partition $\underset{\sim}{X}_k$, $\underset{\sim}{\mu}$, and Σ into m components as

$\underset{\sim}{X}_k=\begin{pmatrix}\underset{\sim}{X}_k^{(1)}\\ \underset{\sim}{X}_k^{(2)}\\ \vdots\\ \underset{\sim}{X}_k^{(m)}\end{pmatrix}$, $\underset{\sim}{\mu}=\begin{pmatrix}\underset{\sim}{\mu}^{(1)}\\ \underset{\sim}{\mu}^{(2)}\\ \vdots\\ \underset{\sim}{\mu}^{(m)}\end{pmatrix}$, where $\underset{\sim}{X}_k^{(i)}$ and $\underset{\sim}{\mu}^{(i)}$ are $p_i\times 1$ vector, and

$\Sigma=\begin{pmatrix}\Sigma_{11}&\Sigma_{12}&\cdots&\Sigma_{1m}\\ \Sigma_{21}&\Sigma_{22}&\cdots&\Sigma_{2m}\\ \vdots&\vdots&\ddots&\vdots\\ \Sigma_{m1}&\Sigma_{m2}&\cdots&\Sigma_{mm}\end{pmatrix}$, where Σ_{ij} is a $p_i\times p_j$ submatrix, $i,j=1,2,...,m$.

Note that $\underset{\sim}{X}_k^{(1)},\underset{\sim}{X}_k^{(2)},...\underset{\sim}{X}_k^{(m)}$ represents a partitioning of $\underset{\sim}{X}_k$, a random sample of independent vectors and $E(\underset{\sim}{X}_k^{(i)})=\underset{\sim}{\mu}^{(i)}$, $E(\underset{\sim}{X}_k^{(i)}-\underset{\sim}{\mu}^{(i)})(\underset{\sim}{X}_k^{(j)}-\underset{\sim}{\mu}^{(j)})^T=\Sigma_{ij}$.

As $\underset{\sim}{\mu}$ and Σ are unknown, the unbiased estimators of these parameters are, respectively, $\bar{\underset{\sim}{X}}=\frac{1}{n}\sum_{k=1}^{n}\underset{\sim}{X}_k$,

$$S=\frac{1}{N}\sum_{k=1}^{N+1}(\underset{\sim}{X}_k-\bar{\underset{\sim}{X}})(\underset{\sim}{X}_k-\bar{\underset{\sim}{X}})^T=\begin{pmatrix}s_{11}&s_{12}&\cdots&s_{1p}\\ s_{12}&s_{22}&\cdots&s_{2p}\\ \vdots&\vdots&\ddots&\vdots\\ s_{1p}&s_{2p}&\cdots&s_{pp}\end{pmatrix}\text{, where } n=N+1.$$

We can partition $\bar{\underset{\sim}{X}}$ and S in the same manner as $\underset{\sim}{\mu}$ and Σ:

$$\bar{\underset{\sim}{X}}=\begin{pmatrix}\bar{\underset{\sim}{X}}^{(1)}\\ \bar{\underset{\sim}{X}}^{(2)}\\ \vdots\\ \bar{\underset{\sim}{X}}^{(m)}\end{pmatrix},\ S=\begin{pmatrix}S_{11}&S_{12}&\cdots&S_{1m}\\ S_{21}&S_{22}&\cdots&S_{2m}\\ \vdots&\vdots&\ddots&\vdots\\ S_{m1}&S_{m2}&\cdots&S_{mm}\end{pmatrix},$$

where $\underset{\sim}{\bar{X}}^{(i)}$ are $p_i \times 1$ vector, and S_{ij} is a $p_i \times p_j$ submatrix, $i, j = 1, 2, ..., m$ and

$\underset{\sim}{\bar{X}}^{(i)} = \frac{1}{n}\sum_{k=1}^{n} \underset{\sim}{\bar{X}}_k^{(i)}$, $S_{ij} = \frac{1}{N}\sum_{k=1}^{N+1} (\underset{\sim}{\bar{X}}_k^{(i)} - \underset{\sim}{\bar{X}}^{(i)})(\underset{\sim}{\bar{X}}_k^{(j)} - \underset{\sim}{\bar{X}}^{(j)})^T$.

The above notations are used in entire this dissertation. The next two subsections describe the classical approach and the high-dimensional approach in testing for a block diagonal covariance matrix.

2.1.1 The Classical Approach

In order to test for a block diagonal covariance structure, the hypothesis is constructed as follows:

$$H_0 : \Sigma = D_\Sigma \text{ vs } H_a : \Sigma \neq D_\Sigma , \tag{2.1}$$

where $D_\Sigma = diag(\Sigma_{11}, \Sigma_{22}, \ldots, \Sigma_{mm})$, Σ_{ii} is $p_i \times p_i$, $i = 1, 2, ..., m$ square matrix on the main diagonal and all submatrix off-diagonal is zero matrix of $p \times p$ matrix. Suppose a sample of size $n, \underset{\sim}{x}_1, \underset{\sim}{x}_2, \cdots, \underset{\sim}{x}_n$ are observation on $\underset{\sim}{X}_k$, $k = 1, 2, ..., n$, then the likelihood ratio is

$$\Lambda_n = \frac{\max\limits_{\{\underset{\sim}{\mu}, D_\Sigma\}} L(\underset{\sim}{\mu}, D_\Sigma)}{\max\limits_{\{\underset{\sim}{\mu}, \Sigma\}} L(\underset{\sim}{\mu}, \Sigma)} , \tag{2.2}$$

where $L\left(\underset{\sim}{\mu}, \Sigma\right) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} e^{-\sum_{k=1}^{n}\left(\underset{\sim}{x}_k - \underset{\sim}{\mu}\right)^T \Sigma^{-1}\left(\underset{\sim}{x}_k - \underset{\sim}{\mu}\right) / 2}$.

$L(\underset{\sim}{\mu}, D_\Sigma)$ is $L(\underset{\sim}{\mu}, \Sigma)$ with $\Sigma_{ij} = 0, i \neq j$, for all $0 \leq i, j \leq m;$ and the maximum is taken with respect to all vectors $\underset{\sim}{\mu}$ and with positive definite Σ and D_Σ. Let $\hat{\Sigma}$ and $\hat{\Sigma}_{ii}$ are maximum likelihood estimators of Σ and Σ_{ii}, respectively. According to Theorem 11.2.2 from Muirhead (1982), we have

$$\max_{\{\mu, \Sigma\}} L(\mu, \Sigma) = L\left(\hat{\mu}, \hat{\Sigma}\right) = \frac{1}{(2\pi)^{np/2} |\hat{\Sigma}|^{n/2}} e^{-np/2}$$

Under the null hypothesis,

$$\max_{\{\mu, D_\Sigma\}} L(\mu, D_\Sigma) = \prod_{i=1}^{m} L_i(\underset{\sim}{\mu}^{(i)}, \hat{\Sigma}_{ii}),$$

$$= \prod_{i=1}^{m} \frac{1}{(2\pi)^{np_i/2} \left|\hat{\Sigma}_{ii}\right|^{n/2}} e^{-np_i/2},$$

$$= \frac{1}{(2\pi)^{np/2} \prod_{i=1}^{m} \left|\hat{\Sigma}_{ii}\right|^{n/2}} e^{-np/2},$$

then (2.2) becomes

$$\Lambda_n = \frac{\max_{\{\mu, D_\Sigma\}} L(\mu, D_\Sigma)}{\max_{\{\mu, \Sigma\}} L(\mu, \Sigma)} = \frac{\left|\hat{\Sigma}\right|^{n/2}}{\prod_{i=1}^{m} \left|\hat{\Sigma}_{ii}\right|^{n/2}}.$$

The value of likelihood ratio Λ_n is between 0 and 1. Low values of the likelihood ratio Λ_n mean that the observed result was less probable to occur under the hypothesis H_0 as compared to the hypothesis H_a. The likelihood ratio test rejects the hypothesis H_0 if the value of Λ_n is too small. How small is too small depends on the size of the test. Thus, the likelihood ratio test rejects the hypothesis H_0 if $\Lambda_n < c_\alpha$, where c_α is chosen so that the size of the test is α. For a large n, the asymptotic distribution of $-w \log \Lambda_n$ under H_0 is a chi-squared distribution with t degrees of freedom, where $w = n - 2u$, $u = \dfrac{\left(p^3 - \sum_{i=1}^{m} p_i^3\right) + 9\left(p^2 - \sum_{i=1}^{m} p_i^2\right)}{6\left(p^2 - \sum_{i=1}^{m} p_i^2\right)}$, and $t = \dfrac{1}{2}\left(p^2 - \sum_{i=1}^{m} p_i^2\right)$ (this result can be found in Srivastava (2002, p. 492). Under the assumption that the dimension p is smaller than the sample size n, the asymptotic results perform well, but when the dimension p is larger than the sample size n, the asymptotic results cannot be applied.

2.1.2 The High-Dimensional Approach

For high-dimensional data, Srivastava (2005) proposed a test statistic for testing the hypothesis

$$H_{01}:\Sigma = diag(\sigma_{11},...,\sigma_{pp}) \text{ vs } H_{a1}:\Sigma \neq diag(\sigma_{11},...,\sigma_{pp}), \tag{2.3}$$

where σ_{ii}, $i=1,2,...,p$ are the diagonal elements of the population covariance matrix. This hypothesis can be considered as a special case of H_0, where $p_i = 1$. Let

$$a_{20} = \frac{1}{p}\sum_{i=1}^{p}\sigma_{ii}^2,$$

$$a_{40} = \frac{1}{p}\sum_{i=1}^{p}\sigma_{ii}^4,$$

Then,

$$a_2 = \frac{1}{p}tr\Sigma^2 = \frac{1}{p}\left(\sum_{i=1}^{p}\sigma_{ii}^2 + \sum_{i\neq j}^{p}\sigma_{ij}^2\right),$$

$$= a_{20} + \frac{1}{p}\sum_{i\neq j}^{p}\sigma_{ij}^2.$$

Srivastava (2005) consider the parametric function

$$T_{S1} = \frac{a_2}{a_{20}}.$$

Obviously $T_{S1} = 1$ if and only if H_{01} is true, and if H_{01} is false, $T_{S1} \geq 1$. Thus, (2.3) can be based on the consistent estimator $\hat{T}_{S1}$ given by

$$\hat{T}_{S1} = \frac{\hat{a}_2}{\hat{a}_{20}},$$

where
$$\hat{a}_2 = \frac{N^2}{(N-1)(N+2)p}\left\{trS^2 - \frac{1}{N}(trS)^2\right\} \tag{2.4}$$

and
$$\hat{a}_{20} = \frac{N^2}{(N+2)p}\frac{1}{p}\sum_{i=1}^{p}s_{ii}^2 \tag{2.5}$$

are the unbiased and consistent estimator of a_2 and a_{20}, respectively. Since (2.3) is equivalent to

$$H_{01}^*: T_{S1} = 1 \text{ vs } H_{a1}^*: T_{S1} \geq 1, \tag{2.6}$$

which is a one-sided test, then testing H_{01}^{*} can be considered. Srivastava (2005) proposed the test statistic for testing (2.6) as

$$T_{S1}^{*}=\left(\frac{N}{2}\right)\frac{(\hat{T}_{S1}-1)}{\left[1-\left(\frac{1}{p}\right)\left(\frac{\hat{a}_{40}}{\hat{a}_{20}^{2}}\right)\right]^{\frac{1}{2}}},$$

where $\hat{a}_{40}=\frac{1}{p}\sum_{i=1}^{p}s_{ii}^{4}$.

The asymptotic distribution of T_{S1}^{*} is derived under the following assumptions:

(A1) As $p\to\infty,\ a_k\to a_k^0, 0<a_k^0<\infty, k=1,2,...,8,$ where $a_k=tr\Sigma^k/p$.

(A2) $N=O(p^{\delta}),\ 0<\delta<1.$

Under assumptions (A1) and (A2), the assumed distribution of T_{S1}^{*} is given by $T_{S1}^{*}\sim N\left[\gamma,\tau^2\right]$, where $\gamma=\left(\frac{N}{2}\right)\frac{(T_{S1}-1)}{\left[1-\left(\frac{1}{p}\right)\left(\frac{a_{40}}{a_{20}^{2}}\right)\right]^{\frac{1}{2}}}$ and $\tau^2=\frac{a_2^2-p^{-1}a_4}{a_{20}^2-p^{-1}a_{40}}$. If the null hypothesis is true, $\gamma=0$, and $\tau^2=1$, then T_{S1}^{*} has an asymptotic standard normal distribution.

Srivastava and Reid (2012) considered the hypothesis of independence of two subvectors by partitioning the random vector $\underset{\sim}{X}_k$ into two parts: $\underset{\sim}{X}_k=(X_k^{(1)}\quad X_k^{(2)})^T$, of length $p_1,\ p_2$, respectively; thus, this hypothesis can be written as follows:

$$H_{02}:\Sigma=\begin{bmatrix}\Sigma_{11} & O_{12}\\ O_{21} & \Sigma_{22}\end{bmatrix}\ \text{vs}\ H_{a2}:\Sigma\neq\begin{bmatrix}\Sigma_{11} & O_{12}\\ O_{21} & \Sigma_{22}\end{bmatrix}. \qquad (2.7)$$

The proposed test statistic of Srivastava and Reid (2012) is based on the difference function between the null hypothesis H_{02} and the alternative hypothesis H_{a2}, given as

$$\psi^2=\frac{1}{2p\sqrt{2}}tr\left[D^{-1}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}-D^{-1}\begin{pmatrix}\Sigma_{11} & O_{12}\\ O_{21} & \Sigma_{22}\end{pmatrix}\right]^2,$$

where D is a diagonal matrix in which the first p_1 of diagonal elements are $a_{21}^{1/2}$, the last p_2 of diagonal elements are $a_{22}^{1/2}$, and $a_{21} = \frac{tr(\Sigma_{11}^2)}{m}$, $a_{22} = \frac{tr(\Sigma_{22}^2)}{m}$, $a_{(1,2)} = \frac{tr(\Sigma_{12}\Sigma_{21})}{m}$. Thus, ψ^2 can be rewritten as

$$\psi^2 = \frac{1}{2p\sqrt{2}} tr\begin{pmatrix} O_{11} & a_{21}^{-1/2}\Sigma_{12} \\ a_{22}^{-1/2}\Sigma_{21} & O_{22} \end{pmatrix}^2 = \frac{a_{(1,2)}}{\sqrt{a_{21}a_{22}}}.$$

Note that the null hypothesis H_{02} is true if and only if $a_{(1,2)} = 0$. The consistent estimators of a_{21}, a_{22}, and $a_{(1,2)}$ are given by

$$\hat{a}_{2i} = \frac{N^2}{(N-1)(N+2)p}\left[tr(S_{ii}) - \frac{1}{N}tr(S_{ii})^2\right],\ i = 1,2\ , \tag{2.8}$$

$$\hat{a}_{(1,2)} = \frac{N^2}{(N-1)(N+2)p}\left[tr(S_{12}S_{21}) - \frac{1}{N}tr(S_{11})tr(S_{22})\right], \tag{2.9}$$

where S is defined as the partition in the same manner as Σ:

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

Srivastava and Reid (2012) proposed a test statistic for testing (2.7) as

$$\hat{T}_{S2} = \frac{N\hat{a}_{(1,2)}}{\sqrt{\hat{a}_{21}\hat{a}_{22}}}\ .$$

Let new assumption (A3) be $0 < \lim_{p\to\infty}(p_i/p) = c_i < \infty,\ i = 1,2$; if the null hypothesis H_{02} is true, $(p, N) \to \infty$, and assumptions (A1), (A2), and (A3) hold, then the asymptotically distribution of $\hat{T}_{S2}$ is a standard normal distribution.

Motivated by Srivastava (2005), and Srivastava and Reid (2012). Hyodo, Shutoh, Nishiyama, and Pavlenko (2015) defined a test statistic to test hypothesis H_0 using a distance function between the null hypothesis and the alternative hypothesis. This distance function is given by the normalized Frobenious norm of the difference between matrices Σ and D_Σ, which can be expressed by

$$\|D_\Sigma - \Sigma\|^2 = \frac{tr(D_\Sigma - \Sigma)(D_\Sigma - \Sigma)^T}{p} = \frac{2\sum_{j=2}^{m}\sum_{i=1}^{j-1} tr\Sigma_{ji}\Sigma_{ij}}{p}$$

$$= \frac{\sum_{j=1}^{m}\sum_{i=1}^{m} tr\Sigma_{ji}\Sigma_{ij}}{p} - \frac{\sum_{i=1}^{m} tr\Sigma_{ii}^2}{p}$$

$$= \frac{tr\Sigma^2}{p} - \frac{\sum_{i=1}^{m} tr\Sigma_{ii}^2}{p}$$

$$= a_2 - \sum_{i=1}^{m} a_{2i}$$

$$= a_2 - a_{2D} \qquad (2.10)$$

where $a_k = \frac{tr\Sigma^k}{p}$, $a_{ki} = \frac{tr\Sigma_{ii}^k}{p}$, and $a_{kD} = \sum_{i=1}^{m} a_{ki}$.

If the null hypothesis is true, (2.10) is equal to zero and (2.1) can be rewritten as $H_0 : T_b = 0$, where $T_b = a_2 - a_{2D}$. Hyodo, Shutoh, Nishiyama, and Pavlenko (2015) estimated T_b using an unbiased and consistent estimator for high-dimensional data obtained from Srivastava's (2005) results. The estimators of a_2 and a_{2i} are defined as

$$\hat{a}_2 = \frac{N^2}{(N-1)(N+2)p}\left\{trS^2 - \frac{1}{N}(trS)^2\right\}, \qquad (2.11)$$

$$\hat{a}_{2i} = \frac{N^2}{(N-1)(N+2)p}\left\{trS_{ii}^2 - \frac{1}{N}(trS_{ii})^2\right\},\ i = 1,2,\ldots,m . \qquad (2.12)$$

The estimator of T_b is given by $\hat{T}_b = \hat{a}_2 - \hat{a}_{2D}$, where $\hat{a}_{2D} = \sum_{i=1}^{m} \hat{a}_{2i}$. Hyodo, Shutoh, Nishiyama, and Pavlenko (2015) derived higher-order moments of a multivariate normal random vector used to find the distribution of $\hat{T}_b$ using the following assumptions:

(B1) $p_i (i = 1,2,...)$ is fixed and $m \to \infty$,

(B2) $N = O(k^\delta),\ 0 < \delta < 1$,

(B3) $\sum_{i \neq j}^{m} a_{2i}a_{2j} \approx k^2$, (B4) $\sum_{i=1}^{m} a_{2i}^2 = o(k^2)$, $\sum_{i=1}^{m} a_{4i} = o(k^2)$, where $a_{4i} = tr\Sigma_{ii}^4$, $i = 1, 2, ..., m$,

(B5) $a_2 = O(1)$, $a_4 = \frac{tr\Sigma^4}{p} = O(k)$, $\sum_{i \neq j}^{m} tr\Sigma_{ji}\Sigma_{ii}\Sigma_{jj}\Sigma_{ij} = O(k^2)$,

$\sum_{i \neq j}^{m} (tr\Sigma_{ij}\Sigma_{ji})^2 = O.$

Subsequently, under assumptions (B1) – (B5) and that the null hypothesis H_0 is true, the asymptotic distribution of $\hat{T}_b$ which is given by

$$\hat{T}_b \sim N(0, \frac{\hat{\lambda}^2}{N^2}),$$

where $\hat{\lambda}^2 = \frac{\sum_{j=2}^{m}\sum_{i=1}^{j-1} 8(N-1)(N+2)\hat{a}_{2j}\hat{a}_{2i}}{N^2}$. Note that assumptions (B1) – (B5) are stronger than assumptions (A1) – (A3) as Hyodo, Shutoh, Nishiyama, and Pavlenko (2015) measured the performance of $\hat{T}_b$ using a simulation study and showed that this test statistic performed well when p was much larger than n and the correlation between the variables was weak.

For studies on the next relevant point, Bao, Hu, Pan, and Zhou (2014) proposed a statistic to test H_0 developed from Schott's (2005) statistic to test for complete independence. Their idea turned out to be a particular linear spectral statistic of a block correlation, and their statistic is defined as follows:

$$\hat{T}_c = \frac{1}{2}trB^2 - \frac{p}{2},$$

where $B = diag\left[Y^{(i)}Y^{(i)T}\right]^{-1/2} \cdot \left[YY^T\right] \cdot diag\left[Y^{(i)}Y^{(i)T}\right]_{i=1,...,m}^{-1/2}$,

$Y = \left(\underset{\sim}{X}_1 - \underset{\sim}{\bar{X}} \;\; \underset{\sim}{X}_2 - \underset{\sim}{\bar{X}} \; \cdots \underset{\sim}{X}_n - \underset{\sim}{\bar{X}}\right)$, and $Y^{(i)} = \left(\underset{\sim}{X}_1^{(i)} - \underset{\sim}{\bar{X}}^{(i)} \;\; \underset{\sim}{X}_2^{(i)} - \underset{\sim}{\bar{X}}^{(i)} \; \cdots \underset{\sim}{X}_n^{(i)} - \underset{\sim}{\bar{X}}^{(i)}\right)$.

It is assumed that $p_i < n$ for all i holds and $n \to \infty$. The asymptotic distribution of $\hat{T}_c$ under H_0 is given by

$$\hat{T}_c \sim N(a_n, b_n)$$

where $a_n = \frac{1}{2}\frac{\sum_{i\neq j} p_i p_j}{n-1}$, $b_n = \frac{1}{2}\frac{\sum_{i\neq j} p_i p_j (n-1-p_i)(n-1-p_j)}{(n-1)^4}$.

This test performed satisfactory in their simulation study and they also used this test on a real-life dataset.

2.2 Discriminant Analysis

Discriminant analysis is one of the popular multivariate techniques used in sample classification. One objective of discriminant analysis is to construct appropriate rules for assigning new observations to one of g classes or populations.

Let $\underset{\sim}{X}_{1h}, \underset{\sim}{X}_{2h}, ..., \underset{\sim}{X}_{n_h h}$ be i.i.d. as p dimensional random vectors which have multivariate normal distribution with mean $\underset{\sim}{\mu}_h$ and a positive definite covariance matrix Σ_h where h represents the h^{th} classes (populations), given by $\prod_h$, $h = 1, 2, ..., g$; $\underset{\sim}{\mu}_h$ and Σ_h are unknown parameters denoted by $\underset{\sim}{X}_{kh} \sim N_p(\underset{\sim}{\mu}_h, \Sigma_h)$; and k represents the number of observation from a random sample, $k = 1, 2, ..., n_h$. The set of all random vectors constructed as an observations matrix can be written as

$$X_h = \left(\underset{\sim}{X}_{1h}\ \underset{\sim}{X}_{2h}\ \cdots \underset{\sim}{X}_{n_h h}\right)^T = \begin{pmatrix} x_{11h} & x_{21h} & \cdots & x_{p1h} \\ x_{12h} & x_{22h} & \cdots & x_{p2h} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1nh} & x_{2nh} & \cdots & x_{pn_h h} \end{pmatrix}.$$

The pdf for random vector $\underset{\sim}{X}_{kh}$, $k = 1, 2, ..., n_h$, $h = 1, 2, ..., g$ from a multivariate normal distribution is defined as

$$f(\underset{\sim}{X}_{kh}) = \frac{1}{(2\pi)^{p/2}|\Sigma_h|^{1/2}} e^{-\left(\underset{\sim}{X}_{kh}-\underset{\sim}{\mu}_h\right)^T \Sigma^{-1}\left(\underset{\sim}{X}_{kh}-\underset{\sim}{\mu}_h\right)/2},$$

where $E(\underset{\sim}{X}_{kh}) = \underset{\sim}{\mu}_h = \begin{pmatrix} \mu_{1h} \\ \mu_{2h} \\ \vdots \\ \mu_{ph} \end{pmatrix}$, $E(\underset{\sim}{X}_{kh} - \underset{\sim}{\mu}_h)(\underset{\sim}{X}_{kh} - \underset{\sim}{\mu}_h)^T = \Sigma_h = \begin{pmatrix} \sigma_{11h} & \sigma_{12h} & \cdots & \sigma_{1ph} \\ \sigma_{12h} & \sigma_{22h} & \cdots & \sigma_{2ph} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1ph} & \sigma_{2ph} & \cdots & \sigma_{pph} \end{pmatrix}$,

$k = 1, 2, ..., n_h$, and $h = 1, 2, ..., g$.

Now consider for each vector $\underset{\sim}{X}_{kh}$, $p\times 1$, we partition it into m components with each group of size $p_i \times 1$, $i=1,2,...,m$. We partition $\underset{\sim}{X}_{kh}$, $\underset{\sim}{\mu}_h$, and Σ_h into m components as

$$\underset{\sim}{X}_{kh} = \begin{pmatrix} \underset{\sim}{X}_{kh}^{(1)} \\ \underset{\sim}{X}_{kh}^{(2)} \\ \vdots \\ \underset{\sim}{X}_{kh}^{(m)} \end{pmatrix},\ \underset{\sim}{\mu}_h = \begin{pmatrix} \underset{\sim}{\mu}_h^{(1)} \\ \underset{\sim}{\mu}_h^{(2)} \\ \vdots \\ \underset{\sim}{\mu}_h^{(m)} \end{pmatrix}, \text{ where } \underset{\sim}{X}_k^{(i)} \text{ and } \underset{\sim}{\mu}^{(i)} \text{ are } p_i \times 1 \text{ vector, and}$$

$$\Sigma_h = \begin{pmatrix} \Sigma_{11h} & \Sigma_{12h} & \cdots & \Sigma_{1mh} \\ \Sigma_{21h} & \Sigma_{22h} & \cdots & \Sigma_{2mh} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m1h} & \Sigma_{m2h} & \cdots & \Sigma_{mmh} \end{pmatrix}, \text{ where } \Sigma_{ijh} \text{ is a } p_i \times p_j \text{ submatrix, } i,j=1,2,...,m.$$

Note that $\underset{\sim}{X}_{kh}^{(1)}, \underset{\sim}{X}_{kh}^{(2)},..., \underset{\sim}{X}_{kh}^{(m)}$ represents a partitioning of $\underset{\sim}{X}_{kh}$, not a random sample of independent vectors and $E(\underset{\sim}{X}_{kh}^{(i)}) = \underset{\sim}{\mu}_h^{(i)}$, $E(\underset{\sim}{X}_{kh}^{(i)} - \underset{\sim}{\mu}_h^{(i)})(\underset{\sim}{X}_{kh}^{(j)} - \underset{\sim}{\mu}_h^{(j)})^T = \Sigma_{ijh}$. As $\underset{\sim}{\mu}_h$, Σ_h are unknown, the unbiased estimators of these parameters are, respectively,

$$\bar{\underset{\sim}{X}}_h = \frac{1}{n_h}\sum_{k=1}^{n_h} \underset{\sim}{X}_{kh},$$

$$S_h = \frac{1}{n_h - 1}\sum_{k=1}^{n_h} (\underset{\sim}{X}_{kh} - \bar{\underset{\sim}{X}}_h)(\underset{\sim}{X}_{kh} - \bar{\underset{\sim}{X}}_h)^T = \begin{pmatrix} s_{11h} & s_{12h} & \cdots & s_{1ph} \\ s_{12h} & s_{22h} & \cdots & s_{2ph} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1ph} & s_{2ph} & \cdots & s_{pph} \end{pmatrix}.$$

We partition $\bar{\underset{\sim}{X}}_h$, S_h in the same manner as $\underset{\sim}{\mu}_h$, Σ_h:

$$\bar{\underset{\sim}{X}}_h = \begin{pmatrix} \bar{\underset{\sim}{X}}_h^{(1)} \\ \bar{\underset{\sim}{X}}_h^{(2)} \\ \vdots \\ \bar{\underset{\sim}{X}}_h^{(m)} \end{pmatrix},\ S_h = \begin{pmatrix} S_{11h} & S_{12l} & \cdots & S_{1mh} \\ S_{21h} & S_{22h} & \cdots & S_{2mh} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1h} & S_{m2h} & \cdots & S_{mmh} \end{pmatrix},$$

where $\bar{\underset{\sim}{X}}_h^{(i)}$ are $p_i \times 1$ vector, and S_{ijh} is a $p_i \times p_j$ submatrix, $i,j=1,2,...,m$ and

$$\bar{\underset{\sim}{X}}_h^{(i)} = \frac{1}{n_h}\sum_{k=1}^{n_h} \bar{\underset{\sim}{X}}_{kh}^{(i)},\ S_{ijh} = \frac{1}{n_h - 1}\sum_{k=1}^{n_h} (\bar{\underset{\sim}{X}}_{kh}^{(i)} - \bar{\underset{\sim}{X}}_h^{(i)})(\bar{\underset{\sim}{X}}_{kh}^{(j)} - \bar{\underset{\sim}{X}}_h^{(j)})^T.$$

When the covariance matrix of each class is equal $(\Sigma_1 = \Sigma_2 = \cdots \Sigma_g = \Sigma)$, then the population covariance matrix Σ can be estimated by

$$S_{pooled} = \frac{1}{n_1 + \cdots + n_g} \sum_{h=1}^{g} \sum_{k=1}^{n_h} (\underset{\sim}{X}_{kh} - \bar{\underset{\sim}{X}}_h)(\underset{\sim}{X}_{kh} - \bar{\underset{\sim}{X}}_h)^T .$$

In this study, it is assumed that the population covariance matrix has a block diagonal structure, i.e. $\Sigma_{ijh} = 0$ for all h and $i \neq j$.

The above notations are used in entire this dissertation. The classical approach and the high-dimensional approach of discriminant analysis are described in the next two subsections.

2.2.1 The Classical Approach

Two classical discriminant methods are described in this section. The first is the minimum expected cost of misclassification **(**ECM**)** method and the second is Fisher's discriminant method.

2.2.1.1 The minimum ECM method

Let $f(\underset{\sim}{X}_l)$ be the density associated with classes or populations $l^{th}, \prod_l$, for $l = 1, 2, ..., g$**.** For the development of the general theory, it is unnecessary **(**except where specified**)** to assume multivariate normality for $\underset{\sim}{X}$ **.** Let

p_l = the prior probability of $\prod_l$, $l = 1, 2, ..., g$;

$c(h \mid l)$ = the cost of assigning an observation to π_h when in fact, it belongs to $\prod_l$, for $l, h = 1, 2, ..., g$, with $l = h$, $c(l \mid l) = 0$;

$P(h \mid l)$ = the conditional probability of assigning an observation to $\prod_h$ when in fact, it belongs to $\prod_l$.

$= \int_{R_h} f(\underset{\sim}{X}_l) d\underset{\sim}{x}$,where R_h is the set of $\underset{\sim}{x}$ values classified into $\prod_h$, $l, h = 1, 2, ..., g$, with $P(l \mid l) = 1 - \sum_{\substack{h=1 \\ l \neq h}}^{g} P(h \mid l)$.

Classification schemes are often assessed in terms of their misclassification probabilities and also misclassification cost, and so it is reasonable to use a classification rule which minimizes the ECM.

The conditional expected cost of misclassifying an $\underset{\sim}{x}$ from $\prod_1$ into $\prod_2$, or $\prod_3,\ldots$, up to $\prod_g$ is

$$ECM(1) = P(2|1)c(2|1) + P(3|1)c(3|1) + \cdots + P(g|1)c(g|1)$$
$$= \sum_{h=2}^{g} P(h|1)c(h|1)$$

This conditional expected cost occurs with prior probability p_1 (the probability of belonging to class 1).

The other conditional expected cost of misclassification, $ECM(2),...,$ $ECM(g)$ can be obtained in a similar manner. Multiplying each conditional ECM by its prior probability and summing them gives the overall ECM:

$$ECM = p_1 ECM(1) + p_2 ECM(2) + \cdots + p_g ECM(g)$$
$$= p_1 \sum_{h=2}^{g} P(h|1)c(h|1) + p_2 \sum_{\substack{h=1 \\ h\neq 2}}^{g} P(h|2)c(h|2) + \cdots + p_g \sum_{h=1}^{g-1} P(h|g)c(h|g)$$

$$ECM = \sum_{l=1}^{g} p_l \sum_{\substack{h=1 \\ l\neq h}}^{g} P(h|l)c(h|l) \qquad (2.13)$$

The classification regions that minimize (2.13) are defined by assigning $\underset{\sim}{x}$ to population $h,\ h = 1,2,...,g,$ for which

$$\sum_{\substack{l=1 \\ h\neq l}}^{g} p_l f_l(\underset{\sim}{x}) c(h|l) \qquad (2.14)$$

is the smallest. If a tie occurs in (2.14), $\underset{\sim}{x}$ can be assigned to any of the tied populations (for proof, see Anderson (1984)).

Suppose all the misclassification costs are equal, then without loss of generality, all the misclassification costs are set to 1. $\underset{\sim}{x}$ can be assigned to $\prod_h$, $h = 1, 2,..., g$, for which

$$\sum_{\substack{l=1 \\ l\neq h}}^{g} p_l f_l(\underset{\sim}{x}) \qquad (2.15)$$

is the smallest. Now, (2.15) will be the smallest when the omitted term $p_h f_h(\underset{\sim}{x})$ is the largest. Consequently, when all of the misclassification costs are equal, the minimum ECM rule has the following rather simple form.

For the minimum ECM rule with equal misclassification costs:

Assign $\underset{\sim}{x}$ to $\prod_h$ if $p_h f_h(\underset{\sim}{x}) > p_l f_l(\underset{\sim}{x})$ for all $l \neq h$.

or, equivalently,

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\ln p_h f_h(\underset{\sim}{x}) > \ln p_l f_l(\underset{\sim}{x})$ for all $l \neq h$. (2.16)

An essential case occurs when the population distribution is the multivariate normal distribution $\underset{\sim}{X}_h \sim N_p(\underset{\sim}{\mu}_h, \Sigma_h)$ with pdf

$$f(\underset{\sim}{X}_h) = (2\pi)^{-\frac{p}{2}} \left|\Sigma_h\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\underset{\sim}{X}_h - \underset{\sim}{\mu}_h)^T \Sigma_h^{-1} (\underset{\sim}{X}_h - \underset{\sim}{\mu}_h)\right), \quad h = 1, 2, ..., g,$$

and so from (2.16), we obtain

$$\ln p_h f_h(\underset{\sim}{x}) = \ln p_h - \left(\frac{p}{2}\right)\ln(2\pi) - \frac{1}{2}\ln\left|\Sigma_h\right| - \frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu}_h)^T \Sigma_h^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_h),$$

$$-2\ln p_h f_h(\underset{\sim}{x}) = -2\ln p_h + p\ln(2\pi) + \ln\left|\Sigma_h\right| + (\underset{\sim}{x} - \underset{\sim}{\mu}_h)^T \Sigma_h^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_h).$$

$\underset{\sim}{x}$ can be assigned to $\prod_h$, for which

$$\min_l \left\{-2\ln p_l f_l(\underset{\sim}{x})\right\}. \quad (2.17)$$

The constant $p\ln(2\pi)$ in (2.17) can be ignored, since it is the same for all populations. Next, define the quadratic discriminant score for the h^{th} population to be

$$D_h(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\mu}_h)^T \Sigma_h^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_h) + \ln\left|\Sigma_h\right| - 2\ln p_h, \quad h = 1, 2, ..., g.$$

Using the quadratic discriminant scores, classification rule (2.16) becomes

Assign $\underset{\sim}{x}$ to $\prod_h$ if $D_h(\underset{\sim}{x}) < D_l(\underset{\sim}{x})$ for all $l \neq h$. (2.18)

In practice, $\underset{\sim}{\mu}_h$ and $\sum_h$ are unknown and need to be estimated from the data. The most commonly used estimators are their unbiased estimates, thus the classification rule (2.18) becomes:

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h(\underset{\sim}{x}) < \hat{D}_l(\underset{\sim}{x})$ for all $l \neq h$,

where $\hat{D}_h(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\bar{x}}_h)^T S_h^{-1} (\underset{\sim}{x} - \underset{\sim}{\bar{x}}_h) + \ln\left|S_h\right| - 2\ln p_h$.

When all the covariance matrices are the same, that is to say $\Sigma_h = \Sigma$ for all h, the discriminant score can be simplified to

$$D_h(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\mu}_h)^T \Sigma^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_h) + \ln|\Sigma| - 2\ln p_h,$$

$$= \ln|\Sigma| + \underset{\sim}{x}^T \Sigma^{-1} \underset{\sim}{x} - 2\underset{\sim}{\mu}_h^T \Sigma^{-1} \underset{\sim}{x} + \underset{\sim}{\mu}_h^T \Sigma^{-1} \underset{\sim}{\mu}_h - 2\ln p_h .$$

The first two terms are the same for $D_1(\underset{\sim}{x}), D_2(\underset{\sim}{x}), \ldots, D_g(\underset{\sim}{x})$, and, consequently, they can be ignored for assigning purposes. The remaining terms consist of the constant $c_h = \underset{\sim}{\mu}_h^T \Sigma^{-1} \underset{\sim}{\mu}_h - 2\ln p_h$ and a linear combination of the components of $\underset{\sim}{x}$. $D_k(\underset{\sim}{x})$ is estimated by $\hat{D}_h(\underset{\sim}{x}) = -2\overline{\underset{\sim}{x}}_h^T S_{pooled}^{-1} \underset{\sim}{x} + \overline{\underset{\sim}{x}}_h^T S_{pooled}^{-1} \overline{\underset{\sim}{x}}_h^T - 2\ln p_h$. This discriminant score is called the linear discriminant score.

2.2.1.2 Fisher's Discriminant Method

The motivation behind the Fisher (1936) method is the need to obtain a reasonable representation of the populations that involves only a few linear combinations of the observations, such as $\underset{\sim}{c}_1^T \underset{\sim}{x}$, $\underset{\sim}{c}_2^T \underset{\sim}{x}$, and $\underset{\sim}{c}_3^T \underset{\sim}{x}$. This primary purpose of Fisher's discriminant method is to separate the populations, but it can also be used for classification purposes. It is not necessary to assume that the g populations are multivariate normal, but it is assumed that the $p \times p$ population covariance matrices are equal and of full rank $(\Sigma_1 = \Sigma_2 = \cdots \Sigma_g = \Sigma)$.

Let $\overline{\underset{\sim}{\mu}}$ denote the mean vector of the combined populations and B_μ be the sum of squares between the class so that

$$B_\mu = \sum_{h=1}^{g} (\underset{\sim}{\mu}_h - \overline{\underset{\sim}{\mu}})(\underset{\sim}{\mu}_h - \overline{\underset{\sim}{\mu}})^T, \text{ where } \overline{\underset{\sim}{\mu}} = \frac{1}{g} \sum_{h=1}^{g} \underset{\sim}{\mu}_h .$$

We consider the linear combination

$$Y_h = \underset{\sim}{c}^T \underset{\sim}{X}_h$$

which has an expected value

$$E(Y_h) = \underset{\sim}{c}^T E(\underset{\sim}{X}_h) = \underset{\sim}{c}^T \underset{\sim}{\mu}_h \text{ for class } h$$

and variance

$$Var(Y_h) = \underset{\sim}{c}^T Var(\underset{\sim}{X}_h) \underset{\sim}{c} = \underset{\sim}{c}^T \Sigma_h \underset{\sim}{c} = \underset{\sim}{c}^T \Sigma \underset{\sim}{c} \text{ for all of the class.}$$

Consequently, the expected value $\underset{\sim}{\mu}_{hY} = \underset{\sim}{c}^T \underset{\sim}{\mu}_h$ changes as the population from which $\underset{\sim}{X}_h$ is selected changes. We first define the overall mean as

$$\bar{\underset{\sim}{\mu}}_Y = \frac{1}{g}\sum_{h=1}^{g} \underset{\sim}{\mu}_{hY} = \frac{1}{g}\sum_{h=1}^{g} \underset{\sim}{c}^T \underset{\sim}{\mu}_h = \underset{\sim}{c}^T \left(\frac{1}{g}\sum_{h=1}^{g} \underset{\sim}{\mu}_h \right)$$
$$= \underset{\sim}{c}^T \bar{\underset{\sim}{\mu}}$$

and the ratio as

$$\frac{\sum_{h=1}^{g}\left(\underset{\sim}{\mu}_{hY} - \bar{\underset{\sim}{\mu}}_Y\right)^2}{\sigma_Y^2} = \frac{\sum_{h=1}^{g}\left(\underset{\sim}{c}^T \underset{\sim}{\mu}_h - \underset{\sim}{c}^T \bar{\underset{\sim}{\mu}}\right)^2}{\underset{\sim}{c}^T \Sigma \underset{\sim}{c}}$$
$$= \frac{\underset{\sim}{c}^T \left[\sum_{h=1}^{g}\left(\underset{\sim}{\mu}_h - \bar{\underset{\sim}{\mu}}\right)\left(\underset{\sim}{\mu}_h - \bar{\underset{\sim}{\mu}}\right)^T \right] \underset{\sim}{c}}{\underset{\sim}{c}^T \Sigma \underset{\sim}{c}},$$

or

$$\frac{\sum_{h=1}^{g}\left(\underset{\sim}{\mu}_{hY} - \bar{\underset{\sim}{\mu}}_Y\right)^2}{\sigma_Y^2} = \frac{\underset{\sim}{c}^T B_\mu \underset{\sim}{c}}{\underset{\sim}{c}^T \Sigma \underset{\sim}{c}}.$$

This ratio measures the variability between the groups of Y values relative to the common variability within the groups, and an appropriate $\underset{\sim}{c}$ can be selected to maximize this ratio.

Next, the sample sum of squares for class matrix S_b, which includes the sample sizes, is defined. Let

$$S_b = \sum_{h=1}^{g} n_h (\bar{\underset{\sim}{X}}_h - \bar{\underset{\sim}{X}})(\bar{\underset{\sim}{X}}_h - \bar{\underset{\sim}{X}})^T,$$

where $\bar{\underset{\sim}{X}} = \frac{\sum_{h=1}^{g} n_h \bar{\underset{\sim}{X}}_h}{\sum_{h=1}^{g} n_h} = \frac{\sum_{h=1}^{g}\sum_{k=1}^{n_h} \underset{\sim}{X}_{kh}}{\sum_{h=1}^{g} n_h}$.

In addition, an estimate of Σ is based on the sample sum of squares within class matrix S_w:

$$S_w = \sum_{h=1}^{g}\sum_{k=1}^{n_h} (\bar{\underset{\sim}{X}}_{kh} - \bar{\underset{\sim}{X}}_h)(\bar{\underset{\sim}{X}}_{kh} - \bar{\underset{\sim}{X}}_h)^T .$$

Consequently, $S_w/(n_1+n_2+\cdots+n_g-g)=S$ is the estimate of Σ. Before presenting the sample discriminants, note that S_w is constant $(n_1+n_2+\cdots+n_g-g)$ times S_{pooled}, so the same $\underset{\sim}{c}$ that maximizes $\frac{\underset{\sim}{c}^T S_b \underset{\sim}{c}}{\underset{\sim}{c}^T S_{pooled} \underset{\sim}{c}}$ also maximizes $\frac{\underset{\sim}{c}^T S_b \underset{\sim}{c}}{\underset{\sim}{c}^T S_w \underset{\sim}{c}}$. Moreover, an optimized $\underset{\sim}{c}$ can be presented in the more customary form of eigenvectors $\underset{\sim}{e}_h$ of $S_w^{-1}S_b$, because if $S_w^{-1}S_b\underset{\sim}{e}=\lambda\underset{\sim}{e}$, then $S^{-1}S_b\underset{\sim}{e}=\lambda(n_1+n_2+\cdots+n_g-g)\underset{\sim}{e}$.

Theorem 1. Let $\lambda_1,\lambda_2,...,\lambda_s>0$ denote the $s\le\min(g-1,p)$ eigenvalues of $S_w^{-1}S_b$ and $\underset{\sim}{e}_1,\underset{\sim}{e}_2,...,\underset{\sim}{e}_s$ be the corresponding eigenvectors, then the vector of coefficients $\underset{\sim}{c}$ that maximizes Fisher's criterion function $F(\underset{\sim}{c})$ is obtained as

$$F(\underset{\sim}{c})=\frac{\underset{\sim}{c}^T S_b \underset{\sim}{c}}{\underset{\sim}{c}^T S_w \underset{\sim}{c}}=\frac{\underset{\sim}{c}^T \sum_{h=1}^{g} n_h(\underset{\sim}{\bar{x}}_h-\underset{\sim}{\bar{x}})(\underset{\sim}{\bar{x}}_h-\underset{\sim}{\bar{x}})^T \underset{\sim}{c}}{\underset{\sim}{c}^T \sum_{h=1}^{g}\sum_{k=1}^{n_h}(\underset{\sim}{x}_{kh}-\underset{\sim}{\bar{x}}_h)(\underset{\sim}{x}_{kh}-\underset{\sim}{\bar{x}}_h)^T \underset{\sim}{c}}. \tag{2.19}$$

Given that $\underset{\sim}{c}_1=\underset{\sim}{e}_1$, the linear combination $\underset{\sim}{c}_1^T\underset{\sim}{x}$ is called the sample first discriminant, and the choice $\underset{\sim}{c}_2=\underset{\sim}{e}_2$ produces the sample second discriminant $\underset{\sim}{c}_2^T\underset{\sim}{x}$, and so on until we obtain $\underset{\sim}{c}_r^T\underset{\sim}{x}=\underset{\sim}{e}_r^T\underset{\sim}{x}$, where the sample r^{th} discriminant $r\le s$.

Proof. See Appendix A.2 □

Using Fisher's Discriminants to classify observations, we set $Y_{rh}=\underset{\sim}{c}_r^T\underset{\sim}{X}_h=r^{th}$ population discriminant $(r\le s)$, then

$$\underset{\sim}{Y}_h=\begin{bmatrix}Y_{1h}\\Y_{2h}\\\vdots\\Y_{sh}\end{bmatrix} \text{ has a mean vector } \underset{\sim}{\mu}_{hY}=\begin{bmatrix}\mu_{hY_1}\\\mu_{hY_2}\\\vdots\\\mu_{hY_s}\end{bmatrix}=\begin{bmatrix}\underset{\sim}{c}_1^T\underset{\sim}{\mu}_h\\\underset{\sim}{c}_2^T\underset{\sim}{\mu}_h\\\vdots\\\underset{\sim}{c}_s^T\underset{\sim}{\mu}_h\end{bmatrix} \text{ under } \prod_h \quad (h=1,2,...,g)$$

and covariance matrix I_s, for all populations. Because the components of Y have unit variances and zero covariance, the appropriate measure of squared distance from $\underset{\sim}{Y}=\underset{\sim}{y}$ to $\underset{\sim}{\mu}_{hY}$ is

$$(\underset{\sim}{y}-\underset{\sim}{\mu}_{hY})^T(\underset{\sim}{y}-\underset{\sim}{\mu}_{hY})=\sum_{k=1}^{s}(y_k-\mu_{hY_k})^2 \ .$$

A reasonable classification rule is one that assigns $\underset{\sim}{y}$ to $\prod_h$ if the square of the distance from $\underset{\sim}{y}$ to $\underset{\sim}{\mu}_{hY}$ is smaller than the square of the distance from $\underset{\sim}{y}$ to $\underset{\sim}{\mu}_{hY}$ for all $l \neq h$.

If only r of the discriminants are used for allocation, the classification rule based on Fisher's criterion function is:

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\sum_{k=1}^{r}\left[y_k-\bar{y}_{hk})\right]^2=\sum_{k=1}^{r}\left[\underset{\sim}{c}_k^T(\underset{\sim}{x}-\underset{\sim}{\bar{x}}_h)\right]^2 \leq \sum_{k=1}^{r}\left[\underset{\sim}{c}_k^T(\underset{\sim}{x}-\underset{\sim}{\bar{x}}_h)\right]^2$ for all $l \neq h$.

When the prior probabilities are all equal $\left(p_1=p_2=...=p_g=\frac{1}{g}\right)$ and $r=s$, the classification rule based on Fisher' s criterion function is equivalent to the minimum ECM classification rule with equal misclassification costs for normal populations with equal Σ_h.

2.2.2 The High-Dimensional Approach

Recently, a lot of discriminant methods for high-dimensional data have appeared in the literature. Since they form part of the core body of work in the proposed method, they are reviewed here.

Regularization techniques are highly successful for the inversing matrices problem (O'Sullivan, 1986). Because the inverse of the sample covriance matrix does not exist in high-dimensional data, O' Sullivan (1986) attempted to use the sample covariance bias to solve this problem.

The choice between the individual class sample covariance matrices and the sample pooled covariance matrix represents a set of regularization alternatives represented by

$$\hat{S}_h(\lambda)=\frac{S_h(\lambda)}{n_h(\lambda)}, \tag{2.20}$$

where $S_h(\lambda)=(1-\lambda)(n_h-1)S_h+\lambda(n-1)S_{pooled}$ and $n_h(\lambda)=(1-\lambda)(n_h-1)+\lambda(n-1)$.

The regularization parameter λ when taking on values $0 \le \lambda \le 1$ is used to control the degree of shrinkage of the individual class covariance matrix estimates toward the sample pooled covariance matrix**.** The value $\lambda = 0$ gives $\hat{S}_h(\lambda) = S_h$, whereas the value $\lambda = 1$ gives $\hat{S}_h(\lambda) = S_{pooled}$.

The regularization parameter provided by (2.20) is still fairly limited when regularizing. For example, if the population class covariance matrices were all quite different from each other, then shrinkage toward the sample pooled covariance matrix would introduce severe bias (Friedman, 1989). Friedman suggested that shrinking should be carried out toward the identity matrix by multiplying by $tr(S_h)/p$, which has almost no bias, as

$$\hat{S}_h(\lambda,\gamma) = (1-\gamma)\hat{S}_h(\lambda) + \frac{\gamma}{p}[tr(S_h)]I_p, \qquad (2.21)$$

where I_p is the $p \times p$ identity matrix.

For a given value λ, the additional regularization parameter γ $(\gamma \in [0,1])$ controls shrinkage toward a multiple of the identity matrix where the multiplier is just the average eigenvalue of $\hat{S}_h(\lambda)$**.** Hence, this shrinkage has the effect of decreasing the larger eigenvalues and increasing the smaller ones**.** Equations (2.20) and (2.21) represent a two-parameter family of regularized sample class covariance matrix estimators, and the discriminant score is

$$\hat{D}_h^r(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h)^T \hat{S}_h^{-1}(\lambda,\gamma)(\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h) + \ln\left|\hat{S}_h(\lambda,\gamma)\right| - 2\ln p_h .$$

Friedman (1989) gives the classification rule with equal misclassification costs based on the regularized discriminant method as:

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h^r(\underset{\sim}{x}) < \hat{D}_l^r(\underset{\sim}{x})$ for all $l \ne h$.

The additional regularization parameter γ can substantially improve the misclassification error when the population class covariance matrices are not equal or the sample size is too small (Friedman, 1989) The main disadvantage associated with the regularized discriminant method is that the determination of the optimal choice of the regularization parameter is determined by cross-validation, which consumes much time for high-dimensional data (Guo et al., 2007).

Another regularization technique was considered by Ledoit and Wolf (2004). Their goal was to find the linear combination $\Sigma^* = \rho_1 I + \rho_2 S$ of the identity matrix and the sample covariance matrix where the expected quadratic loss $E\left(\left\|\Sigma^* - \Sigma\right\|^2\right)$ is at a minimum, where $\|A\| = \sqrt{tr(AA^T)/p}$.

Consider the optimization problem

$$\min_{\rho_1, \rho_2} E\left(\left\|\Sigma^* - \Sigma\right\|^2\right) \text{ subject to } \Sigma^* = \rho_1 I + \rho_2 S ,$$

where the coefficients ρ_1 and ρ_2 are nonrandom**.** Its solution verifies that

$$\rho_1 = \frac{\beta^2}{\delta^2}\mu \text{ and } \rho_2 = \frac{\alpha^2}{\delta^2}$$

then $\Sigma^* = \frac{\beta^2}{\delta^2}\mu I + \frac{\alpha^2}{\delta^2}S$ and $E\left(\left\|\Sigma^* - \Sigma\right\|^2\right) = \frac{\alpha^2\beta^2}{\delta^2}$,

where $\mu = tr(\Sigma I)/p$, $\alpha^2 = \left\|\Sigma - \mu I\right\|^2$, $\beta^2 = E\left(\left\|S - \Sigma\right\|^2\right)$, and $\delta^2 = E\left(\left\|S - \mu I\right\|^2\right)$.

Since Σ^* depends on the four scalar functions of the true (unobservable) covariance matrix Σ: $\mu, \alpha^2, \beta^2, \delta^2$, they addressed this problem by replacing these functions $\delta, \mu, \alpha, \beta$ with their consistent estimators d, m, a, b where

$d^2 = \left\|S - mI\right\|^2$, $m = tr(SI)/p$, $a^2 = d^2 - b^2$, $\bar{b}^2 = \frac{1}{n^2}\sum_{k=1}^{n}\left\|\underset{\sim}{X}_k \underset{\sim}{X}_k^T - S\right\|^2$, and

$b^2 = \min(\bar{b}^2, d^2)$.

This yields a well-conditioned estimator of covariance matrix

$$\hat{\Sigma}^* = \frac{b^2}{d^2} mI + \frac{a^2}{d^2} S .$$

Ledoit and Wolf (2004) showed that $\hat{\Sigma}^*$ is a consistent estimator of Σ^*, i.e. $\left\|\hat{\Sigma}^* - \Sigma^*\right\| \to 0$. As a consequence, $\hat{\Sigma}^*$ has the same asymptotic expected loss (or risk) as Σ^*, i.e. $E\left(\left\|\hat{\Sigma}^* - \Sigma\right\|^2\right) - E\left(\left\|\Sigma^* - \Sigma\right\|^2\right) \to 0$**.**

Instead of considering the linear combination $\Sigma^* = \rho_1 I + \rho_2 S$, Schäfer and Strimmer (2005) used $S^* = \lambda T + (1-\lambda)S$, which guarantees a minimum mean squared error. They showed that it performed very well in both simulations and with real-life data.

Consider the optimization problem

$$\min_{\lambda} E\left(\left\|S^* - \Sigma\right\|^2\right) \text{ subject to } \left\|S^* - \Sigma\right\|^2 = \left\|\lambda T + (1-\lambda)S - \Sigma\right\|^2,$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p}(\lambda t_{ij} + (1-\lambda)s_{ij} - \sigma_{ij})^2,$$

where S is the sample covariance matrix $(S = [s_{ij}]_{p\times p})$ and T is the shrinkage target for the covariance matrix $(T = [t_{ij}]_{p\times p})$, then its solution verifies

$$\lambda^* = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p}\left[Var(s_{ij}) - C\,\text{ov}(s_{ij}, t_{ij})\right]}{\sum_{i=1}^{p}\sum_{j=1}^{p}E\left[(t_{ij} - s_{ij})^2\right]}. \tag{2.22}$$

For the practical application of (2.22), Schäfer and Strimmer (2005) decided to estimate the optimal shrinkage intensity λ^*. They suggested computing the optimal shrinkage intensity estimator $\hat{\lambda}^*$ by replacing all expectations, variances, and covariance in (2.22) with their unbiased estimates**.** Three commonly used shrinkage targets for the covariance matrix are compiled and the resulting estimate $\hat{\lambda}^*$ is made as follows:

1) T = A: "Diagonal unit variance"; in this case, we do not need an estimated parameter because T is a constant matrix (an identity matrix). Thus,

$$t_{ij} = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j \end{cases} \text{ and } \hat{\lambda}^* = \frac{\sum_{i\neq j} Var(s_{ij}) + \sum_{i=1}^{p} Var(s_{ii})}{\sum_{i\neq j} s_{ij}^2 + \sum_{i=1}^{p}(s_{ii} - 1)^2}. \tag{2.23}$$

2) T = B: "Diagonal common variance"; in this case, we need to estimate the diagonal element of T (the common variance α). Thus,

$$t_{ij} = \begin{cases} \alpha = \left(\sum_{i=1}^{p} s_{ii}\right)\Big/ p & if\ i = j \\ 0 & if\ i \neq j \end{cases} \text{ and } \hat{\lambda}^* = \frac{\sum_{i\neq j} Var(s_{ij}) + \sum_{i=1}^{p} Var(s_{ii})}{\sum_{i\neq j} s_{ij}^2 + \sum_{i=1}^{p}(s_{ii} - \alpha)^2}. \tag{2.24}$$

3) T = C: "Diagonal common variance and common covariance"; because this shrinkage target matrix is provided by the two parameters' common variance α and common covariance β, we need to estimate both parameters. Thus,

$$t_{ij} = \begin{cases} \alpha = \left(\sum_{i=1}^{p} s_{ii} \right) \Big/ p & \text{if } i = j \\ \beta = \left(\sum_{i \neq j} s_{ij} \right) \Big/ p(p-1) & \text{if } i \neq j \end{cases} \quad \text{and} \quad \hat{\lambda}^* = \frac{\sum_{i \neq j} Var(s_{y,ij}) + \sum_{i=1}^{p} Var(s_{ii})}{\sum_{i \neq j} (s_{y,ij} - \beta)^2 + \sum_{i=1}^{p} (s_{ii} - \alpha)^2}. \qquad (2.25)$$

When $\hat{\lambda}^*$ is computed, the well-conditioned estimator of the covariance matrix is calculated by

$$\hat{S}_h^* = \hat{\lambda}^* T + (1 - \hat{\lambda}^*) S_h .$$

The classification rule with equal misclassification costs based on a well-conditioned estimator of covariance matrix is:

Assign $\underset{\sim}{x}$ to Π_h if $\hat{D}_h^w(\underset{\sim}{x}) < \hat{D}_l^w(\underset{\sim}{x})$ for all $h \neq l$,

where $\hat{D}_h^w(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\bar{x}}_h)^T \hat{S}_h^{*-1} (\underset{\sim}{x} - \underset{\sim}{\bar{x}}_h) + \ln\left|S_h^*\right| - 2\ln p_h$.

Xu, Brock, and Parrish (2009) stated that a well-conditioned estimator has a simple explicit formula that is easy to compute and interpret**.** Unlike the regularized discriminant method, a well-conditioned estimator not only solves the singularity problem but also produces a unique optimal solution to the shrinkage parameter without the need to search for the optimal regularization parameter**.**

Dudoit, Fridlyand, and Speed (2002) introduced simplified discriminant rules by assuming independence between variables and replacing all off-diagonal elements of the sample covariance matrix with zero. Specifically, they used only the diagonal elements as $S_{d,h} = diag(s_{11h}, ..., s_{pph})$ and created the DI classification rule as:

Assign $\underset{\sim}{x}$ to Π_h if $\hat{D}_h^d(\underset{\sim}{x}) < \hat{D}_l^d(\underset{\sim}{x})$ for all $h \neq l$,

where $\hat{D}_h^d(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\bar{x}}_h)^T S_{d,h}^{-1} (\underset{\sim}{x} - \underset{\sim}{\bar{x}}_h) + \ln\left|S_{d,h}\right| - 2\ln p_h$.

As $S_{d,k}$ above uses only the diagonal elements of S_h, this method will lose some information from the off-diagonal elements.

Srivastava and Kubokawa (2007) derived the empirical Bayes estimator of Σ^{-1} given by $S_{SK,h}^{-1} = \left(S_h + \frac{tr(S_h)}{\min(n, p)} I \right)^{-1}$ and gave the SK classification rule as:

Assign $\underset{\sim}{x}$ to Π_h if $\hat{D}_h^s(\underset{\sim}{x}) < \hat{D}_l^s(\underset{\sim}{x})$ for all $h \neq l$, where $\hat{D}_h^s(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h)^T S_{SK,h}^{-1} (\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h) + \ln\left|S_{SK,h}\right| - 2\ln p_h$.

Note that S_{SK}^{-1} exists irrespective of whether $n < p$ or $n > p$, and this method performed the best in their study.

The next part of the literature review concentrates on the Fisher’ s Discriminant Method. Liu, Cheng and Yang (1993) proposed a modified Fisher’ s criterion function as

$$F^*(\underset{\sim}{c}) = \frac{\underset{\sim}{c}^T S_b \underset{\sim}{c}}{\underset{\sim}{c}^T S_t \underset{\sim}{c}} = \frac{\underset{\sim}{c}^T S_b \underset{\sim}{c}}{\underset{\sim}{c}^T S_b \underset{\sim}{c} + \underset{\sim}{c}^T S_w \underset{\sim}{c}}\ , \tag{2.26}$$

where $S_t = \sum_{h=1}^{g}\sum_{k=1}^{n_h} (\underset{\sim}{x}_{kh} - \underset{\sim}{\overline{x}})(\underset{\sim}{x}_{kh} - \underset{\sim}{\overline{x}})^T$ is the sample total sum of squares. Let $S_t^{\perp} = \left\{\underset{\sim}{x} \mid S_t \underset{\sim}{x} = 0, x \in \mathbb{R}^p\right\}$ and $S_t^{\perp^C}$ be the complementary subspace of $S_t^{\perp}$, then the algorithm to calculate the vector of coefficients $\underset{\sim}{c}$ subject to max (2.25) is designed as follows:

1) Calculate the first vector of coefficients $\underset{\sim}{c}_1$

Let $S_t^{\perp^C} = span\left\{\underset{\sim}{\varphi}_1^{(1)}, \underset{\sim}{\varphi}_2^{(1)}, ..., \underset{\sim}{\varphi}_{r_1}^{(1)}\right\}$, where $\varphi_1^{(1)}, \varphi_1^{(1)}, ..., \varphi_{r_1}^{(1)}$ are orthogonal unit vectors.

Case1. $r_1 = p$.

Subsequently, $\underset{\sim}{c}_1$ is the unit eigenvector corresponding to the maximal eigenvalue of the matrix $S_t^{-1} S_b$.

Case2. $1 < r_1 < p$.

Let $P_1 = \left(\varphi_1^{(1)}, \varphi_2^{(1)}, ..., \varphi_{r_1}^{(1)}\right)$ and $Z^{(1)}$ be the eigenvector corresponding to the maximal eigenvalue of $\left(P_1^C S_t P_1\right)^{-1} \left(P_1^C S_b P_1\right)$, then $\underset{\sim}{c}_1$ is determined by the following formula:

$$\underset{\sim}{c}_1 = \frac{P_1 Z^{(1)}}{\left|P_1 Z^{(1)}\right|}\ .$$

2) Calculate the i^{th} vector of the $\underset{\sim}{c}_i$ coefficients

Let $S_t^{\perp} = span\left\{\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, ..., \underset{\sim}{\alpha}_{p-r_1}\right\}$, where $\alpha_1, \alpha_2, ..., \alpha_{p-r_1}$ are orthogonal unit vectors.

Suppose $V_i = span\left\{\underset{\sim}{c}_1, \underset{\sim}{c}_2, ..., \underset{\sim}{c}_{i-1}, \underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, ..., \underset{\sim}{\alpha}_{p-r_1}\right\}$ is the subspace spanned by the vector of coefficients $\underset{\sim}{c}_1, \underset{\sim}{c}_2, ..., \underset{\sim}{c}_{i-1}$, which have already been calculated, and the vectors $\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, ..., \underset{\sim}{\alpha}_{p-r_1}$ and $V_i^C = span\left\{\underset{\sim}{\varphi}_1^{(i)}, \underset{\sim}{\varphi}_2^{(i)}, ..., \underset{\sim}{\varphi}_{r_1-i+1}^{(i)}\right\}$ are the complementary subspace of V_i, where $\underset{\sim}{\varphi}_1^{(i)}, \underset{\sim}{\varphi}_2^{(i)}, ..., \underset{\sim}{\varphi}_{r_1-i+1}^{(i)}$ are the orthogonal unit vectors.

Let $P_i = \left(\underset{\sim}{\varphi}_1^{(i)}, \underset{\sim}{\varphi}_2^{(i)}, ..., \underset{\sim}{\varphi}_{r_1-i+1}^{(i)}\right)$ and $Z^{(i)}$ be the eigenvector corresponding to the maximal eigenvalue of $\left(P_i^C S_t P_i\right)^{-1}\left(P_i^C S_b P_i\right)$, then $\underset{\sim}{c}_i$ is determined by the following formula:

$$\underset{\sim}{c}_i = \frac{P_i Z^{(i)}}{\left|P_i Z^{(i)}\right|}.$$

Chen et al. (2000) proposed a more efficient, accurate, and stable method to derive the vector of coefficients $\underset{\sim}{c}$ that maximizes $F^*(\underset{\sim}{c})$. First decompose S_w as $S_w = H\Lambda H^T$, where $H = [\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2 ...., \underset{\sim}{\alpha}_r, \underset{\sim}{\alpha}_{r+1}, ..., \underset{\sim}{\alpha}_p]$ are the eigenvectors of S_w corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_p = 0,$ and $\Lambda = diag(\lambda_1, \lambda_2 ...., \lambda_r, \lambda_{r+1}, ..., \lambda_p).$ Second, compute $\tilde{S}_b = QQ^T S_b \left(QQ^T\right)^T$, where $Q = [\underset{\sim}{\alpha}_{r+1}, ..., \underset{\sim}{\alpha}_p],$ then the vector of coefficients $\underset{\sim}{c}$ are the eigenvectors of $\tilde{S}_b$ corresponding to the nonzero eigenvalues of $\tilde{S}_b$.

The classification rule based on modified Fisher's criterion function is:

Assign $\underset{\sim}{x}$ to Π_h if

$$\sum_{j=1}^{r}\left[y_j - \bar{y}_{hj})\right]^2 = \sum_{j=1}^{r}\left[\underset{\sim}{c}_j^T(\underset{\sim}{x} - \underset{\sim}{\bar{x}}_h)\right]^2 \leq \sum_{j=1}^{r}\left[\underset{\sim}{c}_j^T(\underset{\sim}{x} - \underset{\sim}{\bar{x}}_l)\right]^2 \text{ for all } h \neq l.$$

This method can be applied to high-dimensional data whereas the Fisher's discriminant method based on (2.19) cannot. Experimental results have shown that the method of Chen et al. (2000) is superior to that of Liu et al. (1993) in terms of recognition accuracy, training efficiency, and stability.

Lu et al. (2003) proposed a new regularized discriminant method called the regularized direct discriminant method by incorporating the dimension reduction technique of Chen et al. (2000) into the regularized discriminant method proposed by Friedman (1989). In order to reduce the dimensions, they formed matrix H containing a vector of coefficients $\underset{\sim}{c}$ by letting $U = (\underset{\sim}{u}_1, \underset{\sim}{u}_2, ..., \underset{\sim}{u}_q)$ be the q eigenvectors of S_b corresponding to the q nonzero eigenvalues denoted by $\omega_1, \omega_2, ..., \omega_q$. Thus, $H = UD_b^{-1/2}$ so as to obtain $H^T S_b H = I$, where $D_b = diag(\omega_1, \omega_2, ..., \omega_q)$ and I is the $q \times q$ identity matrix. They defined the classification rule with equal misclassification costs based on regularized direct discriminant method as:

$$\text{Assign } \underset{\sim}{y} = H^T \underset{\sim}{x} \text{ to } \textstyle\prod_h \text{ if } \hat{D}_h^{\omega}(\underset{\sim}{x}) < \hat{D}_l^{\omega}(\underset{\sim}{x}) \text{ for all } h \neq l,$$

where $\hat{D}_h^{\omega}(\underset{\sim}{x}) = (\underset{\sim}{y} - \underset{\sim}{\bar{y}}_h)^T \hat{S}_h^{-1}(\lambda, \gamma)(\underset{\sim}{y} - \underset{\sim}{\bar{y}}_h) + \ln\left|\hat{S}_h(\lambda, \gamma)\right| - 2\ln p_h$,

in which $\hat{S}_h(\lambda, \gamma) = (1-\gamma)\hat{S}_h(\lambda) + \dfrac{\gamma}{p}[tr(S_h)]I_p$, $\hat{S}_h(\lambda) = \dfrac{(1-\lambda)n_h S_h + \lambda n S}{(1-\lambda)n_h + \lambda n}$,

$$S_h = \frac{1}{n_h}\sum_{j=1}^{n_h}(\underset{\sim}{y}_{jh} - \underset{\sim}{\bar{y}}_h)(\underset{\sim}{y}_{jh} - \underset{\sim}{\bar{y}}_h)^T \text{, and } S = \frac{1}{n}\sum_{k=1}^{g} S_k .$$

The regularized direct discriminant method can be used for high-dimensional data and can reduce the time consuming task of obtaining $\hat{S}_h(\lambda, \gamma)$.

CHAPTER 3

THE PROPOSED TEST

The proposed test for testing for a block diagonal covariance matrix for high-dimensional data is presented in Section 3.1, followed by the two new discriminant methods for high-dimensional data in section 3.2.

3.1 Testing for a Block Diagonal Covariance Matrix in High-Dimensional Data

In this section, the problem of testing hypothesis,

$$H_0 : \Sigma = D_\Sigma \text{ vs } H_a : \Sigma \neq D_\Sigma ,$$

for high-dimensional data is of interest. The proposed statistic based on the fact that if H_0 is true $(\Sigma = D_\Sigma)$, then $tr\Sigma^2 = trD_\Sigma^2$ or $tr\Sigma^2 / trD_\Sigma^2 = 1$. Thus, under H_0, an equivalent test is obtained as follows:

$$H_0 : \frac{tr\Sigma^2}{trD_\Sigma^2} = 1 \text{ vs } H_a : \frac{tr\Sigma^2}{trD_\Sigma^2} > 1. \tag{3.1}$$

Consider that

$$\begin{aligned} trD_\Sigma^2 &= tr\begin{pmatrix} \Sigma_{11} & O_{12} & \cdots & O_{1m} \\ O_{21} & \Sigma_{22} & \cdots & O_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ O_{m1} & O_{m2} & \cdots & \Sigma_{mm} \end{pmatrix}\begin{pmatrix} \Sigma_{11} & O_{12} & \cdots & O_{1m} \\ O_{21} & \Sigma_{22} & \cdots & O_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ O_{m1} & O_{m2} & \cdots & \Sigma_{mm} \end{pmatrix} \\ &= tr\begin{pmatrix} \Sigma_{11}^2 & O_{12} & \cdots & O_{1m} \\ O_{21} & \Sigma_{22}^2 & \cdots & O_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ O_{m1} & O_{m2} & \cdots & \Sigma_{mm}^2 \end{pmatrix} \\ &= \sum_{i=1}^{m} tr\Sigma_{ii}^2 \end{aligned}$$

Recall from Hyodo et al. (2015) that we defined $a_k = \frac{tr\Sigma^k}{p}$ and $a_{ki} = \frac{tr\Sigma_{ii}^k}{p}$, then we can rewrite (3.1) as $H_0 : T = 1$, where $T = \frac{a_2}{a_{2D}}$, in which $a_{kD} = \sum_{i=1}^{m} a_{ki}$.

The quantity T can be estimated by the unbiased and consistency estimator for high-dimensional data proposed by Srivastava (2005), who made the following assumptions in order to estimate a_2, a_{2i}:

(C1) $p_i, i = 1, 2, ...$ is fixed and $m \to \infty$.

(C2) $N = O(p^{\delta}),\ 0 < \delta \leq 1$.

(C3) $0 < a_k^0 = \lim_{p\to\infty} a_k < \infty,\ \lim_{p\to\infty} \frac{a_4}{p} \to 0,\ k = 1, 2$.

(C4) $0 < \lim_{p\to\infty} \frac{p_i}{p} < c_i,\ i = 1, 2, ...$

Recall that the unbiased and consistent estimators of a_2 and a_{2D} obtained from the results of Srivastava (2005) are defined as follows:

$$\hat{a}_2 = \frac{N^2}{(N-1)(N+2)p}\left\{trS^2 - \frac{1}{N}(trS)^2\right\}, \tag{3.2}$$

$$\hat{a}_{2i} = \frac{N^2}{(N-1)(N+2)p}\left\{trS_{ii}^2 - \frac{1}{N}(trS_{ii})^2\right\}. \tag{3.3}$$

From these estimators, we can obtain the estimator of a_{2D} from

$$\hat{a}_{2D} = \sum_{i=1}^{m} \hat{a}_{2i} = \sum_{i=1}^{m} \frac{N^2}{(N-1)(N+2)p}\left\{trS_{ii}^2 - \frac{1}{N}(trS_{ii})^2\right\}. \tag{3.4}$$

Hence, these estimators can be used to estimate T as $\hat{T} = \frac{\hat{a}_2}{\hat{a}_{2D}}$.

The follow lemma gives the asymptotic distribution of $\left(\hat{a}_2 \quad \hat{a}_{2D}\right)^T$ used to derive the distribution of the test statistic $\hat{T}$.

Lemma1 Under assumptions (C1) – (C4), the asymptotic distribution of $(\hat{a}_2 \quad \hat{a}_{2D})^T$ is

$$\begin{pmatrix} \hat{a}_2 \\ \hat{a}_{2D} \end{pmatrix} \xrightarrow{D} N_2\left(\begin{pmatrix} a_2 \\ a_{2D} \end{pmatrix}, \begin{pmatrix} \frac{8}{Np}a_4 + \frac{4}{N^2}a_2^2 & Cov(\hat{a}_2, \hat{a}_{2D}) \\ Cov(\hat{a}_2, \hat{a}_{2D}) & \sum_{i=1}^{m}(\frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2) \end{pmatrix} \right).$$

Note that it is not necessary to find $Cov(\hat{a}_2, \hat{a}_{2D})$ in order to derive the distribution of $\hat{T}$ under H_0.

Proof. See Appendix A.1. □

In order to test $H_0 : T = 1$ against $H_a : T > 1$, it is necessary to find the distribution of $\hat{T}$. Since the test statistic $\hat{T}$ is a function of random variables, then we can use Lemma 1 incorporating the Delta method to obtain the distribution of $\hat{T}$.

Lemma 2 (the Delta method) Suppose $\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n$ are random vectors in the $\mathbb{R}^q$ Euclidean space and assume that $c_n(\underset{\sim}{x}_n - \underset{\sim}{\mu}) \sim N_q(\underset{\sim}{0}, \Sigma)$, where $\underset{\sim}{\mu}$ is a constant vector and $\{c_n\}$ is a sequence of constants $c_n \to \infty$. In addition, it is assumed that $g(\bullet)$ is a function from $\mathbb{R}^q$ to $\mathbb{R}$ which is differentiable at $\underset{\sim}{\mu}$ with a gradient (the vector of first partial derivatives) of $1 \times k$ dimensions at $\underset{\sim}{\mu}$ equal to $g'(\underset{\sim}{\mu})$, then

$$c_n\left[g(\underset{\sim}{x}_n) - g(\underset{\sim}{\mu})\right] \sim N(\underset{\sim}{0}, g'(\underset{\sim}{\mu})\Sigma g'(\underset{\sim}{\mu})^T).$$

Proof. See Lehmann and Romano (2006, p. 436). □

The next theorem shows that the distribution of test statistic $\hat{T}$ is normal. We also find the distribution of test statistic $\hat{T}$ under H_0, which is stated in the corollary after the next theorem.

Theorem 2 Under assumptions (C1) – (C4), let $T = \frac{a_2}{a_{2D}}$ and $\hat{T} = \frac{\hat{a}_2}{\hat{a}_{2D}}$, then the asymptotic distribution of $\hat{T}$ is

$$\hat{T} - T \xrightarrow{D} N(0, \theta^2),$$

where $\theta^2 = \frac{1}{a_{2D}^2}\left(\frac{8}{Np}a_4 + \frac{4}{N^2}a_2^2\right) - \frac{2a_2}{a_{2D}^3}Cov(\hat{a}_2, \hat{a}_{2D}) + \left(\frac{a_2}{a_{2D}^2}\right)^2 \sum_{i=1}^{m}\left(\frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2\right)$.

Proof. Let $T = g(a_2, a_{2D}) = \frac{a_2}{a_{2D}}$, then $\hat{T} = g(\hat{a}_2, \hat{a}_{2D}) = \frac{\hat{a}_2}{\hat{a}_{2D}}$.

The first partial derivatives of $g(a_2, a_{2D})$ with respect to a_2 and a_{2D} are, respectively,

$$\frac{\partial g(a_2, a_{2D})}{\partial a_2} = \frac{1}{a_{2D}} \text{ and } \frac{\partial g(a_2, a_{2D})}{\partial a_{2D}} = -\frac{a_2}{a_{2D}^2}.$$

By applying the Delta method, we obtain $\hat{T} - T \xrightarrow{D} N(0, \theta^2)$, where

$$\theta^2 = \begin{pmatrix} \frac{1}{a_{2D}} & -\frac{a_2}{a_{2D}^2} \end{pmatrix} \begin{pmatrix} \frac{8}{Np}a_4 + \frac{4}{N^2}a_2^2 & Cov(\hat{a}_2, \hat{a}_{2D}) \\ Cov(\hat{a}_2, \hat{a}_{2D}) & \sum_{i=1}^{m}(\frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2) \end{pmatrix} \begin{pmatrix} \frac{1}{a_{2D}} \\ -\frac{a_2}{a_{2D}^2} \end{pmatrix}$$

$$= \frac{1}{a_{2D}^2}\left(\frac{8}{Np}a_4 + \frac{4}{N^2}a_2^2\right) - \frac{2a_2}{a_{2D}^3}Cov(\hat{a}_2, \hat{a}_{2D}) + \left(\frac{a_2}{a_{2D}^2}\right)^2 \sum_{i=1}^{m}\left(\frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2\right). \quad \square$$

Corollary 1 Under the null hypothesis $H_0 : T = 1$, we obtain

$$T_p = \frac{\hat{T} - 1}{\theta} \xrightarrow{D} N(0,1),$$

where $\theta^2 = \frac{4}{a_{2D}^2 N^2}\sum_{i \neq j}^{m} a_{2i}a_{2j}$.

Proof. When H_0 is true, $a_2 = a_{2D}$,

$$\theta^2 = \frac{1}{a_{2D}^2}\left(\frac{8}{Np}a_{4D} + \frac{4}{N^2}a_{2D}^2\right) - \frac{2}{a_{2D}^2}Cov(\hat{a}_{2D}, \hat{a}_{2D}) + \frac{1}{a_{2D}^2}\sum_{i=1}^{m}\left(\frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2\right),$$

$$= \frac{1}{a_{2D}^2}\left(\frac{8}{Np}a_{4D} + \frac{4}{N^2}a_{2D}^2\right) - \frac{2}{a_{2D}^2}\sum_{i=1}^{m}\left(\frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2\right) + \frac{1}{a_{2D}^2}\sum_{i=1}^{m}\left(\frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2\right),$$

$$= \frac{1}{a_{2D}^2}\left(\frac{8}{Np}a_{4D} + \frac{4}{N^2}a_{2D}^2\right) - \frac{1}{a_{2D}^2}\sum_{i=1}^{m}\left(\frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2\right),$$

$$= \frac{4}{a_{2D}^2 N^2}\left(a_{2D}^2 - \sum_{i=1}^{m}a_{2i}^2\right),$$

$$= \frac{4}{a_{2D}^2 N^2}\left((\sum_{i=1}^{m}a_{2i})^2 - \sum_{i=1}^{m}a_{2i}^2\right),$$

$$= \frac{4}{a_{2D}^2 N^2}\left((\sum_{i=1}^{m}a_{2i})^2 - \sum_{i=1}^{m}a_{2i}^2\right),$$

$$= \frac{4\sum_{i\neq j}^{m}a_{2i}a_{2j}}{a_{2D}^2 N^2}.$$

Therefore, $\theta^2 = \frac{4}{a_{2D}^2 N^2}\sum_{i\neq j}^{m}a_{2i}a_{2j}$. □

Apparently, to use T_p in practice, it is necessary to estimate θ^2 by replacing a_{2i} and a_{2D} by $\hat{a}_{2i}$ and $\hat{a}_{2D}$ respectively, i.e. $\hat{\theta}^2 = \frac{4}{\hat{a}_{2D}^2 N^2}\sum_{i\neq j}^{m}\hat{a}_{2i}\hat{a}_{2j}$.

Thus, a test of H_0 can be constructed as the following statistic:

$$\hat{T}_p = \frac{\hat{T} - 1}{\hat{\theta}} \xrightarrow{D} N(0,1),$$

where $\hat{\theta}^2 = \frac{4}{\hat{a}_{2D}^2 N^2}\sum_{i\neq j}^{m}\hat{a}_{2i}\hat{a}_{2j}$.

Testing the hypothesis $H_0 : T = 1$ against $H_a : T > 1$ is a one-tailed test. At significance level α, H_0 is rejected if $\hat{T}_p > z_\alpha$, where z_α is the $100\alpha^{th}$ percentile of a standard normal distribution.

3.2 Discriminant Analysis in High-Dimensional Data

In this section, two new discriminant methods are proposed to construct a method for dealing with high-dimensional data. Only the classification of two classes where the population covariance of each class is equal $(\Sigma_1 = \Sigma_2 = \Sigma)$ with a block diagonal structure are considered. In this situation, high-dimensional data can occur when the degrees of freedom exceeds the dimensions instead of the sample size.

3.2.1 The First Proposed Method

A well-conditioned estimator of the covariance matrix (Schäfer & Strimmer, 2005) and the regularized direct discriminant method (Lu et al., 2003) are considered to be extremely beneficial in discriminant analysis**.** One advantage of the regularized direct discriminant method is that it can reduce the dimensionality of the data, and so in this study, these advantages are combined in a new technique by incorporating a well-conditioned estimator of the covariance matrix with the regularized direct discriminant method to classify high-dimensional data**.** The definition of a well-conditioned estimator of the covariance matrix is the linear combination $S^* = \lambda T + (1-\lambda)S$ of the shrinkage target matrix and the sample covariance matrix where the expected quadratic loss $E\left(\left\|S^* - \Sigma\right\|^2\right)$ is at a minimum which defined by Schäfer and Strimmer (2005). Additionally, it is always positive definite. The proposed technique uses a well-conditioned estimator of the covariance matrix instead of $\hat{S}_h^{-1}(\lambda,\gamma)$ in the regularized direct discriminant method in order to avoid searching for an optimal regularization parameter resulting in a unique optimal solution. First, Lu et al.'s (2003) technique to reduce the dimensions is carried out.

Step 1 Diagonalize S_b: Find matrix V such that

$$V^T S_b V = \Lambda ,$$

where V is a matrix of the eigenvectors of S_b $(S_b = \sum_{h=1}^{g} n_h(\underset{\sim}{\bar{x}}_h - \underset{\sim}{\bar{x}})(\underset{\sim}{\bar{x}}_h - \underset{\sim}{\bar{x}})^T)$ and Λ is a diagonal matrix which contains all the eigenvalues corresponding with V. As S_b

might be singular, some of the eigenvalues will be 0, and it is necessary to discard these along with any eigenvectors that contain them.

Let $\omega_1, \omega_2, ..., \omega_q$ be nonzero eigenvalues of S_b and U be the first q eigenvectors of S_b corresponding to nonzero eigenvalues. After that, we write

$$U^T S_b U = D_b ,$$

where $D_b = diag(\omega_1, \omega_2, ..., \omega_q)$ is the $q \times q$ submatrix of Λ **.**

Step 2 Unitize S_b: Let $H = UD_b^{-\frac{1}{2}}$, then

$$\left(UD_b^{-\frac{1}{2}}\right)^T S_b \left(UD_b^{-\frac{1}{2}}\right) = I \Rightarrow H^T S_b H = I \ .$$

Thus, H unitizes S_b and reduces the dimensionality from p to q .

Step 3 Create a well-conditioned covariance matrix estimator S^* in the low dimension subspace spanned by H by projecting the original observations into it to obtain $\underset{\sim}{y}_{kh} = H^T \underset{\sim}{x}_{kh}$, where $h = 1, 2$, $k = 1, 2, ..., n_h$, then consider the optimization problem

$$\min_{\lambda} E\left(\left\|S^* - \Sigma\right\|^2\right) \text{ subject to } \left\|S^* - \Sigma\right\|^2 = \left\|\lambda T + (1-\lambda) S_{y,pooled} - \Sigma\right\|^2$$

$$= \sum_{i=1}^{q} \sum_{j=1}^{q} (\lambda t_{ij} + (1-\lambda) s_{y,ij} - \sigma_{ij})^2 ,$$

where $S^* = \lambda T + (1-\lambda) S_{y,pooled}$ is a well-conditioned covariance matrix estimator,

$$S_{y,pooled} = \frac{\sum_{h=1}^{g} \sum_{k=1}^{n_h} (\underset{\sim}{y}_{kh} - \underset{\sim}{\bar{y}}_h)(\underset{\sim}{y}_{kh} - \underset{\sim}{\bar{y}}_h)^T}{n-g}$$ is the pooled variance of $\underset{\sim}{y}$,

$s_{y,ij}$ represents the element at the i^{th} row and j^{th} column of $S_{y,pooled}$, and

g is the number of classes.

In this study, three shrinkage target matrices for $S_{y,pooled}$ (Schäfer & Strimmer, 2005) and the resulting estimate $\hat{\lambda}^*$ (as show in Chapter 2 (2. 23-2. 25)) are investigated.

Step 4 Calculated the well-conditioned estimator of the covariance matrix by

$$\hat{S}^{*} = \hat{\lambda}^{*}T + (1-\hat{\lambda}^{*})S_{y,pooled} .$$

Step 5 Define the first proposed classification rule**:**

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h(\underset{\sim}{x}) < \hat{D}_l(\underset{\sim}{x})$ for all $l \neq h$,

where $\hat{D}_h(\underset{\sim}{x}) = (H^T\underset{\sim}{x} - H^T\underset{\sim}{\bar{x}}_h)^T \hat{S}^{*-1}(H^T\underset{\sim}{x} - H^T\underset{\sim}{\bar{x}}_h) - 2\ln p_h$.

From here on, symbols TA, TB, and TC are used for the classification rules for shrinkage targets matrices $T =$ A, B, and C, respectively.

3.2.2 The Second Proposed Method

Recall that the population covariance matrices are assumed to be block diagonal structures, thus the proposed method is based on constructing the sample covariance matrix to be the same pattern. For a pooled sample covariance matrix S_{pooled}, we partition it in blocks as

$$S_{pooled} = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1m} \\ S_{21} & S_{22} & \cdots & S_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{mm} \end{pmatrix} = [S_{ij}]_{p\times p} ,$$

where S_{ij} are submatrices of S_{pooled}, for $i, j = 1,2,...,m$, and the dimensions of S_{ij} are $p_i \times p_j$ and $\sum_{i=1}^{m} p_i = p$. S_{pooled} is partitioned in the same manner as Σ which the block size of S_{ij} is equal to the block size of Σ_{ij} for all $i, j,$ thus we define block diagonal matrix sample covariance matrix S_{block} as

$$S_{block} = diag(S_{11}, S_{22},..., S_{(m-1)(m-1)}, S_{mm}) = \begin{pmatrix} S_{11} & 0 & \cdots & \cdots & 0 \\ 0 & S_{22} & & & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & S_{(m-1)(m-1)} & 0 \\ 0 & \cdots & \cdots & 0 & S_{mm} \end{pmatrix}_{p\times p} .$$

When classifying two classes, the degrees of freedom for S_{pooled} is $n_1 + n_2 - 2$, and so $S_{ii}, i = 1,2,...,m$ are submatrices of p_i dimensions with ν degrees of freedom. If we specify that $p_i < \nu$, then $S_{ii}, i = 1, 2, ..., m$ are all invertible (Dempster, 1958). As a result, S_{block} is also invertible, and the inverse of S_{block} is given by

$$S_{block}^{-1} = diag(S_{11}^{-1}, S_{22}^{-1}, ..., S_{(m-1)(m-1)}^{-1}, S_{mm}^{-1}) = \begin{pmatrix} S_{11}^{-1} & 0 & \cdots & \cdots & 0 \\ 0 & S_{22}^{-1} & & & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & S_{(m-1)(m-1)}^{-1} & 0 \\ 0 & \cdots & \cdots & 0 & S_{mm}^{-1} \end{pmatrix}_{p \times p} .$$

Hence, S_{block}^{-1} is used instead of S_{pooled}^{-1} because the latter does not exist for high-dimensional data.

The second proposed classification rule is:

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h(\underset{\sim}{x}) < \hat{D}_l(\underset{\sim}{x})$ for all $l \neq h$,

where $\hat{D}_h(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h)^T S_{block}^{-1} (\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h) - 2 \ln p_h$.

From here on, symbol BD is used for the classification rule that uses a block diagonal sample covariance matrix.

CHAPTER 4

SIMULATION STUDY

In this section, the performance of the proposed test statistic $(\hat{T}_p)$ is evaluated using a simulation study under with various parameter setting and a comparison with some of the previously reported tests are reported. Furthermore, the proposed methods (TA, TB, TC, and BD) are evaluated using a simulation study with various parameter settings and compared with some previously reported methods.

4.1 Simulation Study for Testing Block Diagonal Covariance Matrices in High-Dimensional Data

To test the hypothesis $H_0 : \Sigma = D_\Sigma$ against $H_a : \Sigma \neq D_\Sigma$, the proposed test statistic $(\hat{T}_p)$ is investigated via a simulation study with 10,000 iterations under various parameter settings of population covariance matrix by considering its empirical Type I error rate and empirical power. A comparison of the performance of $\hat{T}_p$ with $\hat{T}_b$ (Hyodo, Shutoh, Nishiyama, & Pavlenko, 2015) and $\hat{T}_c$ (Bao et al., 2014) which described in section 2.1.2 is also carried out**.**

Recall the three test statistics which were compared in this study:

1) The proposed test statistic $\hat{T}_p$ is given as

$$\hat{T}_p = \frac{\dfrac{\hat{a}_2}{\hat{a}_{2D}} - 1}{\hat{\theta}} \xrightarrow{D} N(0,1),$$

where $\hat{\theta}^2 = \dfrac{4}{\hat{a}_{2D}^2 N^2} \sum_{i \neq j}^{m} \hat{a}_{2i} \hat{a}_{2j}$,

$$\hat{a}_2 = \frac{N^2}{(N-1)(N+2)p}\left\{trS^2 - \frac{1}{N}(trS)^2\right\},$$

$$\hat{a}_{2i} = \frac{N^2}{(N-1)(N+2)p}\left\{trS_{ii}^2 - \frac{1}{N}(trS_{ii})^2\right\}, \text{ and}$$

$$\hat{a}_{2D} = \sum_{i=1}^{m}\hat{a}_{2i} = \sum_{i=1}^{m}\frac{N^2}{(N-1)(N+2)p}\left\{trS_{ii}^2 - \frac{1}{N}(trS_{ii})^2\right\}.$$

2) The Hyodo, Shutoh, Nishiyama, and Pavlenko (2015)'s test statistic $\hat{T}_b$ is given as

$$\hat{T}_b = \hat{a}_2 - \hat{a}_{2D} \sim N(0, \frac{\hat{\lambda}^2}{N^2}),$$

where $\hat{\lambda}^2 = \dfrac{\sum_{j=2}^{m}\sum_{i=1}^{j-1} 8(N-1)(N+2)\hat{a}_{2j}\hat{a}_{2i}}{N^2}$ and $\hat{a}_2$, $\hat{a}_{2i}$, and $\hat{a}_{2D}$ are define above.

3) The Bao, Hu, Pan, and Zhou (2014)'s test statistic $\hat{T}_c$ is given as

$$\hat{T}_c = \frac{1}{2}trB^2 - \frac{p}{2},$$

where $B = diag\left[Y^{(i)}Y^{(i)T}\right]^{-1/2} \cdot \left[YY^T\right] \cdot diag\left[Y^{(i)}Y^{(i)T}\right]^{-1/2}_{i=1,\ldots,m}$,

$Y = \left(\underset{\sim}{X}_1 - \bar{\underset{\sim}{X}} \;\; \underset{\sim}{X}_2 - \bar{\underset{\sim}{X}} \;\cdots\; \underset{\sim}{X}_n - \bar{\underset{\sim}{X}}\right)$, $Y^{(i)} = \left(\underset{\sim}{X}_1^{(i)} - \bar{\underset{\sim}{X}}^{(i)} \;\; \underset{\sim}{X}_2^{(i)} - \bar{\underset{\sim}{X}}^{(i)} \;\cdots\; \underset{\sim}{X}_n^{(i)} - \bar{\underset{\sim}{X}}^{(i)}\right)$

with $$\hat{T}_c \sim N(a_n, b_n),$$

where $a_n = \dfrac{1}{2}\dfrac{\sum_{i\neq j} p_i p_j}{n-1}$, $b_n = \dfrac{1}{2}\dfrac{\sum_{i\neq j} p_i p_j (n-1-p_i)(n-1-p_j)}{(n-1)^4}$.

4.1.1 The Performance Evaluation Methods for Test Statistics

The empirical type I error rate (ε_1) and the empirical power (ε_2) are obtained by generating a sample of n independent observations from $N_p\left(\underset{\sim}{0}, \Sigma\right)$ and repeating 10000 times using either $\hat{T}_p$, $\hat{T}_b$, and $\hat{T}_c$ to calculate

$$\varepsilon_1 = \frac{(\#\,\text{test statistics under } \mathrm{H}_0 > z_\alpha)}{10000}$$

$$\text{and } \varepsilon_2 = \frac{(\#\,\text{test statistics under H}_a > z_\alpha)}{10000},$$

where z_α is the $100\alpha\%$ quantile of the standard normal distribution and α is the significance level (fixed at $\alpha = 0.05$ in this study).

4.1.2 Parameter Settings to Test for a Block Diagonal Covariance Matrix in High-Dimensional Data

The empirical type I error rate is calculated under the null hypothesis with four different forms of population covariance matrix as follows:

1) The first form of covariance matrix is $\Sigma_1 = diag(\Sigma_{11}, \Sigma_{22}, ..., \Sigma_{mm})$ and $\Sigma_{ii} = (1-\theta)I_{p_i} + \theta J_{p_i}$, $\theta = 0.1, 0.5, 0.9,$ $i = 1, 2, ..., m,$ in which J is a matrix where all elements are 1's, and the dimensions of Σ_{ii} are $p_i \times p_i$ and $\sum_{i=1}^{m} p_i = p.$

2) The second form of covariance matrix is $\Sigma_2 = diag(\Sigma_{11}, \Sigma_{22}, ..., \Sigma_{mm})$ and $\Sigma_{ii} = [\rho_{kl}],$ $\rho_{kl} = \theta^{|k-l|},$ $\theta = 0.9,$ $i = 1, 2, ..., m,$ and $k, l = 1, 2, ..., p_i,$ in which the dimensions of Σ_{ii} are $p_i \times p_i$ and $\sum_{i=1}^{m} p_i = p.$

3) The third form of covariance matrix is Σ_3, which the same as Σ_1 except that + and – are alternately assigned to the elements of Σ_1.

4) The fourth form of covariance matrix is $\Sigma_4 = diag(\Sigma_{11}, \Sigma_{22}, ..., \Sigma_{mm})$ and $\Sigma_{ii} = [\rho_{kl}],$ $\rho_{kl} = (-\theta)^{|k-l|},$ $\theta = 0.9,$ $i = 1, 2, ..., m,$ and $k, l = 1, 2, ..., p_i,$ in which the dimensions of Σ_{ii} are $p_i \times p_i$ and $\sum_{i=1}^{m} p_i = p.$

Four different forms of population are used to calculate the empirical power under the alternative hypothesis as follows:

1) The first form of covariance matrix is $\Sigma_5 : \Sigma_{ii} = (1-\theta)I_{p_i} + \theta J_{p_i}$ and the off-block elements are $0.5\theta,$ $\theta = 0.1, 0.5, 0.9,$ $i = 1, 2, ..., m,$ in which J is a matrix where all the elements are 1's and the dimensions of Σ_{ii} are $p_i \times p_i$ and $\sum_{i=1}^{m} p_i = p.$

2) The second form of covariance matrix is $\Sigma_6 = [\rho_{kl}]$, $\rho_{kl} = \theta^{|k-l|}$, $\theta = 0.9$, $k,l = 1,2,...,p$.

3) The third form of covariance matrix is Σ_7, which is the same as Σ_5 except that + and – are alternately assigned to the elements of Σ_5.

4) The fourth form of covariance matrix is $\Sigma_8 = [\rho_{kl}]$, $\rho_{kl} = (-\theta)^{|k-l|}$, $\theta = 0.9$, $k,l = 1,2,...,p$.

The simulations are conducted at $p \in \{100, 200, 300, 400\}$ with $n \in \{50,100\}$. For each combination of (p,n), the block sizes are either equal or mixed. For the equal block size case, all the Σ_{ii} are of equal size $p_i = 5,10,25$ containing p/p_i blocks, and for the mixed block size case, there are two different block sizes in the matrix. The two block sizes of submatrix Σ_{ii} are chosen from $p_i, p_j = 5,10,25$, in which size p_i has $p/2p_i$ blocks and size p_j has $p/2p_j$ blocks. The number of blocks rather than the block size is considered in order to reach conclusions in the same direction.

4.1.3 Simulation Results

The test statistics are compared in terms of their empirical type I error rates (ε_1) and empirical powers (ε_2) (these values are reported in Tables 4.1-4.8 and additional report of ε_1 and ε_2 are shown in Tables B.1-B.8 in Appendix B).

The values of ε_1 for $\hat{T}_p$, $\hat{T}_b$ and $\hat{T}_c$ when $\Sigma = \Sigma_1$ with equal and mixed block size are presented in Tables 4.1 and 4.2, respectively. It is evident that ε_1 of $\hat{T}_p$ are close to $\alpha = 0.05$ and the maximum difference between ε_1 and α is 0.0165 (i.e. not much different). When the value of θ is 0.1, the absolute values of difference between ε_1 and α, $|\varepsilon_1 - \alpha|$, are not much different for any p and the number of blocks. The values of $|\varepsilon_1 - \alpha|$ for $\hat{T}_p$ increase when θ increases and the number of blocks decreases and the values of $|\varepsilon_1 - \alpha|$ decrease as the value of p is made larger.

The values of ε_1 for $\hat{T}_p$ are compared with those of $\hat{T}_b$ and $\hat{T}_c$. When comparing the test statistic $\hat{T}_p$ with $\hat{T}_b$, the values of ε_1 for $\hat{T}_p$ and $\hat{T}_b$ are slightly different which the values of ε_1 for $\hat{T}_b$ are minor less than the values of ε_1 for $\hat{T}_p$ in all set of parameters. When comparing the test statistic $\hat{T}_p$ with $\hat{T}_c$, the values of $|\varepsilon_1 - \alpha|$ for $\hat{T}_p$ and $\hat{T}_c$ are close together when θ is 0.1 or the number of blocks is large (for equal block size case $p_i = 5$ and for mixed block size case $p_i = 5$ and $p_j = 10$). However, the values of $|\varepsilon_1 - \alpha|$ for $\hat{T}_p$ are higher than $\hat{T}_c$ and become slightly different as p increases for the small number of blocks (for equal block size case $p_i = 10, 25$ and for mixed block size case $p_i = 5$ and $p_j = 25$ or $p_i = 10$ and $p_j = 25$) with a large value of θ (0.5, 0.9).

Tables 4.3 and 4.4 present the values of ε_1 for $\hat{T}_p$, $\hat{T}_b$ and $\hat{T}_c$ when $\Sigma = \Sigma_2$ with equal and mixed block sizes, respectively. The values of ε_1 for $\hat{T}_p$ are close to $\alpha = 0.05$ with the maximum difference between ε_1 and α being 0.0165. The different number of blocks affects ε_1 for $p = 100, 200$, whereas a larger p does not affect ε_1 with any number of blocks.

Table 4.1 The empirical type I error rate when $\Sigma = \Sigma_1$, $\theta = 0.1$, 0.5, and 0.9 with equal block sizes

n	p	p_i	$\theta = 0.1$			$\theta = 0.5$			$\theta = 0.9$		
			$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	0.0486	0.0464	0.0500	0.0534	0.0519	0.0500	0.0534	0.0517	0.0500
		10	0.0490	0.0476	0.0475	0.0578	0.0566	0.0475	0.0596	0.0579	0.0475
		25	0.0533	0.0516	0.0484	0.0647	0.0625	0.0484	0.0649	0.0642	0.0484
	200	5	0.0454	0.0445	0.0484	0.0502	0.0479	0.0484	0.0509	0.0491	0.0484
		10	0.0463	0.0450	0.0481	0.0541	0.0529	0.0481	0.0572	0.0558	0.0481
		25	0.0516	0.0497	0.0495	0.0625	0.0610	0.0495	0.0632	0.0624	0.0495
	300	5	0.0524	0.0506	0.0535	0.0510	0.0502	0.0535	0.0512	0.0496	0.0535
		10	0.0504	0.0489	0.0522	0.0516	0.0501	0.0522	0.0537	0.0528	0.0522
		25	0.0523	0.0512	0.0483	0.0583	0.0561	0.0483	0.0581	0.0558	0.0483
	400	5	0.0477	0.0464	0.0487	0.0497	0.0479	0.0487	0.0507	0.0488	0.0487
		10	0.0515	0.0494	0.0494	0.0503	0.0490	0.0494	0.0518	0.0499	0.0494
		25	0.0520	0.0508	0.0536	0.0559	0.0547	0.0536	0.0561	0.0538	0.0536
100	100	5	0.0489	0.0480	0.0514	0.0510	0.0501	0.0514	0.0527	0.0519	0.0514
		10	0.0461	0.0455	0.0502	0.0546	0.0537	0.0502	0.0571	0.0562	0.0502
		25	0.0535	0.0528	0.0505	0.0638	0.0628	0.0505	0.0648	0.0641	0.0505
	200	5	0.0513	0.0507	0.0510	0.0517	0.0506	0.0510	0.0540	0.0530	0.0510
		10	0.0495	0.0486	0.0500	0.0557	0.0550	0.0500	0.0556	0.0548	0.0500
		25	0.0519	0.0509	0.0542	0.0615	0.0607	0.0542	0.0635	0.0629	0.0542
	300	5	0.0441	0.0428	0.0458	0.0507	0.0497	0.0458	0.0510	0.0507	0.0458
		10	0.0481	0.0476	0.0471	0.0539	0.0525	0.0471	0.0551	0.0544	0.0471
		25	0.0505	0.0495	0.0486	0.0599	0.0590	0.0486	0.0568	0.0560	0.0486
	400	5	0.0469	0.0461	0.0516	0.0493	0.0484	0.0516	0.0509	0.0503	0.0516
		10	0.0472	0.0467	0.0508	0.0506	0.0497	0.0508	0.0519	0.0515	0.0508
		25	0.0471	0.0459	0.0522	0.0561	0.0552	0.0522	0.0564	0.0559	0.0522

Table 4.2 The empirical type I error rate when $\Sigma = \Sigma_1$, $\theta = 0.1$, 0.5, and 0.9 with mixed block sizes

n	p	p_i	p_j	$\theta = 0.1$			$\theta = 0.5$			$\theta = 0.9$		
				$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	10	0.0534	0.0520	0.0522	0.0593	0.0575	0.0522	0.0576	0.0561	0.0522
		5	25	0.0482	0.0467	0.0521	0.0600	0.0585	0.0521	0.0638	0.0624	0.0521
		10	25	0.0525	0.0506	0.0490	0.0634	0.0621	0.0490	0.0643	0.0630	0.0490
	200	5	10	0.0453	0.0437	0.0479	0.0513	0.0495	0.0479	0.0545	0.0530	0.0479
		5	25	0.0461	0.0447	0.0473	0.0574	0.0563	0.0473	0.0608	0.0595	0.0473
		10	25	0.0487	0.0473	0.0501	0.0549	0.0533	0.0501	0.0575	0.0558	0.0501
	300	5	10	0.0518	0.0505	0.0530	0.0524	0.0510	0.0530	0.0549	0.0535	0.0530
		5	25	0.0480	0.0460	0.0507	0.0558	0.0551	0.0507	0.0556	0.0550	0.0507
		10	25	0.0517	0.0492	0.0491	0.0573	0.0555	0.0491	0.0569	0.0551	0.0491
	400	5	10	0.0501	0.0490	0.0490	0.0500	0.0487	0.0490	0.0509	0.0495	0.0490
		5	25	0.0496	0.0477	0.0503	0.0527	0.0520	0.0503	0.0525	0.0523	0.0503
		10	25	0.0493	0.0475	0.0530	0.0552	0.0536	0.0530	0.0543	0.0522	0.0530
100	100	5	10	0.0546	0.0537	0.0525	0.0555	0.0550	0.0525	0.0575	0.0568	0.0525
		5	25	0.0520	0.0508	0.0467	0.0622	0.0622	0.0467	0.0636	0.0628	0.0467
		10	25	0.0540	0.0530	0.0518	0.0653	0.0647	0.0518	0.0665	0.0658	0.0518
	200	5	10	0.0527	0.0516	0.0525	0.0549	0.0540	0.0525	0.0552	0.0549	0.0525
		5	25	0.0532	0.0520	0.0528	0.0564	0.0560	0.0528	0.0591	0.0580	0.0528
		10	25	0.0515	0.0509	0.0507	0.0564	0.0558	0.0507	0.0566	0.0560	0.0507
	300	5	10	0.0510	0.0504	0.0489	0.0538	0.0532	0.0489	0.0559	0.0547	0.0489
		5	25	0.0506	0.0498	0.0493	0.0564	0.0551	0.0493	0.0578	0.0571	0.0493
		10	25	0.0496	0.0484	0.0505	0.0573	0.0568	0.0505	0.0595	0.0588	0.0505
	400	5	10	0.0471	0.0462	0.0483	0.0539	0.0533	0.0483	0.0554	0.0547	0.0483
		5	25	0.0450	0.0440	0.0470	0.0548	0.0538	0.0470	0.0584	0.0575	0.0470
		10	25	0.0483	0.0476	0.0478	0.0524	0.0521	0.0478	0.0536	0.0531	0.0478

In a comparison of the values of ε_1 for $\hat{T}_p$ and $\hat{T}_b$, the results are similar to $\Sigma = \Sigma_1$, i.e. the values of ε_1 for $\hat{T}_p$ and $\hat{T}_b$ are slightly different which the values ε_1 for $\hat{T}_b$ are minor less than the value ε_1 for $\hat{T}_p$ in all set of parameter. In a comparison of the values of ε_1 for $\hat{T}_p$ and $\hat{T}_c$, the $|\varepsilon_1 - \alpha|$ are not much different for the large number of blocks (for equal block size case $p_i = 5$ and for mixed block size case $p_i = 5$ and $p_j = 10$) and the values of ε_1 for $\hat{T}_c$ are closer to $\alpha = 0.05$ than $\hat{T}_p$ when the number of blocks decreases (for equal block size case $p_i = 10, 25$ and for mixed block size case $p_i = 5$ and $p_j = 25$ or $p_i = 10$ and $p_j = 25$) with $p = 100, 200$ Additionally the values of $|\varepsilon_1 - \alpha|$ for $\hat{T}_p$ and $\hat{T}_c$ are similary when $p = 300, 400$.

An investigation into the impact of alternately assigning + and – in Σ_1, Σ_2 as Σ_3, Σ_4, respectively, is also carried out, the results of which are shown in Tables B.1-B.4 in Appendix B. It is found that the results had almost the same pattern as in Tables 4.1-4.4, respectively. In addition, it should be noted that the sample size does not affect ε_1 in all forms of the population covariance matrix.

The values of ε_2 for $\hat{T}_p$, $\hat{T}_b$ and $\hat{T}_c$ for $\Sigma = \Sigma_5$ with equal and mixed block sizes are given in Tables 4.5 and 4.6. In almost all cases, the results show that $\hat{T}_p$ obtains the values of ε_2 nearly 1 with a minimum value of 0.8699, which is still acceptable. When comparing the ε_2 of $\hat{T}_p$ with $\hat{T}_b$, they show that the ε_2 of $\hat{T}_p$ and $\hat{T}_b$ are not different. When comparing the ε_2 of $\hat{T}_p$ with $\hat{T}_c$, they showed that the ε_2 of $\hat{T}_c$ are small (far from 1) when the number of blocks is small, an effect which is dominant when $\theta = 0.1$ and smaller p. Even though the values of ε_2 of $\hat{T}_p$ and $\hat{T}_c$ increases when the sample size increases, the values of ε_2 for $\hat{T}_c$ are still far from 1.

Tables 4.7 and 4.8 report the values of ε_2 for $\hat{T}_p$, $\hat{T}_b$ and $\hat{T}_c$ for $\Sigma = \Sigma_6$ with equal and mixed block sizes, respectively. The values of ε_2 for $\hat{T}_p$ are once again nearly 1, while the minimum value is 0.9837. In a comparison of the values of ε_2 for $\hat{T}_p$ and $\hat{T}_b$, the values ε_2 for $\hat{T}_p$ and $\hat{T}_b$ are not different. Whereas the performance of $\hat{T}_c$ is poor when the number of blocks is small in equal block size case, it performs better in the mixed block size case. For any p and a small number of block, the values of ε_2 for $\hat{T}_c$ are smaller than those of $\hat{T}_p$, and they become close to each other when the number of blocks increases.

In Tables B.5-B.8 in Appendix B, the results when alternately assigning + and – in Σ_5, Σ_6 as Σ_7, Σ_8, respectively, are presented. They show once again that the results show almost the same pattern as Σ_5, Σ_6, i.e. assigning + and – sign does not affect either test.

From this simulation study, we observe that the performance of the proposed test statistic $\hat{T}_p$ is similar to the test statistic $\hat{T}_b$. When θ is small or the number of blocks is large, the values of empirical type I error of the proposed test statistic $\hat{T}_p$ are not different with those of $\hat{T}_c$. The absolute values of difference between empirical type I error and α of the proposed test statistic $\hat{T}_p$ are higher than the test statistic $\hat{T}_c$ when θ is large with p is small. When the empirical powers are considered, the proposed test statistic $\hat{T}_p$ produces these values close to 1, while the test statistic $\hat{T}_c$ produces these values far from 1.

Table 4.3 The empirical type I error rate when $\Sigma = \Sigma_2$ and $\theta = 0.9$ with equal block sizes

n	p	p_i	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	0.0531	0.0517	0.0500
		10	0.0603	0.0585	0.0475
		25	0.0599	0.0590	0.0484
	200	5	0.0509	0.0494	0.0484
		10	0.0550	0.0538	0.0481
		25	0.0612	0.0598	0.0495
	300	5	0.0507	0.0494	0.0535
		10	0.0534	0.0516	0.0522
		25	0.0546	0.0530	0.0483
	400	5	0.0506	0.0485	0.0487
		10	0.0515	0.0500	0.0494
		25	0.0518	0.0514	0.0536
100	100	5	0.0559	0.0547	0.0514
		10	0.0592	0.0582	0.0522
		25	0.0620	0.0609	0.0525
	200	5	0.0520	0.0505	0.0510
		10	0.0559	0.0555	0.0500
		25	0.0597	0.0590	0.0542
	300	5	0.0512	0.0505	0.0458
		10	0.0531	0.0524	0.0471
		25	0.0543	0.0533	0.0486
	400	5	0.0504	0.0493	0.0516
		10	0.0508	0.0497	0.0508
		25	0.0513	0.0510	0.0522

Table 4.4 The empirical type I error rate when $\Sigma = \Sigma_2$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	10	0.0554	0.0542	0.0480
		5	25	0.0627	0.0614	0.0459
		10	25	0.0665	0.0652	0.0500
	200	5	10	0.0536	0.0521	0.0479
		5	25	0.0564	0.0541	0.0473
		10	25	0.0561	0.0551	0.0501
	300	5	10	0.0548	0.0532	0.0530
		5	25	0.0556	0.0542	0.0507
		10	25	0.0556	0.0544	0.0491
	400	5	10	0.0533	0.0513	0.0490
		5	25	0.0511	0.0497	0.0503
		10	25	0.0522	0.0513	0.0530
100	100	5	10	0.0568	0.0564	0.0525
		5	25	0.0592	0.0585	0.0467
		10	25	0.0647	0.0635	0.0518
	200	5	10	0.0543	0.0532	0.0525
		5	25	0.0555	0.0547	0.0528
		10	25	0.0577	0.0568	0.0507
	300	5	10	0.0540	0.0532	0.0489
		5	25	0.0529	0.0524	0.0493
		10	25	0.0546	0.0541	0.0505
	400	5	10	0.0547	0.0536	0.0483
		5	25	0.0526	0.0516	0.0470
		10	25	0.0507	0.0494	0.0478

Table 4.5 The empirical power when $\Sigma = \Sigma_5$, $\theta = 0.1$, 0.5, and 0.9 with equal block sizes

n	p	p_i	$\theta = 0.1$			$\theta = 0.5$			$\theta = 0.9$		
			$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	**5**	0.9507	0.9496	0.7420	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	0.9368	0.9357	0.3649	1.0000	1.0000	0.9367	1.0000	1.0000	0.9762
		25	0.8699	0.8679	0.0767	0.9999	0.9999	0.1024	1.0000	1.0000	0.1074
	200	5	0.9988	0.9988	0.9693	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	0.9981	0.9981	0.7480	1.0000	1.0000	0.9993	1.0000	1.0000	0.9999
		25	0.9930	0.9927	0.1216	1.0000	1.0000	0.2050	1.0000	1.0000	0.2227
	300	5	0.9999	0.0999	0.9957	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	0.9997	0.0997	0.9210	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	0.9993	0.9993	0.1859	1.0000	1.0000	0.3535	1.0000	1.0000	0.3867
	400	5	1.0000	1.0000	0.9993	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	0.9999	0.0999	0.9777	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	0.9997	0.9997	0.2620	1.0000	1.0000	0.5165	1.0000	1.0000	0.5619
100	100	5	0.9993	0.9993	0.9886	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	0.9991	0.9991	0.8341	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	0.9961	0.9961	0.1762	1.0000	1.0000	0.3274	1.0000	1.0000	0.3567
	200	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9975	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.4415	1.0000	1.0000	0.7967	1.0000	1.0000	0.8400
	300	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.7122	1.0000	1.0000	0.9736	1.0000	1.0000	0.9859
	400	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.8846	1.0000	1.0000	0.9972	1.0000	1.0000	0.9989

Table 4.6 The empirical power when $\Sigma = \Sigma_5$, $\theta = 0.1$, 0.5, and 0.9 with mixed block sizes

n	p	p_i	p_j	$\theta = 0.1$			$\theta = 0.5$			$\theta = 0.9$		
				$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	10	0.9413	0.9403	0.5732	1.0000	1.0000	0.9990	1.0000	1.0000	0.9999
		5	25	0.9229	0.9217	0.3712	1.0000	1.0000	0.9923	1.0000	1.0000	0.9995
		10	25	0.9052	0.9030	0.1766	1.0000	1.0000	0.5527	1.0000	1.0000	0.6551
	200	5	10	0.9974	0.9972	0.9112	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	0.9964	0.9964	0.7434	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000
		10	25	0.9950	0.9950	0.4122	1.0000	1.0000	0.9410	1.0000	1.0000	0.9732
	300	5	10	0.9999	0.9999	0.9819	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	0.9998	0.9998	0.9148	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	0.9996	0.9996	0.6600	1.0000	1.0000	0.9948	1.0000	1.0000	0.9983
	400	5	10	1.0000	1.0000	0.9962	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	0.9712	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	0.9999	0.9999	0.8169	1.0000	1.0000	0.9994	1.0000	1.0000	0.9999
100	100	5	10	0.9994	0.9994	0.9558	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	0.9986	0.9986	0.8137	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	0.9982	0.9982	0.5072	1.0000	1.0000	0.9810	1.0000	1.0000	0.9934
	200	5	10	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	0.9935	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9203	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	300	5	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9932	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	400	5	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9989	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4.7 The empirical power when $\Sigma = \Sigma_6$ and $\theta = 0.9$ with equal block sizes

n	p	p_i	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9473
		25	0.9837	0.9828	0.1667
	200	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9171
		25	0.9973	0.9973	0.1834
	300	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9520
		25	0.9985	0.9984	0.1877
	400	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9584
		25	0.9988	0.9987	0.1921
100	100	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.6684
	200	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.7261
	300	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.7407
	400	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.7432

Table 4.8 The empirical power when $\Sigma = \Sigma_6$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.6291
	200	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.7348
	300	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.7498
	400	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	0.9999
		10	25	1.0000	1.0000	0.7748
100	100	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9989
	200	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	1.0000
	300	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9997
	400	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	1.0000

4.2 Simulation Study for Testing Discriminant Analysis in High-Dimensional Data

The proposed methods (TA, TB, TC, and BD) are investigated via a simulation study with 1,000 iterations and various parameter settings of the population covariance matrix. The performance of the TA, TB, TC,and BD methods are also compared with the DI (Dudoit et al., 2002) and SK (Srivastava & Kubokawa, 2007) methods which represented in section 2.2.2.

Recall that the four methods which are evaluated in this study:

1) The first proposed method TA, TB, and TC is given as

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h(\underset{\sim}{x}) < \hat{D}_l(\underset{\sim}{x})$ for all $l \neq h$,

where $\hat{D}_h(\underset{\sim}{x}) = (H^T\underset{\sim}{x} - H^T\underset{\sim}{\overline{x}}_h)^T \hat{S}^{*-1}(H^T\underset{\sim}{x} - H^T\underset{\sim}{\overline{x}}_h) - 2\ln p_h$ and

$\hat{S}^* = \hat{\lambda}^* T + (1-\hat{\lambda}^*)S_{pooled,y}$ with T is the shrinkage target matrix.

2) The second proposed method BD is given as

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h(\underset{\sim}{x}) < \hat{D}_l(\underset{\sim}{x})$ for all $l \neq h$,

where $\hat{D}_h(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h)^T S_{block}^{-1}(\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h) - 2\ln p_h$ and

$S_{block} = diag(S_{11}, S_{22}, ..., S_{mm})$.

3) The Dudoit et al. (2002)'s method DI is given as

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h^d(\underset{\sim}{x}) < \hat{D}_l^d(\underset{\sim}{x})$ for all $l \neq h$,

where $\hat{D}_h^d(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h)^T S_d^{-1}(\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h) - 2\ln p_h$ .and

$S_d = diag(s_{11}, ..., s_{pp})$, $s_{ii}, i = 1, ..., p$ are the diagonal element of the pooled sample covariance matrix.

4) The Srivastava and Kubokawa (2007)'s method SK is given as

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h^s(\underset{\sim}{x}) < \hat{D}_l^s(\underset{\sim}{x})$ for all $h \neq l$,

where $\hat{D}_h^s(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h)^T S_{SK}^{-1}(\underset{\sim}{x} - \underset{\sim}{\overline{x}}_h) - 2\ln p_h$ .and $S_{SK} = \left(S_{pooled} + \frac{tr(S_{pooled})}{\min(n, p)} I \right)$.

4.2.1 The Performance Evaluation Methods for Discriminant Analysis

To assess the performance of the two proposed methods, they are compared with the DI and SK methods by considering the misclassification rates (M), sensitity (SE), and specificity (SP), as defined below.

Let us consider a 2×2 contingency tale of confusion matrix as follows:

Table 4.9 The 2×2 confusion matrix

Actual class	**Predicted class**	
	1	**2**
1	A	B
2	C	D

where A be the number of observations from the actual class 1 assigned to the predicted class 1,

B be the number of observations from the actual class 1 assigned to the predicted class 2,

C be the number of observations from the actual class 2 assigned to the predicted class 1,

D be the number of observations from the actual class 2 assigned to the predicted class 2.

1) The Misclassification Rate (M)

The misclassification rate is defined as

$$M=1-\frac{A+D}{A+B+C+D}.$$

Its values range from 0 to 1 with the minimum value being 0, which means that all observations are assigned to their correct classes, and the maximum value is 1, which means that all new observations are assigned to incorrect classes. Therefore, the higher the misclassification rate (near to 1), the poorer the method.

2) The Sensitivity (SE)

The sensitivity is defined as

$$\text{SE}=\frac{\text{A}}{\text{A+B}}.$$

Its values range from 0 to 1 with the minimum value being 0, which means that all observations from class 1 are assigned to class 2, and the maximum value is 1, which means that all observations from class 1 are assigned to class 1. Therefore, the higher the sensitity (near to 1), the better the method.

3) Specificity (SP)

The specificity is defined as

$$\text{SP}=\frac{\text{D}}{\text{C+D}}.$$

Its values range from 0 to 1 with the minimum value being 0, which means that all observations from class 2 are assigned to class 1, and the maximum value is 1, which means that all observations from class 2 are assigned to class 2. Therefore, the higher the specificity (near to 1), the better the method.

Sensitivity and specificity are useful in a medical diagnosis which used to classify a sick people. From confusion matrix, actual class 1 and 2 are defined as the class of sick people and the class of healthy people respectively and. predicted class 1 and 2 are defined as the class of people who are identified as sick and the class of people who are identified as healthy respectively. That is, sensitivity is the proportion of sick people who are correctly identified as sick and specificity is the proportion of healthy people who are correctly identified as healthy.

4.2.2 Parameter Settings for Discriminant Analysis in High Dimensional Data

In this section, the performance of the two proposed methods are compared with the DI and SK methods via a simulation study by considering their misclassification rates, sensitivity, and specificity with 1,000 iterations.

The datasets are generated as follows: $\underset{\sim}{x}_{j1} \sim i.i.d.N_P(\underset{\sim}{\mu}_1, \Sigma)$ and $\underset{\sim}{x}_{j2} \sim i.i.d.N_P(\underset{\sim}{\mu}_2, \Sigma)$, $j = 1,...,n$, where, $\underset{\sim}{\mu}_1 = (\underset{\sim}{m}, 0,...,0)^T$, $\underset{\sim}{m}$ is a r dimensional vector generated from uniform(-1. 5,1. 5), $r = 0.05p$ and $\underset{\sim}{\mu}_2 = (0,0,...,0)^T$. In this study, only two classes are investigated and the prior probability of $\prod_1$ and $\prod_2$ are equal, i.e. the chance of an observation coming from $\prod_1$ or $\prod_2$ is equal.

Four different forms of population covariance are used to calculate the misclassification rates, sensitivity, and specificity are defined as follows:

1) The first form of covariance matrix is $\Sigma_9 = diag(\Sigma_{11}, \Sigma_{22}, ..., \Sigma_{mm})$ and $\Sigma_{ii} = (1-\theta)I_{p_i} + \theta J_{p_i}$, $\theta = 0.1, 0.5, 0.9$, $i = 1, 2, ..., m$, in which J is a matrix where all elements are 1's and the dimensions of Σ_{ii} are $p_i \times p_i$ and $\sum_{i=1}^{m} p_i = p.$

2) The second form of covariance matrix is $\Sigma_{10} = diag(\Sigma_{11}, \Sigma_{22}, ..., \Sigma_{mm})$ and $\Sigma_{ii} = [\rho_{kl}]$, $\rho_{kl} = \theta^{|k-l|}$, $\theta = 0.9$, $i = 1, 2, ..., m$, and $k, l = 1, 2, ..., p_i$, in which the dimensions of Σ_{ii} are $p_i \times p_i$ and $\sum_{i=1}^{m} p_i = p.$

3) The third form of covariance matrix is Σ_{11}, which is the same as Σ_9 except that + and – are alternately assigned to the elements of Σ_9.

4) The fourth form of covariance matrix is $\Sigma_{12} = diag(\Sigma_{11}, \Sigma_{22}, ..., \Sigma_{mm})$ and $\Sigma_{ii} = [\rho_{kl}]$, $\rho_{kl} = (-\theta)^{|k-l|}$, $\theta = 0.9$, $i = 1, 2, ..., m$, and $k, l = 1, 2, ..., p_i$, in which the dimensions of Σ_{ii} are $p_i \times p_i$ and $\sum_{i=1}^{m} p_i = p.$

The simulations are conducted at $p \in \{100, 200, 300, 400\}$ with $n \in \{35, 70\}$. Each experiment consist of a training dataset with 25,50 observations corresponding with a testing dataset with 10,20 observations from each class. The classification rules are built with the parameters estimated using the training dataset after which the classification procedure is performed on the testing dataset. For each combination of (p, n), both equal and mixed block sizes are considered. For the equal block size case, all the Σ_{ii} are of equal size $p_i = 5, 10, 25$ with p/p_i blocks, and in the mixed block size case, there are two different block sizes in the matrix. The two block sizes of submatrix Σ_{ii} are chosen from $p_i, p_j = 5, 10, 25$, in which size p_i has $p/2p_i$

blocks and size p_j has $p/2p_j$ blocks. The number of blocks rather than block size is considered in order to reach conclusions in the same direction.

For each of the simulations, 1,000 iterations are generated and the performance of each method is evaluated according to their misclassification rate, sensitivity, and specificity.

4.2.3 Simulation Results

The methods are compared in terms of the misclassification rate (M), sensitivity (SE), and specificity (SP) reported in Tables 4.9-4.18 and additional report of M, SE, and SP are reported in Tables B.9-B.48 in Appendix B. When $\Sigma = \Sigma_9$, the values of M are shown in Tables 4.9-4.14 and the values of SE and SP are shown in Tables B.9-B.20 in Appendix B. For any θ, when p and n increase, the values of M for TA, TB, TC, and BD decrease and the values of SE and SP increase. For fixed p and n, when θ increases, the TA, TB, and TC methods achieve higher values of M than the BD method and the TA, TB, and TC methods achieve lower values of SE and SP than the BD method.

When $\theta = 0.1$ and p and n are fixed with a decrease in the number of blocks, the values of M for the BD method increase for n is small while those of TA, TB, and TC methods only slightly increase for any n and the values of SE and SP for the BD method decrease for n is small while those of TA, TB, and TC methods only slightly decrease for any n. When comparing the two proposed methods with the DI and SK methods, all of them obtain similar values of M, SE, and SP except for BD, which are slightly higher for M and are slightly lower for SE and SP than the others method when the number of block decreases.

When $\theta = 0.5$ and p and n a re fixed with a decrease in the number of blocks, the values of M for the TA, TB, and TC methods increase whereas those of the BD method only increase when n is small and the values of SE and SP for the TA, TB, and TC methods decrease whereas those of the BD method only decreased when n is small.

Table 4.10 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.1$ with equal block sizes

n	*p*	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.3096	0.3146	0.3142	0.3166
		10	0.3115	0.3227	0.3172	0.3163
		25	0.3162	0.3518	0.3198	0.3155
	200	5	0.2379	0.2462	0.2434	0.2402
		10	0.2405	0.2562	0.2462	0.2415
		25	0.2484	0.3052	0.2534	0.2454
	300	5	0.1803	0.1866	0.1844	0.1831
		10	0.1819	0.2011	0.1863	0.1858
		25	0.1913	0.2539	0.1937	0.1882
	400	5	0.1494	0.1584	0.1558	0.1515
		10	0.1557	0.1686	0.1585	0.1554
		25	0.1653	0.2292	0.1693	0.1617
70	100	5	0.2621	0.2619	0.2642	0.2730
		10	0.2644	0.2661	0.2663	0.2710
		25	0.2691	0.2794	0.2716	0.2684
	200	5	0.1823	0.1848	0.1850	0.1940
		10	0.1865	0.1873	0.1882	0.1931
		25	0.1952	0.2059	0.1978	0.1904
	300	5	0.1346	0.1328	0.1360	0.1385
		10	0.1364	0.1369	0.1383	0.1397
		25	0.1462	0.1528	0.1480	0.1400
	400	5	0.0983	0.0980	0.1008	0.1032
		10	0.1001	0.1000	0.1023	0.1043
		25	0.1080	0.1187	0.1111	0.1054

Table 4.11 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.1$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.3077	0.3129	0.3152	0.3159
		5	25	0.3117	0.3305	0.3143	0.3129
		10	25	0.3119	0.3325	0.3176	0.3132
	200	5	10	0.2304	0.2388	0.2370	0.2352
		5	25	0.2391	0.2655	0.2427	0.2383
		10	25	0.2446	0.2750	0.2493	0.2438
	300	5	10	0.1887	0.1881	0.1921	0.1882
		5	25	0.1912	0.2183	0.1958	0.1902
		10	25	0.1900	0.2197	0.1946	0.1876
	400	5	10	0.1506	0.1582	0.1506	0.1556
		5	25	0.1554	0.1875	0.1554	0.1598
		10	25	0.1577	0.1929	0.1577	0.1632
70	100	5	10	0.2636	0.2609	0.2663	0.2735
		5	25	0.2640	0.2643	0.2652	0.2668
		10	25	0.2655	0.2672	0.2673	0.2684
	200	5	10	0.1848	0.1823	0.1870	0.1902
		5	25	0.1843	0.1874	0.1869	0.1855
		10	25	0.1901	0.1891	0.1915	0.1897
	300	5	10	0.1325	0.1283	0.1359	0.1360
		5	25	0.1391	0.1394	0.1418	0.1378
		10	25	0.1380	0.1402	0.1400	0.1368
	400	5	10	0.0942	0.0928	0.0966	0.0985
		5	25	0.1053	0.1050	0.1079	0.1064
		10	25	0.1029	0.1030	0.1051	0.1029

Table 4.12 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.5$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.3423	0.2456	0.3438	0.3095
		10	0.3622	0.2396	0.3662	0.2910
		25	0.3921	0.2791	0.3944	0.2594
	200	5	0.2857	0.1612	0.2911	0.2597
		10	0.3146	0.1597	0.3187	0.2586
		25	0.3599	0.2107	0.3642	0.2387
	300	5	0.2386	0.1029	0.2399	0.2180
		10	0.2780	0.1013	0.2779	0.2285
		25	0.3326	0.1513	0.3323	0.2178
	400	5	0.2055	0.0742	0.2119	0.1887
		10	0.2470	0.0735	0.2516	0.2099
		25	0.3146	0.1201	0.3139	0.2186
70	100	5	0.2935	0.1912	0.2948	0.2335
		10	0.3176	0.1796	0.3194	0.2080
		25	0.3561	0.1927	0.3566	0.1847
	200	5	0.2263	0.1098	0.2268	0.1833
		10	0.2615	0.1019	0.2633	0.1662
		25	0.3108	0.1124	0.3113	0.1309
	300	5	0.1799	0.0649	0.1801	0.1450
		10	0.2180	0.0565	0.2197	0.1338
		25	0.2797	0.0663	0.2807	0.1070
	400	5	0.1415	0.0376	0.1430	0.1146
		10	0.1834	0.0321	0.1836	0.1148
		25	0.2536	0.0397	0.2562	0.0947

Table 4.13 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.5$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.3579	0.2312	0.3633	0.2993
		5	25	0.3712	0.2516	0.3723	0.2786
		10	25	0.3831	0.2521	0.3833	0.2742
	200	5	10	0.2949	0.1441	0.2990	0.2519
		5	25	0.3320	0.1716	0.3340	0.2401
		10	25	0.3507	0.1793	0.3495	0.2511
	300	5	10	0.2607	0.0949	0.2606	0.2251
		5	25	0.2978	0.1210	0.3020	0.2189
		10	25	0.3063	0.1221	0.3101	0.2215
	400	5	10	0.2269	0.0638	0.2307	0.1973
		5	25	0.2691	0.0892	0.2717	0.1994
		10	25	0.2868	0.0910	0.2879	0.2114
70	100	5	10	0.3065	0.1787	0.3081	0.2188
		5	25	0.3273	0.1786	0.3300	0.2016
		10	25	0.3342	0.1808	0.3349	0.1945
	200	5	10	0.2427	0.0951	0.2441	0.1664
		5	25	0.2722	0.0951	0.2738	0.1427
		10	25	0.2862	0.0972	0.2867	0.1373
	300	5	10	0.1979	0.0508	0.2001	0.1307
		5	25	0.2385	0.0575	0.2403	0.1171
		10	25	0.2501	0.0584	0.2509	0.1138
	400	5	10	0.1600	0.0285	0.1621	0.1087
		5	25	0.2095	0.0315	0.2108	0.1018
		10	25	0.2174	0.0320	0.2186	0.0983

Table 4.14 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.9$ with equal block sizes

n	***p***	$\boldsymbol{p_i}$	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.3777	0.0430	0.3849	0.2837
		10	0.4029	0.0307	0.4050	0.2264
		25	0.4288	0.0544	0.4314	0.1337
	200	5	0.3402	0.0070	0.3420	0.2783
		10	0.3768	0.0050	0.3781	0.2541
		25	0.4157	0.0124	0.4169	0.1786
	300	5	0.3044	0.0008	0.3048	0.2593
		10	0.3485	0.0005	0.3481	0.2560
		25	0.3982	0.0028	0.3986	0.2003
	400	5	0.2693	0.0001	0.2738	0.2379
		10	0.3283	0.0001	0.3313	0.2580
		25	0.3858	0.0006	0.3871	0.2278
70	100	5	0.3357	0.0282	0.3367	0.1363
		10	0.3688	0.0182	0.3708	0.0780
		25	0.4043	0.0216	0.4053	0.0371
	200	5	0.2821	0.0032	0.2850	0.1462
		10	0.3291	0.0021	0.3314	0.0908
		25	0.3763	0.0023	0.3791	0.0309
	300	5	0.2457	0.0003	0.2467	0.1420
		10	0.2998	0.0001	0.3012	0.0971
		25	0.3596	0.0002	0.3625	0.0368
	400	5	0.2051	0.0000	0.2084	0.1283
		10	0.2680	0.0000	0.2693	0.1036
		25	0.3430	0.0001	0.3421	0.0442

Table 4.15 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.4032	0.0283	0.4040	0.2508
		5	25	0.4095	0.0473	0.4109	0.2051
		10	25	0.4250	0.0471	0.4254	0.1825
	200	5	10	0.3295	0.0006	0.3325	0.2535
		5	25	0.3693	0.0019	0.3725	0.2320
		10	25	0.3775	0.0021	0.3771	0.2259
	300	5	10	0.3295	0.0006	0.3325	0.2535
		5	25	0.3693	0.0019	0.3725	0.2320
		10	25	0.3775	0.0021	0.3771	0.2259
	400	5	10	0.3055	0.0001	0.3065	0.2437
		5	25	0.3471	0.0005	0.3491	0.2301
		10	25	0.3619	0.0003	0.3627	0.2417
70	100	5	10	0.3538	0.0166	0.3552	0.0946
		5	25	0.3783	0.0175	0.3782	0.0698
		10	25	0.3862	0.0176	0.3846	0.0493
	200	5	10	0.3071	0.0017	0.3101	0.1019
		5	25	0.3412	0.0013	0.3432	0.0668
		10	25	0.3574	0.0013	0.3561	0.0483
	300	5	10	0.2745	0.0002	0.2747	0.1094
		5	25	0.3192	0.0002	0.3226	0.0727
		10	25	0.3319	0.0001	0.3315	0.0575
	400	5	10	0.2420	0.0000	0.2450	0.1095
		5	25	0.2970	0.0000	0.2973	0.0782
		10	25	0.3094	0.0000	0.3105	0.0653

For n is large, the BD method achieves similar values of M, SE, and SP in any number of blocks. When the proposed methods are compared with the DI and SK methods, the BD method performs the best in alomost cases by obtaining the lowest values of M and the highest value of SE and SP, while the SK method performs better than the DI, TA, TB, and TC method, since these give the highest values of M and the lowest value of SE and SP, reflecting poor performance.

When $\theta = 0.9$ and p and n are fixed with a decrease in the number of blocks, the values of M for the TA,TB, and TC methods increase and the BD method obtains the lowest values compared with the other methods. The values of SE and SP for the TA, TB, and TC methods decrease and the BD method obtains the highest values compared with the other methods. In particular, the BD method is able to classify the test set nearly 100% correctly when p and n are high. The SK method performs better than the DI, TA, TB, and TC methods, and the latter three methods obtain the highest values of M and lowest values of SE and SP (similar to the DI method).

The results from simulation study when $\Sigma = \Sigma_{10}$ are give in Tables 4.15-4.18 and in Tables B.21-B.24 in Appendix B. For any θ, when p and n increase, the values of M of the proposed methods decrease and the values of SE and SP of them increase. The values of M of the TA,TB, and TC methods increase and the values of SE and SP of the TA,TB, and TC methods decrease when the number of blocks decreases with any p . For the BD method, the values of M increase and the values of SE and SP decrease when the number of blocks decreases and p and n are small. When comparing the proposed methods with the previously reported ones, the results are almost the same as the results from $\Sigma = \Sigma_9$ with $\theta = 0.9$, for which the BD method performs the best with this form of population covariance matrix.

In the case of $\Sigma = \Sigma_{11}, \Sigma_{12}$ (alternately assigning + and – to the elements of Σ_9, Σ_{10}) , the results given in Tables B.25-B.48 in Appendix B show that the performance of all methods are almost the same as Σ_9, Σ_{10}, i.e. alternating + and – has no effect.

Note that, the misclassification rate, sensitivity, and specificity of TA, TB and TC methods are equal for the same combination of the dimensions p and the number of observations n, i.e. the different of shrinkage target matrices in the first proposed method are not affect performance of this method in this simulation study.

Table 4.16 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{10}$ and $\theta = 0.9$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.3716	0.0561	0.3777	0.2906
		10	0.3879	0.0497	0.3897	0.2661
		25	0.3962	0.0824	0.3988	0.2507
	200	5	0.3304	0.0127	0.3345	0.2765
		10	0.3572	0.0113	0.3574	0.2664
		25	0.3762	0.0238	0.3770	0.2583
	300	5	0.2915	0.0017	0.2936	0.2545
		10	0.3194	0.0018	0.3215	0.2518
		25	0.3397	0.0053	0.3401	0.2434
	400	5	0.2585	0.0005	0.2626	0.2291
		10	0.2951	0.0004	0.2977	0.2402
		25	0.3214	0.0020	0.3226	0.2418
70	100	5	0.3279	0.0386	0.3292	0.1612
		10	0.3490	0.0316	0.3506	0.1334
		25	0.3632	0.0399	0.3641	0.1256
	200	5	0.2716	0.0062	0.2725	0.1565
		10	0.3033	0.0049	0.3046	0.1299
		25	0.3191	0.0060	0.3190	0.1126
	300	5	0.2321	0.0007	0.2344	0.1447
		10	0.2643	0.0002	0.2665	0.1211
		25	0.2881	0.0006	0.2904	0.1063
	400	5	0.1934	0.0001	0.1960	0.1271
		10	0.2321	0.0001	0.2327	0.1163
		25	0.2621	0.0000	0.2637	0.1028

Table 4.17 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{10}$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.3920	0.1252	0.3949	0.3008
		5	25	0.3883	0.1600	0.3902	0.2912
		10	25	0.4025	0.1577	0.4004	0.2921
	200	5	10	0.3410	0.0450	0.3448	0.2778
		5	25	0.3561	0.0709	0.3596	0.2801
		10	25	0.3691	0.0759	0.3727	0.2825
	300	5	10	0.3124	0.0198	0.3154	0.2633
		5	25	0.3236	0.0342	0.3244	0.2611
		10	25	0.3314	0.0374	0.3353	0.2634
	400	5	10	0.2819	0.0078	0.2854	0.2417
		5	25	0.3027	0.0188	0.3022	0.2419
		10	25	0.3149	0.0172	0.3165	0.2504
70	100	5	10	0.3422	0.0876	0.3436	0.1849
		5	25	0.3518	0.0962	0.3520	0.1824
		10	25	0.3589	0.0961	0.3610	0.1729
	200	5	10	0.2900	0.0256	0.2923	0.1630
		5	25	0.2985	0.0295	0.3008	0.1513
		10	25	0.3132	0.0289	0.3140	0.1449
	300	5	10	0.2536	0.0073	0.2562	0.1486
		5	25	0.2673	0.0101	0.2699	0.1436
		10	25	0.2795	0.0103	0.2804	0.1353
	400	5	10	0.2199	0.0021	0.2216	0.1350
		5	25	0.2409	0.0028	0.2418	0.1309
		10	25	0.2498	0.0028	0.2509	0.1263

From this simulation study, we observe that the TA, TB, and TC methods perform well and similar to DI and SK methods when θ is small. The BD method performs the best when θ is greater than 0.5.

4.3 Application to a Real-life Dataset

In this section, the test statistics $\hat{T}_p$, $\hat{T}_b$, $\hat{T}_c$ and the TA, TB, TC, BD, DI, and SK methods are applied to a real-life dataset. The Notterman Carcinoma dataset uses for this study is taken from a gene expression project at Princeton University, New Jersey by Notterman, Alon, Sierk, and Levine (2001). These data consist of 7,457 expression genes p in 18 paired colon tissue samples (18 tumor tissues n_1 and 18 normal tissues n_2) publicly available at http://genomics-pubs.princeton.edu/oncology/.

4.3.1 Testing for a Block Diagonal Covariance Matrix

From the dataset, 100 genes with sample size 10 from tumor and normal tissues are selected to test for a block diagonal covariance matrix; these data are assumed to be multivariate normal. Recall from the simulation study, the BD method performs well when the correlation coefficient between variables in the same block are higher than 0.5. Thus, the variables of this dataset are arranged in order that the correlation coefficient between any two adjacent variables in the same block is greater than or equal to 0.5. Procedure for arranging the variables in each block is as follows:

1) Select the first two variables which have maximum correlation to contain in the same block

2) Select the variable that has the maximum correlation with the first two variables in Step 1 from remaining variables

3) Select the variable that has the maximum correlation with the variable in previous step from remaining variables

4) Repeat Step 3 until all correlation with the variable in previous step of each remaining variable has less than 0.5

5) Move to new block and repeat Step 1-4

6) Do it until no variable left

The block sizes for the null hypothesis are of mixed size with a maximum of 17 and a minimum of 1, and the number of block is 14. There are 6 blocks which are of dimension one. Analysis of the dataset led to the proposed test statistic $\hat{T}_p$ producing a value for tumor tissues of 0.8837 (p-value ≈ 0.1885) and 0.5721 (p-value ≈ 0.2836) for normal tissues. The test statistic $\hat{T}_b$ produces a value for tumor tissues of 0.8479 (p-value ≈ 0.1983) and 0.5490 (p-value ≈ 0.2915) for normal tissues. Since the maximum block size is greater than the sample size, the test statistic $\hat{T}_c$ cannot be applied with this dataset. From the two test statistics, it can be concluded that the covariance matrix for the two groups is a block diagonal structure.

4.3.2 Discriminant Analysis

Before performing discriminant analysis on this dataset, the assumption that there is equality in the covariance of both classes needed to be checked, for which the test statistic proposed by Saowapa Chaipitak & Samruam Chongcharoen (2013) is used. The test statistic is -0.9383 (p-value ≈ 0.3481), which indicates that the covariance of both classes are equal. In this study, 10 tumor and normal tissues are selected for the training set and 5 tumor and normal tissues for the testing set.

The TA, TB, TC, BD, DI, and SK methods are applied to this dataset. The results are presented in confusion matrix as follows:

The results when the TA, TB, TC, DI, and SK methods are used for classification are showed in Table 4.18 and the results when the BD methods is used for classification are showed in Table 4.19.

Table 4.18 The 2×2 confusion matrix of TA, TB, TC, DI and SK methods

Actual class	**Predicted class**	
	Tumor tissues	**Normal tissues**
Tumor tissues	5	0
Normal tissues	0	5

Table 4.19 The 2×2 confusion matrix of BD method

Actual class	Predicted class	
	Tumor tissues	**Normal tissues**
Tumor tissues	5	0
Normal tissues	3	2

TA, TB, TC, DI, and SK methods produced zero values for M and the values of SE and SP are equal to 1.0000, i.e. 100% correct classification rate, while the BD method achieves values for M, SE, and SP of 0.3000, 1.0000, and 0.4000, respectively, indicating that the TA, TB, TC, DI, and SK methods perform better than the BD method with this dataset. Note that, the covariance matrix of this dataset can be constructed in block diagonal matrix with many blocks are of dimension one which may be the cause of poor performance of the BD method.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

Conclusions reached on this work are presented in this chapter for both testing for a block diagonal covariance matrix and discriminant analysis for high-dimensional data under a multivariate normal distribution. Some recommendations for future work are also suggested at the end of chapter.

5.1 Conclusions

In this dissertation, data are assumed to be independent multivariate normal distribution and high-dimensional, where the dimension p is larger than the sample size n. A new test for testing the hypothesis $H_0 : \Sigma = D_\Sigma$ against $H_a : \Sigma \neq D_\Sigma$, where D_Σ is the population covariance matrix with a block diagonal structure is proposed. The proposed test statistic $\hat{T}_p$ based on the ratio of the unbiased and consistent estimators proposed by Srivastava (2005) in Chapter 3 is presented under the null hypothesis as

$$\hat{T}_p = \frac{\hat{T} - 1}{\hat{\theta}},$$

where

$$\hat{T} = \frac{\left\{ trS^2 - \frac{1}{N}(trS)^2 \right\}}{\sum_{i=1}^{m} \left\{ trS_{ii}^2 - \frac{1}{N}(trS_{ii})^2 \right\}},$$

$$\hat{\theta}^2 = \frac{4}{\hat{a}_{2D}^2 N^2} \sum_{i \neq j}^{m} \hat{a}_{2i} \hat{a}_{2j}.$$

The distribution of $\hat{T}_p$ is achieved by applying the Delta method with the joint distribution of $\hat{a}_2$ and $\hat{a}_{2D}$. The asymptotic distribution of the proposed test statistic under the null hypothesis is derived and found to be standard normal. A simulation study to investigate the performance of the proposed test statistic and to compare the performance with other previously reported tests is carried out. The results show that the proposed test statistic performed desirably, i.e. its empirical type I error rate was close to the significance level $(\alpha = 0.05)$ and the empirical power was close to 1. The proposed test statistic was compared with that of Hyodo, Shutoh, Nishiyama, and Pavlenko (2015) and Bao et al. (2014) under the same conditions. Although the values of empirical type I error of the proposed test statistic were higher than the comparative test statistic in some cases, its performance was not significantly different. Moreover, the values of the empirical power of the proposed test statistic were closer to 1 than the previously reported test statistic in almost every case.

Two new discriminant methods for classifying data with high dimensions from two groups under two covariance matrices equal to the population covariance matrix with a block diagonal structure was also proposed.

The block diagonal structure of the population covariance can be tested by using the proposed test $\hat{T}_p$, and the two proposed discriminant methods guarantee that the inverse of the sample covariance matrix always exists. In the first method, the dimensionality of the observations is reduced and a well-conditioned covariance matrix used that guarantees minimum mean squared error (the TA, TB, and TC methods)

The first proposed classification rule is:

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h(\underset{\sim}{x}) < \hat{D}_l(\underset{\sim}{x})$ for all $l \neq h$,

where $\hat{D}_h(\underset{\sim}{x}) = (H^T\underset{\sim}{x} - H^T\underset{\sim}{\overline{x}}_h)^T \hat{S}^{*-1}(H^T\underset{\sim}{x} - H^T\underset{\sim}{\overline{x}}_h) - 2\ln p_h$ and

$\hat{S}^* = \hat{\lambda}^* T + (1-\hat{\lambda}^*) S_{pooled,y}$ with T is the shrinkage target matrix.

In the second method (BD), the block diagonal sample covariance matrix is used instead of the sample covariance matrix.

The second proposed classification rule is:

Assign $\underset{\sim}{x}$ to $\prod_h$ if $\hat{D}_h(\underset{\sim}{x}) < \hat{D}_l(\underset{\sim}{x})$ for all $l \neq h$,

where $\hat{D}_h(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\bar{x}}_h)^T S_{block}^{-1}(\underset{\sim}{x} - \underset{\sim}{\bar{x}}_h) - 2\ln p_h$. and $S_{block} = diag(S_{11}, S_{22}, ..., S_{mm})$.

The discriminant method in this study was used to consider only 2 classes of classification with an equal prior probability, i.e. the chance that an observation came from either class 1 or class 2 is equal. A simulation study to investigate the performance of the proposed methods and to compare their performance with other previously reported tests was carried out. The TA, TB, and TC methods performed well when the correlation among the variables in a block was weak but was inappropriate for classification when the correlation among the variables in a block was strong. Nonetheless, the BD method was superior when the correlation among variables in a block was strong.

Finally, the proposed test statistic $\hat{T}_p$ and two proposed discriminant methods (TA, TB, TC, and BD) are able to be applied in real dataset, the Notterman Carcinoma dataset.

5.2 Recommendations for Future Works

At this point, extending this study is suggested as follows:

1) In this study, the data are assumed to be a multivariate normal distribution. Instead, non-normal data could be considered to further develop the test statistic and discrimination methods.

2) Since the population covariance matrix with a block diagonal structure was tested for, a block sample covariance matrix could be considered for use in another multivariate task such as normality testing.

3) In this study, the new two discriminant methods are proposed for a two classes problem with equal covariance matrix, thus problems with *k* classes and/or an unequal covariance matrix could be examined.

4) Cluster analysis could also be considered for arranging variables within a block in a covariance matrix in real-life dataset.

BIBLIOGRAPHY

Anderson, T. W. (1984). *Multivariate statistical analysis.* New York: John Wiley & Sons.

Bao, Z., Hu, J., Pan, G., & Zhou, W. (2014). Test of independence for high-dimensional random vectors based on block correlation matrices. Retrieved from https://arxiv.org/pdf/1410.5082.pdf

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 711-720.

Chen, L.-F., Liao, H.-Y. M., Ko, M.-T., Lin, J.-C., & Yu, G.-J. (2000). A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, *33*(10), 1713-1726.

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

Di Pillo, P. J. (1976). The application of bias to discriminant analysis. *Communications in Statistics-Theory and Methods*, *5*(9), 843-854.

Dudoit, S., Fridlyand, J., & Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, *97*(457), 77-87.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179-188.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, *84*(405), 165-175.

Gini, C. (1912). Variabilità e mutabilità. In E. Pizetti & T. Salvemini (Eds.), *Reprinted in memorie di metodologica statistica.* Rome: Libreria Eredi Virgilio Veschi.

Guo, Y., Hastie, T., & Tibshirani, R. (2007). Regularized linear discriminant analysis and its application in microarrays. *Biostatistics*, *8*(1), 86-100.

Hyodo, M., Shutoh, N., Nishiyama, T., & Pavlenko, T. (2015). Testing block-diagonal covariance structure for high-dimensional data. *Statistica Neerlandica*, *69*(4), 460-482.

Jiang, D., Bai, Z., & Zheng, S. (2013). Testing the independence of sets of large-dimensional variables. *Science China Mathematics*, *56*(1), 135-147.

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (vol. 5). Upper Saddle River, NJ.: Prentice Hall.

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

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, *10*(5), 603-621.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365-411.

Lehmann, E. L., & Romano, J. P. (2006). *Testing statistical hypotheses*. New York: Springer-Verlag.

Liu, K., Cheng, Y.-Q., & Yang, J.-Y. (1993). Algebraic feature extraction for image recognition based on an optimal discriminant criterion. *Pattern Recognition*, *26*(6), 903-911.

Lu, J., Plataniotis, K. N., & Venetsanopoulos, A. N. (2003). Regularized discriminant analysis for the small sample size problem in face recognition. *Pattern Recognition Letters*, *24*(16), 3079-3087.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. New York: John Wiley & Sons.

Notterman, D. A., Alon, U., Sierk, A. J., & Levine, A. J. (2001). Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Research*, *61*(7), 3124-3130.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, 502-518.

Saowapha Chaipitak, & Samruam Chongcharoen. (2013). A test for testing the equality of two covariance matrices for high-dimensional data. *Journal of Applied Sciences*, *13*(2), 270-277.

Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, *4*(1), 1-30. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.165.3620&rep=rep1&type=pdf

Schmidt-Heck, W., Guthke, R., Toepfer, S., Reischer, H., Duerrschmid, K., & Bayer, K. (2004). Reverse engineering of the stress response during expression of a recombinant protein. In *Proceedings of the EUNITE symposium* (pp. 10-12).

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, *92*(4), 951-956.

Shannon, C. E. (1949). Communication theory of secrecy systems. *Bell Labs Technical Journal*, *28*(4), 656-715.

Srivastava, M. S. (2002). *Methods of multivariate statistics*. New York: Wiley-Interscience.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society*, *35*(2), 251-272.

Srivastava, M. S., & Kubokawa, T. (2007). Comparison of discrimination methods for high dimensional data. *Journal of the Japan Statistical Society*, *37*(1), 123-134.

Srivastava, M. S., & Reid, N. (2012). Testing the structure of the covariance matrix with fewer observations than the dimension. *Journal of Multivariate Analysis*, *112*, 156-171.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Harlow: Addison Wesley.

Tian, Q., Barbero, M., Gu, Z.-H., & Lee, S. H. (1986). Image classification by the Foley-Sammon transform. *Optical Engineering*, *25*(7), 257834.

Xu, P., Brock, G. N., & Parrish, R. S. (2009). Modified linear discriminant analysis approaches for classification of high-dimensional microarray data. *Computational Statistics & Data Analysis*, *53*(5), 1674-1687.

APPENDICES

Appendix A

Proof of Theorem Used in Study

A.1 The distribution of $\left(\hat{a}_2 \quad \hat{a}_{2D}\right)^T$

In order to find distribution of $\left(\hat{a}_2 \quad \hat{a}_{2D}\right)^T$, we need to find the expectation and variance of $\hat{a}_2$ and $\hat{a}_{2D}$ by express them in term of chi-square. The distribution of $\left(\hat{a}_2 \quad \hat{a}_{2D}\right)^T$ was obtained in this section by following step:

1. Expression of $\hat{a}_2$ and $\hat{a}_{2D}$ in term of chi-square

1.1 Recall from the result of Srivastava (2005), we obtain the expression of trS, $(trS)^2$, and trS^2 in term of chi-square random variable as follows:

Let $NS = YY^T \sim W_p(\Sigma, N)$, where $Y = (\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, ..., \underset{\sim}{Y}_N)$ and $\underset{\sim}{Y}_i \sim N_p(0, \Sigma)$. Let $\Gamma = (\underset{\sim}{\gamma}_1, \underset{\sim}{\gamma}_2, ..., \underset{\sim}{\gamma}_p)$ be the matrix contain p eigenvectors of Σ corresponding to the p eigenvalues, denoted by $\lambda_1, \lambda_2, ..., \lambda_p$ such that $\Gamma\Sigma\Gamma^T = \Lambda$, $\Gamma\Gamma^T = I_P$ where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_p)$. Then, if $U = (\underset{\sim}{u}_1, \underset{\sim}{u}_2, ..., \underset{\sim}{u}_N)$, where $\underset{\sim}{u}_i$ are $i.i.d.\ N_p(0, I)$, $Y = \Sigma^{\frac{1}{2}}U$, and $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma$,

$$
\begin{aligned}
tr(S) &= \frac{1}{N} tr(YY^T) \\
&= \frac{1}{N} tr(\Sigma^{\frac{1}{2}} U U^T \Sigma^{\frac{1}{2}}) \\
&= \frac{1}{N} tr(U^T \Sigma U) \\
&= \frac{1}{N} tr(U^T \Gamma^T \Lambda \Gamma U)
\end{aligned}
$$

$$tr(S) = \frac{1}{N} tr(\mathrm{A}^T \Lambda \mathrm{A})$$

$$= \frac{1}{N} \sum_{i=1}^{p} \lambda_i \underset{\sim}{\alpha}_i^T \underset{\sim}{\alpha}_i \,,$$

where $U^T \Gamma^T = \mathrm{A}^T = (\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, ..., \underset{\sim}{\alpha}_p)$, and $\underset{\sim}{\alpha}_i$ are $i.i.d.\ N_N(0, I)$, Thus, if $v_{ii} = \underset{\sim}{\alpha}_i^T \underset{\sim}{\alpha}_i$, v_{ii} are $i.i.d.\ \chi^2(n)$.

$$tr(S) = \frac{1}{N} \sum_{i=1}^{p} \lambda_i v_{ii}$$

$$(trS)^2 = \left(\frac{1}{N} \sum_{i=1}^{p} \lambda_i v_{ii} \right)^2$$

$$= \frac{1}{N^2} \left(\sum_{i=1}^{p} \lambda_i v_{ii}^2 + 2 \sum_{i<j}^{p} \lambda_i \lambda_j v_{ii} v_{jj} \right)$$

$$trS^2 = \frac{1}{N^2} tr(\mathrm{A}^T \Lambda \mathrm{A})(\mathrm{A}^T \Lambda \mathrm{A})$$

$$= \frac{1}{N^2} tr \left(\sum_{i=1}^{p} \lambda_i \underset{\sim}{\alpha}_i \underset{\sim}{\alpha}_i^T \right) \left(\sum_{i=1}^{p} \lambda_i \underset{\sim}{\alpha}_i \underset{\sim}{\alpha}_i^T \right)$$

$$= \frac{1}{N^2} tr \left(\sum_{i=1}^{p} \lambda_i^2 \underset{\sim}{\alpha}_i \underset{\sim}{\alpha}_i^T \underset{\sim}{\alpha}_i \underset{\sim}{\alpha}_i^T + 2 \sum_{i<j}^{p} \lambda_i \lambda_j \underset{\sim}{\alpha}_i \underset{\sim}{\alpha}_i^T \underset{\sim}{\alpha}_j \underset{\sim}{\alpha}_j^T \right)$$

$$= \frac{1}{N^2} \sum_{i=1}^{p} \lambda_i^2 (\underset{\sim}{\alpha}_i^T \underset{\sim}{\alpha}_i)^2 + \sum_{i<j}^{p} \lambda_i \lambda_j (\underset{\sim}{\alpha}_i^T \underset{\sim}{\alpha}_j)^2$$

$$= \frac{1}{N^2} \left(\sum_{i=1}^{p} \lambda_i^2 v_{ii}^2 + \sum_{i<j}^{p} \lambda_i \lambda_j v_{ij}^2 \right), \text{ where } v_{ij} = \underset{\sim}{\alpha}_i^T \underset{\sim}{\alpha}_j$$

Since $\dfrac{N^2}{(N-1)(N+2)} = c$ in (3.1) then,

$$\hat{a}_2 = \frac{c}{p} \left\{ trS^2 - \frac{1}{N} (trS)^2 \right\}$$

$$= \frac{c}{N^2 p} \left[\sum_{i=1}^{p} \lambda_i^2 v_{ii}^2 + \sum_{i<j}^{p} \lambda_i \lambda_j v_{ij}^2 - \frac{1}{N} \left(\sum_{i=1}^{p} \lambda_i v_{ii}^2 + 2 \sum_{i<j}^{p} \lambda_i \lambda_j v_{ii} v_{jj} \right) \right]$$

$$= \frac{c}{N^2 p} \left[\frac{N-1}{N} \sum_{i=1}^{p} \lambda_i^2 v_{ii}^2 + 2 \sum_{i<j}^{p} \lambda_i \lambda_j \left(v_{ij}^2 - \frac{1}{N} v_{ii} v_{jj} \right) \right]$$

$$= c\left[\frac{N-1}{N^3 p}\sum_{i=1}^{p}\lambda_i^2 v_{ii}^2\right] + c\left[\frac{2}{N^2 p}\sum_{i<j}^{p}\lambda_i\lambda_j\phi_{ij}\right]$$

Then, $\hat{a}_2 = c(q_1 + q_2)$, where $q_1 = \frac{N-1}{N^3 p}\sum_{i=1}^{p}\lambda_i^2 v_{ii}^2$, $q_2 = \frac{2}{N^2 p}\sum_{i<j}^{p}\lambda_i\lambda_j\phi_{ij}$, and

$$\phi_{ij} = v_{ij}^2 - \frac{1}{N}v_{ii}v_{jj}$$

$$= (\underset{\sim}{\alpha}_i^T \underset{\sim}{\alpha}_j)^2 - \frac{1}{N}(\underset{\sim}{\alpha}_i^T \underset{\sim}{\alpha}_i)(\underset{\sim}{\alpha}_j^T \underset{\sim}{\alpha}_j) \tag{A.1}$$

Lemma A.1 For ϕ_{ij} defined above, we have

$$E(\phi_{ij}) = 0, \tag{A.2}$$

$$E(\phi_{ij}\phi_{ik}) = 0 \text{ for all distinct } i, j, k \tag{A.3}$$

$$Var(\phi_{ij}) = 2(N+2)(N-1) \tag{A.4}$$

Proof see Srivastava (2015) □

1.2 From Srivastava (2015) result, we obtain the expression of trS_{ii}, $(trS_{ii})^2$, and trS_{ii}^2 in term of chi-square random variable as follows:

Under H_0, let $NS = YY^T \sim W_p(D_\Sigma, N)$, where $Y = (\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, ..., \underset{\sim}{Y}_N)$ and $\underset{\sim}{Y}_i \sim N_p(0, D_\Sigma)$. Let $\Gamma_D = Diag(\Gamma_1, \Gamma_2, ..., \Gamma_m)$, where $\Gamma_i = (\underset{\sim}{\gamma}_1, \underset{\sim}{\gamma}_2, ..., \underset{\sim}{\gamma}_{p_i})$ be the matrix contain p_i eigenvectors of Σ_{ii} corresponding to the p_i eigenvalues, denoted by $\omega_1^{(i)}, \omega_2^{(i)}, ..., \omega_{p_i}^{(i)}$ such that $\Gamma_i\Sigma_i\Gamma_i^T = \Omega_i$, $\Gamma_i\Gamma_i^T = I_{p_i}$, $i = 1, 2, ..., m$, where $\Omega_i = diag(\omega_1^{(i)}, \omega_2^{(i)}, ..., \omega_{p_i}^{(i)})$ and $\Omega = diag(\Omega_1, \Omega_2, ..., \Omega_m)$. Then, if $W = (\underset{\sim}{w}_1, \underset{\sim}{w}_2, ..., \underset{\sim}{w}_N)$, where $\underset{\sim}{w}_i$ are $i.i.d.\ N_p(0, I)$, $Y = D_\Sigma^{\frac{1}{2}}W$, and $D_\Sigma^{\frac{1}{2}}D_\Sigma^{\frac{1}{2}} = D_\Sigma$,

$$\begin{aligned} tr(S) &= \frac{1}{N}tr(YY^T) \\ &= \frac{1}{N}tr(D_\Sigma^{\frac{1}{2}}WW^T D_\Sigma^{\frac{1}{2}}) \\ &= \frac{1}{N}tr(W^T D_\Sigma W) \end{aligned}$$

$$= \frac{1}{N} tr(W^T \Gamma_D^T \Omega \Gamma_D W)$$

$$tr(S) = \frac{1}{N} tr(\mathrm{B}^T \Omega \mathrm{B})$$

$$= \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{p_i} \omega_j^{(i)} \underset{\sim}{\beta}_j^{(i)^T} \underset{\sim}{\beta}_j^{(i)},$$

where $W^T \Gamma_D^T = \mathrm{B}^T = (\underset{\sim}{\beta}_1^{(1)}, \underset{\sim}{\beta}_2^{(1)}, ..., \underset{\sim}{\beta}_{p_1}^{(1)}, \underset{\sim}{\beta}_{p_1+1}^{(2)} ..., \underset{\sim}{\beta}_{p_2}^{(2)}, ..., \underset{\sim}{\beta}_{p_m}^{(m)})$, and $\underset{\sim}{\beta}_i$ are $i.i.d.\ N_N(0, I)$,

Thus, if $w_{jj}^{(i)} = \underset{\sim}{\beta}_j^{(i)^T} \underset{\sim}{\beta}_j^{(i)}$, $w_{ii}^{(j)}$ are $i.i.d.\ \chi^2(N)$.

Note, we can consider S_{ii} is the sample covariance matrix of $\underset{\sim}{x}^{(i)}$, $i = 1, ..., m$.

In the same manner as trS, $(trS)^2$, and trS^2, we replace λ_i, v_{ij} with ω_i, w_{ij} respectively

$$tr(S_{ii}) = \frac{1}{N} \sum_{j=1}^{p_i} \omega_j^{(i)} w_{jj}^{(i)}$$

$$(trS_{ii})^2 = \frac{1}{N^2} \left(\sum_{j=1}^{p_i} \omega_j^{(i)} w_{jj}^{(i)^2} + 2 \sum_{j<k}^{p_i} \omega_j^{(i)} \omega_k^{(i)} w_{jj}^{(i)} w_{kk}^{(i)} \right)$$

$$trS_{ii}^2 = \frac{1}{N^2} \left(\sum_{j=1}^{p_i} \omega_j^{(i)^2} w_{jj}^{(i)^2} + \sum_{j<k}^{p_i} \omega_j^{(i)} \omega_k^{(i)} w_{jk}^{(i)^2} \right), \text{ where } w_{jk}^{(i)} = \underset{\sim}{\beta}_j^{(i)^T} \underset{\sim}{\beta}_k^{(i)}$$

$$\hat{a}_{2i} = \frac{c}{p} \left\{ trS_{ii}^2 - \frac{1}{N} (trS_{ii})^2 \right\}$$

$$= \frac{c}{N^2 p} \left[\sum_{j=1}^{p_i} \omega_j^{(i)^2} w_{jj}^{(i)^2} + \sum_{j<k}^{p_i} \omega_j^{(i)} \omega_k^{(i)} w_{jk}^{(i)^2} - \frac{1}{N} \left(\sum_{j=1}^{p_i} \omega_j^{(i)} w_{jj}^{(i)^2} + 2 \sum_{j<k}^{p_i} \omega_j^{(i)} \omega_k^{(i)} w_{jj}^{(i)} w_{kk}^{(i)} \right) \right]$$

$$= \frac{c}{N^2 p} \left[\frac{N-1}{N} \sum_{j=1}^{p_i} \omega_j^{(i)^2} w_{jj}^{(i)^2} + 2 \sum_{j<k}^{p_i} \omega_j^{(i)} \omega_k^{(i)} \left(w_{jk}^{(i)^2} - \frac{1}{N} w_{jj}^{(i)} w_{kk}^{(i)} \right) \right]$$

$$= c \left[\frac{N-1}{N^3 p} \sum_{j=1}^{p_i} \omega_j^{(i)} w_{jj}^{(i)^2} \right] + c \left[\frac{2}{N^2 p} \sum_{j<k}^{p_i} \omega_j^{(i)} \omega_k^{(i)} \eta_{jk}^{(i)} \right].$$

Then, $\hat{a}_{2i} = c(r_{1i} + r_{2i})$, where $r_{1i} = \frac{N-1}{N^3 p} \sum_{j=1}^{p_i} \omega_j^{(i)} w_{jj}^{(i)^2}$, $r_{2i} = \frac{2}{N^2 p} \sum_{j<k}^{p_i} \omega_j^{(i)} \omega_k^{(i)} \eta_{jk}^{(i)}$, and

$$\eta_{jk}^{(i)} = w_{jk}^{(i)^2} - \frac{1}{N} w_{jj}^{(i)} w_{kk}^{(i)}$$

$$= (\underset{\sim}{\beta}_j^{(i)^T} \underset{\sim}{\beta}_k^{(i)})^2 - \frac{1}{N}(\underset{\sim}{\beta}_j^{(i)^T} \underset{\sim}{\beta}_j^{(i)})(\underset{\sim}{\beta}_k^{(i)^T} \underset{\sim}{\beta}_k^{(i)})$$

$$\hat{a}_{2D} = \sum_{i=1}^{m} \hat{a}_{2i} = \sum_{i=1}^{m} \left(c\left[\frac{N-1}{N^3 p} \sum_{j=1}^{p_i} \omega_j^{(i)} w_{jj}^{(i)^2} \right] + c\left[\frac{2}{N^2 p} \sum_{j<k}^{p_i} \omega_j^{(i)} \omega_k^{(i)} \eta_{jk}^{(i)} \right] \right)$$

Consider, $\Omega = diag(\Omega_1, \Omega_2, ..., \Omega_m) = diag(\omega_1^{(1)}, \omega_2^{(1)}, ..., \omega_{p_1}^{(1)}, \omega_{p_1+1}^{(2)}, ..., \omega_{p_2}^{(2)}, ..., \omega_{p_m}^{(m)})$,

$W^T \Gamma_D^T = \mathrm{B}^T = (\underset{\sim}{\beta}_1^{(1)}, \underset{\sim}{\beta}_2^{(1)}, ..., \underset{\sim}{\beta}_{p_1}^{(1)}, \underset{\sim}{\beta}_{p_1+1}^{(2)} ..., \underset{\sim}{\beta}_{p_2}^{(2)}, ..., \underset{\sim}{\beta}_{p_m}^{(m)})$, for convenience we may rewrite

$\Omega = diag(\omega_1, \omega_2, .., \omega_p)$ and $\mathrm{B}^T = (\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, ..., \underset{\sim}{\beta}_p)$, respectively. So as we have

$w_{ii} = \underset{\sim}{\beta}_i^T \underset{\sim}{\beta}_i$, w_{ii} are $i.i.d.\ \chi^2(N)$.

$$\hat{a}_{2D} = c\left[\frac{N-1}{N^3 p} \sum_{j=1}^{p} \omega_j^2 w_{jj}^2 \right] + c\left[\frac{2}{N^2 p} \sum_{j<k}^{p} \omega_j \omega_k \eta_{jk} \right]$$

$$= c(r_1 + r_2),$$

where $r_1 = \frac{N-1}{N^3 p} \sum_{j=1}^{p} \omega_j^2 w_{jj}^2$, $r_2 = \frac{2}{N^2 p} \sum_{j<k}^{p} \omega_j \omega_k \eta_{jk}$, and

$$\eta_{jk} = w_{jk}^2 - \frac{1}{N} w_{jj} w_{kk}$$

$$= (\underset{\sim}{\beta}_j^T \underset{\sim}{\beta}_k)^2 - \frac{1}{N}(\underset{\sim}{\beta}_j^T \underset{\sim}{\beta}_j)(\underset{\sim}{\beta}_k^T \underset{\sim}{\beta}_k)$$

Lemma A.2 For η_{jk} defined above, we have

$$E(\eta_{jk}) = 0, \tag{A.5}$$

$$E(\eta_{jk}\eta_{jl}) = 0 \text{ for all distinct } j,k,l \tag{A.6}$$

$$Var(\eta_{jk}) = 2(N+2)(N-1) \tag{A.7}$$

Proof see Srivastava (2015) □

Lemma A. 3 Let v_{ii} and v_{ij} be a chi-square random variable with N degrees of freedom. Then

$$E(v_{ii}^r) = N(N+2)\cdots(N+2r-2), \quad r = 1, 2, ..., \tag{A.8}$$

$$Var(v_{ii}) = 2N, \tag{A.9}$$

$$Var(v_{ii}^2) = 8N(N+2)(N+3), \tag{A.10}$$

$$E(v_{ii} - N)^3 = 8N, \tag{A.11}$$

$$E(v_{ii} - N)^4 = 12N(N+4), \tag{A.12}$$

$$E(v_{ii}^2 - N(N+2))^4 = 3N(N+2)[272N^4 + O(N^3)], \tag{A.13}$$

$$E(v_{ij}^2) = N, \tag{A.14}$$

$$E(v_{ij}^4) = 3N(N+2), \tag{A.15}$$

$$E(v_{ii}v_{ij}^2) = N(N+2), \tag{A.16}$$

$$E(v_{ii}^2 v_{ij}^2) = N(N+2)(N+4), \tag{A.17}$$

$$E(v_{ii}v_{jj}v_{ij}^2) = N(N+2)^2 \tag{A.18}$$

Proof see Fisher et al. (2010) and Srivastava (2005) □

2. Find the expectation and variance of $\hat{a}_2$ and $\hat{a}_{2D}$

Lemma A.4 For $\hat{a}_2, \hat{a}_{2i}$, and $\hat{a}_{2D}$ are defined above, we have

1. $E(\hat{a}_2) = a_2$,
2. $E(\hat{a}_{2i}) = a_{2i}$,
3. $E(\hat{a}_{2D}) = a_{2D}$

Proof

1. $$E(\hat{a}_2) = \frac{N^2}{(N-1)(N+2)p} E\left\{ trS^2 - \frac{1}{N}(trS)^2 \right\}$$

$$= \frac{N^2}{(N-1)(N+2)} E\left(\left[\frac{N-1}{N^3 p} \sum_{i=1}^{p} \lambda_i^2 v_{ii}^2 \right] + \left[\frac{2}{N^2 p} \sum_{i<j}^{p} \lambda_i \lambda_j \phi_{ij} \right] \right)$$

From (A.2), $E(\phi_{ij}) = 0$, then

$$E(\hat{a}_2) = \frac{N^2}{(N-1)(N+2)} \left[\frac{N-1}{N^3 p} \sum_{i=1}^{p} \lambda_i^2 E(v_{ii}^2) \right]$$

From (A.8), $E(v_{ii}^2) = N(N+2)$, then

$$E(\hat{a}_2) = \frac{N^2}{(N-1)(N+2)}\left[\frac{N-1}{N^3 p}\sum_{i=1}^{p}\lambda_i^2 N(N+2)\right]$$

$$= \frac{1}{p}\sum_{i=1}^{p}\lambda_i^2$$

$$= \frac{tr\Sigma^2}{p} = a_2$$ □

2. We can prove in the same way as 1. □

3. Since $\hat{a}_{2D} = \sum_{i=1}^{m}\hat{a}_{2i} = \sum_{i=1}^{m}\frac{N^2}{(N-1)(N+2)p}\left\{trS_{ii}^2 - \frac{1}{N}(trS_{ii})^2\right\}$, then

$$E(\hat{a}_{2D}) = E\left(\sum_{i=1}^{m}\hat{a}_{2i}\right) = \sum_{i=1}^{m}a_{2i} = a_{2D}$$ □

Lemma A.5 For $\hat{a}_2, \hat{a}_{2i}$, and $\hat{a}_{2D}$ are defined above, we have

$$1. Var(\hat{a}_2) \simeq \frac{8}{Np}a_4 + \frac{4}{N^2}a_2^2$$

$$2. Var(\hat{a}_{2i}) \simeq \frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2$$

$$3. Var(\hat{a}_{2D}) \simeq \sum_{i=1}^{m}(\frac{8}{Np}a_{4i} + \frac{4}{N^2}a_{2i}^2)$$

Proof

1. $Var(\hat{a}_2) = Var\left(c(q_1 + q_2)\right)$

$$= c^2\left[Var(q_1) + Var(q_2) + 2Cov(q_1, q_2)\right]$$

Consider $Var(q_1)$ term,

$$Var(q_1) = \left(\frac{N-1}{N^3 p}\right)^2 Var\left(\sum_{i=1}^{p}\lambda_i^2 v_{ii}^2\right)$$

$$= \left(\frac{N-1}{N^3 p}\right)^2 \sum_{i=1}^{p}\lambda_i^4 Var(v_{ii}^2)$$

From (A.10), $Var(v_{ii}^2) = 8N(N+2)(N+3),$ then

$$Var(q_1) = \left(\frac{N-1}{N^3 p}\right)^2 \sum_{i=1}^{p} \lambda_i^4 8N(N+2)(N+3)$$

$$= \left(\frac{N-1}{N^3}\right)^2 \frac{8N(N+2)(N+3)}{p} \left(\frac{1}{p}\sum_{i=1}^{p} \lambda_i^4\right)$$

$$\simeq \frac{8}{Np}\left(tr\Sigma^4\right) = \frac{8a_4}{Np}$$

Consider $Var(q_2)$ term,

$$Var(q_2) \;= Var\left(\frac{2}{N^2 p}\sum_{i<j}^{p} \lambda_i \lambda_j \phi_{ij}\right)$$

$$= \left(\frac{2}{N^2 p}\right)^2 \sum_{i<j}^{p} \lambda_i^2 \lambda_j^2 Var(\phi_{ij})$$

From (A.4), $Var(\phi_{ij}) = 2(N+2)(N-1)$ then

$$Var(q_2) \qquad = \left(\frac{2}{N^2 p}\right)^2 2(N+2)(N-1)\sum_{i<j}^{p} \lambda_i^2 \lambda_j^2$$

$$= \left(\frac{2}{N^2}\right)^2 (N+2)(N-1)\left(2p^{-2}\sum_{i<j}^{p} \lambda_i^2 \lambda_j^2\right)$$

$$= \left(\frac{2}{N^2}\right)^2 (N+2)(N-1)\left(p^{-2}\left(\sum_{i=1}^{p} \lambda_i^2\right)^2 - p^{-2}\sum_{i=1}^{p} \lambda_i^4\right)$$

$$= \left(\frac{2}{N^2}\right)^2 (N+2)(N-1)\left(\left(\frac{\sum_{i=1}^{p} \lambda_i^2}{p}\right)^2 - \frac{1}{p}\left(\frac{\sum_{i=1}^{p} \lambda_i^4}{p}\right)\right)$$

$$= \left(\frac{2}{N^2}\right)^2 (N+2)(N-1)\left(a_2^2 - \frac{a_4}{p}\right)$$

$$\simeq \frac{4}{N^2}\left(a_2^2 - \frac{a_4}{p}\right)$$

Consider $Cov(q_1,q_2)$ term,

$$Cov(q_1,q_2)=E(q_1q_2)-E(q_1)E(q_2)$$

From (A.1), $E(\phi_{ij})=0$ then $E(q_2)=0$, so that

$$\begin{aligned}
Cov(q_1,q_2) &= E(q_1q_2)\\
&= \frac{2(N-1)}{N^5p^2}E\left(\sum_{i=1}^{p}\lambda_i^2v_{ii}^2\right)\left(\sum_{j<k}^{p}\lambda_j\lambda_k\phi_{jk}\right)\\
&= \frac{2(N-1)}{N^5p^2}\left\{\lambda_1^2E\left[v_{11}^2\left(\sum_{j<k}^{p}\lambda_j\lambda_k\phi_{jk}\right)\right]+\lambda_2^2E\left[v_{22}^2\left(\sum_{j<k}^{p}\lambda_j\lambda_k\phi_{jk}\right)\right]+\right.\\
&\left.\cdots+\lambda_p^2E\left[v_{pp}^2\left(\sum_{j<k}^{p}\lambda_j\lambda_k\phi_{jk}\right)\right]\right\}
\end{aligned}$$

We note that, from (A.1), (A.8), and (A.17)

$$\begin{aligned}
E(v_{ii}^2\phi_{jk}) &= E\left[v_{ii}^2\left(v_{jk}^2-\frac{1}{N}v_{jj}v_{kk}\right)\right]\\
&= E\left(v_{ii}^2v_{jk}^2-\frac{1}{N}v_{ii}^2v_{jj}v_{kk}\right)\\
&= E\left(v_{ii}^2v_{ik}^2-\frac{1}{N}v_{ii}^3v_{kk}\right)=0 \text{ for } i=j\neq k\\
&= E\left(v_{ii}^2v_{ij}^2-\frac{1}{N}v_{ii}^3v_{jj}\right)=0 \text{ for } i=k\neq j\\
&= E\left(v_{ii}^2v_{jk}^2-\frac{1}{N}v_{ii}^2v_{jj}v_{kk}\right)=0 \text{ for } i\neq j\neq k
\end{aligned}$$

Thus, $Cov(q_1,q_2)=0$.

Therefore, we have

$$Var(\hat{a}_2)=c^2\left[Var(q_1)+Var(q_2)+2Cov(q_1,q_2)\right].$$

Since, $c^2\simeq 1$ as well as $N\to\infty$, then

$$\begin{aligned}
Var(\hat{a}_2) &\simeq \frac{8a_4}{Np}+\frac{4}{N^2}\left(a_2^2-\frac{a_4}{p}\right)\\
&\simeq \frac{8}{Np}a_4+\frac{4}{N^2}a_2^2
\end{aligned}$$

□

2. We can prove in the same way as 1. □

3. Since $\hat{a}_{2D}=\sum_{i=1}^{m}\hat{a}_{2i}=\sum_{i=1}^{m}\frac{N^2}{(N-1)(N+2)p}\left\{trS_{ii}^2-\frac{1}{N}(trS_{ii})^2\right\}$, then

$$Var(\hat{a}_{2D})=Var\left(\sum_{i=1}^{m}\hat{a}_{2i}\right)=\sum_{i=1}^{m}Var(\hat{a}_{2i})+\sum_{i\neq j}Cov(\hat{a}_{2i},\hat{a}_{2j})$$

Since $\underset{\sim}{Y}^{(i)}$ and $\underset{\sim}{Y}^{(j)}$ are independent and $\hat{a}_{2i}$ and $\hat{a}_{2j}$ are the function of $\underset{\sim}{Y}^{(i)}$ and $\underset{\sim}{Y}^{(j)}$ respectively, then $Cov(\hat{a}_{2i},\hat{a}_{2j})=0,\ i\neq j$.

$$Var(\hat{a}_{2D})=\sum_{i=1}^{m}Var(\hat{a}_{2i})$$

$$\simeq\sum_{i=1}^{m}\left(\frac{8}{Np}a_{4i}+\frac{4}{N^2}a_{2i}^2\right)$$ □

3. Find the distribution of $\left(\hat{a}_2\quad \hat{a}_{2D}\right)^T$

Lemma A.6 Under the assumption (A1)-(A4), the asymptotically distribution of $\left(\hat{a}_2\quad \hat{a}_{2D}\right)^T$ is

$$\begin{pmatrix}\hat{a}_2\\ \hat{a}_{2D}\end{pmatrix}\xrightarrow{D}N_2\left(\begin{pmatrix}a_2\\ a_{2D}\end{pmatrix},\begin{pmatrix}\frac{8}{Np}a_4+\frac{4}{N^2}a_2^2 & Cov(\hat{a}_2,\hat{a}_{2D})\\ Cov(\hat{a}_2,\hat{a}_{2D}) & \sum_{i=1}^{m}\left(\frac{8}{Np}a_{4i}+\frac{4}{N^2}a_{2i}^2\right)\end{pmatrix}\right)$$

Proof

Recall, we have

$$\hat{a}_2=c(q_1+q_2)$$

$$\hat{a}_{2D}=c(r_1+r_2)$$

where $c=\frac{N^2}{(N-1)(N+2)}$, $q_1=\frac{N-1}{N^3p}\sum_{i=1}^{p}\lambda_i^2v_{ii}^2$, $q_2=\frac{2}{N^2p}\sum_{i<j}^{p}\lambda_i\lambda_j\phi_{ij}$, $r_1=\frac{N-1}{N^3p}\sum_{j=1}^{p}\omega_j^2w_{jj}^2$, and $r_2=\frac{2}{N^2p}\sum_{j<k}^{p}\omega_j\omega_k\eta_{jk}$. In this study we consider when $N\to\infty$, thus $c\simeq 1$, $\hat{a}_2\simeq q_1+q_2$, and $\hat{a}_{2D}\simeq r_1+r_2$.

Before finding distribution of $\left(\hat{a}_2\quad \hat{a}_{2D}\right)^T$. We need to find distribution of $\left(q_1\quad r_1\right)^T$, q_2, and r_2.

3.1 Find the distribution of $\left(q_1 \quad r_1\right)^T$

Lemma A.7 For q_1 and r_1 defined above**,** the asymptotic distribution of $\left(q_1 \quad r_1\right)^T$ is $\begin{pmatrix} q_1 \\ r_1 \end{pmatrix} \xrightarrow{D} N_2\left(\begin{pmatrix} a_2 \\ a_{2D} \end{pmatrix}, H\right)$ where H is a covariance matrix of $\left(q_1 \quad r_1\right)^T$.

Proof

Let

$$\delta_{1i} = \frac{\lambda_i^2 (v_{ii}^2 - N(N+2))}{\sqrt{N(N+2)(N+3)}}, \quad \delta_{2i} = \frac{\omega_i^2 (w_{ii}^2 - N(N+2))}{\sqrt{N(N+2)(N+3)}}.$$

From (A.7), $E(v_{ii}^2) = N(N+2)$ and (9), $Var(v_{ii}^2) = 8N(N+2)(N+3),$ then $E(\delta_{1i}) = 0,\ E(\delta_{2i}) = 0,\ Var(\delta_{1i}) = 8\lambda_i^4,\ Var(\delta_{2i}) = 8\omega_i^4,\ Cov(\delta_{1i}, \delta_{2i}) = \Xi_i.$

Thus $\underset{\sim}{\delta}_i = \begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix}$ are independently distributed random vectors, $i = 1, 2, ..., p,$ with mean vectors as zero vectors and the covariance matrices G_{iN} given by

$$G_{iN} = \begin{pmatrix} 8\lambda_i^4 & \Xi_i \\ \Xi_i & 8\omega_i^4 \end{pmatrix}, \ i = 1, 2, ..., p.$$

Now, as $p \to \infty$ then

$$G_N = \frac{1}{p}(G_{1N} + \cdots + G_{pN})$$

$$= \begin{pmatrix} 8a_4 & \Xi^* \\ \Xi^* & 8\sum_{i=1}^{m} a_{4i} \end{pmatrix} \to G_N^0 \neq 0, \text{ for any } N$$

and $G_N^0 \to \begin{pmatrix} 8a_4^0 & \Xi^0 \\ \Xi^0 & 8a_{4D}^0 \end{pmatrix} \equiv G^0$ as $N \to \infty$.

Also, if F_i is the distribution function of $\underset{\sim}{\delta}_i$ then

$$\frac{1}{p}\sum_{i=1}^{p}\int_{(\underset{\sim}{\delta}^T \underset{\sim}{\delta}) > p\varepsilon^2} \underset{\sim}{\delta}^T \underset{\sim}{\delta} dF_i \le \frac{1}{p}\sum_{i=1}^{p}(p\varepsilon^2)^{-1}\int (\underset{\sim}{\delta}^T \underset{\sim}{\delta}) dF_i$$

$$= \frac{1}{p^2\varepsilon^2}\sum_{i=1}^{p} E(\delta_{1i}^2 + \delta_{2i}^2)^2$$

$$\le \frac{1}{p^2\varepsilon^2}\sum_{i=1}^{p} E(\delta_{1i}^4 + \delta_{2i}^4).$$

From c_r-inequality, see Rao (1973, p. 149). Now

$$\frac{1}{p^2}\sum_{i=1}^{p} E(\delta_{1i}^4) = \frac{1}{p^2}\sum_{i=1}^{p} \lambda_i^8 \frac{E(v_{ii}^2 - N(N+2))^4}{(N(N+2)(N+3))^2}$$

$$= O(p^{-1}) \to 0 \text{ as } p \to \infty$$

and

$$\frac{1}{p^2}\sum_{i=1}^{p} E(\delta_{2i}^4) = \frac{1}{p^2}\sum_{i=1}^{p} \omega_i^8 \frac{E(w_{ii}^2 - N(N+2))^4}{(N(N+2)(N+3))^2}$$

$$= O(p^{-1}) \to 0 \text{ as } p \to \infty .$$

Then from the multivariate central limit theorem of Liapunov type given in Roa (1973, p. 147, Problem 4.7), it follows that as $p \to \infty$, and for any N,

$$\frac{1}{\sqrt{p}}\sum_{i=1}^{p} \underset{\sim}{\delta}_i = \frac{1}{\sqrt{(N(N+2)(N+3))p}} \begin{pmatrix} \sum_{i=1}^{p} \lambda_i^2 (v_{ii}^2 - N(N+2)) \\ \sum_{i=1}^{p} \omega_i^2 (w_{ii}^2 - N(N+2)) \end{pmatrix} \sim N_2(\underset{\sim}{0}, G_N^0).$$

Thus, it follows that as $p \to \infty$ and then $N \to \infty$,

$$\frac{1}{\sqrt{p}}\sum_{i=1}^{p} \underset{\sim}{\delta}_i \sim N_2(\underset{\sim}{0}, G^0).$$

On the other hand, as $N \to \infty$, we get from the multivariate central limit theorem that

$$\underset{\sim}{\delta}_i \sim N_2(\underset{\sim}{0}, G_i), \;\; i = 1, 2, ..., p.$$

For any p, where G_i is the limit of G_{iN} given by

$$G_i = \begin{pmatrix} 8\lambda_i^4 & \Xi^{**} \\ \Xi^{**} & 8\omega_i^4 \end{pmatrix}.$$

Let

$$G = \frac{1}{p}(G_1 + \cdots + G_p) = \begin{pmatrix} 8a_4 & \Xi^{***} \\ \Xi^{***} & 8\sum_{i=1}^{m} a_{4i} \end{pmatrix},$$ which go to G^0 as $p \to \infty$. Since $\underset{\sim}{\delta}_i$

are asymptotically independently distributed random vectors, it follows from the argument given above as $N \to \infty$ and then $p \to \infty$

$$\frac{1}{\sqrt{p}} \sum_{i=1}^{p} \underset{\sim}{\delta}_i \sim N_2(\underset{\sim}{0}, G^0).$$

Without any loss of generality, we may replace G^0 by G. Noting that

$$q_1 = \frac{N-1}{N^3 p} \sum_{i=1}^{p} \lambda_i^2 v_{ii}^2 \simeq \frac{1}{N^2 p} \sum_{i=1}^{p} \lambda_i^2 v_{ii}^2,$$

$$r_1 = \frac{N-1}{N^3 p} \sum_{j=1}^{p} \omega_j^2 w_{jj}^2 \simeq \frac{1}{N^2 p} \sum_{j=1}^{p} \omega_j^2 w_{jj}^2,$$

$$a_2 = \frac{tr\Sigma^2}{p} = \frac{\sum_{i=1}^{p} \lambda_i^2}{p} \text{ and } a_{2D} = \frac{trD_\Sigma^2}{p} = \frac{\sum_{i=1}^{p} \omega_i^2}{p}.$$

Consider

$$\frac{1}{\sqrt{p}} \sum_{i=1}^{p} \delta_{1i} = \frac{1}{\sqrt{p}} \frac{\sum_{i=1}^{p} \lambda_i^2 (v_{ii}^2 - N(N+2))}{\sqrt{N(N+2)(N+3)}}$$

$$= \frac{1}{\sqrt{p}} \left[\sum_{i=1}^{p} \frac{\lambda_i^2 v_{ii}^2}{\sqrt{N(N+2)(N+3)}} - \sum_{i=1}^{p} \frac{\lambda_i^2 N(N+2)}{\sqrt{N(N+2)(N+3)}} \right]$$

$$= \frac{1}{\sqrt{p}} \left[\frac{N^2 p q_1}{\sqrt{N(N+2)(N+3)}} - \frac{N(N+2) p a_2}{\sqrt{N(N+2)(N+3)}} \right]$$

Since, as $N \to \infty$,

$$\frac{1}{\sqrt{p}} \left[\frac{N^2 p q_1}{\sqrt{N(N+2)(N+3)}} - \frac{N(N+2) p a_2}{\sqrt{N(N+2)(N+3)}} \right] \approx \frac{1}{\sqrt{Np}} (Npq_1 - Npa_2)$$

$$= \sqrt{Np}(q_1 - a_2)$$

then $\sqrt{Np}(q_1 - a_2) \xrightarrow{D} N(0, 8a_4)$ and with a linear transformation, we have

$$q_1 \xrightarrow{D} N(a_2, \frac{8a_4}{Np}).$$

Next, we can fine the distribution of r_1 in same manner as q_1, thus we have

$$r_1 \xrightarrow{D} N(a_{2D}, \frac{8\sum_{i=1}^{p} a_{4i}}{Np}).$$

From these results, we can have the asymptotic distribution of $\begin{pmatrix} q_1 & r_1 \end{pmatrix}^T$ is

$$\begin{pmatrix} q_1 \\ r_1 \end{pmatrix} \xrightarrow{D} N_2\left(\begin{pmatrix} a_2 \\ a_{2D} \end{pmatrix}, H \right) \text{ where } H \text{ is a covariance matrix of } \begin{pmatrix} q_1 & r_1 \end{pmatrix}^T \quad \square$$

3.2 Find the distribution of q_2 and r_2

Lemma A.8 For q_2 and r_2 are defined above**,** the asymptotic distribution of q_2 and r_2 is

$$q_2 \xrightarrow{D} N(0, \frac{4}{N^2}(a_2^2 - p^{-1}a_4)),$$

$$r_2 \xrightarrow{D} N(0, \frac{4}{N^2}(\sum_{i=1}^{m} a_{2i}^2 - p^{-1}\sum_{i=1}^{m} a_{4i})).$$

Proof

We find the distribution of q_2 and r_2 by the Lindeberg Central Limit Theorem From Billingsley (1995, p.359).

Theorem A.1 Lindeberg Central Limit Theorem From Billingsley (1995, p.359)

Let $x_1, ..., x_n$ be a sequence of independent random variables which satisfies

1) $E(x_i) = 0$
2) $\sigma^2 = E(x_i^2)$

Let $S_n^2 = \sum_{i=1}^{n} \sigma_i^2 > 0$ and P_i be the distribution function of x_i

If $\sum_{i=1}^{n} \frac{1}{S_n^2} \int_{|x_i| \geq \varepsilon S_n} x_i^2 dP_i \to 0,$ for $\varepsilon > 0,$ then

$$\frac{\sum_{i=1}^{n} x_i}{S_n} \xrightarrow{D} N(0,1).$$

Since $N^{-1/2} v_{ij} \sim N(0,1)$ as $N \to \infty,$ it follows that $N^{-1} v_{ij}^2 \sim \chi^2(1),$ which asymptotically independently distributed for all distinct i and j.

Let $\kappa_{ij} = \frac{2\lambda_i \lambda_j \phi_{ij}}{N^2 p},$ we have $E(\kappa_{ij}) = 0$ and let

$$S_p^2 = \sum_{i<j}^{p} Var(\kappa_{ij}) = Var(q_2) \simeq \frac{4}{N^2}(a_2^2 - p^{-1}a_4), \text{ as } p, N \to \infty$$

and $\sum_{i<j}^{p} \kappa_{ij} = q_2$. If P_{ij} is the distribution function of κ_{ij} , for $\varepsilon > 0,$ then

$$\begin{aligned}
\sum_{i<j}^{p} \frac{1}{S_p^2} \int_{|\kappa_{ij}| \geq \varepsilon S_p} \kappa_{ij}^2 dP_{ij} &< \sum_{i<j}^{p} \frac{1}{\varepsilon^2 S_p^2} \int \kappa_{ij}^2 dP_{ij} \\
&= \sum_{i<j}^{p} \frac{1}{\varepsilon^2 S_p^2} E(\kappa_{ij}^2) \\
&= \sum_{i<j}^{p} \frac{1}{\varepsilon^2 S_p^2} E\left(\frac{2\lambda_i \lambda_j \phi_{ij}}{N^2 p} \right)^2 \\
&= \sum_{i<j}^{p} \frac{4\lambda_i^2 \lambda_j^2}{N^4 p^2 \varepsilon^2 S_p^2} E(\phi_{ij}^2),
\end{aligned}$$

$E(\phi_{ij}^2)$ is given in Srivastava (2005)

$$= \sum_{i<j}^{p} \frac{4\lambda_i^2 \lambda_j^2}{N^4 p^2 \varepsilon^2 S_p^2} 2(N+2)(N-1)$$

$$\sum_{i<j}^{p} \frac{1}{S_p^2} \int_{|\kappa_{ij}| \geq \varepsilon S_p} \kappa_{ij}^2 dP_{ij} \approx \sum_{i<j}^{p} \frac{4\lambda_i^2 \lambda_j^2}{N^2 p^2 \varepsilon^2 S_p^2} \to 0, \text{ as } p \to \infty.$$

Therefore, we can apply the Lindeberg Central Limit Theorem i.e.,

$$\frac{q_2}{\sqrt{\frac{4}{N^2}(a_2^2 - p^{-1}a_4)}} \xrightarrow{D} N(0,1),$$

and with a linear transformation, we have $q_2 \xrightarrow{D} N(0, \frac{4}{N^2}(a_2^2 - p^{-1}a_4))$. □

Similarly, we can fine the distribution of r_2 in same manner as q_2, thus we have $r_2 \xrightarrow{D} N(0, \frac{4}{N^2}(\sum_{i=1}^{m} a_{2i}^2 - p^{-1}\sum_{i=1}^{m} a_{4i}))$. □

Consider $\hat{a}_2$ is a linear function of random variable q_1 and q_2 i.e. $\hat{a}_2 \simeq q_1 + q_2$ as $N \to \infty$. Now, we obtain the asymptotically distribution of $\hat{a}_2$ from these results as follows

$$\hat{a}_2 \xrightarrow{D} N\left(a_2, \frac{8}{Np} a_4 + \frac{4}{N^2} a_2^2 \right).$$

Similarly with $\hat{a}_2$, we obtain the asymptotically distribution of $\hat{a}_{2D}$ as follows

$$\hat{a}_{2D} \xrightarrow{D} N\left(a_{2D}, \sum_{i=1}^{m} (\frac{8}{Np} a_{4i} + \frac{4}{N^2} a_{2i}^2) \right).$$

Therefore, we obtain the asymptotically distribution of $\left(\hat{a}_2 \quad \hat{a}_{2D}\right)^T$ is

$$\begin{pmatrix} \hat{a}_2 \\ \hat{a}_{2D} \end{pmatrix} \xrightarrow{D} N_2 \left(\begin{pmatrix} a_2 \\ a_{2D} \end{pmatrix}, \begin{pmatrix} \frac{8}{Np} a_4 + \frac{4}{N^2} a_2^2 & Cov(\hat{a}_2, \hat{a}_{2D}) \\ Cov(\hat{a}_2, \hat{a}_{2D}) & \sum_{i=1}^{m} (\frac{8}{Np} a_{4i} + \frac{4}{N^2} a_{2i}^2) \end{pmatrix} \right)$$ □

A.2 Fisher Discriminant Method

Theorem A.2. Let $\lambda_1, \lambda_2, ..., \lambda_s$ denote the $s < \min(g-1, p)$ eigenvalues of $S_w^{-1} S_b$ and $\underset{\sim}{e}_1, \underset{\sim}{e}_2, ..., \underset{\sim}{e}_s$ be the corresponding eigenvectors. Then the vector of coefficients $\underset{\sim}{c}$ that maximizes Fisher's criterion function, $F(\underset{\sim}{c})$;

$$F(\underset{\sim}{c}) = \frac{\underset{\sim}{c}^T S_b \underset{\sim}{c}}{\underset{\sim}{c}^T S_w \underset{\sim}{c}} = \frac{\underset{\sim}{c}^T \sum_{h=1}^{g} n_h (\underset{\sim}{\overline{x}}_h - \underset{\sim}{\overline{x}})(\underset{\sim}{\overline{x}}_h - \underset{\sim}{\overline{x}})^T \underset{\sim}{c}}{\underset{\sim}{c}^T \sum_{h=1}^{g} \sum_{k=1}^{n_h} (\underset{\sim}{\overline{x}}_{kh} - \underset{\sim}{\overline{x}}_h)(\underset{\sim}{\overline{x}}_{kh} - \underset{\sim}{\overline{x}}_h)^T \underset{\sim}{c}}$$

is given by $\underset{\sim}{c}_1 = \underset{\sim}{e}_1$ The linear combination $\underset{\sim}{c}_1^T \underset{\sim}{x}$ is called the sample first discriminant. The choice $\underset{\sim}{c}_2 = \underset{\sim}{e}_2$ produces the sample second discriminant, $\underset{\sim}{c}_2^T \underset{\sim}{x}$, and continuing, we obtain $\underset{\sim}{c}_r^T \underset{\sim}{x} = \underset{\sim}{e}_r^T \underset{\sim}{x}$, the sample r^{th} discriminant, $r \le s$.

Proof. Let $\underset{\sim}{u} = \Sigma^{\frac{1}{2}} \underset{\sim}{c}$, so $\underset{\sim}{u}^T \underset{\sim}{u} = \underset{\sim}{c}^T \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}} \underset{\sim}{c} = \underset{\sim}{c}^T \Sigma \underset{\sim}{c}$ and

$$\underset{\sim}{u}^T \Sigma^{-\frac{1}{2}} B_\mu \Sigma^{-\frac{1}{2}} \underset{\sim}{u} = \underset{\sim}{c}^T \Sigma^{\frac{1}{2}} \Sigma^{-\frac{1}{2}} B_\mu \Sigma^{-\frac{1}{2}} \Sigma^{\frac{1}{2}} \underset{\sim}{c} = \underset{\sim}{c}^T B_\mu \underset{\sim}{c}$$

Consequently, the maximization of the ratio $\dfrac{\underset{\sim}{c}^T B_\mu \underset{\sim}{c}}{\underset{\sim}{c}^T \Sigma \underset{\sim}{c}}$ and the ratio $\dfrac{\underset{\sim}{u}^T \Sigma^{-\frac{1}{2}} B_\mu \Sigma^{-\frac{1}{2}} \underset{\sim}{u}}{\underset{\sim}{u}^T \underset{\sim}{u}}$ are equivalently in terms of solving the vector of coefficients $\underset{\sim}{c}$ and $\underset{\sim}{u}$.

From Johnson, A. R. and Wichen, D. W. (2002) , if the vector $\underset{\sim}{x}$ satisfying $\max\limits_{\underset{\sim}{x} \ne 0} \dfrac{\underset{\sim}{x}^T B \underset{\sim}{x}}{\underset{\sim}{x}^T \underset{\sim}{x}} = \lambda_1$ is attained when $\underset{\sim}{x} = \underset{\sim}{e}_1$ and $\min\limits_{\underset{\sim}{x} \ne 0} \dfrac{\underset{\sim}{x}^T B \underset{\sim}{x}}{\underset{\sim}{x}^T \underset{\sim}{x}} = \lambda_p$ is attained when $\underset{\sim}{x} = \underset{\sim}{e}_p$. Moreover, $\max\limits_{\underset{\sim}{x} \perp \underset{\sim}{e}_1, \dots, \underset{\sim}{e}_k} \dfrac{\underset{\sim}{x}^T B \underset{\sim}{x}}{\underset{\sim}{x}^T \underset{\sim}{x}} = \lambda_{k+1}$ is attained when $\underset{\sim}{x} = \underset{\sim}{e}_{k+1}$, $k = 1, \dots, p-1$, where B is a $p \times p$ positive definite matrix with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$ associated normalized eigenvectors $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \dots, \underset{\sim}{e}_p$ and the symbol $\perp$ is "perpendicular to" such that $\underset{\sim}{e}_i^T \underset{\sim}{e}_j = 0,\ i \ne j$. The maximum of the ratio $\dfrac{\underset{\sim}{u}^T \Sigma^{-\frac{1}{2}} B_\mu \Sigma^{-\frac{1}{2}} \underset{\sim}{u}}{\underset{\sim}{u}^T \underset{\sim}{u}}$ equals to λ_1 which is the largest eigenvalue of $\Sigma^{-\frac{1}{2}} B_\mu \Sigma^{-\frac{1}{2}}$. This maximum occurs when $\underset{\sim}{u} = \underset{\sim}{e}_1$, the normalized eigenvector associated with λ_1. Since $\underset{\sim}{e}_1 = \underset{\sim}{u} = \Sigma^{\frac{1}{2}} \underset{\sim}{c}_1$ or $\underset{\sim}{c}_1 = \Sigma^{-\frac{1}{2}} \underset{\sim}{e}_1$, then $Var(\underset{\sim}{c}_1^T \underset{\sim}{x}) = \underset{\sim}{c}_1^T \Sigma \underset{\sim}{c}_1 = \underset{\sim}{e}_1^T \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} \underset{\sim}{e}_1 = \underset{\sim}{e}_1^T \underset{\sim}{e}_1 = 1$ and $\underset{\sim}{u} \perp \underset{\sim}{e}_1$ maximizes the proceeding ratio when $\underset{\sim}{u} = \underset{\sim}{e}_2$, the normalized eigenvector associated with λ_2. For this choice, $\underset{\sim}{c}_2 = \Sigma^{-\frac{1}{2}} \underset{\sim}{e}_2$ and $Cov(\underset{\sim}{c}_1^T \underset{\sim}{x}, \underset{\sim}{c}_2^T \underset{\sim}{x}) = \underset{\sim}{c}_2^T \Sigma \underset{\sim}{c}_1 = \underset{\sim}{e}_2^T \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} \underset{\sim}{e}_1 = \underset{\sim}{e}_2^T \underset{\sim}{e}_1 = 0$ since $\underset{\sim}{e}_1 \perp \underset{\sim}{e}_2$. Similarly, $Var(\underset{\sim}{c}_2^T \underset{\sim}{x}) = \underset{\sim}{c}_2^T \Sigma \underset{\sim}{c}_2 = \underset{\sim}{e}_2^T \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} \underset{\sim}{e}_2 = \underset{\sim}{e}_2^T \underset{\sim}{e}_2 = 1$. Continue in this fashion for the remaining discriminants. Note that if λ and $\underset{\sim}{e}$ are an eigenvalue-eigenvector pair of $\Sigma^{-\frac{1}{2}}B_{\mu}\Sigma^{-\frac{1}{2}}$, then

$$\Sigma^{-\frac{1}{2}}B_{\mu}\Sigma^{-\frac{1}{2}}\underset{\sim}{e} = \lambda\underset{\sim}{e}$$

and multiplication on the left by $\Sigma^{-\frac{1}{2}}$ gives

$$\Sigma^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}B_{\mu}\Sigma^{-\frac{1}{2}}\underset{\sim}{e} = \lambda\Sigma^{-\frac{1}{2}}\underset{\sim}{e} \text{ or } \Sigma^{-1}B_{\mu}(\Sigma^{-\frac{1}{2}}\underset{\sim}{e}) = \lambda(\Sigma^{-\frac{1}{2}}\underset{\sim}{e})$$

Thus, $\Sigma^{-1}B_{\mu}$ has the same eigenvalues as $\Sigma^{-\frac{1}{2}}B_{\mu}\Sigma^{-\frac{1}{2}}$, but corresponding eigenvector is $\Sigma^{-\frac{1}{2}}\underset{\sim}{e} = \underset{\sim}{c}$. In sample counterpart, the eigenvectors of $S_w^{-1}S_b$ maximize $F(\underset{\sim}{c})$.

Appendix B

Tables

Table of Test Block Diagonal Covariance Matrix

Table B.1 The empirical type I error rate when $\Sigma = \Sigma_3$, $\theta = 0.1$, 0.5, and 0.9 with equal block sizes

n	p	p_i	$\theta = 0.1$			$\theta = 0.5$			$\theta = 0.9$		
			$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	0.0470	0.0456	0.0508	0.0528	0.0519	0.0508	0.0554	0.0536	0.0508
		10	0.0495	0.0483	0.0475	0.0581	0.0575	0.0475	0.0588	0.0564	0.0475
		25	0.0515	0.0492	0.0484	0.0686	0.0673	0.0484	0.0697	0.0682	0.0484
	200	5	0.0465	0.0450	0.0484	0.0529	0.0515	0.0484	0.0523	0.0510	0.0484
		10	0.0489	0.0472	0.0481	0.0561	0.0544	0.0481	0.0566	0.0551	0.0481
		25	0.0466	0.0456	0.0495	0.0605	0.0596	0.0495	0.0612	0.0594	0.0495
	300	5	0.0482	0.0470	0.0535	0.0468	0.0448	0.0535	0.0507	0.0498	0.0535
		10	0.0488	0.0469	0.0522	0.0517	0.0502	0.0522	0.0519	0.0509	0.0522
		25	0.0512	0.0493	0.0483	0.0586	0.0569	0.0483	0.0604	0.0585	0.0483
	400	5	0.0493	0.0476	0.0487	0.0509	0.0498	0.0487	0.0514	0.0498	0.0487
		10	0.0493	0.0478	0.0494	0.0494	0.0483	0.0494	0.0498	0.0486	0.0494
		25	0.0498	0.0480	0.0536	0.0529	0.0518	0.0536	0.0527	0.0508	0.0536

Table B.1 (Continued)

n	p	p_i	$\theta = 0.1$			$\theta = 0.5$			$\theta = 0.9$		
			$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
100	100	5	0.0532	0.0523	0.0514	0.0543	0.0533	0.0514	0.0565	0.0553	0.0514
		10	0.0520	0.0513	0.0502	0.0593	0.0588	0.0502	0.0623	0.0617	0.0502
		25	0.0562	0.0554	0.0505	0.0637	0.0629	0.0505	0.0646	0.0641	0.0505
	200	5	0.0519	0.0507	0.0510	0.0509	0.0506	0.0510	0.0540	0.0529	0.0510
		10	0.0535	0.0526	0.0507	0.0579	0.0572	0.0507	0.0583	0.0580	0.0507
		25	0.0506	0.0499	0.0542	0.0587	0.0583	0.0542	0.0613	0.0605	0.0542
	300	5	0.0444	0.0440	0.0458	0.0506	0.0502	0.0458	0.0533	0.0523	0.0458
		10	0.0439	0.0433	0.0471	0.0557	0.0552	0.0471	0.0569	0.0562	0.0471
		25	0.0512	0.0500	0.0486	0.0579	0.0572	0.0486	0.0611	0.0599	0.0486
	400	5	0.0474	0.0461	0.0516	0.0499	0.0489	0.0516	0.0499	0.0488	0.0516
		10	0.0454	0.0446	0.0508	0.0487	0.0478	0.0508	0.0516	0.0511	0.0508
		25	0.0475	0.0466	0.0522	0.0553	0.0547	0.0522	0.0565	0.0556	0.0522

Table B.2 The empirical type I error rate when $\Sigma = \Sigma_3$, $\theta = 0.1$, 0.5, and 0.9 with mixed block sizes

n	p	p_i	p_j	$\theta = 0.1$			$\theta = 0.5$			$\theta = 0.9$		
				$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	10	0.0544	0.0530	0.0480	0.0584	0.0571	0.0480	0.0580	0.0571	0.0480
		5	25	0.0517	0.0499	0.0459	0.0616	0.0602	0.0459	0.0640	0.0635	0.0459
		10	25	0.0536	0.0518	0.0505	0.0597	0.0586	0.0505	0.0600	0.0587	0.0505
	200	5	10	0.0515	0.0492	0.0527	0.0519	0.0505	0.0527	0.0527	0.0513	0.0527
		5	25	0.0498	0.0484	0.0515	0.0589	0.0574	0.0515	0.0618	0.0603	0.0515
		10	25	0.0526	0.0509	0.0484	0.0621	0.0606	0.0484	0.0638	0.0620	0.0484
	300	5	10	0.0529	0.0513	0.0530	0.0542	0.0521	0.0530	0.0554	0.0540	0.0530
		5	25	0.0519	0.0496	0.0507	0.0558	0.0549	0.0507	0.0556	0.0547	0.0507
		10	25	0.0519	0.0500	0.0491	0.0552	0.0538	0.0491	0.0545	0.0531	0.0491
	400	5	10	0.0490	0.0472	0.0490	0.0484	0.0471	0.0490	0.0529	0.0515	0.0490
		5	25	0.0501	0.0490	0.0503	0.0512	0.0507	0.0503	0.0513	0.0509	0.0503
		10	25	0.0489	0.0474	0.0530	0.0534	0.0520	0.0530	0.0541	0.0529	0.0530
100	100	5	10	0.0547	0.0541	0.0525	0.0605	0.0600	0.0525	0.0627	0.0616	0.0525
		5	25	0.0506	0.0500	0.0467	0.0639	0.0631	0.0467	0.0643	0.0641	0.0467
		10	25	0.0533	0.0526	0.0518	0.0597	0.0590	0.0518	0.0615	0.0610	0.0518
	200	5	10	0.0512	0.0500	0.0511	0.0556	0.0547	0.0511	0.0564	0.0556	0.0511
		5	25	0.0534	0.0532	0.0531	0.0606	0.0602	0.0531	0.0649	0.0641	0.0531
		10	25	0.0535	0.0525	0.0540	0.0632	0.0623	0.0540	0.0631	0.0620	0.0540
	300	5	10	0.0492	0.0482	0.0489	0.0533	0.0524	0.0489	0.0574	0.0567	0.0489
		5	25	0.0509	0.0505	0.0493	0.0605	0.0599	0.0493	0.0616	0.0610	0.0493
		10	25	0.0498	0.0493	0.0505	0.0565	0.0557	0.0505	0.0581	0.0575	0.0505
	400	5	10	0.0459	0.0453	0.0483	0.0484	0.0478	0.0483	0.0503	0.0498	0.0483
		5	25	0.0451	0.0447	0.0470	0.0536	0.0530	0.0470	0.0541	0.0532	0.0470
		10	25	0.0457	0.0450	0.0478	0.0525	0.0519	0.0478	0.0543	0.0535	0.0478

Table B.3 The empirical type I error rate when $\Sigma = \Sigma_4$ and $\theta = 0.9$ with equal block sizes

$\boldsymbol{n}$	$\boldsymbol{p}$	$\boldsymbol{p_i}$	$\hat{\boldsymbol{T}}_{\boldsymbol{p}}$	$\hat{\boldsymbol{T}}_{\boldsymbol{b}}$	$\hat{\boldsymbol{T}}_{\boldsymbol{c}}$
50	100	5	0.0550	0.0539	0.0500
		10	0.0580	0.0568	0.0475
		25	0.0647	0.0632	0.0484
	200	5	0.0529	0.0514	0.0484
		10	0.0562	0.0549	0.0481
		25	0.0593	0.0581	0.0495
	300	5	0.0491	0.0476	0.0535
		10	0.0496	0.0485	0.0522
		25	0.0533	0.0515	0.0483
	400	5	0.0518	0.0500	0.0487
		10	0.0492	0.0476	0.0494
		25	0.0522	0.0511	0.0536
100	100	5	0.0560	0.0552	0.0514
		10	0.0611	0.0602	0.0502
		25	0.0634	0.0626	0.0505
	200	5	0.0519	0.0511	0.0510
		10	0.0552	0.0542	0.0503
		25	0.0582	0.0574	0.0542
	300	5	0.0520	0.0512	0.0458
		10	0.0521	0.0517	0.0471
		25	0.0562	0.0555	0.0486
	400	5	0.0507	0.0492	0.0516
		10	0.0514	0.0508	0.0508
		25	0.0509	0.0496	0.0522

Table B.4 The empirical type I error rate when $\Sigma = \Sigma_4$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	10	0.0539	0.0534	0.0480
		5	25	0.0633	0.0618	0.0459
		10	25	0.0589	0.0568	0.0500
	200	5	10	0.0538	0.0519	0.0479
		5	25	0.0552	0.0539	0.0473
		10	25	0.0563	0.0549	0.0501
	300	5	10	0.0556	0.0534	0.0530
		5	25	0.0542	0.0526	0.0507
		10	25	0.0530	0.0515	0.0491
	400	5	10	0.0524	0.0511	0.0490
		5	25	0.0537	0.0525	0.0503
		10	25	0.0512	0.0505	0.0530
100	100	5	10	0.0553	0.0545	0.0525
		5	25	0.0609	0.0598	0.0467
		10	25	0.0625	0.0616	0.0518
	200	5	10	0.0528	0.0528	0.0525
		5	25	0.0588	0.0578	0.0528
		10	25	0.0577	0.0571	0.0507
	300	5	10	0.0563	0.0550	0.0489
		5	25	0.0545	0.0535	0.0493
		10	25	0.0545	0.0537	0.0505
	400	5	10	0.0471	0.0463	0.0483
		5	25	0.0524	0.0519	0.0470
		10	25	0.0495	0.0489	0.0478

Table B.5 The empirical power when $\Sigma = \Sigma_7$, $\theta = 0.1$, 0.5, and 0.9 with equal block sizes

n	p	p_i	$\theta = 0.1$			$\theta = 0.5$			$\theta = 0.9$		
			$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	0.9504	0.9494	0.7432	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000
		10	0.9349	0.9339	0.3652	1.0000	1.0000	0.9322	1.0000	1.0000	0.9733
		25	0.8640	0.8623	0.0756	0.9999	0.9999	0.1029	0.9999	0.9999	0.1078
	200	5	0.9978	0.9977	0.9671	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	0.9965	0.9963	0.7436	1.0000	1.0000	0.9988	1.0000	1.0000	0.9998
		25	0.9907	0.9905	0.1216	1.0000	1.0000	0.2030	1.0000	1.0000	0.2208
	300	5	0.9998	0.9998	0.9953	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	0.9997	0.9997	0.9188	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000
		25	0.9993	0.9993	0.1817	1.0000	1.0000	0.3466	1.0000	1.0000	0.3791
	400	5	1.0000	1.0000	0.9991	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9755	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.2592	1.0000	1.0000	0.5195	1.0000	1.0000	0.5638
100	100	5	0.9996	0.9995	0.9910	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	0.9991	0.9991	0.8257	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	0.9958	0.9957	0.1759	1.0000	1.0000	0.3207	1.0000	1.0000	0.3533
	200	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9952	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.4385	1.0000	1.0000	0.7968	1.0000	1.0000	0.8396
	300	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.7081	1.0000	1.0000	0.9729	1.0000	1.0000	0.9833
	400	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.8837	1.0000	1.0000	0.9966	1.0000	1.0000	0.9980

Table B.6 The empirical power when $\Sigma = \Sigma_7$, $\theta = 0.1$, 0.5, and 0.9 with mixed block sizes

n	p	p_i	p_j	$\theta = 0.1$			$\theta = 0.5$			$\theta = 0.9$		
				$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	10	0.9465	0.9455	0.5743	1.0000	1.0000	0.9992	1.0000	1.0000	1.0000
		5	25	0.9167	0.9153	0.3737	1.0000	1.0000	0.9918	1.0000	1.0000	0.9989
		10	25	0.9117	0.9104	0.1762	1.0000	1.0000	0.5608	1.0000	1.0000	0.6584
	200	5	10	0.9965	0.9964	0.9042	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	0.9948	0.9948	0.7397	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	0.9935	0.9933	0.4061	1.0000	1.0000	0.9400	1.0000	1.0000	0.9709
	300	5	10	0.9997	0.9997	0.9811	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	0.9999	0.9999	0.9163	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	0.9995	0.9995	0.6584	1.0000	1.0000	0.9937	1.0000	1.0000	0.9982
	400	5	10	1.0000	1.0000	0.9958	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	0.9999	0.9999	0.9740	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	0.9999	0.9999	0.8153	1.0000	1.0000	0.9991	1.0000	1.0000	0.9998
100	100	5	10	0.9995	0.9995	0.9599	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	0.9988	0.9988	0.8133	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	0.9981	0.9981	0.5059	1.0000	1.0000	0.9817	1.0000	1.0000	0.9943
	200	5	10	1.0000	1.0000	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	0.9935	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9180	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	300	5	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9932	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	400	5	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9992	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table B.7 The empirical powers when $\Sigma = \Sigma_8$ and $\theta = 0.9$ with equal block sizes

n	p	p_i	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9491
		25	0.9824	0.9821	0.1693
	200	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9543
		25	0.9967	0.9966	0.1793
	300	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9583
		25	0.9986	0.9986	0.1867
	400	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	0.9606
		25	0.9993	0.9991	0.1929
100	100	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.6655
	200	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.7280
	300	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.7402
	400	5	1.0000	1.0000	1.0000
		10	1.0000	1.0000	1.0000
		25	1.0000	1.0000	0.7470

Table B.8 The empirical powers when $\Sigma = \Sigma_8$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	$\hat{T}_p$	$\hat{T}_b$	$\hat{T}_c$
50	100	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	0.9999
		10	25	1.0000	1.0000	0.6239
	200	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.7439
	300	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.7570
	400	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.8440
100	100	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9985
	200	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9998
	300	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9999
	400	5	10	1.0000	1.0000	1.0000
		5	25	1.0000	1.0000	1.0000
		10	25	1.0000	1.0000	0.9999

Table of discriminant analysis

Table B.9 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.1$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.6905	0.6867	0.6860	0.6859
		10	0.6895	0.6786	0.6835	0.6850
		25	0.6833	0.6451	0.6790	0.6843
	200	5	0.7643	0.7595	0.7607	0.7630
		10	0.7647	0.7507	0.7602	0.7601
		25	0.7592	0.6946	0.7533	0.7604
	300	5	0.8198	0.8127	0.8156	0.8153
		10	0.8185	0.8001	0.8147	0.8145
		25	0.8110	0.7484	0.8072	0.8124
	400	5	0.8537	0.8436	0.8456	0.8515
		10	0.8488	0.8329	0.8423	0.8484
		25	0.8376	0.7694	0.8332	0.8417
70	100	5	0.7409	0.7413	0.7380	0.7295
		10	0.7378	0.7367	0.7366	0.7319
		25	0.7351	0.7228	0.7315	0.7340
	200	5	0.8182	0.8140	0.8147	0.8055
		10	0.8142	0.8109	0.8121	0.8083
		25	0.8079	0.7917	0.8055	0.8100
	300	5	0.8645	0.8641	0.8630	0.8593
		10	0.8628	0.8617	0.8606	0.8581
		25	0.8530	0.8465	0.8524	0.8594
	400	5	0.9004	0.8991	0.8979	0.8957
		10	0.8990	0.8963	0.8964	0.8949
		25	0.8908	0.8784	0.8865	0.8939

Table B.10 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.1$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6944	0.6916	0.6834	0.6880
		5	25	0.6875	0.6726	0.6849	0.6847
		10	25	0.6884	0.6713	0.6834	0.6905
	200	5	10	0.7733	0.7676	0.7688	0.7691
		5	25	0.7673	0.7396	0.7635	0.7687
		10	25	0.7594	0.7257	0.7555	0.7605
	300	5	10	0.8142	0.8173	0.8130	0.8144
		5	25	0.8094	0.7807	0.8055	0.8112
		10	25	0.8149	0.7852	0.8084	0.8168
	400	5	10	0.8495	0.8391	0.8495	0.8424
		5	25	0.8429	0.8162	0.8429	0.8362
		10	25	0.8445	0.8048	0.8445	0.8394
70	100	5	10	0.7325	0.7352	0.7291	0.7228
		5	25	0.7367	0.7385	0.7379	0.7372
		10	25	0.7318	0.7299	0.7306	0.7298
	200	5	10	0.8164	0.8168	0.8126	0.8105
		5	25	0.8173	0.8107	0.8133	0.8134
		10	25	0.8103	0.8124	0.8102	0.8105
	300	5	10	0.8681	0.8712	0.8639	0.8652
		5	25	0.8626	0.8620	0.8621	0.8645
		10	25	0.8640	0.8606	0.8622	0.8635
	400	5	10	0.9055	0.9067	0.9027	0.9020
		5	25	0.8960	0.8966	0.8933	0.8942
		10	25	0.8984	0.8963	0.8946	0.8986

Table B.11 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.1$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.6903	0.6842	0.6857	0.6810
		10	0.6875	0.6761	0.6822	0.6824
		25	0.6844	0.6513	0.6814	0.6847
	200	5	0.7600	0.7482	0.7525	0.7567
		10	0.7543	0.7370	0.7474	0.7569
		25	0.7441	0.6950	0.7400	0.7488
	300	5	0.8197	0.8142	0.8157	0.8186
		10	0.8178	0.7977	0.8128	0.8139
		25	0.8065	0.7439	0.8055	0.8113
	400	5	0.8476	0.8397	0.8428	0.8455
		10	0.8398	0.8299	0.8408	0.8409
		25	0.8319	0.7722	0.8282	0.8350
70	100	5	0.7349	0.7350	0.7337	0.7246
		10	0.7334	0.7312	0.7308	0.7263
		25	0.7269	0.7186	0.7253	0.7293
	200	5	0.8173	0.8166	0.8154	0.8066
		10	0.8128	0.8146	0.8116	0.8057
		25	0.8017	0.7966	0.7990	0.8093
	300	5	0.8665	0.8704	0.8652	0.8638
		10	0.8645	0.8645	0.8629	0.8625
		25	0.8547	0.8480	0.8518	0.8607
	400	5	0.9030	0.9050	0.9005	0.8979
		10	0.9009	0.9039	0.8991	0.8965
		25	0.8932	0.8842	0.8915	0.8953

Table B.12 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.1$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6902	0.6826	0.6863	0.6802
		5	25	0.6892	0.6664	0.6866	0.6895
		10	25	0.6879	0.6638	0.6815	0.6832
	200	5	10	0.7659	0.7549	0.7573	0.7605
		5	25	0.7546	0.7294	0.7511	0.7547
		10	25	0.7515	0.7244	0.7460	0.7519
	300	5	10	0.8085	0.8065	0.8028	0.8092
		5	25	0.8082	0.7828	0.8029	0.8084
		10	25	0.8052	0.7754	0.8024	0.8080
	400	5	10	0.8494	0.8445	0.8494	0.8465
		5	25	0.8464	0.8088	0.8464	0.8442
		10	25	0.8402	0.8095	0.8402	0.8343
70	100	5	10	0.7404	0.7431	0.7384	0.7303
		5	25	0.7353	0.7329	0.7318	0.7292
		10	25	0.7372	0.7357	0.7349	0.7334
	200	5	10	0.8141	0.8187	0.8134	0.8092
		5	25	0.8141	0.8145	0.8129	0.8157
		10	25	0.8096	0.8095	0.8069	0.8102
	300	5	10	0.8670	0.8722	0.8644	0.8629
		5	25	0.8593	0.8592	0.8543	0.8600
		10	25	0.8601	0.8590	0.8579	0.8629
	400	5	10	0.9061	0.9078	0.9042	0.9010
		5	25	0.8935	0.8935	0.8910	0.8930
		10	25	0.8959	0.8978	0.8953	0.8957

Table B.13 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.5$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.6607	0.7535	0.6601	0.6939
		10	0.6397	0.7602	0.6365	0.7073
		25	0.6101	0.7195	0.6088	0.7411
	200	5	0.7184	0.8447	0.7147	0.7434
		10	0.6930	0.8453	0.6890	0.7472
		25	0.6407	0.7880	0.6359	0.7691
	300	5	0.7628	0.8944	0.7602	0.7800
		10	0.7263	0.8971	0.7256	0.7748
		25	0.6735	0.8494	0.6729	0.7831
	400	5	0.7988	0.9259	0.7906	0.8149
		10	0.7572	0.9274	0.7523	0.7950
		25	0.6848	0.8817	0.6858	0.7810
70	100	5	0.7108	0.8110	0.7098	0.7672
		10	0.6853	0.8230	0.6840	0.7916
		25	0.6438	0.8094	0.6432	0.8157
	200	5	0.7768	0.8887	0.7756	0.8178
		10	0.7435	0.8963	0.7416	0.8356
		25	0.6929	0.8858	0.6938	0.8698
	300	5	0.8193	0.9351	0.8200	0.8539
		10	0.7805	0.9434	0.7796	0.8660
		25	0.7222	0.9355	0.7200	0.8920
	400	5	0.8569	0.9617	0.8562	0.8851
		10	0.8178	0.9663	0.8171	0.8850
		25	0.7444	0.9591	0.7423	0.9038

Table B.14 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.5$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6443	0.7753	0.6402	0.7048
		5	25	0.6247	0.7494	0.6251	0.7193
		10	25	0.6164	0.7533	0.6161	0.7254
	200	5	10	0.7070	0.8584	0.7025	0.7506
		5	25	0.6723	0.8272	0.6690	0.7648
		10	25	0.6482	0.8246	0.6485	0.7522
	300	5	10	0.7378	0.9072	0.7380	0.7738
		5	25	0.7006	0.8778	0.6964	0.7815
		10	25	0.6954	0.8808	0.6937	0.7794
	400	5	10	0.7731	0.9367	0.7705	0.8031
		5	25	0.7292	0.9143	0.7243	0.7959
		10	25	0.7127	0.9106	0.7120	0.7924
70	100	5	10	0.6931	0.8161	0.6923	0.7817
		5	25	0.6727	0.8250	0.6712	0.8006
		10	25	0.6642	0.8167	0.6631	0.8045
	200	5	10	0.7580	0.9041	0.7559	0.8342
		5	25	0.7282	0.9027	0.7267	0.8581
		10	25	0.7131	0.9020	0.7127	0.8618
	300	5	10	0.8023	0.9506	0.7994	0.8694
		5	25	0.7633	0.9430	0.7617	0.8846
		10	25	0.7467	0.9415	0.7444	0.8856
	400	5	10	0.8410	0.9715	0.8389	0.8936
		5	25	0.7900	0.9680	0.7878	0.8993
		10	25	0.7850	0.9677	0.7831	0.9018

Table B.15 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.5$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.6547	0.7553	0.6524	0.6871
		10	0.6360	0.7606	0.6311	0.7108
		25	0.6058	0.7224	0.6025	0.7402
	200	5	0.7103	0.8330	0.7031	0.7372
		10	0.6778	0.8354	0.6737	0.7357
		25	0.6395	0.7907	0.6357	0.7536
	300	5	0.7601	0.8998	0.7600	0.7840
		10	0.7178	0.9004	0.7187	0.7683
		25	0.6614	0.8481	0.6625	0.7814
	400	5	0.7903	0.9258	0.7857	0.8077
		10	0.7489	0.9257	0.7445	0.7853
		25	0.6861	0.8781	0.6865	0.7819
70	100	5	0.7023	0.8067	0.7007	0.7659
		10	0.6796	0.8178	0.6773	0.7924
		25	0.6442	0.8054	0.6437	0.8150
	200	5	0.7706	0.8918	0.7709	0.8157
		10	0.7336	0.9000	0.7318	0.8321
		25	0.6856	0.8895	0.6837	0.8685
	300	5	0.8209	0.9352	0.8199	0.8561
		10	0.7837	0.9437	0.7811	0.8664
		25	0.7186	0.9319	0.7186	0.8940
	400	5	0.8602	0.9632	0.8579	0.8857
		10	0.8154	0.9695	0.8157	0.8856
		25	0.7485	0.9616	0.7454	0.9068

Table B.16 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.5$ with mixed block sizes

n	p	p_i	p_j	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	10	0.6400	0.7624	0.6333	0.6966
		5	25	0.6329	0.7474	0.6304	0.7235
		10	25	0.6175	0.7426	0.6174	0.7262
	200	5	10	0.7032	0.8534	0.6996	0.7457
		5	25	0.6638	0.8297	0.6631	0.7550
		10	25	0.6504	0.8168	0.6525	0.7457
	300	5	10	0.7408	0.9030	0.7408	0.7761
		5	25	0.7039	0.8802	0.6996	0.7807
		10	25	0.6921	0.8751	0.6861	0.7776
	400	5	10	0.7732	0.9358	0.7681	0.8023
		5	25	0.7327	0.9073	0.7323	0.8054
		10	25	0.7138	0.9075	0.7122	0.7849
70	100	5	10	0.6940	0.8266	0.6916	0.7808
		5	25	0.6727	0.8179	0.6688	0.7962
		10	25	0.6675	0.8217	0.6671	0.8066
	200	5	10	0.7566	0.9057	0.7560	0.8331
		5	25	0.7274	0.9072	0.7259	0.8565
		10	25	0.7146	0.9036	0.7141	0.8637
	300	5	10	0.8019	0.9480	0.8005	0.8692
		5	25	0.7599	0.9420	0.7578	0.8812
		10	25	0.7532	0.9418	0.7538	0.8869
	400	5	10	0.8391	0.9715	0.8369	0.8891
		5	25	0.7911	0.9690	0.7906	0.8973
		10	25	0.7803	0.9683	0.7797	0.9017

Table B.17 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.9$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.6261	0.9577	0.6207	0.7184
		10	0.5997	0.9686	0.5975	0.7721
		25	0.5712	0.9450	0.5690	0.8659
	200	5	0.6611	0.9933	0.6580	0.7231
		10	0.6277	0.9959	0.6267	0.7469
		25	0.5812	0.9882	0.5810	0.8251
	300	5	0.6956	0.9989	0.6994	0.7421
		10	0.6558	0.9995	0.6566	0.7474
		25	0.6064	0.9972	0.6040	0.8002
	400	5	0.7317	0.9998	0.7278	0.7640
		10	0.6761	0.9999	0.6752	0.7460
		25	0.6150	0.9995	0.6122	0.7705
70	100	5	0.6657	0.9711	0.6644	0.8642
		10	0.6341	0.9810	0.6321	0.9204
		25	0.5933	0.9789	0.5921	0.9636
	200	5	0.7220	0.9969	0.7182	0.8550
		10	0.6771	0.9977	0.6744	0.9122
		25	0.6291	0.9977	0.6260	0.9695
	300	5	0.7550	0.9999	0.7527	0.8560
		10	0.6985	1.0000	0.6992	0.9023
		25	0.6433	0.9998	0.6400	0.9624
	400	5	0.7925	1.0000	0.7904	0.8711
		10	0.7349	1.0000	0.7327	0.8970
		25	0.6559	0.9999	0.6554	0.9548

Table B.18 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.5978	0.9711	0.5961	0.7588
		5	25	0.5884	0.9517	0.5852	0.7949
		10	25	0.5732	0.9536	0.5736	0.8147
	200	5	10	0.6677	0.9994	0.6641	0.7454
		5	25	0.6252	0.9982	0.6215	0.7652
		10	25	0.6259	0.9981	0.6269	0.7757
	300	5	10	0.6677	0.9994	0.6641	0.7454
		5	25	0.6252	0.9982	0.6215	0.7652
		10	25	0.6259	0.9981	0.6269	0.7757
	400	5	10	0.6983	0.9999	0.6965	0.7578
		5	25	0.6498	0.9998	0.6498	0.7651
		10	25	0.6380	0.9999	0.6386	0.7607
70	100	5	10	0.6484	0.9822	0.6470	0.9062
		5	25	0.6205	0.9826	0.6201	0.9291
		10	25	0.6107	0.9811	0.6131	0.9507
	200	5	10	0.6959	0.9987	0.6926	0.8994
		5	25	0.6615	0.9984	0.6592	0.9355
		10	25	0.6448	0.9991	0.6472	0.9508
	300	5	10	0.7247	0.9999	0.7221	0.8901
		5	25	0.6800	0.9999	0.6779	0.9264
		10	25	0.6626	1.0000	0.6624	0.9421
	400	5	10	0.7608	1.0000	0.7575	0.8919
		5	25	0.7024	1.0000	0.7036	0.9234
		10	25	0.6925	1.0000	0.6923	0.9346

Table B.19 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.9$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.6186	0.9564	0.6096	0.7142
		10	0.5946	0.9701	0.5925	0.7751
		25	0.5713	0.9462	0.5683	0.8668
	200	5	0.6585	0.9928	0.6581	0.7203
		10	0.6187	0.9941	0.6172	0.7449
		25	0.5874	0.9871	0.5853	0.8178
	300	5	0.6957	0.9995	0.6910	0.7394
		10	0.6472	0.9995	0.6473	0.7407
		25	0.5972	0.9973	0.5988	0.7992
	400	5	0.7298	1.0000	0.7247	0.7603
		10	0.6674	1.0000	0.6622	0.7380
		25	0.6135	0.9994	0.6137	0.7740
70	100	5	0.6629	0.9726	0.6622	0.8633
		10	0.6283	0.9826	0.6264	0.9237
		25	0.5982	0.9780	0.5973	0.9624
	200	5	0.7139	0.9967	0.7118	0.8526
		10	0.6648	0.9982	0.6629	0.9064
		25	0.6184	0.9978	0.6158	0.9687
	300	5	0.7538	0.9996	0.7540	0.8602
		10	0.7020	0.9999	0.6986	0.9037
		25	0.6377	0.9998	0.6351	0.9641
	400	5	0.7973	1.0000	0.7928	0.8724
		10	0.7292	1.0000	0.7288	0.8958
		25	0.6582	1.0000	0.6605	0.9568

Table B.20 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_9$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.5958	0.9724	0.5959	0.7396
		5	25	0.5927	0.9538	0.5931	0.7950
		10	25	0.5768	0.9523	0.5757	0.8203
	200	5	10	0.6733	0.9995	0.6709	0.7477
		5	25	0.6362	0.9981	0.6335	0.7709
		10	25	0.6192	0.9978	0.6190	0.7726
	300	5	10	0.6733	0.9995	0.6709	0.7477
		5	25	0.6362	0.9981	0.6335	0.7709
		10	25	0.6192	0.9978	0.6190	0.7726
	400	5	10	0.6907	1.0000	0.6906	0.7549
		5	25	0.6560	0.9993	0.6521	0.7748
		10	25	0.6383	0.9995	0.6361	0.7559
70	100	5	10	0.6441	0.9847	0.6427	0.9047
		5	25	0.6230	0.9825	0.6236	0.9313
		10	25	0.6171	0.9838	0.6178	0.9508
	200	5	10	0.6899	0.9981	0.6873	0.8969
		5	25	0.6562	0.9990	0.6544	0.9309
		10	25	0.6405	0.9984	0.6407	0.9526
	300	5	10	0.7264	0.9999	0.7286	0.8912
		5	25	0.6817	0.9997	0.6770	0.9283
		10	25	0.6736	0.9999	0.6746	0.9430
	400	5	10	0.7552	1.0000	0.7526	0.8892
		5	25	0.7037	1.0000	0.7018	0.9202
		10	25	0.6888	1.0000	0.6868	0.9348

Table B.21 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{10}$ and $\theta = 0.9$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.6307	0.9452	0.6256	0.7113
		10	0.6146	0.9509	0.6134	0.7387
		25	0.6037	0.9185	0.6021	0.7484
	200	5	0.6723	0.9888	0.6691	0.7275
		10	0.6470	0.9905	0.6485	0.7364
		25	0.6258	0.9790	0.6224	0.7406
	300	5	0.7099	0.9981	0.7094	0.7457
		10	0.6844	0.9982	0.6818	0.7491
		25	0.6616	0.9939	0.6630	0.7591
	400	5	0.7422	0.9997	0.7393	0.7719
		10	0.7086	0.9996	0.7049	0.7635
		25	0.6780	0.9978	0.6767	0.7565
70	100	5	0.6731	0.9624	0.6717	0.8396
		10	0.6519	0.9683	0.6507	0.8664
		25	0.6381	0.9600	0.6362	0.8747
	200	5	0.7328	0.9934	0.7305	0.8440
		10	0.7034	0.9944	0.7013	0.8728
		25	0.6874	0.9935	0.6882	0.8903
	300	5	0.7675	0.9995	0.7649	0.8533
		10	0.7345	0.9999	0.7324	0.8778
		25	0.7132	0.9996	0.7098	0.8917
	400	5	0.8052	1.0000	0.8033	0.8728
		10	0.7702	0.9999	0.7691	0.8831
		25	0.7353	1.0000	0.7312	0.8947

Table B.22 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{10}$ and $\theta = 0.9$ with mixed block sizes

n	***p***	$\boldsymbol{p_i}$	$\boldsymbol{p_j}$	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	10	0.6111	0.8769	0.6067	0.7080
		5	25	0.6077	0.8410	0.6069	0.7080
		10	25	0.5981	0.8475	0.5997	0.7095
	200	5	10	0.6610	0.9562	0.6559	0.7223
		5	25	0.6478	0.9281	0.6454	0.7255
		10	25	0.6314	0.9274	0.6266	0.7182
	300	5	10	0.6847	0.9791	0.6817	0.7345
		5	25	0.6729	0.9669	0.6726	0.7384
		10	25	0.6697	0.9631	0.6676	0.7389
	400	5	10	0.7195	0.9928	0.7167	0.7609
		5	25	0.6927	0.9830	0.6923	0.7558
		10	25	0.6869	0.9846	0.6832	0.7563
70	100	5	10	0.6603	0.9090	0.6593	0.8155
		5	25	0.6477	0.9045	0.6471	0.8198
		10	25	0.6396	0.9014	0.6376	0.8268
	200	5	10	0.7142	0.9742	0.7113	0.8397
		5	25	0.7067	0.9710	0.7032	0.8526
		10	25	0.6890	0.9699	0.6872	0.8555
	300	5	10	0.7446	0.9924	0.7408	0.8502
		5	25	0.7336	0.9900	0.7304	0.8576
		10	25	0.7176	0.9907	0.7165	0.8631
	400	5	10	0.7843	0.9978	0.7810	0.8687
		5	25	0.7576	0.9976	0.7570	0.8693
		10	25	0.7504	0.9972	0.7482	0.8724

Table B.23 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{10}$ and $\theta = 0.9$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.6262	0.9426	0.6191	0.7076
		10	0.6096	0.9497	0.6073	0.7292
		25	0.6040	0.9167	0.6004	0.7503
	200	5	0.6669	0.9859	0.6619	0.7195
		10	0.6386	0.9870	0.6368	0.7308
		25	0.6219	0.9735	0.6237	0.7429
	300	5	0.7072	0.9985	0.7034	0.7453
		10	0.6769	0.9983	0.6752	0.7473
		25	0.6591	0.9956	0.6569	0.7542
	400	5	0.7408	0.9994	0.7355	0.7700
		10	0.7012	0.9997	0.6998	0.7561
		25	0.6792	0.9983	0.6781	0.7600
70	100	5	0.6711	0.9605	0.6700	0.8382
		10	0.6503	0.9686	0.6481	0.8668
		25	0.6356	0.9602	0.6356	0.8741
	200	5	0.7241	0.9943	0.7246	0.8431
		10	0.6900	0.9958	0.6896	0.8675
		25	0.6746	0.9946	0.6739	0.8846
	300	5	0.7683	0.9992	0.7664	0.8574
		10	0.7370	0.9997	0.7346	0.8801
		25	0.7107	0.9992	0.7094	0.8958
	400	5	0.8081	0.9999	0.8047	0.8732
		10	0.7656	0.9999	0.7656	0.8844
		25	0.7407	1.0000	0.7416	0.8998

Table B.24 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{10}$ and $\theta = 0.9$ with mixed block sizes

n	***p***	***p**$_i$*	***p**$_j$*	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	10	0.6049	0.8728	0.6036	0.6905
		5	25	0.6158	0.8391	0.6128	0.7096
		10	25	0.5969	0.8372	0.5995	0.7064
	200	5	10	0.6570	0.9538	0.6545	0.7222
		5	25	0.6400	0.9302	0.6355	0.7144
		10	25	0.6305	0.9208	0.6280	0.7168
	300	5	10	0.6905	0.9813	0.6875	0.7389
		5	25	0.6800	0.9647	0.6786	0.7395
		10	25	0.6676	0.9621	0.6618	0.7343
	400	5	10	0.7168	0.9916	0.7126	0.7558
		5	25	0.7020	0.9794	0.7034	0.7605
		10	25	0.6833	0.9810	0.6838	0.7430
70	100	5	10	0.6553	0.9159	0.6537	0.8148
		5	25	0.6489	0.9032	0.6489	0.8155
		10	25	0.6426	0.9065	0.6404	0.8275
	200	5	10	0.7059	0.9747	0.7041	0.8344
		5	25	0.6963	0.9700	0.6953	0.8448
		10	25	0.6848	0.9724	0.6849	0.8547
	300	5	10	0.7483	0.9930	0.7469	0.8527
		5	25	0.7318	0.9899	0.7298	0.8552
		10	25	0.7235	0.9887	0.7227	0.8663
	400	5	10	0.7760	0.9980	0.7758	0.8614
		5	25	0.7606	0.9968	0.7595	0.8690
		10	25	0.7501	0.9973	0.7501	0.8750

Table B.25 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.1$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.3098	0.3148	0.3152	0.3154
		10	0.3133	0.3225	0.3174	0.3155
		25	0.3193	0.3535	0.3246	0.3144
	200	5	0.2376	0.2456	0.2430	0.2427
		10	0.2405	0.2557	0.2448	0.2417
		25	0.2483	0.3054	0.2562	0.2435
	300	5	0.1772	0.1872	0.1835	0.1826
		10	0.1807	0.1998	0.1878	0.1848
		25	0.1937	0.2530	0.2001	0.1925
	400	5	0.1488	0.1583	0.1552	0.1534
		10	0.1552	0.1683	0.1597	0.1561
		25	0.1644	0.2282	0.1712	0.1610
70	100	5	0.2611	0.2615	0.2630	0.2734
		10	0.2637	0.2660	0.2670	0.2716
		25	0.2686	0.2804	0.2708	0.2689
	200	5	0.1834	0.1858	0.1864	0.1939
		10	0.1867	0.1876	0.1887	0.1939
		25	0.1943	0.2050	0.1969	0.1908
	300	5	0.1348	0.1344	0.1359	0.1386
		10	0.1366	0.1359	0.1382	0.1384
		25	0.1457	0.1532	0.1471	0.1387
	400	5	0.0986	0.0981	0.1012	0.1043
		10	0.1011	0.1001	0.1030	0.1037
		25	0.1080	0.1187	0.1103	0.1047

Table B.26 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.1$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.3084	0.3127	0.3138	0.3148
		5	25	0.3106	0.3304	0.3165	0.3155
		10	25	0.3150	0.3304	0.3175	0.3095
	200	5	10	0.2295	0.2372	0.2376	0.2340
		5	25	0.2373	0.2659	0.2420	0.2383
		10	25	0.2456	0.2753	0.2483	0.2438
	300	5	10	0.1854	0.1892	0.1908	0.1884
		5	25	0.1891	0.2179	0.1900	0.1893
		10	25	0.1927	0.2195	0.1958	0.1893
	400	5	10	0.1527	0.1576	0.1569	0.1533
		5	25	0.1551	0.1873	0.1596	0.1537
		10	25	0.1539	0.1920	0.1594	0.1558
70	100	5	10	0.2631	0.2617	0.2653	0.2732
		5	25	0.2648	0.2642	0.2667	0.2675
		10	25	0.2661	0.2665	0.2702	0.2687
	200	5	10	0.1831	0.1827	0.1869	0.1909
		5	25	0.1830	0.1881	0.1874	0.1855
		10	25	0.1900	0.1895	0.1911	0.1891
	300	5	10	0.1326	0.1287	0.1363	0.1379
		5	25	0.1380	0.1380	0.1406	0.1375
		10	25	0.1377	0.1397	0.1398	0.1363
	400	5	10	0.0960	0.0927	0.0987	0.0991
		5	25	0.1037	0.1045	0.1069	0.1066
		10	25	0.1027	0.1034	0.1039	0.1035

Table B.27 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.1$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.6894	0.6865	0.6856	0.6863
		10	0.6854	0.6786	0.6803	0.6856
		25	0.6786	0.6431	0.6758	0.6857
	200	5	0.7691	0.7597	0.7648	0.7623
		10	0.7677	0.7499	0.7642	0.7657
		25	0.7592	0.6940	0.7498	0.7629
	300	5	0.8233	0.8124	0.8165	0.8156
		10	0.8208	0.8009	0.8126	0.8153
		25	0.8043	0.7488	0.8007	0.8076
	400	5	0.8549	0.8424	0.8464	0.8496
		10	0.8481	0.8341	0.8422	0.8479
		25	0.8385	0.7703	0.8317	0.8424
70	100	5	0.7416	0.7407	0.7407	0.7312
		10	0.7401	0.7364	0.7377	0.7321
		25	0.7343	0.7226	0.7329	0.7346
	200	5	0.8162	0.8134	0.8123	0.8055
		10	0.8124	0.8105	0.8101	0.8056
		25	0.8050	0.7916	0.8026	0.8082
	300	5	0.8633	0.8633	0.8612	0.8598
		10	0.8620	0.8634	0.8596	0.8598
		25	0.8520	0.8457	0.8512	0.8595
	400	5	0.8990	0.8987	0.8954	0.8934
		10	0.8965	0.8979	0.8936	0.8936
		25	0.8886	0.8803	0.8869	0.8915

Table B.28 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.1$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6937	0.6891	0.6868	0.6902
		5	25	0.6885	0.6735	0.6848	0.6848
		10	25	0.6863	0.6764	0.6834	0.6933
	200	5	10	0.7748	0.7685	0.7685	0.7719
		5	25	0.7704	0.7401	0.7662	0.7673
		10	25	0.7647	0.7279	0.7625	0.7650
	300	5	10	0.8186	0.8169	0.8142	0.8150
		5	25	0.8108	0.7811	0.8131	0.8126
		10	25	0.8086	0.7847	0.8064	0.8142
	400	5	10	0.8468	0.8407	0.8424	0.8456
		5	25	0.8445	0.8181	0.8381	0.8465
		10	25	0.8501	0.8070	0.8452	0.8483
70	100	5	10	0.7324	0.7352	0.7296	0.7225
		5	25	0.7372	0.7393	0.7354	0.7357
		10	25	0.7301	0.7310	0.7253	0.7291
	200	5	10	0.8184	0.8161	0.8135	0.8092
		5	25	0.8155	0.8102	0.8118	0.8123
		10	25	0.8105	0.8103	0.8092	0.8109
	300	5	10	0.8681	0.8699	0.8646	0.8641
		5	25	0.8642	0.8639	0.8628	0.8668
		10	25	0.8634	0.8610	0.8622	0.8639
	400	5	10	0.9025	0.9059	0.8988	0.9001
		5	25	0.8972	0.8965	0.8942	0.8947
		10	25	0.8959	0.8955	0.8946	0.8966

Table B.29 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.1$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.6911	0.6840	0.6840	0.6829
		10	0.6881	0.6764	0.6850	0.6835
		25	0.6828	0.6499	0.6751	0.6856
	200	5	0.7557	0.7492	0.7492	0.7524
		10	0.7513	0.7388	0.7463	0.7510
		25	0.7443	0.6952	0.7378	0.7501
	300	5	0.8224	0.8132	0.8165	0.8193
		10	0.8179	0.7995	0.8119	0.8152
		25	0.8084	0.7453	0.7992	0.8074
	400	5	0.8476	0.8410	0.8433	0.8436
		10	0.8415	0.8294	0.8385	0.8400
		25	0.8328	0.7733	0.8260	0.8356
70	100	5	0.7362	0.7363	0.7334	0.7220
		10	0.7326	0.7317	0.7284	0.7247
		25	0.7285	0.7166	0.7256	0.7277
	200	5	0.8171	0.8150	0.8150	0.8067
		10	0.8142	0.8144	0.8125	0.8066
		25	0.8064	0.7985	0.8037	0.8102
	300	5	0.8672	0.8680	0.8671	0.8631
		10	0.8649	0.8649	0.8642	0.8634
		25	0.8566	0.8480	0.8547	0.8631
	400	5	0.9038	0.9051	0.9022	0.8980
		10	0.9014	0.9020	0.9004	0.8990
		25	0.8954	0.8824	0.8925	0.8992

Table B.30 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.1$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6895	0.6856	0.6856	0.6802
		5	25	0.6904	0.6657	0.6822	0.6842
		10	25	0.6837	0.6629	0.6816	0.6878
	200	5	10	0.7662	0.7571	0.7563	0.7602
		5	25	0.7551	0.7282	0.7498	0.7562
		10	25	0.7442	0.7215	0.7410	0.7474
	300	5	10	0.8106	0.8048	0.8042	0.8083
		5	25	0.8111	0.7832	0.8070	0.8088
		10	25	0.8061	0.7764	0.8021	0.8072
	400	5	10	0.8478	0.8441	0.8438	0.8479
		5	25	0.8454	0.8074	0.8427	0.8462
		10	25	0.8421	0.8090	0.8360	0.8402
70	100	5	10	0.7415	0.7414	0.7398	0.7312
		5	25	0.7332	0.7323	0.7313	0.7294
		10	25	0.7377	0.7360	0.7344	0.7335
	200	5	10	0.8154	0.8187	0.8128	0.8091
		5	25	0.8186	0.8137	0.8134	0.8168
		10	25	0.8096	0.8107	0.8086	0.8110
	300	5	10	0.8668	0.8727	0.8628	0.8602
		5	25	0.8599	0.8601	0.8561	0.8582
		10	25	0.8614	0.8596	0.8582	0.8637
	400	5	10	0.9055	0.9089	0.9038	0.9017
		5	25	0.8955	0.8945	0.8921	0.8923
		10	25	0.8988	0.8978	0.8977	0.8966

Table B.31 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.5$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.3442	0.2470	0.3483	0.3093
		10	0.3670	0.2413	0.3714	0.2943
		25	0.3970	0.2803	0.3987	0.2667
	200	5	0.2839	0.1605	0.2873	0.2600
		10	0.3177	0.1574	0.3208	0.2593
		25	0.3639	0.2095	0.3635	0.2422
	300	5	0.2347	0.1013	0.2366	0.2176
		10	0.2747	0.1027	0.2777	0.2309
		25	0.3365	0.1514	0.3376	0.2251
	400	5	0.2055	0.0755	0.2095	0.1938
		10	0.2474	0.0743	0.2502	0.2115
		25	0.3105	0.1200	0.3106	0.2186
70	100	5	0.2915	0.1910	0.2933	0.2359
		10	0.3164	0.1804	0.3184	0.2116
		25	0.3531	0.1937	0.3539	0.1866
	200	5	0.2254	0.1128	0.2290	0.1850
		10	0.2589	0.1042	0.2608	0.1647
		25	0.3141	0.1108	0.3147	0.1317
	300	5	0.1776	0.0644	0.1801	0.1452
		10	0.2169	0.0568	0.2193	0.1345
		25	0.2757	0.0664	0.2760	0.1092
	400	5	0.1412	0.0378	0.1446	0.1188
		10	0.1852	0.0325	0.1871	0.1151
		25	0.2506	0.0403	0.2509	0.0957

Table B.32 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.5$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.3527	0.2300	0.3553	0.2982
		5	25	0.3772	0.2484	0.3782	0.2831
		10	25	0.3784	0.2526	0.3801	0.2723
	200	5	10	0.2976	0.1404	0.2985	0.2525
		5	25	0.3280	0.1715	0.3289	0.2423
		10	25	0.3434	0.1811	0.3467	0.2493
	300	5	10	0.2634	0.0962	0.2652	0.2244
		5	25	0.3003	0.1214	0.3008	0.2213
		10	25	0.3044	0.1234	0.3082	0.2223
	400	5	10	0.2325	0.0631	0.2354	0.2038
		5	25	0.2714	0.0896	0.2731	0.2024
		10	25	0.2809	0.0925	0.2815	0.2077
70	100	5	10	0.3060	0.1804	0.3075	0.2191
		5	25	0.3236	0.1798	0.3251	0.2014
		10	25	0.3329	0.1797	0.3352	0.1963
	200	5	10	0.2438	0.0955	0.2469	0.1668
		5	25	0.2726	0.0973	0.2737	0.1431
		10	25	0.2859	0.0995	0.2879	0.1418
	300	5	10	0.1968	0.0516	0.1991	0.1337
		5	25	0.2373	0.0562	0.2398	0.1139
		10	25	0.2487	0.0585	0.2495	0.1139
	400	5	10	0.1594	0.0294	0.1621	0.1097
		5	25	0.2100	0.0322	0.2131	0.1006
		10	25	0.2156	0.0319	0.2160	0.0989

Table B.33 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.5$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.6532	0.7521	0.6492	0.6918
		10	0.6355	0.7574	0.6301	0.7060
		25	0.6046	0.7173	0.6033	0.7337
	200	5	0.7213	0.8454	0.7203	0.7439
		10	0.6869	0.8499	0.6852	0.7468
		25	0.6405	0.7907	0.6423	0.7625
	300	5	0.7655	0.8959	0.7660	0.7841
		10	0.7247	0.8941	0.7220	0.7713
		25	0.6608	0.8474	0.6595	0.7764
	400	5	0.7960	0.9263	0.7925	0.8100
		10	0.7530	0.9266	0.7521	0.7934
		25	0.6897	0.8810	0.6918	0.7812
70	100	5	0.7117	0.8108	0.7096	0.7694
		10	0.6850	0.8220	0.6835	0.7924
		25	0.6505	0.8079	0.6505	0.8163
	200	5	0.7766	0.8849	0.7733	0.8159
		10	0.7419	0.8935	0.7402	0.8346
		25	0.6869	0.8872	0.6872	0.8673
	300	5	0.8196	0.9357	0.8161	0.8519
		10	0.7801	0.9437	0.7786	0.8638
		25	0.7215	0.9352	0.7213	0.8898
	400	5	0.8540	0.9612	0.8514	0.8773
		10	0.8119	0.9657	0.8105	0.8824
		25	0.7460	0.9580	0.7460	0.9015

Table B.34 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.5$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6494	0.7733	0.6466	0.7038
		5	25	0.6234	0.7528	0.6225	0.7202
		10	25	0.6212	0.7538	0.6171	0.7310
	200	5	10	0.7115	0.8610	0.7104	0.7544
		5	25	0.6765	0.8269	0.6771	0.7621
		10	25	0.6600	0.8232	0.6582	0.7581
	300	5	10	0.7400	0.9057	0.7370	0.7802
		5	25	0.6981	0.8748	0.7004	0.7757
		10	25	0.6945	0.8792	0.6882	0.7753
	400	5	10	0.7630	0.9361	0.7637	0.7967
		5	25	0.7306	0.9138	0.7298	0.7990
		10	25	0.7236	0.9086	0.7222	0.7970
70	100	5	10	0.6939	0.8157	0.6917	0.7799
		5	25	0.6772	0.8246	0.6758	0.8026
		10	25	0.6670	0.8186	0.6656	0.8027
	200	5	10	0.7572	0.9041	0.7541	0.8327
		5	25	0.7291	0.9019	0.7277	0.8534
		10	25	0.7163	0.9007	0.7136	0.8584
	300	5	10	0.8051	0.9479	0.8026	0.8673
		5	25	0.7650	0.9453	0.7631	0.8886
		10	25	0.7525	0.9411	0.7508	0.8846
	400	5	10	0.8377	0.9699	0.8339	0.8873
		5	25	0.7892	0.9681	0.7853	0.8994
		10	25	0.7842	0.9677	0.7842	0.9001

Table B.35 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.5$ with equal block sizes

n	***p***	***p_i***	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.6584	0.7539	0.6543	0.6896
		10	0.6306	0.7600	0.6271	0.7054
		25	0.6015	0.7221	0.5994	0.7330
	200	5	0.7109	0.8337	0.7052	0.7361
		10	0.6777	0.8354	0.6732	0.7347
		25	0.6317	0.7904	0.6308	0.7531
	300	5	0.7651	0.9015	0.7609	0.7808
		10	0.7260	0.9006	0.7227	0.7670
		25	0.6662	0.8499	0.6654	0.7734
	400	5	0.7931	0.9227	0.7885	0.8025
		10	0.7523	0.9248	0.7475	0.7836
		25	0.6894	0.8790	0.6870	0.7816
70	100	5	0.7054	0.8072	0.7038	0.7588
		10	0.6823	0.8174	0.6798	0.7845
		25	0.6433	0.8048	0.6417	0.8107
	200	5	0.7727	0.8897	0.7687	0.8142
		10	0.7404	0.8983	0.7383	0.8361
		25	0.6850	0.8913	0.6835	0.8694
	300	5	0.8253	0.9356	0.8238	0.8577
		10	0.7861	0.9428	0.7828	0.8673
		25	0.7271	0.9321	0.7268	0.8918
	400	5	0.8637	0.9632	0.8594	0.8853
		10	0.8177	0.9694	0.8153	0.8875
		25	0.7530	0.9616	0.7523	0.9071

Table B.36 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.5$ with mixed block sizes

n	***p***	p_i	p_j	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	10	0.6452	0.7668	0.6429	0.6999
		5	25	0.6222	0.7504	0.6211	0.7137
		10	25	0.6221	0.7411	0.6228	0.7245
	200	5	10	0.6934	0.8582	0.6927	0.7407
		5	25	0.6676	0.8301	0.6652	0.7533
		10	25	0.6533	0.8146	0.6484	0.7433
	300	5	10	0.7333	0.9019	0.7326	0.7711
		5	25	0.7013	0.8824	0.6980	0.7817
		10	25	0.6967	0.8741	0.6954	0.7802
	400	5	10	0.7720	0.9377	0.7656	0.7958
		5	25	0.7267	0.9071	0.7240	0.7962
		10	25	0.7146	0.9064	0.7149	0.7877
70	100	5	10	0.6942	0.8237	0.6933	0.7820
		5	25	0.6756	0.8159	0.6740	0.7946
		10	25	0.6672	0.8221	0.6642	0.8048
	200	5	10	0.7552	0.9049	0.7521	0.8337
		5	25	0.7258	0.9036	0.7251	0.8604
		10	25	0.7120	0.9004	0.7107	0.8580
	300	5	10	0.8014	0.9489	0.7994	0.8654
		5	25	0.7606	0.9424	0.7574	0.8838
		10	25	0.7502	0.9420	0.7503	0.8878
	400	5	10	0.8435	0.9714	0.8421	0.8933
		5	25	0.7909	0.9675	0.7887	0.8995
		10	25	0.7847	0.9685	0.7838	0.9022

Table B.37 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.9$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.3833	0.0441	0.3879	0.2835
		10	0.4065	0.0303	0.4109	0.2276
		25	0.4334	0.0542	0.4339	0.1441
	200	5	0.3407	0.0063	0.3422	0.2790
		10	0.3762	0.0047	0.3807	0.2573
		25	0.4155	0.0122	0.4151	0.1790
	300	5	0.2992	0.0009	0.3003	0.2558
		10	0.3460	0.0005	0.3510	0.2585
		25	0.4022	0.0025	0.4025	0.2034
	400	5	0.2768	0.0003	0.2789	0.2413
		10	0.3235	0.0002	0.3263	0.2535
		25	0.3801	0.0008	0.3805	0.2215
70	100	5	0.3351	0.0291	0.3377	0.1365
		10	0.3682	0.0189	0.3682	0.0774
		25	0.4057	0.0217	0.4054	0.0363
	200	5	0.2818	0.0038	0.2854	0.1490
		10	0.3270	0.0020	0.3283	0.0921
		25	0.3783	0.0021	0.3792	0.0323
	300	5	0.2411	0.0003	0.2450	0.1411
		10	0.2959	0.0002	0.2971	0.0955
		25	0.3573	0.0002	0.3591	0.0364
	400	5	0.2061	0.0001	0.2079	0.1301
		10	0.2699	0.0000	0.2718	0.1033
		25	0.3383	0.0000	0.3404	0.0440

Table B.38 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.3932	0.0280	0.3988	0.2453
		5	25	0.4132	0.0474	0.4164	0.2117
		10	25	0.4187	0.0458	0.4213	0.1749
	200	5	10	0.3570	0.0031	0.3601	0.2598
		5	25	0.3861	0.0093	0.3911	0.2273
		10	25	0.3985	0.0088	0.4024	0.2156
	300	5	10	0.3366	0.0004	0.3404	0.2609
		5	25	0.3732	0.0019	0.3717	0.2364
		10	25	0.3723	0.0019	0.3760	0.2293
	400	5	10	0.3047	0.0001	0.3100	0.2468
		5	25	0.3484	0.0006	0.3515	0.2299
		10	25	0.3589	0.0003	0.3591	0.2348
70	100	5	10	0.3561	0.0161	0.3579	0.0943
		5	25	0.3750	0.0177	0.3734	0.0706
		10	25	0.3847	0.0188	0.3871	0.0517
	200	5	10	0.3108	0.0019	0.3118	0.1054
		5	25	0.3426	0.0015	0.3425	0.0659
		10	25	0.3549	0.0012	0.3570	0.0488
	300	5	10	0.2724	0.0001	0.2738	0.1114
		5	25	0.3185	0.0002	0.3205	0.0713
		10	25	0.3302	0.0001	0.3308	0.0561
	400	5	10	0.2385	0.0000	0.2426	0.1078
		5	25	0.2972	0.0000	0.2989	0.0764
		10	25	0.3040	0.0000	0.3063	0.0643

Table B.39 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.9$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.6113	0.9551	0.6086	0.7169
		10	0.5971	0.9673	0.5925	0.7731
		25	0.5653	0.9432	0.5654	0.8562
	200	5	0.6644	0.9935	0.6631	0.7253
		10	0.6308	0.9960	0.6239	0.7446
		25	0.5877	0.9891	0.5889	0.8199
	300	5	0.7026	0.9988	0.7022	0.7501
		10	0.6527	0.9994	0.6474	0.7424
		25	0.5960	0.9974	0.5965	0.7970
	400	5	0.7279	0.9998	0.7257	0.7614
		10	0.6768	0.9999	0.6750	0.7476
		25	0.6181	0.9992	0.6169	0.7767
70	100	5	0.6700	0.9709	0.6669	0.8660
		10	0.6338	0.9803	0.6344	0.9242
		25	0.5965	0.9782	0.5986	0.9635
	200	5	0.7183	0.9970	0.7151	0.8530
		10	0.6737	0.9984	0.6731	0.9083
		25	0.6209	0.9982	0.6210	0.9669
	300	5	0.7568	0.9997	0.7532	0.8576
		10	0.7011	0.9999	0.7017	0.9045
		25	0.6423	0.9999	0.6409	0.9632
	400	5	0.7895	0.9999	0.7874	0.8660
		10	0.7262	1.0000	0.7247	0.8967
		25	0.6615	1.0000	0.6590	0.9542

Table B.40 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6107	0.9730	0.6067	0.7568
		5	25	0.5885	0.9525	0.5835	0.7888
		10	25	0.5806	0.9547	0.5776	0.8303
	200	5	10	0.6506	0.9965	0.6467	0.7430
		5	25	0.6163	0.9901	0.6120	0.7732
		10	25	0.6039	0.9922	0.6014	0.7894
	300	5	10	0.6652	0.9994	0.6629	0.7394
		5	25	0.6284	0.9980	0.6284	0.7616
		10	25	0.6239	0.9981	0.6200	0.7664
	400	5	10	0.6952	0.9998	0.6911	0.7528
		5	25	0.6528	0.9996	0.6503	0.7724
		10	25	0.6393	0.9999	0.6370	0.7690
70	100	5	10	0.6437	0.9830	0.6432	0.9046
		5	25	0.6251	0.9823	0.6270	0.9334
		10	25	0.6157	0.9799	0.6135	0.9497
	200	5	10	0.6895	0.9980	0.6894	0.8944
		5	25	0.6605	0.9985	0.6623	0.9323
		10	25	0.6454	0.9990	0.6431	0.9537
	300	5	10	0.7279	0.9999	0.7261	0.8881
		5	25	0.6828	1.0000	0.6810	0.9300
		10	25	0.6683	1.0000	0.6676	0.9426
	400	5	10	0.7593	1.0000	0.7541	0.8905
		5	25	0.6985	1.0000	0.6970	0.9222
		10	25	0.6933	1.0000	0.6903	0.9363

Table B.41 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.9$ with equal block sizes

n	p	p_i	TA,TB,TC	BD	DI	SK
35	100	5	0.6221	0.9568	0.6156	0.7162
		10	0.5899	0.9721	0.5858	0.7718
		25	0.5680	0.9484	0.5669	0.8556
	200	5	0.6543	0.9939	0.6525	0.7168
		10	0.6168	0.9946	0.6148	0.7409
		25	0.5813	0.9865	0.5810	0.8221
	300	5	0.6990	0.9995	0.6973	0.7384
		10	0.6553	0.9996	0.6506	0.7406
		25	0.5996	0.9976	0.5986	0.7963
	400	5	0.7186	0.9997	0.7165	0.7560
		10	0.6762	0.9997	0.6724	0.7455
		25	0.6218	0.9992	0.6222	0.7804
70	100	5	0.6600	0.9710	0.6578	0.8610
		10	0.6299	0.9820	0.6293	0.9211
		25	0.5921	0.9786	0.5906	0.9640
	200	5	0.7183	0.9955	0.7143	0.8491
		10	0.6725	0.9978	0.6704	0.9076
		25	0.6225	0.9976	0.6207	0.9686
	300	5	0.7611	0.9997	0.7568	0.8603
		10	0.7072	0.9999	0.7041	0.9045
		25	0.6432	0.9999	0.6410	0.9641
	400	5	0.7984	1.0000	0.7969	0.8739
		10	0.7340	1.0000	0.7319	0.8967
		25	0.6619	1.0000	0.6602	0.9578

Table B.42 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{11}$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6030	0.9711	0.5957	0.7526
		5	25	0.5852	0.9528	0.5838	0.7879
		10	25	0.5820	0.9537	0.5798	0.8199
	200	5	10	0.6354	0.9974	0.6332	0.7375
		5	25	0.6115	0.9914	0.6059	0.7723
		10	25	0.5991	0.9902	0.5939	0.7794
	300	5	10	0.6617	0.9998	0.6564	0.7389
		5	25	0.6252	0.9982	0.6283	0.7657
		10	25	0.6316	0.9981	0.6280	0.7751
	400	5	10	0.6955	1.0000	0.6890	0.7536
		5	25	0.6504	0.9993	0.6467	0.7678
		10	25	0.6429	0.9995	0.6449	0.7614
70	100	5	10	0.6442	0.9850	0.6410	0.9070
		5	25	0.6249	0.9824	0.6262	0.9254
		10	25	0.6150	0.9826	0.6124	0.9470
	200	5	10	0.6891	0.9983	0.6870	0.8948
		5	25	0.6543	0.9986	0.6528	0.9359
		10	25	0.6449	0.9986	0.6429	0.9488
	300	5	10	0.7273	0.9999	0.7265	0.8893
		5	25	0.6803	0.9996	0.6780	0.9275
		10	25	0.6713	0.9999	0.6709	0.9452
	400	5	10	0.7638	1.0000	0.7608	0.8939
		5	25	0.7071	1.0000	0.7054	0.9250
		10	25	0.6988	1.0000	0.6971	0.9351

Table B.43 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{12}$ and $\theta = 0.9$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.3780	0.0584	0.3795	0.2920
		10	0.3941	0.0510	0.3973	0.2689
		25	0.4065	0.0840	0.4085	0.2590
	200	5	0.3283	0.0113	0.3287	0.2758
		10	0.3552	0.0105	0.3554	0.2678
		25	0.3699	0.0216	0.3725	0.2554
	300	5	0.2866	0.0019	0.2886	0.2495
		10	0.3179	0.0015	0.3215	0.2537
		25	0.3450	0.0059	0.3468	0.2515
	400	5	0.2641	0.0007	0.2664	0.2326
		10	0.2935	0.0005	0.2967	0.2407
		25	0.3193	0.0019	0.3196	0.2405
70	100	5	0.3267	0.0407	0.3283	0.1619
		10	0.3466	0.0330	0.3463	0.1371
		25	0.3641	0.0407	0.3635	0.1285
	200	5	0.2705	0.0065	0.2749	0.1592
		10	0.2985	0.0049	0.3008	0.1326
		25	0.3211	0.0063	0.3231	0.1138
	300	5	0.2283	0.0009	0.2319	0.1431
		10	0.2628	0.0003	0.2647	0.1243
		25	0.2879	0.0006	0.2884	0.1082
	400	5	0.1943	0.0002	0.1967	0.1294
		10	0.2328	0.0001	0.2354	0.1168
		25	0.2601	0.0001	0.2616	0.1026

Table B.44 The misclassification rate (M) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{12}$ and $\theta = 0.9$ with mixed block sizes

n	***p***	p_i	p_j	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	10	0.3842	0.1204	0.3883	0.2992
		5	25	0.3969	0.1567	0.3986	0.2981
		10	25	0.3922	0.1565	0.3929	0.2863
	200	5	10	0.3398	0.0426	0.3432	0.2807
		5	25	0.3530	0.0733	0.3544	0.2773
		10	25	0.3647	0.0781	0.3636	0.2770
	300	5	10	0.3176	0.0190	0.3199	0.2686
		5	25	0.3254	0.0362	0.3273	0.2632
		10	25	0.3318	0.0346	0.3355	0.2653
	400	5	10	0.2848	0.0075	0.2893	0.2466
		5	25	0.2989	0.0191	0.3015	0.2449
		10	25	0.3066	0.0187	0.3104	0.2496
70	100	5	10	0.3444	0.0875	0.3452	0.1814
		5	25	0.3496	0.1001	0.3493	0.1816
		10	25	0.3575	0.0986	0.3594	0.1724
	200	5	10	0.2921	0.0251	0.2929	0.1647
		5	25	0.3011	0.0305	0.3028	0.1567
		10	25	0.3178	0.0296	0.3181	0.1502
	300	5	10	0.2528	0.0080	0.2523	0.1504
		5	25	0.2644	0.0094	0.2661	0.1372
		10	25	0.2789	0.0095	0.2799	0.1364
	400	5	10	0.2177	0.0027	0.2198	0.1358
		5	25	0.2375	0.0031	0.2397	0.1314
		10	25	0.2449	0.0025	0.2476	0.1245

Table B.45 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{12}$ and $\theta = 0.9$ with equal block sizes

n	***p***	p_i	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.6190	0.9401	0.6155	0.7045
		10	0.6067	0.9477	0.6060	0.7313
		25	0.5923	0.9142	0.5905	0.7398
	200	5	0.6767	0.9890	0.6776	0.7290
		10	0.6498	0.9895	0.6480	0.7371
		25	0.6365	0.9782	0.6341	0.7510
	300	5	0.7156	0.9975	0.7139	0.7563
		10	0.6830	0.9981	0.6780	0.7481
		25	0.6553	0.9940	0.6512	0.7502
	400	5	0.7394	0.9995	0.7381	0.7701
		10	0.7063	0.9997	0.7046	0.7606
		25	0.6804	0.9981	0.6777	0.7596
70	100	5	0.6784	0.9587	0.6774	0.8437
		10	0.6558	0.9668	0.6573	0.8669
		25	0.6387	0.9584	0.6393	0.8737
	200	5	0.7307	0.9943	0.7263	0.8420
		10	0.7013	0.9955	0.7003	0.8687
		25	0.6804	0.9936	0.6782	0.8856
	300	5	0.7698	0.9992	0.7652	0.8548
		10	0.7349	0.9997	0.7344	0.8752
		25	0.7112	0.9995	0.7109	0.8919
	400	5	0.8019	0.9998	0.7991	0.8677
		10	0.7644	1.0000	0.7627	0.8823
		25	0.7376	0.9999	0.7367	0.8954

Table B.46 The sensitivity (SE) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{12}$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6186	0.8807	0.6132	0.7018
		5	25	0.5988	0.8428	0.5974	0.6993
		10	25	0.6085	0.8486	0.6072	0.7155
	200	5	10	0.6702	0.9568	0.6646	0.7248
		5	25	0.6520	0.9243	0.6523	0.7277
		10	25	0.6372	0.9228	0.6373	0.7265
	300	5	10	0.6844	0.9793	0.6829	0.7342
		5	25	0.6724	0.9635	0.6704	0.7355
		10	25	0.6655	0.9659	0.6610	0.7309
	400	5	10	0.7150	0.9921	0.7109	0.7554
		5	25	0.7028	0.9818	0.7011	0.7565
		10	25	0.6930	0.9812	0.6897	0.7491
70	100	5	10	0.6551	0.9101	0.6550	0.8165
		5	25	0.6527	0.9014	0.6534	0.8189
		10	25	0.6426	0.9001	0.6413	0.8282
	200	5	10	0.7091	0.9745	0.7075	0.8351
		5	25	0.6997	0.9693	0.6982	0.8416
		10	25	0.6833	0.9695	0.6836	0.8497
	300	5	10	0.7469	0.9922	0.7473	0.8507
		5	25	0.7376	0.9918	0.7363	0.8646
		10	25	0.7212	0.9902	0.7201	0.8624
	400	5	10	0.7803	0.9971	0.7777	0.8623
		5	25	0.7606	0.9970	0.7578	0.8684
		10	25	0.7530	0.9978	0.7498	0.8717

Table B.47 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{12}$ and $\theta = 0.9$ with equal block sizes

n	***p***	***p_i***	**TA,TB,TC**	**BD**	**DI**	**SK**
35	100	5	0.6251	0.9432	0.6255	0.7116
		10	0.6051	0.9504	0.5995	0.7309
		25	0.5947	0.9178	0.5925	0.7422
	200	5	0.6668	0.9885	0.6650	0.7194
		10	0.6399	0.9896	0.6413	0.7273
		25	0.6238	0.9787	0.6209	0.7383
	300	5	0.7112	0.9988	0.7090	0.7448
		10	0.6813	0.9989	0.6790	0.7445
		25	0.6548	0.9942	0.6553	0.7468
	400	5	0.7324	0.9992	0.7291	0.7648
		10	0.7068	0.9994	0.7021	0.7581
		25	0.6810	0.9981	0.6832	0.7595
70	100	5	0.6682	0.9600	0.6660	0.8325
		10	0.6510	0.9673	0.6501	0.8590
		25	0.6331	0.9603	0.6337	0.8693
	200	5	0.7285	0.9927	0.7241	0.8396
		10	0.7019	0.9948	0.6982	0.8662
		25	0.6775	0.9938	0.6757	0.8868
	300	5	0.7737	0.9990	0.7710	0.8591
		10	0.7396	0.9997	0.7363	0.8762
		25	0.7131	0.9994	0.7124	0.8917
	400	5	0.8095	0.9998	0.8075	0.8737
		10	0.7701	1.0000	0.7666	0.8843
		25	0.7424	1.0000	0.7402	0.8995

Table B.48 The specificity (SP) of TA, TB, TC, BD, SK, and DI when $\Sigma = \Sigma_{12}$ and $\theta = 0.9$ with mixed block sizes

n	p	p_i	p_j	TA,TB,TC	BD	DI	SK
35	100	5	10	0.6130	0.8785	0.6103	0.6999
		5	25	0.6074	0.8438	0.6055	0.7046
		10	25	0.6071	0.8385	0.6071	0.7120
	200	5	10	0.6502	0.9580	0.6490	0.7138
		5	25	0.6420	0.9292	0.6389	0.7177
		10	25	0.6334	0.9211	0.6356	0.7196
	300	5	10	0.6805	0.9828	0.6773	0.7287
		5	25	0.6769	0.9641	0.6750	0.7381
		10	25	0.6710	0.9649	0.6681	0.7386
	400	5	10	0.7155	0.9929	0.7106	0.7514
		5	25	0.6994	0.9801	0.6960	0.7537
		10	25	0.6938	0.9814	0.6896	0.7517
70	100	5	10	0.6562	0.9150	0.6546	0.8208
		5	25	0.6483	0.8985	0.6481	0.8179
		10	25	0.6424	0.9029	0.6401	0.8270
	200	5	10	0.7068	0.9755	0.7067	0.8355
		5	25	0.6981	0.9699	0.6962	0.8451
		10	25	0.6811	0.9715	0.6802	0.8500
	300	5	10	0.7475	0.9919	0.7481	0.8486
		5	25	0.7337	0.9895	0.7315	0.8611
		10	25	0.7211	0.9909	0.7202	0.8650
	400	5	10	0.7845	0.9976	0.7828	0.8662
		5	25	0.7644	0.9968	0.7628	0.8689
		10	25	0.7573	0.9973	0.7550	0.8794

BIOGRAPHY

NAME	Mr. Poompong Kaewumpai
ACADEMIC BACKGROUND	B.Sc. (Mathematics), Mahidol University, Thailand, 2005 M.S. (Statistics), Chulalongkorn University, Thailand, 2008
PRESENT POSITION	Data Scientist, True Corporation Public Company Limited
EXPERIENCE	2014 - 2017 Consultant, Predictive Analytics, Sift Analytics Group (Thailand) Co., Ltd. 2009 Teaching Assistant, Walailuk University