〈Research Paper〉

Journal of the Korean
Statistical Society
vol. 18, No.1, 1989

Discriminant Analysis under a Patterned Missing Values+

Hea Jung Kim*

ABSTRACT

This paper suggests a classification rule with unequal covariance matrices when a patterned incomplete data are involved in the discriminant analysis. This is an extension of Geisser's (1966) result to the case of missing observations. For the classification rule, we introduce an algorithm which contains data augmentation step and Monte Carlo integration step and show that the algorithm yields a consistant estimator of true classification probability. The proposed method is compared to the complete observation vector method through a Monte Carlo study. The results show that the suggested method, in general, performs better than the complete observation vector method which ignores those vectors of observation with one or more missing values from the analysis. The results also verify the consistency of the algorithm.

1. Introduction

Incomplete observation vectors cause an important problem when many measurements are made on each individual in a study. The classification analysis naturally bears this problem. Little (1978) and Chan, Gilman and Dunn (1976) as well as two earlier studies by Chan and Dunn (1972, 1974) suggest several methods of handling incomplete data vectors in the training sample, ranging from ignoring them to estimating the missing values by some means and then including them in the classification rule.

+ This work was supported by a grant form the Korea Research Foundation

* Department of Statistics, Dongguk University, Seoul 100, Korea.

Their studies were in conjunction with the normal based Fisher's (1936) classification rule with equal covariance matrices, and thus they do not give a general solution for the problem.

Our motivation for this study is to give a method of handling incomplete data vectors in the classification analysis with equal or unequal covariance matrices. To formulate this method, the Bayesian approach to the classification analysis taken by Geisser(1966) is extended to the case of missing ovservations. Hence, as he has done, we shall restrict ourselves to the case when the group prior probabilities are known. The extension to the case where they are unknown, on which Geisser (1964) has already commented in detail, can be easily managed. Here we will only deal with patterned incomplete p-variate sample which can be classified into three types;

Type 1: $p_2(< p)$ components of validation sample Z is missing,

Type 2: $p_2(< p)$ components of some observed vectors among training sample are missing,

Type 3: both Type 1 and Type 2 cases are combined.

Under the Bayesian approach, constructing a classification rule for Type 1 missing case is straightforward by applying marginal density of the multivariate Student t-distribution (cf, Press, 1982, p. 137), so that the remaining part of this paper will be contributed to exploit classification rules for Type 2 and Type 3 missing cases.

The plan of this paper is as follows. In Section 2 basic algorithm for constructing a classification rule for Type 2 is suggested. This utilizes the sample mean Monte Carlo method in Rubinstein(1981, p. 113). Section 3 is devoted to derive predictive densities of missing vector and matrix of observations which are eassential for the algorithm. We demonstrate, in Section 4, that the suggested algorithm yields a consistent estimator of predictive density for the validation sample Z, and that formulation of a classification rule for Type 3 is straight-forward from the algorithm. In Section 5 a result of Monte Carlo simulaton study is given for the comparison of the algorithm. Section 6 contains remarks on a number of related points.

2. Basic Algorithm

Suppose Z be a p-variate observation from one of the populations $\Pi_j = N(\theta_j, \Sigma_j)$, $j = 1,\cdots,k$, where the parameters θ_j, Σ_j are unknown. Also suppose that the independent p-variate observations $x_1(j), x_2(j), \cdots, x_{N_j}(j)$, say, training sample. from j-th population are available, $j=1,\cdots,k$. If the prior distribution of the parameters is diffuse, Geisser(1966) showed that the predictive probability density for classifying Z into Π_j is the multivariate

Student t-density given by

$$P(Z \mid \text{data}, j) = p_j P(Z \mid x_1(j), \cdots, x_{N_j}(j)) = C_j[1+\frac{N_j}{N_j^2-1}(Z-\bar{X}_j)'S_j^{-1}(Z-\bar{X}_j)]^{-N_j/2} \quad (2.1)$$

where $C_j = [N_j/(N_j+1)\pi]^{p/2} p_j \mid (N_j-1)S_j \mid^{-\frac{1}{2}} \Gamma(N_j/2)/\Gamma(((N_j-p)/2)$,

p_j is the prior probability of Z classifying into $\prod_j$, and $\bar{X}_j$ and S_j denote the sample mean and the unbiased covariance matrix, respectively.

This has been derived by multiplying the prior group probability p_j to the predictive density of the form

$$\int\int P(Z \mid \theta_j, \Sigma_j) P(\theta_j, \Sigma_j \mid \bar{X}_j, S_j)\, d\theta_j\, d\Sigma_j, \quad (2.2)$$

where $P(\theta_j, \Sigma_j \mid \bar{X}_j, S_j)$ denotes the joint posterior density of θ_j and Σ_j under the diffuse prior . It follows, from (2.1), that the predictive odds ratio for classifying Z into $\prod_i$, as compared with $\prod_j$, becomes the ratio of the associated multivariate Student t-densities

$$P(Z \mid \text{data}, i) \;/\; P(Z \mid \text{data}, j), \quad i, j = 1, \cdots, k. \quad (2.3)$$

An individual, validation sample, will be classified into a population $\prod_i$ if the predictive odds ratio in (2.3) is larger than 1 for all j.

When missing values occur, the predictive odds ratio in (2.3) may be modified to some missing value method. The present study suggests a missing value method which can be formulated by an algorithm. This algorithm contains data augmentation step and Monte Carlo integration step, and is motivated by the following simple representation of the desired predictive density

$$P(Z \mid X) = \int P(Z \mid M, X)\, dF(M \mid X), \quad (2.4)$$

where $P(Z \mid X)$ denotes the predictive density of the new p-variate observation Z given the incomplete data matrix X, $F(M \mid X)$ denotes the cumulative predictive distribution of the missing observation matrix M given X, and $P(Z \mid M, X)$ denotes the conditional density of Z given the augmented data (M, X).

Suppose $p \times N_j$ matrix $X(j)$ denotes the training sample from j-th population and suppose n_1 observations in the j-th sample contains the p_2 missing components. That is

$X(j)$ can be partitioned as

$$X(j)=\begin{bmatrix} X_{11}(j) & X_{12}(j) \\ - & X_{22}(j) \end{bmatrix}.$$

Here $[X_{11}(j)',\ -\ ]'$ denotes p x n_1 matrix which contains block of missing observations and $X_2(j)= [X_{12}(j)', X_{22}(j)']'$ represents the set of complete observation vectors. Then, the predictive density of the missing observation block, say $M(j)$, given $X(j)$ can, in turn, be expressed by

$$P(M(j) \mid X(j))= P(M(j), X_{11}(j) \mid X_2(j)) / P(X_{11}(j) \mid X_2(j)), \tag{2.5}$$

where

$$P(M(j), X_{11}(j) \mid X_2(j))= \int\int P(M(j)), X_{11}(j) \mid X_2(j), \theta_j, \Sigma_j)\ P(\theta_j, \Sigma_j \mid X_2(j))\ d\theta_j d\Sigma_j.$$

If the integration (2.4) can be calculated analytically, then the implementation of deriving $P(Z \mid X(j))$ is straightforward. Unfortunately, this runs into severe computational difficulties. It is possible, however, by the Monte Carlo method, to perform the integration. Equation (2.4) then motivates the following algorithm scheme.

Given the predictive density of $M(j)$ given $X(j)$ in (2.5),

step 1. generate a sample $M_1(j),\cdots, M_m(j)$ from the predictive density $P(M(j) \mid X(j))$, then construct the augmented data patterns $(M_i(j), X(j))$, $i= 1,\cdots,m$.

step 2. use the sample mean Monte Carlo method(cf. Rubinstein. 1981) to perform the integration in (2.4). This gives that the approximation to $P(Z \mid X(j))$ is the mixture of conditional densities of Z given the augmented data patterns generated in step 1 such that

$$P(Z \mid X(j)) \cong m^{-1} \sum_{i=1}^{m} P(Z \mid M_i(j), X(j)), \tag{2.6}$$

In the above, we must be able to calculate $P(Z \mid M_i(j),\ X(j))$ for any augmented data $(M_i(j), X(j))$ and we must be able to know $P(M(j) \mid X(j))$ in (2.5).

Since we know, from (2.1), $P(Z \mid M_i(j), X(j))$ to be a multivariate Student t-density, the problem of the algorithm is to derive $P(M(j) \mid X(j))$ in the explicit form under the assumption that the prior distribution of the parameters is diffuse.

3. Predictive Density of The Missing Data

In the previous section, we have seen that the algorithm needs $P(M(j) \mid X(j))$ to augment the incomplete data.

We now develop the particular normal predictive machinery for the missing observation matrix in j-th normal population training sample $x_1(j),\cdots,x_{N_j}(j)$, $j= 1,\cdots,k$.

Here we assume that all the coordinates of missing observation vectors are the same. That is, the data matrix is the monotonic or nested pattern such that

$$[x_1(j), x_2(j),\cdots,x_{N_j}(j)] = \begin{bmatrix} X_{11}(j) & X_{12}(j) \\ M(j) & X_{22}(j) \end{bmatrix}, \tag{3.1}$$

where $M(j)$ indicates block of missing observations. X_{uv} is an $p_u(j)$ x $n_v(j)$ submatrix, where $\sum_{v=1}^{2} n_v(j)= N_j$, the total number of independent sampling units, and $\sum_{u=1}^{2} p_u(j) = p$, the number of responses. Hereafter, for the notational convenience, we denote $p_u(j)$ and $n_v(j)$ as p_u and n_v, respectively. The data matrix may be written in a more compact form in the following way. Define

$$X_1(j)= [x_1(j), x_2(j),\cdots,x_{n_1}(j)] = [X_{11}(j)', M(j)']', \tag{3.2}$$
$$X_2(j)= [x_{n_1+1}(j), x_{n_1+2}(j),\cdots,x_{N_j}(j)] = [X_{12}(j)', X_{22}(j)']', \tag{3.3}$$

Lemma 1. Given the independent p-variate observation vectors $X_2(j)$ from j-th normal population $\Pi_j= N(\theta_j, \Sigma_j)$, $j= 1,\cdots, k$, the predictive probability distribution of $X_1(j)$ is a martrix T-distribution with N_j-1 degrees of freedom, with density

$$P(X_1(j) \mid X_2(j)) \propto | V_j+ (X_1(j)- \bar{X}_j^* e_j') \Omega_j (X_1(j)- \bar{X}_j^* e_1')' |^{-(N_j-1)/2}, \tag{3.4}$$

where

$\bar{X}_j^*= X_2(j)e_2 / n_2$, $V_j= X_2(j)\ X_2(j)'-n_2\bar{X}_j^*\ \bar{X}_j^{*\prime}$,
$\Omega_j= I_{n_1} - e_1e_1' / N_j$, $N_j > p + n_1$, and e_k is the n_k x 1 vector all of whose elements are unity, k= 1,2.

Proof. Define $\bar{X}_j^{**}= X_1(j)\ e_1 / n_1$ and $U_j= X_1(j)X_1(j)'-n_1\bar{X}_j^{**}\ \bar{X}_j^{**\prime}$.

Then the predictive density of $X_1(j)$ given $X_2(j)$ is

$$
\begin{aligned}
&P(X_1(j) \mid X_2(j)) \\
&= \int\int P(x_1(j), \cdots, x_{n_1}(j) \mid \theta_j, \Sigma_j) P(\theta_j, \Sigma_j \mid x_{n_1+1}(j), \cdots, x_{Nj}(j))\, d\theta_j d\Sigma_j \\
&\propto \int |\Sigma_j|^{-(N_j+P)/2} \exp\{-\frac{1}{2} \operatorname{tr}[\Sigma_j^{-1}(V_j + U_j + n_1(N_j-n_1)(\bar{X}_j^*-\bar{X}^{**}_j)(\bar{X}_j^*-\bar{X}^{**}_j)'/N_j]\} d\Sigma_j \\
&\propto |V_j + U_j + n_1(N_j-n_1)(\bar{X}_j^*-\bar{X}^{**}_j)(\bar{X}_j^*-\bar{X}^{**}_j)'/N_j|^{-(N_j-1)/2} \qquad (3.5)
\end{aligned}
$$

Constructing (3.5) in terms of random matrix $X_1(j)$ gives

$$
P(X_1(j) \mid X_2(j)) \propto |V_j + (X_1(j) - \bar{X}_j^* e_1')\, \Omega_j (X_1(j) - \bar{X}_j^* e_1')'|^{-(N_j-1)/2},
$$

which is the Kernal of a matrix T-distribution with N_j-1 degrees of freedom. This completes the proof. ■

Thus, via conditional distribution theorem by Dickey(1967), it is possible to derive predictive distribution of M(j) from Lemma 1. To apply the Theorem, we need to partition $\bar{X}_j^* e_1'$ and V_j in (3.4) as

$$
\bar{X}_j^* e_1' = \begin{bmatrix} (\bar{X}_j^* e_1')_1 \\ (\bar{X}_j^* e_1')_2 \end{bmatrix}, \quad V_j = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \qquad (3.6)
$$

where $(\bar{X}_j^* e_1')_i$: p_i x n_1 and V_{jk}: p_j x p_k, i, k= 1,2. Then the following Theorem gives the desired distribution.

Theorem 2. Suppose missing observation matrix $M(j)$ is of the form (3.1). Then predictive distribution of $M(j) \mid X_{11}(j), X_2(j)$ is a matrix T-distribution with N_j - 1 degrees of freedom, with density

$$
\begin{aligned}
&P(M(j) \mid X_{11}(j), X_2(j)) \\
&= \frac{|V_{22\cdot1}|^{(N_j-n_1-1)/2}}{K(N_j-1, n_1, P_2)\,|Q_j|^{P_2/2}} |V_{22\cdot1} + (M(j)-G_j)Q_j^{-1}(M(j)-G_j)'|^{(N_j-1)/2}, \qquad (3.7)
\end{aligned}
$$

where $N_j > p_2 + n_1$, $V_{22\cdot1} = V_{22} - V_{21} V_{11}^{-1} V_{12}$,

$G_j = (\bar{X}_j^* e_1')_2 + V_{21} V_{11}^{-1} (X_{11}(j) - (\bar{X}_j^* e_1')_1)$,

$Q_j = \Omega_j^{-1} + (X_{11}(j) - (\bar{X}_j^* e_1')_1)' V_{11}^{-1} (X_{11}(j) - (\bar{X}_j^* e_1')_1)$,

and

$$K(N_j\text{-}1,\ n_1,\ p_2) = \pi^{n_1 p_2/2}\ \Gamma_{p_2}\left(\frac{N_j\text{-}n_1\text{-}1}{2}\right) \Big/ \Gamma_{p_2}\left(\frac{N_j\text{-}1}{2}\right) \text{ with}$$
$$\Gamma_{p_2}(\lambda) = \pi^{p_2(p_2-1)/4}\ \Gamma(\lambda)\ \Gamma(\lambda\text{-}1/2)\ \cdots\cdots\ \Gamma(\lambda\text{-}p_2/2+1/2).$$

Proof. with the result in Lemma 1 and the following relations of the partitioned matrices

$$V_{22\cdot1}\,V_{22}^{-1}\ V_{21}\,V_{11\cdot2}^{-1} = V_{21}\,V_{11}^{-1},$$
$$V_{11\cdot2}^{-1} - V_{11}^{-1}\ V_{12}\ V_{22}^{-1}\ V_{21}\ V_{11\cdot2}^{-1} = V_{11}^{-1},$$

the predictive density of $M(j)$ is readily obtained by Dickey's conditional matrix T-distribution Theorem (cf. Press 1982, p. 139). ■

Corollary 2.1. The predictive probability distribution of a missing observation vector

$$M(j) \mid X_{11}(j),\ X_2(j),\quad \text{where } M(j):\ p_2 \times 1,\ X_{11}(j):\ p_1 \times 1,$$

is p_2- variate Student t-distribution with $N_j - p_2 - 1$ degrees of freedom, with density given by

$$P(M(j) \mid X_{11}(j),\ X_2(j))$$
$$= \frac{\Gamma\left(\frac{N_j-1}{2}\right) |A_j|^{-\frac{1}{2}}}{\{\pi(N_j-P_2-1)\}^{p_2/2}\,\Gamma\left(\frac{N_j-P_2-1}{2}\right)} [1+(M(j)-a_j)'A_j^{-1}(M(j)-a_j)/(N_j-P_2-1)]^{-(N_j-1)/2}, \tag{3·8}$$

where $a_j = (\bar{X}_j^*)_2 + V_{21}\,V_{11}^{-1}\,(X_{11}(j)-(\bar{X}_j^*)_1)$,
$A_j = N_j V_{22\cdot1}\,W_j / [(N_j-1)(N_j-p_2-1)]$,
$W_j = 1 + (N_j-1)\,[X_{11}(j)-(\bar{X}_j^*)_1]'\,V_{11}^{-1}\,[X_{11}(j)-(\bar{X}_j^*)_1]$.

Proof. This result is a special case of Theorem 2 when $n_1 = 1$. Since, for $n_1 = 1$, notations in (3.7) become

$$G_j = (\bar{X}_j^*)_2 + V_{21}\,V_{11}^{-1}(X_{11}(j)-(\bar{X}_j^*)_1),$$
$$Q_j = N_j/(N_j-1) + [X_{11}(j)-(\bar{X}_j^*)_1]'V_{11}^{-1}\,[X_{11}(j)-(\bar{X}_j^*)_1],$$

and $K(N_j-1, 1, p_2)= \pi^{p_2/2} \Gamma(\frac{N_j-p_2-1}{2}) / \Gamma(\frac{N_j-1}{2})$. ■

4. Property of the Algorithm and Classification Rules

In this section, we return to the study of the algorithm outlined in Section 2. Let us rewrite the integral of interest in (2.4) for the j-th population sample case such that

$$P(Z \mid X(j))= \int P(Z \mid M(j), X(j))\, dF(M(j) \mid X(j)), \tag{4.1}$$

where $X(j)= (X_{11}(j), X_2(j))$.
Assuming that $P(M(j) \mid X(j))$ is the predictive density of the missing observation matrix $M(j)$ derived in Theorem 2

$$P(Z \mid X(j))= E[P(Z \mid M(j), X(j))\ ], \tag{4.2}$$

where the random matrix $M(j)$ is distributed according to $P(M(j) \mid X(j))$.
It is obvious from (4.2) that an unbiased estimator of $P(Z \mid X(j))$ is its sample mean

$$\delta_j= \sum_{i=1}^{m_j} P(Z \mid M_i(j), X(j)) / m_j, \tag{4.3}$$

where $M_i(j)$, $i= 1,\cdots, m_j$ denotes a generated sample from the predictive distribution $M(j) \mid X(j)$.
The variance of δ_j is

$$\begin{aligned} Var(\delta_j)= [\int \{P(Z \mid M(j)\ X(j))\}^2\, dF(M(j)) \mid X(j)) - \{P(Z \mid X(j))\}^2] / m_j \\ < 1 / m_j. \end{aligned} \tag{4.4}$$

Hence, from (4.3) and (4.4),

$$\sum_{i=1}^{m_j} P(Z \mid M_i(j), X(j)) / m_j \xrightarrow{p} P(Z \mid X(j)). \tag{4.5}$$

In other words, the algorithm converges to the true predictive density $P(Z \mid X(j))$.

Theorem 3. Let Z: p x 1 be an observation from one of the populations $\Pi_j= N(\theta_j, \Sigma_j)$, $j=1,\cdots,k$, where the parameters are unknown. If the prior distribution of the parameters is diffuse and j-th training sample has the nested pattern missing data $M(j)$, a consistent estimator of the predictive probability for classifying Z into Π_j given the incomplete data $X(j)$ is

$$\hat{P}(Z \mid \text{incomplete data}, j) = \frac{p_j}{m_j} \sum_{i=1}^{m_j} P(Z \mid M_i(j), X(j)), \tag{4.6}$$

where $M_i(j)$ is i-th generated observation matrix from the distribution (3.7).

Proof. If we notate the augmented data matrix as

$$\begin{bmatrix} X_{11}(j) & X_{12}(j) \\ M_i(j) & X_{22}(j) \end{bmatrix} = [x_1, x_2, \cdots, x_{N_j}], \tag{4.7}$$

and denote their sample mean and unbiased sample covariance matrix to be $\bar{X}_j$ and S_j, respectively. Then $p_j P(Z \mid M_i(j), X(j))$ becomes the same as $p(Z \mid \text{data}, j)$ in (2.1). Hence, composition of the result (4.5) and Geisser's result (2.1) completes the proof. ■

Corollary 3.1. A classification rule for Type 2 missing case is to classify Z into a population which satisfies

$$\max_{1 \le j \le k} \frac{p_j}{m_j} \sum_{i=1}^{m_j} P(Z \mid M_i(j), X(j)). \tag{4.8}$$

Proof. Theorem 3 and rationale of the predictive odds ratio in (2.3) give the result. ■

Given the classification rule in (4.8), a classification rule for Type 1 and Type 3 missing cases are immediate from the following Lemma.

Lemma 4. If in the density(2.1) Z, $\bar{X}_j$, and S_j are partitioned so that $Z' = (Z_1', Z_2')$, $\bar{X} = (\bar{X}_{j1}', \bar{X}_{j2}')$ and $S_j = (S_{jik})$, $i,k=1,2$, where Z_1: $p_1 \times 1$ and S_{j11}: $p_1 \times p_1$, the marginal predictive probability density for classifying Z_1 into $\prod_j$ is given by the multivariate Student t-density

$$P(Z_1 \mid \text{data}, j) = C_j / [1 + \frac{N_j}{N_j^2 - 1} (Z_1 - \bar{X}_{j1})' S_{j11}^{-1} (Z_1 - \bar{X}_{j1})]^{(N_j - P_2)/2}, \tag{4.9}$$

where C_j denotes corresponding constant term as in (2.1).

Proof. Expand the quadratic form in (2.1), complete the square in Z_2, and integrate with respect to Z_2(cf. Press, 1982, p.137).

Corollary 4.1. A classification rule for Type 1 missing case is to classify Z_1 into a population which satisfies

$$\max_{1\leq j\leq k} \quad P(Z_1 \mid \text{data}, j). \tag{4.10}$$

Proof. This is a direct result from Lemma 4. ■

Corollary 4.2. A classification rule for Type 3 missing case is to classify Z_1: $p_1 \times 1$ into a population which satisfies

$$\max_{1\leq j\leq k} \quad \frac{p_j}{m_j} \sum_{i=1}^{m_j} P(Z_1 \mid M_i(j), X(j)), \tag{4.11}$$

where $p_jP(Z_1 \mid M_i(j), X(j))$ is the same expression in (4.9) when we substitute X_{j1} and S_{j11} appeared in (4.9) by corresponding values obtained from the augmented data matrix in (4.7).

Proof. This is a consequence of Theorem 3 and Lemma 4. ■

Note that Corollary 4.2. can also be applicable to Type 3 case when validation sample and training sample have different numbers of missing components.

5. A Monte Carlo Study

Purpose of our Monte Carlo study is to see whether the suggested method performs well. To implement this purpose, we consider three different probabilities of classifying Z into a correct population based on a single training sample. The three probabilities considered in this study can be described as follows:

PP1= posterior probability of Z classified into a correct population calculated by a complete training sample. This can be obtained from Geisser's result in (2.1).

PP2= approximate posterior probability of Z classified into a correct population calculated by the suggested algorithm.

PP3= posterior probability of Z classified into a correct population calculated by the complete observation vector method. All vectors of observation with one or more missing values are omitted from analysis. The Geisser's result in (2.1) is calculated as usual with sample size reduced.

In this study, we generate N random vectors(complete training sample) from a p-variate normal population with given mean $\theta = (p, p, \cdots, p)$ and covariance matrix Σ, and then, to construct incomplete training sample, intentionally assume that p_2 ($< p$) components of first n ($< N$) vector observations generated are missing. With the choice of the varidation sample Z to be θ, we calculate the three posterior probabilities based on two generated training samples-pp1 based on the complete training sample, and pp2 and pp3 based on the incomplete training sample. Since pp1 obviously attains true probability of correct classification, relative probabilities pp2 / pp1 and pp3 / pp1 are compared here. This comparison may show which one performs better between two missing value methods-suggested method and complete observation vector method. Without loss of generality we take the variance equal to one and all correlation coefficients equal: hence $\Sigma = R = \{\rho\}$ and ρ values are restricted to the value such that $\rho \geq -1/(p-1)$ for the equicorrelation matrix R being positive definite.

Table below presents average values of PP2 / PP1 and PP3 / PP1 for various values of α, β, and m, where $\alpha = n/N$ is the proportion of missing values, and $\beta = P_2/p$ is the

Table 5.1. Average Relative Posterior Probilities of Correct Classification

P	α \ m / β	0.1 5	0.1 25	0.1 60	0.1 100	0.05 5	0.05 25	0.05 60	0.05 100	0.03 5	0.03 25	0.03 60	0.03 100
3	0.33	.929*	.910	.913	.933	.961	.967	.965*	.969	.969	.989*	.980	.979*
		(.797)	(.797)	(.797)	(.797)	(.811)	(.811)	(.811)	(.811)	(.944)	(.944)	(.944)	(.944)
	0.66	.821	.840	.837*	.861	.956*	.936	.939	.930*	.942	.942	.947	.950
		(.797)	(.797)	(.797)	(.797)	(.811)	(.811)	(.811)	(.811)	(.944)	(.944)	(.944)	(.944)
6	0.17	.925	.884*	.933	.937	.967*	.952	.958	.962	.969	.982	.983	.983
		(.518)	(.518)	(.518)	(.518)	(.814)	(.814)	(.814)	(.814)	(.884)	(.884)	(.884)	(.844)
	0.33	.542	.592	.823	.818*	.962*	.927	.934	.937	.960	.962*	.960	.961
		(.518)	(.518)	(.518)	(.518)	(.814)	(.814)	(.814)	(.814)	(.884)	(.884)	(.884)	(.884)
	0.50	.670	.775	.739	.810	.898	.919*	.909	.911	.919	.936	.944	.940*
		(.581)	(.518)	(.518)	(.518)	(.814)	(.814)	(.814)	(.814)	(.884)	(.884)	(.884)	(.884)
9	0.11	.872	.912	.923	.925	.978*	.965	.963*	.973	.985*	.981*	.975	.978
		(.491)	(.491)	(.491)	(.491)	(.693)	(.693)	(.693)	(.693)	(.822)	(.822)	(.822)	(.822)
	0.33	.691	.798	.874	.881	.864	.911	.920	.922	.957	.958	.962	.963
		(.491)	(.491)	(.491)	(.491)	(.693)	(.693)	(.693)	(.693)	(.822)	(.822)	(.822)	(.822)
	0.55	.623	.713	.812	.811*	.844	.873	.857*	.896	.926	.944*	.936	.936
		(.491)	(.491)	(.491)	(.491)	(.693)	(.693)	(.693)	(.693)	(.822)	(.822)	(.822)	(.822)

Note each cell contains average value of pp2 / pp1 in the above and pp3 / pp1 in the below with parentheses. "*" denotes the case where average value of pp2 / pp1 does not follow monotonic increase.

proportion of missing components, and m is the number of data augmentations in the suggested algorithm. For each sets of parameters(p, α, β, m), 9 cases were randomly chosen, i.e., 3R(ρ= 0.1, 0.5, 0.9) and 3(N, n) with N and n determined to give the desired α. The entire procedure was performed 864 times, and PP2 / PP1 and PP3 / PP1 were obtained at each time. To summarize our results. we divided the 864 pairs of relative probabilities into 96 sets of 9 values according to the parameter values (p, α, β, m). For each set of 9 pairs of relative probabilities, we averaged them and we present these averages in the table.

As expected, we find in the two methods a tendency for poorer performance as p, α, and β increase. Though these are limited informations from our Monte Carlo study, the differences between PP2 / PP1 and PP3 / PP1 are substantial. Hence the table does show that the suggested algorithm performs better than the complete observation vector method. Especially, for large p and α, the performance is prominent. We note also a tendency for PP2 / PP1 to increase toward 1 with m. This fact gives an indirect verification for the consistency property of the algorithm. Since we can not know the true posterior probability in (2.6), direct verification for the property is not available.

Some of PP2 / PP1 have been starred (*), indicating the case where average value of PP2 / PP1 does not follow monotonic increase. In this study, the average value is based on 9 randomly chosen values of ρ, N, and n without repetition. Hence, we can conjecture that the starred cases occur mainly due to lack of repetition.

6. Concluding Remarks

The area of discriminant analysis still contains many issues that have not been fully resolved yet. These include the question of how to construct a classification rule with unequal covariance matrices when incomplete observation vectors are involved in the analysis. A Bayesian approach to the classification problem in the case of normally distributed observations gives a way of dealing with this problem.

In this article we have suggested a general classification rule by extending Geisser's(1966) result to the case of missing observation. This can be used for handling incomplete data with equal or unequal covariance matrices. General indication, based on the author's limited experiance, is that the performance of the suggested method is better than the complete observation vector method. Especially, for large dimension(p) and proportion of missing values(n / N), the performance of the suggested method over the complete observation vector method is prominent.

In the present paper, our concern was confined to constructing a classification rule for the nested patterned missing data. The issues of more general developments pertaining to a classification rule which covers random patterned missing data are not unimportant but are left for continuing study.

Acknowledgement

The detailed and constructive comments of refrees were very useful in making the paper more readable and in correcting some errors. Their efforts are greatly appreciated.

References

(1) Chan, L.S. and Dunn, O. J. (1972), The Treatment of Missing Values in Discriminant Analysis-1. The Sampling Experiment, *Journal of American Statistical Association,* Vol. 67, 473-477.

(2) Chan, L.S. and Dunn. O. J. (1974), A Note on the Asymptotical Aspect of the Treatment of Missing Values in Discriminant Analysis, *Journal of American Statistical Association,* Vol. 69, 672-673.

(3) Chan, L.S., Gilman, J.A. and Dunn, O. J. (1976), Alternative Approaches to Missing Values in Discriminant Analysis, *Journal of American Statistical Association,* Vol. 71, 841-844.

(4) Dickey, J.M. (1967), Multivariate Generalization of the Multivariae t-Distribution and the Inverted Multivariate t-Distribution, *Annals of Mathematical Statistics* Vol. 38, 511-518.

(5) Fisher, R.A. (1936), The Use of Multiple Measurements in Taxonomic Problems, *Annals of Eugenics,* Vol. 7. 179-188.

(6) Geisser, S. (1964), Posterior Odds for Multivariate Normal Classifications, *Journal of the Royal Statistical Society,* Series B, Vol. 26, 69-76.

(7) Geisser, S. (1966), *Predictive Discrimination, Multivariate Analysis,* Ed. by P. R. Krishnaiah, Academic press, new York, 149-163.

(8) Kim, H.J. (1987), Empirical Bayes Posterior Odds Ratio for Heteroscedastic Classification, *Journal of Korean Statistical Society,* Vol. 16, No. 2, 92-101.

(9) Little, R. J. A. (1978), Consistent Regression Methods for Discriminant Analysis, *Journal of American Statistical Association,* Vol. 73, 391-322.

(10) Press, S. J. (1982), *Applied Multivariate Analysis,* Second Ed., Keieger Co., Florida.

(11) Rubinstein, R.Y. (1981), *Simulation and the Monte Carlo Method,* John Wiley, New York.