On Linear Discriminant Analysis with Adaptive Ridge Classification Rules

by

Wei-Liem Loh
Purdue University

Technical Report #93-7

Department of Statistics
Purdue University

January 1993

ON LINEAR DISCRIMINANT ANALYSIS WITH ADAPTIVE RIDGE CLASSIFICATION RULES

BY WEI-LIEM LOH

Department of Statistics, Purdue University, West Lafayette, Indiana

Consider the problem in which we have a sample from each of two multivariate normal populations with equal covariance matrices and we wish to classify a new observation as coming from one of the two populations. It is widely regarded that the most commonly used discriminant procedure for this purpose was introduced by Anderson (1951). During the last fifteen years, promising linear adaptive ridge classification rules have been proposed as alternatives where the ridge parameters are chosen by sample reuse methods like cross-validation and bootstrap.

This article undertakes a large sample study of the relationship between the optimal ridge parameter and the population parameters. The results suggest a new linear adaptive ridge classification procedure which has a simple closed form expression for the ridge parameter. A Monte Carlo study is used to compare its error rate with that of the other classification rules previously mentioned.

Some key words: Linear discriminant analysis; Monte Carlo experiments; Multivariate normal distribution; Error rate; Ridge classification rules; Cross-validation ; Bootstrap.

1 Introduction

Consider the problem in which we have a sample from each of two multivariate normal populations and we wish to classify another observation as coming from one of the two populations. More precisely, suppose we have a training sample $X_1^{(1)}, \ldots, X_{n_1}^{(1)}$ from $N_p(\mu^{(1)}, \Sigma)$ and a training sample $X_1^{(2)}, \ldots, X_{n_2}^{(2)}$ from $N_p(\mu^{(2)}, \Sigma)$. Without loss of generality we shall assume that Σ is nonsingular since singular cases can always be made nonsingular by an appropriate reduction of dimension. We wish to classify another observation X as coming from one of these two distributions where we assume apriori that it is equally likely that X comes from either $N_p(\mu^{(1)}, \Sigma)$ or $N_p(\mu^{(2)}, \Sigma)$.

In the case where the two distributions are completely known, Wald (1944) proved that the classification procedure which minimizes the expected error (misclassification) rate is given by the Fisher's discriminant function, namely: Classify X into $N_p(\mu^{(1)}, \Sigma)$ if

$$[X - (\mu^{(1)} + \mu^{(2)})/2]'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) \geq 0, \tag{1}$$

and into $N_p(\mu^{(2)}, \Sigma)$ otherwise.

However very often, the parameters of the two distributions $N_p(\mu^{(1)}, \Sigma)$ and $N_p(\mu^{(2)}, \Sigma)$ are unknown and need to be estimated from the training samples. Anderson (1951) proposed using the unbiased estimates of $\mu^{(1)}$, $\mu^{(2)}$ and Σ in (1). In this case we obtain the usual (Anderson's) linear discriminant rule, that is, classify an observation X into $N_p(\mu^{(1)}, \Sigma)$ if

$$[X - (\bar{X}^{(1)} + \bar{X}^{(2)})/2]'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) \geq 0, \tag{2}$$

and into $N_p(\mu^{(2)}, \Sigma)$ otherwise, where

$$\bar{X}^{(i)} = \sum_{j=1}^{n_i} X_j^{(i)}/n_i, \qquad i = 1, 2,$$

$$S = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_j^{(i)} - \bar{X}^{(i)})(X_j^{(i)} - \bar{X}^{(i)})'/(n_1 + n_2 - 2).$$

As noted by Gnanadesikan, et. al. (1989), the above procedure is arguably the most widely used rule at present for classifying an observation X into one of two multivariate populations. The main reasons for its wide spread use are its simplicity, the ready availability of computer package programs and reasonable robustness against model violations especially for moderate sample sizes. Excellent accounts of this procedure can be found, for example, in Anderson (1984) and McLachlan (1992).

However as is well known, the usual linear discriminant rule does not share the same optimality properties as (1). Indeed except for asymptotic optimality and in special circumstances (Das Gupta 1965), no finite sample optimality property has yet been found (Friedman 1989). There are at least two heuristics in which we can hope to improve on the usual linear discriminant rule. The first is to observe that there is significant distortion in the eigenvalues of the sample covariance matrix as estimates of the eigenvalues of Σ. This is most apparent when the eigenvalues of Σ are all equal. This phenomenon was first noticed by Stein (1956) and since then there has been an enormous amount of effort in getting better estimates for Σ. James and Stein (1961), Stein (1975), Efron and Morris (1976), Olkin and Selliah (1977), Haff (1986), (1991), Lin and Perlman (1984) and Dey and Srinivasan (1985), (1991) have studied this approach by minimizing some particular loss criterion (which is often some form of squared error loss) on the eigenvalue estimates. Unfortunately as far as we are aware of, none of the loss criteria that have been studied analytically is directly related to the error rate of a classification rule.

The second heuristic is more intimately related to the classification problem of interest. It has been observed (DiPillo 1976, 1977, 1979; Peck and Van Ness 1982; Friedman 1989) that the relationship between $\mu^{(1)} - \mu^{(2)}$ and the eigenvectors of Σ has a significant influence on the estimates of the eigenvalues of Σ. In particular, suppose that the eigenvalues of Σ are well dispersed. In this case the first heuristic would lead us to believe that S should be a good estimator of Σ. However the error rate of the usual rule can still be improved by using an estimate of Σ that is more spherical than S if $\mu^{(1)} - \mu^{(2)}$ lies near the subspace generated by the eigenvector corresponding to the largest eigenvalue of Σ. This can be substantiated by Monte Carlo simulations.

To essentially take advantage of these ideas, linear adaptive ridge classification rules have been proposed in the literature. To be specific, we shall focus on the following ridge classification rules: Namely classify an observation X into $N_p(\mu^{(1)}, \Sigma)$ if

$$[X - (\bar{X}^{(1)} + \bar{X}^{(2)})/2]'[(1-\hat{\lambda})S + \frac{\hat{\lambda}}{p}(\mathrm{tr}S)I]^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) \geq 0, \tag{3}$$

and into $N_p(\mu^{(2)}, \Sigma)$ otherwise. Here I is the $p \times p$ identity matrix, tr denotes the trace of a matrix and $0 \leq \hat{\lambda} \leq 1$ is a ridge parameter that depends on the training samples. To choose a good value for $\hat{\lambda}$, two sample reuse methods have been suggested: cross-validation (Geisser 1977; Lachenbruch 1975) and bootstrap (Efron 1983; Peck and Van Ness 1982). We remark that (3) is a special case of the regularized discriminant rule of Friedman (1989) in which he used leave-one-out cross-validation to determine the value of the ridge parameter $\hat{\lambda}$.

This article undertakes a large sample study of the relationship between the optimal ridge parameter, λ_{OPT}, and the population parameters $\mu^{(1)}$, $\mu^{(2)}$ and Σ. (Here the optimal ridge parameter is that value of $0 \leq \hat{\lambda} \leq 1$ which minimizes the error rate of the ridge classification rule.) Let

$$\Delta_j^2 = (\mu^{(1)} - \mu^{(2)})'\Sigma^{-j}(\mu^{(1)} - \mu^{(2)}), \quad j = 1, 2, \cdots, \tag{4}$$

where Δ_1^2 is the usual Mahalanobis squared distance between the multivariate normal distributions $N_p(\mu^{(1)}, \Sigma)$ and $N_p(\mu^{(2)}, \Sigma)$. Then λ_{OPT} is found in Section 2 to be of the form

$$\begin{aligned}\lambda_{OPT} &= \frac{p}{\mathrm{tr}\Sigma}\{(\frac{n_1+n_2}{n_1 n_2})[(\mathrm{tr}\Sigma^{-1}) - p\Delta_1^{-2}\Delta_2^2] \\ &+ \frac{1}{n_1+n_2-2}[\Delta_1^2(\mathrm{tr}\Sigma^{-1}) - \Delta_2^2]\}/(\Delta_3^2 - \Delta_1^{-2}\Delta_2^4) + O_{3/2}, \end{aligned} \tag{5}$$

where $O_{3/2}$ represents a generic term of the 3/2th order with respect to $1/n_1$ and $1/n_2$. The above expression for λ_{OPT} cannot be used directly in practice since it is given in terms of the unknown population parameters. However as in (2), the unknown parameters $\mu^{(1)}$, $\mu^{(2)}$ and Σ in (5) can be estimated by the unbiased estimators $\bar{X}^{(1)}$, $\bar{X}^{(2)}$ and S respectively. This suggest the following linear adaptive ridge classification rule: Define

$$D_j^2 = (\bar{X}^{(1)} - \bar{X}^{(2)})' S^{-j} (\bar{X}^{(1)} - \bar{X}^{(2)}), \quad j = 1, 2, \cdots,$$

and

$$\begin{aligned}\hat{\lambda}^*_{ASYMP} &= \frac{p}{\mathrm{tr}S}\{(\frac{n_1+n_2}{n_1 n_2})[(\mathrm{tr}S^{-1}) - pD_1^{-2}D_2^2] \\ &+ \frac{1}{n_1+n_2-2}[D_1^2(\mathrm{tr}S^{-1}) - D_2^2]\}/(D_3^2 - D_1^{-2}D_2^4).\end{aligned}$$

Since the natural range of $\hat{\lambda}$ is $[0, 1]$, we truncate $\hat{\lambda}^*_{ASYMPT}$ at 0 and 1 which leads us to

$$\hat{\lambda}_{ASYMP} = \begin{cases} \hat{\lambda}^*_{ASYMP} & \text{if } 0 \le \hat{\lambda}^*_{ASYMP} \le 1, \\ 0 & \text{if } \hat{\lambda}^*_{ASYMP} < 0, \\ 1 & \text{if } \hat{\lambda}^*_{ASYMP} > 1. \end{cases}$$

Now classify an observation X into $N_p(\mu^{(1)}, \Sigma)$ if

$$[X - (\bar{X}^{(1)} + \bar{X}^{(2)})/2]'[(1 - \hat{\lambda}_{ASYMP})S + \frac{\hat{\lambda}_{ASYMP}}{p}(\mathrm{tr}S)I]^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) \ge 0, \tag{6}$$

and into $N_p(\mu^{(2)}, \Sigma)$ otherwise. Even though in this case $\hat{\lambda}_{ASYMP}$ is available in closed form (which should provide a better chance of getting a better understanding of the procedure than the rules that use cross-validation and bootstrap in determining $\hat{\lambda}$), the error rate of this rule appears to be still analytically intractable in terms of a nonasymptotic theoretical treatment. This is not surprising as the usual linear discriminant rule suffers from the same difficulty as well.

As such a Monte Carlo study is used instead to compare the expected error rate of the linear classification rules (2), (3) and (6) discussed above. This simulation study is described in detail in Section 3.

We end this section with the following remark. As it stands, the justification of rule (6) is asymptotic, i.e. we assume that n_1 and n_2 are large relative to a fixed set of parameters $\mu^{(1)}$, $\mu^{(2)}$ and Σ. On closer examination of the argument in Section 2, we observe that the asymptotics break down when Δ_1 is either very large or very small compared to n_1 and n_2. In the case where Δ_1 is very large, it is clear that almost perfect discrimination can be achieved by all reasonable rules and that in the case where Δ_1 is very small, all reasonable rules shall have error rates near 50%. Thus we do not expect the error rates of the usual rule and rule (6) to differ by too much when the asymptotic justification for (6) is invalid.

2 Asymptotic Error Rate Expansion

We begin with the ridge classification procedure: Classify an observation X into $N_p(\mu^{(1)}, \Sigma)$ if

$$W_\lambda = [X - (\bar{X}^{(1)} + \bar{X}^{(2)})/2]'[(1-\lambda)S + \frac{\lambda}{p}(\mathrm{tr}S)I]^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) \geq 0, \tag{7}$$

and into $N_p(\mu^{(2)}, \Sigma)$ otherwise, where $0 \leq \lambda \leq 1$ is a constant. We remark that (7) reduces to the usual linear discriminant rule when $\lambda = 0$ and that Rodriguez (1988) studied the admissibility and unbiasedness of these rules. Following the method of Okamoto (1963), we shall establish an asymptotic expansion for the expected error rate difference between the above rule and the usual discriminant rule (2). Let Δ_j, $j = 1, 2, \ldots$, be given by (4),

$$\begin{aligned} \xi^{(j)} &= (\xi_1^{(j)}, \ldots, \xi_p^{(j)})', \quad j = 1, 2, \\ \partial^{(j)} &= (\partial/\partial\xi_1^{(j)}, \ldots, \partial/\partial\xi_p^{(j)})', \quad j = 1, 2, \end{aligned}$$

and ∂^* be the $p \times p$ matrix whose (j,k)th element is given by $(1/2)(1 + \delta_{j,k})\partial/\partial\theta_{j,k}$ where $\theta_{j,k}$ is the (j,k)th element of a $p \times p$ positive definite matrix Θ and $\delta_{j,k}$ the Kronecker delta. We consider the characteristic function

$$\psi_\lambda(t) = E\{\exp[it\Delta_1^{-1}(W_\lambda - \Delta_1^2/2)]\}$$

of the random variable $\Delta_1^{-1}(W_\lambda - \Delta_1^2/2)$ when $X \sim N_p(\mu^{(1)}, \Sigma)$. Then by conditioning on $\bar{X}^{(1)}$, $\bar{X}^{(2)}$ and S, we get

$$\begin{aligned} \psi_\lambda(t) &= E\exp\{-\frac{it}{2}\Delta_1 + it\Delta_1^{-1}[\mu^{(1)} - \frac{1}{2}(\bar{X}^{(1)} + \bar{X}^{(2)})]'[(1-\lambda)S + \frac{\lambda}{p}(\mathrm{tr}S)I]^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) \\ &\quad -\frac{t^2}{2}\Delta_1^{-2}(\bar{X}^{(1)} - \bar{X}^{(2)})'[(1-\lambda)S + \frac{\lambda}{p}(\mathrm{tr}S)I]^{-1}\Sigma[(1-\lambda)S + \frac{\lambda}{p}(\mathrm{tr}S)I]^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)})\} \\ &= E\Psi(\bar{X}^{(1)}, \bar{X}^{(2)}, (1-\lambda)S + \frac{\lambda}{p}(\mathrm{tr}S)I), \quad \text{say.} \end{aligned}$$

Since Ψ is an analytic function, we expand Ψ as a Taylor series about the point $(\mu^{(1)}, \mu^{(2)}, \Sigma)$. Thus

$$\begin{aligned} \psi_\lambda(t) &= E\exp\{(\bar{X}^{(1)} - \mu^{(1)})'\partial^{(1)} + (\bar{X}^{(2)} - \mu^{(2)})'\partial^{(2)} \\ &\quad + \mathrm{tr}[(1-\lambda)S + \frac{\lambda}{p}(\mathrm{tr}S)I - \Sigma]\partial^*\}\Psi(\xi^{(1)}, \xi^{(2)}, \Theta)|_{(\xi^{(1)}, \xi^{(2)}, \Theta) = (\mu^{(1)}, \mu^{(2)}, \Sigma)}. \end{aligned}$$

Using the moment generating function formulas for $\bar{X}^{(1)}$, $\bar{X}^{(2)}$ and S, we have

$$\begin{aligned} \psi_\lambda(t) &= \exp\{\frac{1}{2n_1}\partial^{(1)\prime}\Sigma\partial^{(1)} + \frac{1}{n_2}\partial^{(2)\prime}\Sigma\partial^{(2)} - (\mathrm{tr}\Sigma\partial^*) - \frac{n_1 + n_2 - 2}{2}\log|I - \frac{2(1-\lambda)\Sigma\partial^*}{n_1 + n_1 - 2}| \\ &\quad - \frac{n_1 + n_2 - 2}{2}\log|I - \frac{2\Sigma\lambda\mathrm{tr}\partial^*}{p(n_1 + n_2 - 2)}|\}\Psi(\xi^{(1)}, \xi^{(2)}, \Theta)|_{(\xi^{(1)}, \xi^{(2)}, \Theta) = (\mu^{(1)}, \mu^{(2)}, \Sigma)}. \end{aligned} \tag{8}$$

Substituting the expansion

$$-\log|I - \Theta| = \mathrm{tr}(\Theta) + \frac{1}{2}\mathrm{tr}(\Theta^2) + \frac{1}{3}\mathrm{tr}(\Theta^3) + \cdots,$$

in (8), we get after some simplification

$$\begin{aligned}
&\psi_\lambda(t) - \psi_0(t) \\
= &\ \{\frac{\lambda}{p}(\mathrm{tr}\Sigma)(\mathrm{tr}\partial^*) - \lambda(\mathrm{tr}\Sigma\partial^*) - \frac{2\lambda}{n_1+n_2-2}(\mathrm{tr}\Sigma\partial^*\Sigma\partial^*)\}\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)} \\
&+\frac{1}{2}\{[\frac{\lambda}{p}(\mathrm{tr}\Sigma)(\mathrm{tr}\partial^*) - \lambda(\mathrm{tr}\Sigma\partial^*)]^2 + 2[\frac{\lambda}{p}(\mathrm{tr}\Sigma)(\mathrm{tr}\partial^*) - \lambda(\mathrm{tr}\Sigma\partial^*)][\frac{1}{2n_1}\partial^{(1)'}\Sigma\partial^{(1)} \\
&+\frac{1}{2n_2}\partial^{(2)'}\Sigma\partial^{(2)} + \frac{1}{n_1+n_2-2}(\mathrm{tr}\Sigma\partial^*\Sigma\partial^*)]\}\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)} + O_3, \quad (9)
\end{aligned}$$

where O_3 stands for a generic term of the third order with respect to $1/n_1$ and $1/n_2$. Now it follows from the Fourier inversion formula, (9) and Lemma 1 (see Appendix) that

$$\begin{aligned}
&P(W_\lambda < 0|X \sim N_p(\mu^{(1)},\Sigma)) - P(W_0 < 0|X \sim N_p(\mu^{(1)},\Sigma)) \\
= &\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}[\psi_\lambda(t) - \psi_0(t)]dtdx \\
= &\ \phi(\frac{\Delta_1}{2})(\mathrm{tr}\Sigma)\{\frac{\lambda^2}{4p^2}(\mathrm{tr}\Sigma)(\Delta_1^{-1}\Delta_3^2 - \Delta_1^{-3}\Delta_2^4) + \frac{\lambda}{pn_1}[p\Delta_1^{-3}\Delta_2^2 - \Delta_1^{-1}(\mathrm{tr}\Sigma^{-1})] \\
&+\frac{\lambda}{2p(n_1+n_2-2)}[\Delta_1^{-1}\Delta_2^2 - \Delta_1(\mathrm{tr}\Sigma^{-1})]\} + O_3, \quad (10)
\end{aligned}$$

where ϕ denotes the probability density function of a standard normal random variable. Similarly if $X \sim N_p(\mu^{(2)},\Sigma)$, we have

$$\begin{aligned}
&P(W_\lambda > 0|X \sim N_p(\mu^{(2)},\Sigma)) - P(W_0 > 0|X \sim N_p(\mu^{(2)},\Sigma)) \\
= &\ \phi(\frac{\Delta_1}{2})(\mathrm{tr}\Sigma)\{\frac{\lambda^2}{4p^2}(\mathrm{tr}\Sigma)(\Delta_1^{-1}\Delta_3^2 - \Delta_1^{-3}\Delta_2^4) + \frac{\lambda}{pn_2}[p\Delta_1^{-3}\Delta_2^2 - \Delta_1^{-1}(\mathrm{tr}\Sigma^{-1})] \\
&+\frac{\lambda}{2p(n_1+n_2-2)}[\Delta_1^{-1}\Delta_2^2 - \Delta_1(\mathrm{tr}\Sigma^{-1})]\} + O_3. \quad (11)
\end{aligned}$$

Thus we conclude from (10) and (11) that the expected error rate difference between rule (7) and the usual linear discriminant rule (2) is given by

$$\begin{aligned}
&\frac{1}{2}\{P(W_\lambda < 0|X \sim N_p(\mu^{(1)},\Sigma)) - P(W_0 < 0|X \sim N_p(\mu^{(1)},\Sigma)) \\
&+P(W_\lambda > 0|X \sim N_p(\mu^{(2)},\Sigma)) - P(W_0 > 0|X \sim N_p(\mu^{(2)},\Sigma))\} \\
= &\ \phi(\frac{\Delta_1}{2})(\mathrm{tr}\Sigma)\{\frac{\lambda^2}{2p^2}(\mathrm{tr}\Sigma)(\Delta_1^{-1}\Delta_3^2 - \Delta_1^{-3}\Delta_2^4) + \frac{\lambda(n_1+n_2)}{pn_1n_2}[p\Delta_1^{-3}\Delta_2^2 - \Delta_1^{-1}(\mathrm{tr}\Sigma^{-1})] \\
&+\frac{\lambda}{p(n_1+n_2-2)}[\Delta_1^{-1}\Delta_2^2 - \Delta_1(\mathrm{tr}\Sigma^{-1})]\} + O_3. \quad (12)
\end{aligned}$$

The λ that minimizes the right hand side of (12) is

$$\begin{aligned}
\lambda_{OPT} = &\ \frac{p}{\mathrm{tr}\Sigma}\{\frac{n_1+n_2}{n_1n_2}[(\mathrm{tr}\Sigma^{-1}) - p\Delta_1^{-2}\Delta_2^2] \\
&+\frac{1}{n_1+n_2-2}[\Delta_1^2(\mathrm{tr}\Sigma^{-1}) - \Delta_2^2]\}/(\Delta_3^2 - \Delta_1^{-2}\Delta_2^4) + O_{3/2}.
\end{aligned}$$

This proves (5).

3 Monte Carlo Study

For simplicity of notation, we use $USUAL$ and $ASYMP$ to denote the usual discriminant rule (2) and the adaptive ridge classification rule given by (6) respectively. In the case of rule (3), we shall use the following two computer intensive sample reuse techniques to estimate the optimal ridge parameter: We assume that training samples (of sizes n_1 and n_2) are available from each of the two multivariate normal populations and that we restrict the ridge parameter $\hat{\lambda}$ to take on only values in the grid $\{0.0, 0.1, \ldots, 1.0\}$.

Leave-one-out cross-validation. Consider rules of the form (3) where $\hat{\lambda}$ is a constant taking on a value in the above grid. The error rate of each of these rules is estimated by leave-one-out cross-validation. The $\hat{\lambda}$ which minimizes the cross-validated error rate is then used as the estimate for the optimal ridge parameter.

Bootstrap. Consider rules of the form (3) where $\hat{\lambda}$ is a constant taking on a value in the above grid. For each of these rules, we estimate the error rate with the following bootstrap algorithm.

1. First compute $\hat{F}_j$, $j = 1, 2$, the empirical distribution function for each of the two training data sets.

2. Randomly generate n_1 and n_2 bootstrap samples from $\hat{F}_1$ and $\hat{F}_2$ respectively. Estimate the error rate of this rule by classify the original training data by using the rule with the bootstrap sample as the training data.

Step 2 is performed 25 times and the error rates are then averaged and used as an estimate of the true error rate. The $\hat{\lambda}$ which minimizes the bootstrapped error rate is then used as the estimate for the optimal ridge parameter.

For simplicity, we shall use LOO and $BOOT$ to denote the procedure as given by (3) where $\hat{\lambda}$ is determined by the cross-validation and the bootstrap algorithm respectively.

We observe that there is a considerable amount of invariance in the four discriminant procedures $USUAL$, $ASYMP$, LOO and $BOOT$. All these rules are location, scale and orthogonal equivariant. Thus without loss of generality, in our simulations we shall take $\mu^{(2)} = (0, \cdots, 0)'$ and Σ to be a diagonal matrix $\text{diag}(\sigma_1^2, \cdots, \sigma_p^2)$ with $\sigma_1 \geq \cdots \geq \sigma_p = 1.0$.

In this section we shall report on two Monte Carlo experiments which compare the expected error rates of the rules $USUAL$, $ASYMP$, LOO and $BOOT$. In the first experiment, we take $n_1 = n_2 = 25$, $p = 10$ and in the second experiment we take $n_1 = n_2 = 30$, $p = 5$. Each experiment consists of 500 independent replications of the following procedure. First, training samples of sizes n_1 and n_2 are generated from $N_p(\mu^{(1)}, \Sigma)$ and $N_p(\mu^{(2)}, \Sigma)$ respectively via the IMSL subroutine DRNNOA. We observe that conditional on $\bar{X}^{(1)}$, $\bar{X}^{(2)}$ and S, the error rate for any of the four procedures is given by

$$(1/2)\Phi(z_1) + (1/2)[1 - \Phi(z_2)], \tag{13}$$

where Φ is the distribution function of the standard normal distribution and for $j = 1, 2$,

$$z_j = [\mu^{(j)} - (\bar{X}^{(1)} + \bar{X}^{(2)})/2]'\hat{\Sigma}^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)})/[(\bar{X}^{(1)} - \bar{X}^{(2)})'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)})]^{1/2}.$$

Here $\hat{\Sigma}$ is an appropriate estimate of Σ which depends on the particular classification rule. The IMSL function DNORDF is used to evaluate (13). This reduces the Monte Carlo variance of the simulation. The average error rates and their standard deviations of the four procedures are computed over the 500 replications. These are reported in Tables 1 to 3 (for experiment 1) and Tables 4 to 6 (for experiment 2). We also observe that the average error rates of these four procedures should be positively correlated as they share the same training samples. Thus the estimated standard deviation (as given in Tables 1 to 6) is probably a conservative indicator of the variability of the relative magnitude of the average error rates.

The two Monte Carlo experiments focus on three configurations of the parameter space, namely (i) $\mu^{(1)} - \mu^{(2)} = (0, \ldots, 0, x)$, (ii) $\mu^{(1)} - \mu^{(2)} = (x, 0, \ldots, 0)$ and (iii) $\mu^{(1)} - \mu^{(2)} = x(\sigma_1, \ldots, \sigma_p)$, where x is a suitable numerical constant and Σ as previously mentioned is a diagonal matrix with eigenvalues $\sigma_1^2 \geq \cdots \geq \sigma_p^2 = 1.0$.

(i) *First configuration*: $\mu^{(1)} - \mu^{(2)} = (0, \ldots, 0, x)$. In this case the difference of the two population mean vectors lies in the subspace associated with the smallest eigenvalue of the population covariance matrix. This is a situation that is widely recognized to be most favorable to *USUAL*. This is substantiated by Tables 1 and 4 which reports on the simulations of Experiment 1 and 2 respectively. These tables indicate that no one classification rule dominates another in terms of error rate. *LOO* does significantly worse than *USUAL* in those situations when Σ is highly elliptical, e.g. in Table 1 when $\Sigma = \text{diag}(10^9, 10^8, \ldots, 10, 1)$ and $(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 1.00$, the error rate of *LOO* is 39.06% and that of *USUAL* is 36.97%. In contrast, *ASYMP* and *BOOT* does slightly worse than *USUAL* when Σ is non-spherical with *BOOT* having a slight edge over *ASYMP*. However when $\Sigma = I$ Tables 1 and 4 show that all three alternative classification rules do dramatically better than *USUAL* and that generally *ASYMP* does best among the different rules; e.g. when $(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 9.00$ in Table 1, the error rates for *USUAL*, *ASYMP*, *LOO* and *BOOT* are respectively 9.98%, 8.31%, 8.68% and 8.46%.

(ii) *Second configuration*: $\mu^{(1)} - \mu^{(2)} = (x, 0, \ldots, 0)$. In this case the difference between the two population mean vectors lies in the subspace associated with the largest eigenvalue of Σ. This is a situation widely recognized to be most unfavorable to *USUAL*. Indeed, Tables 2 and 5 show this to be the case. All three alternative classification rules *ASYMP*, *LOO* and *BOOT* do much better than *USUAL* in terms of error rate with *LOO* having generally the smallest error rate and *BOOT* having a slight edge over *ASYMP*.

(iii) *Third configuration*: $\mu^{(1)} - \mu^{(2)} = x(\sigma_1, \ldots, \sigma_p)$. This is a case intermediate between the two previous cases and the error rates of *USUAL*, *ASYMP*, *LOO* and *BOOT* are reported in Tables 3 and 6. Indeed in this case the *ASYMP* error rate generally (up to Monte Carlo error) falls in between those reported for the two previous situations. However, due to the discretization of the possible values of $\hat{\lambda}$, this is not the case for *LOO* and *BOOT*. There are situations in which *LOO* and *BOOT* do significantly worse than *USUAL*; e.g. in Table 3 when $\Sigma = \text{diag}(512, 256, \ldots, 2, 1)$ with $(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 16.00$, the error rates for *USUAL*, *ASYMP*, *LOO* and *BOOT* are respectively 4.10%, 4.03%, 4.59% and 4.49%. The discretization effect can be reduced by making the grid for $\hat{\lambda}$ finer. However further simulations indicate that this leads to a deterioration of the performance of *LOO* and *BOOT* in (i).

4 Conclusion

This paper gives an asymptotic expression for the optimal ridge parameter within the class of ridge classification rules. This in turn suggests a new adaptive ridge classification rule *ASYMP* which has a closed form expression for its ridge parameter. The error rate of this rule is compared to that of the usual discriminant rule *USUAL* and the ridge classification rules *LOO* and *BOOT* (which were described in Section 3) via a Monte Carlo study. The simulations indicate that *ASYMP* performs reasonably well with respect to *USUAL*, *LOO* and *BOOT* without having the discretization drawbacks associated with the latter two rules.

Table 1

$p = 10 \quad n_1 = 25 \quad n_2 = 25 \quad \mu^{(1)} - \mu^{(2)} = (0, \ldots, 0, x)'$

Expected Error Rates of *USUAL*, *ASYMP*, *LOO*, *BOOT* (standard errors are in parenthesis)

Eigenvalues of $\boldsymbol{\Sigma}$	*USUAL*	*ASYMP*	*LOO*	*BOOT*
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 0.25$				
(1,1,1,1,1,1,1,1,1,1)	45.68 (0.11)	45.52 (0.11)	45.45 (0.11)	45.59 (0.11)
(50, 1, 1, 1, 1, 1, 1, 1, 1, 1)	45.68 (0.11)	45.74 (0.11)	46.68 (0.12)	45.72 (0.11)
(10,10,10,10,10,1,1,1,1,1)	45.68 (0.11)	45.91 (0.11)	46.75 (0.11)	45.83 (0.11)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	45.68 (0.11)	45.78 (0.11)	46.78 (0.13)	45.83 (0.12)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	45.68 (0.11)	45.87 (0.11)	46.88 (0.12)	45.81 (0.11)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	45.68 (0.11)	45.81 (0.11)	46.66 (0.12)	45.88 (0.11)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	45.68 (0.11)	45.80 (0.11)	47.39 (0.13)	45.76 (0.11)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	45.68 (0.11)	45.78 (0.11)	47.56 (0.13)	45.70 (0.11)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 1.00$				
(1,1,1,1,1,1,1,1,1,1)	36.97 (0.13)	36.28 (0.13)	36.19 (0.13)	36.56 (0.13)
(50,1,1,1,1,1,1,1,1,1)	36.97 (0.13)	36.86 (0.13)	37.99 (0.18)	36.94 (0.13)
(10,10,10,10,10,1,1,1,1,1)	36.97 (0.13)	37.24 (0.14)	38.38 (0.19)	37.11 (0.14)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	36.97 (0.13)	37.01 (0.14)	38.20 (0.21)	37.04 (0.14)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	36.97 (0.13)	37.14 (0.14)	38.58 (0.20)	37.15 (0.14)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	36.97 (0.13)	37.04 (0.14)	38.36 (0.19)	37.15 (0.14)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	36.97 (0.13)	37.02 (0.14)	38.66 (0.24)	37.01 (0.14)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	36.97 (0.13)	37.01 (0.14)	39.06 (0.25)	36.99 (0.14)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 4.00$				
(1,1,1,1,1,1,1,1,1,1)	20.72 (0.10)	19.02 (0.08)	19.18 (0.09)	19.43 (0.09)
(50,1,1,1,1,1,1,1,1,1)	20.72 (0.10)	20.00 (0.11)	20.15 (0.12)	20.16 (0.10)
(10,10,10,10,10,1,1,1,1,1)	20.72 (0.10)	20.95 (0.11)	21.43 (0.15)	20.83 (0.11)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	20.72 (0.10)	20.72 (0.10)	21.20 (0.17)	20.72 (0.10)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	20.72 (0.10)	20.83 (0.11)	21.54 (0.15)	20.76 (0.11)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	20.72 (0.10)	20.73 (0.10)	21.57 (0.15)	20.80 (0.11)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	20.72 (0.10)	20.72 (0.10)	20.85 (0.14)	20.72 (0.10)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	20.72 (0.10)	20.72 (0.10)	20.87 (0.14)	20.72 (0.10)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 9.00$				
(1,1,1,1,1,1,1,1,1,1)	9.98 (0.07)	8.31 (0.05)	8.68 (0.06)	8.46 (0.05)
(50,1,1,1,1,1,1,1,1,1)	9.98 (0.07)	9.09 (0.09)	9.19 (0.07)	8.93 (0.06)
(10,10,10,10,10,1,1,1,1,1)	9.98 (0.07)	9.99 (0.08)	10.07 (0.09)	9.94 (0.08)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	9.98 (0.07)	9.98 (0.07)	10.10 (0.09)	9.98 (0.07)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	9.98 (0.07)	10.10 (0.08)	10.24 (0.09)	10.05 (0.08)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	9.98 (0.07)	10.03 (0.08)	10.39 (0.10)	10.11 (0.08)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	9.98 (0.07)	9.98 (0.07)	9.98 (0.07)	9.98 (0.07)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	9.98 (0.07)	9.98 (0.07)	9.98 (0.07)	9.98 (0.07)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 16.00$				
(1,1,1,1,1,1,1,1,1,1)	4.11 (0.05)	3.01 (0.03)	3.52 (0.04)	3.06 (0.03)
(50,1,1,1,1,1,1,1,1,1)	4.11 (0.05)	3.51 (0.09)	3.65 (0.04)	3.31 (0.04)
(10,10,10,10,10,1,1,1,1,1)	4.11 (0.05)	4.00 (0.05)	4.05 (0.05)	3.94 (0.04)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	4.11 (0.05)	4.11 (0.05)	4.20 (0.05)	4.11 (0.05)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	4.11 (0.05)	4.14 (0.05)	4.18 (0.05)	4.16 (0.05)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	4.11 (0.05)	4.13 (0.05)	4.23 (0.05)	4.18 (0.05)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	4.11 (0.05)	4.12 (0.05)	4.11 (0.05)	4.11 (0.05)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	4.11 (0.05)	4.11 (0.05)	4.11 (0.05)	4.11 (0.05)

Table 2

$p = 10 \quad n_1 = 25 \quad n_2 = 25 \quad \mu^{(1)} - \mu^{(2)} = (x, 0, \ldots, 0)'$

Expected Error Rates of *USUAL*, *ASYMP*, *LOO*, *BOOT* (standard errors are in parenthesis)

Eigenvalues of $\mathbf{\Sigma}$	*USUAL*	*ASYMP*	*LOO*	*BOOT*
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 0.25$				
(50, 1, 1, 1, 1, 1, 1, 1, 1, 1)	45.72 (0.10)	44.48 (0.14)	42.52 (0.16)	44.77 (0.14)
(10,10,10,10,10,1,1,1,1,1)	45.72 (0.10)	45.25 (0.11)	44.68 (0.12)	45.39 (0.11)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	45.72 (0.10)	45.55 (0.10)	45.33 (0.11)	45.61 (0.11)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	45.72 (0.10)	45.20 (0.12)	44.23 (0.13)	45.31 (0.12)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	45.72 (0.10)	45.12 (0.12)	44.59 (0.12)	45.36 (0.11)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	45.72 (0.10)	45.32 (0.11)	43.37 (0.15)	45.25 (0.13)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	45.72 (0.10)	45.55 (0.11)	42.64 (0.16)	45.39 (0.12)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 1.00$				
(50,1,1,1,1,1,1,1,1,1)	36.91 (0.12)	34.28 (0.13)	32.59 (0.12)	34.52 (0.16)
(10,10,10,10,10,1,1,1,1,1)	36.91 (0.12)	35.67 (0.13)	34.77 (0.12)	35.80 (0.14)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	36.91 (0.12)	36.42 (0.12)	35.88 (0.12)	36.49 (0.12)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	36.91 (0.12)	35.66 (0.13)	34.16 (0.12)	35.66 (0.14)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	36.91 (0.12)	35.47 (0.13)	34.57 (0.12)	35.71 (0.14)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	36.91 (0.12)	35.99 (0.13)	33.22 (0.12)	35.46 (0.16)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	36.91 (0.12)	36.55 (0.12)	32.71 (0.13)	35.90 (0.16)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 4.00$				
(50,1,1,1,1,1,1,1,1,1)	20.70 (0.10)	18.04 (0.08)	17.41 (0.09)	17.92 (0.10)
(10,10,10,10,10,1,1,1,1,1)	20.70 (0.10)	19.01 (0.09)	18.34 (0.09)	18.83 (0.10)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	20.70 (0.10)	19.80 (0.10)	19.01 (0.08)	19.37 (0.09)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	20.70 (0.10)	19.07 (0.09)	18.05 (0.09)	18.67 (0.10)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	20.70 (0.10)	18.77 (0.09)	18.28 (0.09)	18.61 (0.10)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	20.70 (0.10)	19.54 (0.09)	17.61 (0.09)	18.54 (0.12)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	20.70 (0.10)	20.23 (0.10)	17.40 (0.10)	18.93 (0.14)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 9.00$				
(50,1,1,1,1,1,1,1,1,1)	9.98 (0.07)	8.06 (0.05)	7.81 (0.07)	7.71 (0.06)
(10,10,10,10,10,1,1,1,1,1)	9.98 (0.07)	8.63 (0.34)	8.26 (0.07)	8.16 (0.06)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	9.98 (0.07)	9.17 (0.07)	8.52 (0.06)	8.40 (0.05)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	9.98 (0.07)	8.71 (0.06)	8.13 (0.07)	8.08 (0.06)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	9.98 (0.07)	8.44 (0.06)	8.23 (0.07)	8.10 (0.06)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	9.98 (0.07)	9.07 (0.06)	8.05 (0.07)	8.02 (0.07)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	9.98 (0.07)	9.60 (0.07)	7.82 (0.08)	8.12 (0.09)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 16.00$				
(50,1,1,1,1,1,1,1,1,1)	4.11 (0.04)	3.00 (0.03)	3.19 (0.04)	2.79 (0.03)
(10,10,10,10,10,1,1,1,1,1)	4.11 (0.04)	3.29 (0.03)	3.35 (0.04)	2.97 (0.03)
(25, 25, 25, 25, 25, 25, 25, 25, 25, 1)	4.11 (0.04)	3.57 (0.04)	3.48 (0.04)	3.05 (0.03)
(20, 20, 20, 5, 5, 5, 5, 1, 1, 1)	4.11 (0.04)	3.34 (0.03)	3.37 (0.04)	2.94 (0.03)
(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)	4.11 (0.04)	3.18 (0.03)	3.38 (0.04)	2.94 (0.03)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	4.11 (0.04)	3.55 (0.04)	3.26 (0.05)	2.90 (0.04)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	4.11 (0.04)	3.88 (0.04)	3.17 (0.05)	2.97 (0.05)

Table 3

$p = 10 \quad n_1 = 25 \quad n_2 = 25 \quad \mu^{(1)} - \mu^{(2)} = x(\sigma_1, \sigma_2, \ldots, \sigma_p)'$

Expected Error Rates of *USUAL*, *ASYMP*, *LOO*, *BOOT* (standard errors are in parenthesis)

Eigenvalues of $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$	*USUAL*	*ASYMP*	*LOO*	*BOOT*
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 0.25$				
(50,1,1,1,1,1,1,1,1,1)	45.73 (0.12)	45.67 (0.12)	46.31 (0.13)	45.74 (0.12)
(10,10,10,10,10,1,1,1,1,1)	45.73 (0.12)	45.72 (0.11)	45.94 (0.12)	45.72 (0.12)
(25,25,25,25,25,25,25,25,25,1)	45.73 (0.12)	45.65 (0.11)	45.60 (0.12)	45.66 (0.11)
(20,20,20,5,5,5,5,1,1,1)	45.73 (0.12)	45.71 (0.11)	45.92 (0.12)	45.72 (0.11)
(10,9,8,7,6,5,4,3,2,1)	45.73 (0.12)	45.65 (0.11)	45.65 (0.12)	45.68 (0.12)
(512,256,128,64,32,16,8,4,2,1)	45.73 (0.12)	45.73 (0.11)	46.24 (0.12)	45.73 (0.11)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	45.73 (0.12)	45.74 (0.12)	46.84 (0.12)	45.74 (0.12)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 1.00$				
(50,1,1,1,1,1,1,1,1,1)	36.97 (0.14)	36.74 (0.14)	37.67 (0.16)	36.88 (0.14)
(10,10,10,10,10,1,1,1,1,1)	36.97 (0.14)	36.97 (0.14)	37.39 (0.14)	36.99 (0.14)
(25,25,25,25,25,25,25,25,25,1)	36.97 (0.14)	36.78 (0.13)	36.66 (0.13)	36.84 (0.13)
(20,20,20,5,5,5,5,1,1,1)	36.97 (0.14)	36.98 (0.14)	37.43 (0.13)	36.97 (0.14)
(10,9,8,7,6,5,4,3,2,1)	36.97 (0.14)	36.82 (0.13)	36.81 (0.13)	36.83 (0.14)
(512,256,128,64,32,16,8,4,2,1)	36.97 (0.14)	37.03 (0.14)	38.28 (0.15)	37.10 (0.14)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	36.97 (0.14)	37.03 (0.14)	39.11 (0.18)	37.04 (0.14)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 4.00$				
(50,1,1,1,1,1,1,1,1,1)	20.66 (0.11)	19.88 (0.10)	20.48 (0.13)	20.08 (0.10)
(10,10,10,10,10,1,1,1,1,1)	20.66 (0.11)	20.46 (0.10)	21.07 (0.11)	20.75 (0.11)
(25,25,25,25,25,25,25,25,25,1)	20.66 (0.11)	20.49 (0.10)	20.13 (0.09)	20.35 (0.10)
(20,20,20,5,5,5,5,1,1,1)	20.66 (0.11)	20.53 (0.10)	21.06 (0.11)	20.77 (0.11)
(10,9,8,7,6,5,4,3,2,1)	20.66 (0.11)	20.28 (0.10)	20.35 (0.10)	20.32 (0.10)
(512,256,128,64,32,16,8,4,2,1)	20.66 (0.11)	20.66 (0.11)	21.86 (0.13)	20.89 (0.12)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	20.66 (0.11)	20.74 (0.11)	21.69 (0.18)	20.71 (0.11)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 9.00$				
(50,1,1,1,1,1,1,1,1,1)	9.94 (0.08)	9.16 (0.06)	9.49 (0.08)	9.22 (0.07)
(10,10,10,10,10,1,1,1,1,1)	9.94 (0.08)	9.66 (0.07)	10.13 (0.08)	9.93 (0.07)
(25,25,25,25,25,25,25,25,25,1)	9.94 (0.08)	9.80 (0.07)	9.45 (0.06)	9.52 (0.06)
(20,20,20,5,5,5,5,1,1,1)	9.94 (0.08)	9.73 (0.07)	10.11 (0.08)	9.95 (0.07)
(10,9,8,7,6,5,4,3,2,1)	9.94 (0.08)	9.52 (0.07)	9.58 (0.07)	9.48 (0.06)
(512,256,128,64,32,16,8,4,2,1)	9.94 (0.08)	9.88 (0.07)	10.72 (0.10)	10.21 (0.08)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	9.94 (0.08)	10.00 (0.08)	10.25 (0.12)	9.96 (0.08)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 16.00$				
(50,1,1,1,1,1,1,1,1,1)	4.10 (0.05)	3.57 (0.04)	3.87 (0.05)	3.56 (0.04)
(10,10,10,10,10,1,1,1,1,1)	4.10 (0.05)	3.88 (0.04)	4.21 (0.05)	4.13 (0.05)
(25,25,25,25,25,25,25,25,25,1)	4.10 (0.05)	4.00 (0.04)	3.86 (0.04)	3.71 (0.03)
(20,20,20,5,5,5,5,1,1,1)	4.10 (0.05)	3.92 (0.04)	4.19 (0.05)	4.11 (0.04)
(10,9,8,7,6,5,4,3,2,1)	4.10 (0.05)	3.78 (0.04)	3.91 (0.04)	3.76 (0.03)
(512,256,128,64,32,16,8,4,2,1)	4.10 (0.05)	4.03 (0.04)	4.59 (0.06)	4.49 (0.06)
$(10^9, 10^8, 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10, 1)$	4.10 (0.05)	4.12 (0.05)	4.24 (0.07)	4.12 (0.05)

Table 4

$p = 5 \quad n_1 = 30 \quad n_2 = 30 \quad \mu^{(1)} - \mu^{(2)} = (0, \ldots, 0, x)'$

Expected Error Rates of *USUAL*, *ASYMP*, *LOO*, *BOOT* (standard errors are in parenthesis)

Eigenvalues of Σ	*USUAL*	*ASYMP*	*LOO*	*BOOT*
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 0.25$				
(1,1,1,1,1)	44.01 (0.12)	43.90 (0.12)	43.92 (0.12)	43.93 (0.12)
(10,1,1,1,1)	44.01 (0.12)	44.35 (0.12)	44.83 (0.13)	44.31 (0.12)
(25,25,25,25,1)	44.01 (0.12)	44.18 (0.13)	45.22 (0.15)	44.42 (0.14)
(200,120,100,80,1)	44.01 (0.12)	44.17 (0.13)	45.42 (0.16)	44.38 (0.14)
(9,7,5,3,1)	44.01 (0.12)	44.17 (0.13)	45.16 (0.14)	44.48 (0.13)
(16,8,4,2,1)	44.01 (0.12)	44.17 (0.13)	45.33 (0.14)	44.48 (0.13)
$(10^4, 10^3, 10^2, 10, 1)$	44.01 (0.12)	44.17 (0.13)	45.96 (0.17)	44.31 (0.14)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 1.00$				
(1,1,1,1,1)	33.89 (0.09)	33.55 (0.09)	33.61 (0.09)	33.57 (0.08)
(10,1,1,1,1)	33.89 (0.09)	34.20 (0.11)	34.75 (0.14)	34.09 (0.11)
(25,25,25,25,1)	33.89 (0.09)	33.89 (0.09)	34.64 (0.16)	33.98 (0.11)
(200,120,100,80,1)	33.89 (0.09)	33.89 (0.09)	34.63 (0.18)	33.97 (0.11)
(9,7,5,3,1)	33.89 (0.09)	33.90 (0.10)	34.96 (0.15)	34.19 (0.12)
(16,8,4,2,1)	33.89 (0.09)	33.91 (0.10)	35.04 (0.16)	34.23 (0.12)
$(10^4, 10^3, 10^2, 10, 1)$	33.89 (0.09)	33.89 (0.10)	35.15 (0.22)	33.91 (0.10)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 4.00$				
(1,1,1,1,1)	17.84 (0.06)	17.16 (0.04)	17.37 (0.05)	17.30 (0.05)
(10,1,1,1,1)	17.84 (0.06)	18.07 (0.08)	18.02 (0.08)	17.74 (0.06)
(25,25,25,25,1)	17.84 (0.06)	17.84 (0.06)	18.17 (0.09)	17.84 (0.06)
(200,120,100,80,1)	17.84 (0.06)	17.84 (0.06)	17.84 (0.06)	17.84 (0.06)
(9,7,5,3,1)	17.84 (0.06)	17.84 (0.06)	18.41 (0.10)	17.98 (0.07)
(16,8,4,2,1)	17.84 (0.06)	17.84 (0.06)	18.44 (0.10)	17.98 (0.07)
$(10^4, 10^3, 10^2, 10, 1)$	17.84 (0.06)	17.84 (0.06)	17.84 (0.06)	17.84 (0.06)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 9.00$				
(1,1,1,1,1)	7.94 (0.04)	7.34 (0.03)	7.64 (0.03)	7.52 (0.03)
(10,1,1,1,1)	7.94 (0.04)	8.10 (0.07)	7.94 (0.05)	7.78 (0.04)
(25,25,25,25,1)	7.94 (0.04)	7.94 (0.04)	8.20 (0.06)	7.96 (0.04)
(200,120,100,80,1)	7.94 (0.04)	7.94 (0.04)	7.94 (0.04)	7.94 (0.04)
(9,7,5,3,1)	7.94 (0.04)	7.94 (0.04)	8.17 (0.05)	8.02 (0.04)
(16,8,4,2,1)	7.94 (0.04)	7.95 (0.04)	8.20 (0.05)	8.01 (0.04)
$(10^4, 10^3, 10^2, 10, 1)$	7.94 (0.04)	7.94 (0.04)	7.94 (0.04)	7.94 (0.04)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 16.00$				
(1,1,1,1,1)	2.93 (0.02)	2.57 (0.01)	2.81 (0.02)	2.66 (0.02)
(10,1,1,1,1)	2.93 (0.02)	3.00 (0.04)	2.88 (0.02)	2.81 (0.02)
(25,25,25,25,1)	2.93 (0.02)	2.93 (0.02)	3.05 (0.03)	2.96 (0.02)
(200,120,100,80,1)	2.93 (0.02)	2.93 (0.02)	2.93 (0.02)	2.93 (0.02)
(9,7,5,3,1)	2.93 (0.02)	2.95 (0.02)	3.00 (0.02)	2.97 (0.02)
(16,8,4,2,1)	2.93 (0.02)	2.95 (0.02)	2.99 (0.02)	2.96 (0.02)
$(10^4, 10^3, 10^2, 10, 1)$	2.93 (0.02)	2.93 (0.02)	2.93 (0.02)	2.93 (0.02)

Table 5

$p = 5 \quad n_1 = 30 \quad n_2 = 30 \quad \mu^{(1)} - \mu^{(2)} = (x, 0, \ldots, 0)'$

Expected Error Rates of *USUAL*, *ASYMP*, *LOO*, *BOOT* (standard errors are in parenthesis)

Eigenvalues of $\boldsymbol{\Sigma}$	*USUAL*	*ASYMP*	*LOO*	*BOOT*
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 0.25$				
$(10, 1, 1, 1, 1)$	43.85 (0.12)	42.69 (0.14)	42.04 (0.14)	42.73 (0.14)
$(25, 25, 25, 25, 1)$	43.85 (0.12)	43.69 (0.12)	43.53 (0.12)	43.63 (0.12)
$(200, 120, 100, 80, 1)$	43.85 (0.12)	43.54 (0.12)	42.92 (0.13)	43.26 (0.13)
$(9, 7, 5, 3, 1)$	43.85 (0.12)	43.18 (0.13)	42.99 (0.13)	43.22 (0.13)
$(16, 8, 4, 2, 1)$	43.85 (0.12)	43.02 (0.13)	42.48 (0.13)	42.98 (0.13)
$(10^4, 10^3, 10^2, 10, 1)$	43.85 (0.12)	43.47 (0.13)	42.01 (0.15)	43.07 (0.14)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 1.00$				
(10,1,1,1,1)	33.73 (0.08)	32.14 (0.07)	31.97 (0.06)	32.21 (0.07)
$(25, 25, 25, 25, 1)$	33.73 (0.08)	33.37 (0.08)	33.19 (0.08)	33.27 (0.08)
$(200, 120, 100, 80, 1)$	33.73 (0.08)	33.29 (0.08)	32.65 (0.08)	32.77 (0.08)
$(9, 7, 5, 3, 1)$	33.73 (0.08)	32.71 (0.08)	32.62 (0.07)	32.75 (0.08)
$(16, 8, 4, 2, 1)$	33.73 (0.08)	32.55 (0.08)	32.29 (0.07)	32.51 (0.08)
$(10^4, 10^3, 10^2, 10, 1)$	33.73 (0.08)	33.19 (0.08)	31.95 (0.07)	32.63 (0.09)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 4.00$				
(10,1,1,1,1)	17.77 (0.05)	16.63 (0.03)	16.85 (0.04)	16.78 (0.04)
$(25, 25, 25, 25, 1)$	17.77 (0.05)	17.32 (0.05)	17.27 (0.05)	17.23 (0.05)
$(200, 120, 100, 80, 1)$	17.77 (0.05)	17.35 (0.05)	17.01 (0.04)	17.01 (0.05)
$(9, 7, 5, 3, 1)$	17.77 (0.05)	16.90 (0.04)	17.05 (0.04)	16.94 (0.04)
$(16, 8, 4, 2, 1)$	17.77 (0.05)	16.85 (0.04)	16.91 (0.04)	16.88 (0.04)
$(10^4, 10^3, 10^2, 10, 1)$	17.77 (0.05)	17.31 (0.05)	16.79 (0.04)	16.98 (0.06)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 9.00$				
(10,1,1,1,1)	7.90 (0.03)	7.18 (0.02)	7.41 (0.03)	7.26 (0.03)
$(25, 25, 25, 25, 1)$	7.90 (0.03)	7.55 (0.03)	7.56 (0.03)	7.46 (0.03)
$(200, 120, 100, 80, 1)$	7.90 (0.03)	7.63 (0.03)	7.48 (0.03)	7.37 (0.03)
$(9, 7, 5, 3, 1)$	7.90 (0.03)	7.31 (0.03)	7.49 (0.03)	7.34 (0.03)
$(16, 8, 4, 2, 1)$	7.90 (0.03)	7.29 (0.03)	7.44 (0.03)	7.31 (0.03)
$(10^4, 10^3, 10^2, 10, 1)$	7.90 (0.03)	7.60 (0.03)	7.39 (0.04)	7.36 (0.04)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 16.00$				
(10,1,1,1,1)	2.91 (0.02)	2.53 (0.01)	2.75 (0.02)	2.58 (0.02)
$(25, 25, 25, 25, 1)$	2.91 (0.02)	2.71 (0.02)	2.79 (0.02)	2.65 (0.02)
$(200, 120, 100, 80, 1)$	2.91 (0.02)	2.76 (0.17)	2.77 (0.02)	2.62 (0.02)
$(9, 7, 5, 3, 1)$	2.91 (0.02)	2.59 (0.01)	2.77 (0.02)	2.61 (0.02)
$(16, 8, 4, 2, 1)$	2.91 (0.02)	2.59 (0.01)	2.76 (0.02)	2.60 (0.02)
$(10^4, 10^3, 10^2, 10, 1)$	2.91 (0.02)	2.76 (0.02)	2.73 (0.02)	2.63 (0.02)

Table 6

$p = 5 \quad n_1 = 30 \quad n_2 = 30 \quad \mu^{(1)} - \mu^{(2)} = x(\sigma_1, \sigma_2, \ldots, \sigma_p)'$

Expected Error Rates of *USUAL*, *ASYMP*, *LOO*, *BOOT* (standard errors are in parenthesis)

Eigenvalues of $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$	*USUAL*	*ASYMP*	*LOO*	*BOOT*
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 0.25$				
$(10,1,1,1,1)$	43.86 (0.11)	43.95 (0.11)	44.24 (0.12)	43.97 (0.11)
$(25,25,25,25,1)$	43.86 (0.11)	43.86 (0.11)	43.96 (0.11)	43.89 (0.11)
$(200,120,100,80,1)$	43.86 (0.11)	43.89 (0.11)	44.04 (0.10)	43.89 (0.10)
$(9,7,5,3,1)$	43.86 (0.11)	43.87 (0.10)	43.90 (0.11)	43.89 (0.11)
$(16,8,4,2,1)$	43.86 (0.11)	43.97 (0.11)	44.06 (0.10)	43.95 (0.11)
$(10^4,10^3,10^2,10,1)$	43.86 (0.11)	43.95 (0.11)	44.85 (0.13)	44.05 (0.11)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 1.00$				
(10,1,1,1,1)	33.80 (0.09)	34.00 (0.09)	34.35 (0.10)	34.03 (0.09)
$(25,25,25,25,1)$	33.80 (0.09)	33.89 (0.08)	34.16 (0.08)	34.05 (0.09)
$(200,120,100,80,1)$	33.80 (0.09)	33.94 (0.09)	34.44 (0.09)	34.14 (0.09)
$(9,7,5,3,1)$	33.80 (0.09)	33.93 (0.08)	34.11 (0.08)	34.01 (0.09)
$(16,8,4,2,1)$	33.80 (0.09)	34.01 (0.09)	34.36 (0.09)	34.10 (0.09)
$(10^4,10^3,10^2,10,1)$	33.80 (0.09)	34.00 (0.09)	35.40 (0.13)	34.26 (0.11)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 4.00$				
(10,1,1,1,1)	17.82 (0.06)	17.78 (0.06)	17.97 (0.07)	17.86 (0.06)
$(25,25,25,25,1)$	17.82 (0.06)	17.95 (0.06)	18.32 (0.06)	18.22 (0.06)
$(200,120,100,80,1)$	17.82 (0.06)	17.98 (0.06)	18.66 (0.07)	18.25 (0.07)
$(9,7,5,3,1)$	17.82 (0.06)	17.94 (0.06)	18.15 (0.06)	18.04 (0.06)
$(16,8,4,2,1)$	17.82 (0.06)	17.95 (0.06)	18.30 (0.07)	18.09 (0.07)
$(10^4,10^3,10^2,10,1)$	17.82 (0.06)	18.00 (0.07)	18.57 (0.12)	17.92 (0.07)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 9.00$				
(10,1,1,1,1)	7.92 (0.04)	7.85 (0.04)	8.06 (0.05)	7.94 (0.04)
$(25,25,25,25,1)$	7.92 (0.04)	8.00 (0.04)	8.37 (0.04)	8.33 (0.04)
$(200,120,100,80,1)$	7.92 (0.04)	8.00 (0.04)	8.56 (0.05)	8.33 (0.05)
$(9,7,5,3,1)$	7.92 (0.04)	7.98 (0.04)	8.18 (0.04)	8.18 (0.04)
$(16,8,4,2,1)$	7.92 (0.04)	7.97 (0.04)	8.27 (0.06)	8.19 (0.05)
$(10^4,10^3,10^2,10,1)$	7.92 (0.04)	8.01 (0.04)	8.17 (0.07)	7.99 (0.06)
$(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = 16.00$				
(10,1,1,1,1)	2.92 (0.02)	2.88 (0.02)	3.03 (0.03)	2.97 (0.03)
$(25,25,25,25,1)$	2.92 (0.02)	2.96 (0.02)	3.11 (0.02)	3.23 (0.03)
$(200,120,100,80,1)$	2.92 (0.02)	2.96 (0.02)	3.23 (0.03)	3.33 (0.03)
$(9,7,5,3,1)$	2.92 (0.02)	2.94 (0.02)	3.01 (0.02)	3.13 (0.03)
$(16,8,4,2,1)$	2.92 (0.02)	2.94 (0.02)	3.09 (0.03)	3.16 (0.03)
$(10^4,10^3,10^2,10,1)$	2.92 (0.02)	2.96 (0.02)	3.13 (0.05)	3.00 (0.04)

5 Appendix

Lemma 1 *With the notation of Section 2, we have*

$$\begin{aligned}
\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}(\mathrm{tr}\partial^*)\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx &= 0,\\
\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}(\mathrm{tr}\Sigma\partial^*)\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx &= 0,\\
\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}(\mathrm{tr}\Sigma\partial^*)\partial^{(1)\prime}\Sigma\partial^{(1)}\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx &= 0,\\
\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}(\mathrm{tr}\Sigma\partial^*)\partial^{(2)\prime}\Sigma\partial^{(2)}\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx &= 0,\\
\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}(\mathrm{tr}\partial^*)\partial^{(2)\prime}\Sigma\partial^{(2)}\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx &= 0,
\end{aligned}$$

$$\begin{aligned}
&\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}(\mathrm{tr}\Sigma\partial^*\Sigma\partial^*)\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx\\
&= \frac{p-1}{4}\phi(\frac{\Delta_1}{2})\Delta_1,
\end{aligned}$$

$$\begin{aligned}
&\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}(\mathrm{tr}\Sigma\partial^*)(\mathrm{tr}\Sigma\partial^*\Sigma\partial^*)\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx\\
&= -\frac{p-1}{2}\phi(\frac{\Delta_1}{2})\Delta_1,
\end{aligned}$$

$$\begin{aligned}
&\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}(\mathrm{tr}\partial^*)(\mathrm{tr}\Sigma\partial^*\Sigma\partial^*)\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx\\
&= \frac{1}{2}\phi(\frac{\Delta_1}{2})\Delta_1^{-1}[\Delta_2^2-(\mathrm{tr}\Sigma^{-1})\Delta_1^2],
\end{aligned}$$

$$\begin{aligned}
&\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}[\frac{1}{p}(\mathrm{tr}\Sigma)(\mathrm{tr}\partial^*)-(\mathrm{tr}\Sigma\partial^*)]^2\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx\\
&= \frac{1}{2p^2}(\mathrm{tr}\Sigma)^2\phi(\frac{\Delta_1}{2})\Delta_1^{-1}(\Delta_3^2-\Delta_1^{-2}\Delta_2^4),
\end{aligned}$$

and

$$\begin{aligned}
&\int_{-\infty}^{-\Delta_1/2}\int_{-\infty}^{\infty} e^{-itx}(\mathrm{tr}\partial^*)\partial^{(1)\prime}\Sigma\partial^{(1)}\Psi(\xi^{(1)},\xi^{(2)},\Theta)|_{(\xi^{(1)},\xi^{(2)},\Theta)=(\mu^{(1)},\mu^{(2)},\Sigma)}dtdx\\
&= 2\phi(\frac{\Delta_1}{2})[p\Delta_1^{-3}\Delta_2^2-\Delta_1^{-1}(\mathrm{tr}\Sigma^{-1})],
\end{aligned}$$

where ϕ denotes the probability density function of the standard normal distribution.

The proof of Lemma 1 is straightforward though somewhat tedious and hence will be omitted.

References

[1] Anderson, T. W. (1951), "Classification by Multivariate Analysis," *Psychometrika*, 16, 31-50.

[2] Anderson, T. W. (1984), *An Introduction to Multivariate Analysis*, (2nd ed.),New York: Wiley.

[3] Das Gupta, S. (1965)," Optimum Classification Rules for Classification into Two Multivariate Normal Populations," *The Annals of Mathematical Statistics*, 36, 1174-1184.

[4] Dey, D. K., and Srinivasan, C. (1985), "Estimation of a Covariance Matrix under Stein's Loss," *The Annals of Statistics*, 13, 1581-1591.

[5] Dey, D. K., and Srinivasan, C. (1991), "On Estimation of Discriminant Coefficients," *Statistics and Probability Letters*, 11, 189-193.

[6] DiPillo, P. J. (1976), "The Application of Bias to Discriminant Analysis," *Communications in Statistics, Part A–Theory and Methods*, 5, 843-854.

[7] DiPillo, P. J. (1977), "Further Applications of Bias to Discriminant Analysis," *Communications in Statistics, Part A–Theory and Methods*, 6, 933-943.

[8] DiPillo, P. J. (1979), "Biased Discriminant Analysis: Evaluation of the Optimum Probability of Misclassification," *Communications in Statistics, Part A–Theory and Methods*, 8, 1447-1457.

[9] Efron, B. (1983), "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation," *Journal of the American Statistical Association*, 78, 316-331.

[10] Efron, B., and Morris, C. (1976), "Multivariate Empirical Bayes and Estimation of Covariance Matrices," *The Annals of Statistics*, 4, 22-32.

[11] Friedman, J. H. (1989), " Regularized Discriminant Analysis," *Journal of the American Statistical Association*, 84, 165-175.

[12] Geisser, S. (1977), "Discrimination, Allocatory and Separatory, Linear Aspects," in *Classification and Clustering*, ed. J. Van Ryzin, New York: Academic Press, pp. 301-330.

[13] Gnanadesikan, R., et. al. (1989)," Discriminant Analysis and Clustering," *Statistical Science*, 4, 34-69.

[14] Haff, L. R. (1986), "On Linear Log-odds and Estimation of Discriminant Coefficients," *Communications in Statistics, Part A–Theory and Methods*, 15, 2131-2144.

[15] Haff, L. R. (1991), " The Variational Form of Certain Bayes Estimators," *The Annals of Statistics*, 19, 1163-1190.

[16] James, W., and Stein, C. (1961), "Estimation with Quadratic Loss," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1), Berkeley: University of California Press, pp. 361-379.

[17] Lachenbruch, P. A. (1975), *Discriminant Analysis*, New York: Hafner Press.

[18] Lin, S. P., and Perlman, M. D. (1985), "A Monte Carlo Comparison of Four Estimators for a Covariance Matrix," in *Multivariate Analysis* (Vol. 6), ed. P.K. Krishnaiah, Amsterdam: North Holland, pp. 415-429.

[19] McLachlan, G. J. (1992), *Discriminant Analysis and Statistical Pattern Recognition*, New York: Wiley.

[20] Okamoto, M. (1963), "An Asymptotic Expansion for the Distribution of the Linear Discriminant Function," *The Annals of Mathematical Statistics*, 34, 1286-1301.

[21] Olkin, I., and Selliah, J. B. (1977), "Estimating Covariances in a Multivariate Normal Distribution," in *Statistical Decision Theory and Related Topics II*, eds. S. S. Gupta and D. Moore, New York: Academic Press, pp. 313-326.

[22] Peck, R., and Van Ness, J. (1982), " The Use of Shrinkage Estimators in Linear Discriminant Analysis," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 4, 530-537.

[23] Rodriguez, A. F. (1988), " Admissibility and Unbiasedness of the Ridge Classification Rules for Two Normal Populations with Equal Covariance Matrices," *Statistics*, 19, 383-388.

[24] Stein, C. M. (1956), "Some Problems in Multivariate Analysis," Part I, Report 6, Stanford University, Dept. of Statistics.

[25] Stein, C. M. (1975), "Reitz Lecture," unpublished paper presented at the 38th annual meeting of the Institute of Mathematical Statistics, Atlanta.

[26] Wald, A. (1944), " On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups," *The Annals of Mathematical Statistics*, 15, 145-162.