Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 65(3) 2011 317-322

Probability of Misclassification for Multiple Groups Sample Linear Discriminant Function

B. Singh
Indian Veterinary Research Institute, Izatnagar, U.P.

Received 24 October 2008; Revised 30 April 2010; Accepted 05 April 2011

SUMMARY

Expressions for probability of misclassification (PMC) for multiple groups sample linear discriminant function (MSLDF) are obtained and some approximations are suggested in case of three groups. The PMC for MSLDF has also been obtained by leave-one-out method using simulated samples from three multivariate normal populations to examine the performance of proposed approximations. The numerical results based on simulated data revealed that the PMC for MSLDF are closer to those provided by the approximations suggested. The modified Johnson approximation is suggested for practical applications to obtain PMC in case of three groups sample linear discriminant function.

Keywords: Multiple groups population linear discriminant function, Multiple groups sample linear discriminant function, Probability of misclassification.

1. INTRODUCTION

Fisher's linear Discriminant function is a popular technique in the field of Discriminant analysis. This yields optimal results in the sense of smallest probability of misclassification (PMC) when parameters are known. Various methods have been discussed in the literature for discriminating between two populations. The multiple groups population Discriminant function, discussed in most multivariate text books (e.g. Anderson 1984), has been extended to situation of mixed continuous and categorical data by Krzanowski (1986). The multiple groups Discriminant function has many applications in the field of agriculture such as to identify insects to different species, animals to different disease status or production level categories, flowers to different categories, buffaloes into different breeds, etc. The use of population Discriminant function may not be justified in the same way when parameters are not known. To investigate the performance of sample linear Discriminant function, one needs the sampling distribution of classification statistic (W). The exact distribution of W in case of two groups was derived by Sitgreaves (1961) but the expression was too complicated to be used, numerically. A method of computation for cumulative distribution function of W by simulation was discussed by Teichroew and Sitgreaves (1961) but actual simulation was not done due to the then low speed of computers. Singh (2001) has derived approximate sampling distribution and whence the PMC for two groups sample linear Discriminant function.

We here consider the case of multiple groups and obtain approximate expressions for PMC associated with multiple groups sample linear Discriminant function (MSLDF). The PMC for MSLDF has also been obtained by leave-one-out method (Lachenbruch

E-mail address : bsingh.1952@gmail.com

and Mickey 1968) using simulated samples from three multivariate normal populations to examine the performance of the proposed approximations for PMC. The leave-one-out method gives more reliable assessment of performance than the re-substitution method which uses same units both for parameter estimation and assessment of performance and hence an overoptimistic assessment of performance is to be expected. In the leave-one-out method the same general procedure is adopted as for the re-substitution method but the unit to be classified is omitted from the database used to estimate the parameters in the model.

2. MULTIPLE GROUPS POPULATION LINEAR DISCRIMINANT FUNCTION

Consider the problem of classifying an observation into one of several populations. Suppose $\pi_1, \pi_2, \ldots, \pi_m$ be the m multivariate normal populations with N($\mu^{(i)}$, Σ) as the density function of population, π_i, $i = 1, 2, \ldots, m$. Let X be a multivariate random observation to be classified into one of these populations. If population parameters are known then define the random variables;

$$U_{ji}(X) = [X - (1/2)\,(\mu_j + \mu_i)]'\Sigma^{-1}(\mu_j - \mu_i) \qquad (2.1)$$

$$U_{ij}(X) = -\,U_{ji}\,(X),\; i, j = 1, 2, \ldots, m\;,\; i \neq j$$

Let $U_{ji}(x)$ be the value of random variable $U_{ji}(X)$ at some point x assumed by the random variable X. The classification procedure divides the space of observations into m mutually exclusive and exhaustive regions $R_1, R_2, \ldots, R_m$ such that if the observation x falls into region R_j it will be allocated to population π_j. The regions of classification (see, Anderson 1984) are defined as

$$R_j : U_{ji}(x) \geq 0\;, \quad i = 1, 2, \ldots, m\;,\; i \neq j \qquad (2.2)$$

If X belongs to population π_j, then U_{ji} is distributed as $N((1/2)\Delta_{ji}^2, \Delta_{ji}^2)$, where Δ_{ji}^2, the Mahalanobis distance between two multivariate populations π_i and π_j, is defined as

$$\Delta_{ji}^2 = (\mu_j - \mu_i)'\Sigma^{-1}(\mu_j - \mu_i) \qquad (2.3)$$

The covariance of U_{ji}, U_{jk} is given by

$$\Delta_{jk,\,ji} = (\mu_j - \mu_k)'\Sigma^{-1}(\mu_j - \mu_i) \qquad (2.4)$$

The probability of misclassification of x from population π_j to any other population π_i is given by

$$\text{PMC}\,(\pi_j) = 1 - \text{PCC}\,(\pi_j) \qquad (2.5)$$

The expression for PCC(π_j), the probability of correct classification (see Anderson 1984), is defined as

$$\text{PCC}\,(\pi_j) = \int_{R_j} f_j\, du_{j1}\, du_{j2} \ldots du_{j,\,j-1}\, du_{j,\,j+1} \ldots du_{jm} \qquad (2.6)$$

where f_j is the joint density function of U_{ji}, $i = 1, 2, \ldots, m$, $i \neq j$. In case of multivariate normal populations, f_j is density of $(m-1)$-variate normal distribution. The numerical value of PCC(π_j) can be worked out by using the tables of multivariate normal distribution. A simple approximation for bi-variate and tri-variate normal integrals is given by Cox and Wermuth (1991).

3. MULTIPLE GROUPS SAMPLE LINEAR DISCRIMINANT FUNCTION

The multiple groups population linear Discriminant function, described in the previous section, is based on the assumption that the parameters are known. In most applications the parameters are not known but are estimated from samples of each population. Suppose we have a sample $x_\alpha^{(j)}$, ($\alpha = 1, 2, \ldots, N_j$, $j = 1, 2, \ldots, m$), from a p-variate normal population π_j with distribution $N(\mu_j, \Sigma)$. These are taken as training samples to obtain estimates of μ_j and Σ. The estimates are defined as

$$\bar{x}_j = \frac{1}{N_j}\sum_{\alpha=1}^{N_j} x_\alpha^{(j)} \quad \text{for } \mu_j,\; j = 1, 2, \ldots, m$$

and

$$S = \frac{1}{n}\sum_{j=1}^{m}\sum_{\alpha=1}^{N_j}(x_\alpha^{(j)} - \bar{x}^{(j)})(x_\alpha^{(j)} - \bar{x}^{(j)})' \text{ for } \Sigma,$$

where $\quad n = \sum_{j=1}^{m} N_j - m$.

The classification statistics are defined as

$$W_{ji}(x) = [x - (1/2)(\bar{x}_j + \bar{x}_i)]'S^{-1}(\bar{x}_j - \bar{x}_i) \qquad (3.1)$$

The classification procedure assigns the observation x to population π_j if x belongs to the region defined by

$$R_j : W_{ji}(x) \geq 0\;,\; i = 1, 2, \ldots, m\;,\; i \neq j \qquad (3.2)$$

3.1 Probability of Misclassification

PMC (π_j), the probability of misclassifying X from population π_j, is given by

$$\text{PMC}(\pi_j) = 1 - \text{PCC}(\pi_j) \quad (3.3)$$

and the probability of correct classification PCC for population π_j is given by

$$\text{PCC}(\pi_j) = P(W_{ji} \geq 0,\ i = 1, 2, \ldots, m,\ i \neq j \mid \pi_j)$$

The distribution of W_{ji} has been derived in the Appendix. Since the sample covariance matrix S is a positive definite matrix, the chi-square variates V_{ji} in (A.5) will always assume positive values. Hence, PCC (π_j) can be expressed as

$$\text{PCC}(\pi_j) = P(U_1^{ji} \geq U_2^{ji},\ i = 1, 2, \ldots, m,\ i \neq j \mid \pi_j) \quad (3.4)$$

3.2 Approximations

The numerical evaluation of expressions (3.4) is not possible in general situations. We here consider the case of three groups and workout approximations based on two-variate F-distribution (see Johnson and Kotz 1972). These approximations are compared with corresponding results based on simulated samples from three multivariate normal populations. Although this comparison may not give exact answer but it frequently gives results that are sufficiently accurate for most practical purposes.

Assume that U_{1ji} and U_{2ji} distributed approximately as $a_{ji}\chi_p^2$ and $b_{ji}\chi_p^2$, where the constants a_{ji} and b_{ji} are obtained by equating the means of U_{1ji} and U_{2ji} with means of $a_{ji}\chi_p^2$ and $b_{ji}\chi_p^2$, respectively. Following two approximations are obtained as modification to Bulgren and Johnson approximations (see Johnson and Kotz 1972, page 242) for two-variate F-distribution.

Modified Bulgren Approximation: A1

$$\text{PCC}(\pi_1) = P(U_{112} > U_{212},\ U_{113} > U_{213})$$
$$\approx I_{w12}[p_{11}/2, p_{11}/2]\ I_{w13}[p_{11}/2, p_{11}/2] \quad (3.5)$$

where

$w_{12} = a_{12}/(a_{12} + d_{12})$, $w_{13} = a_{13}/(a_{13} + d_{13})$,

$p_{11} = p - \rho_1^2$

$d_{12} = b_{12}\{1 - \rho_1^2\}$, $d_{13} = b_{13}\{1 - \rho_1^2\}$ and

$\rho_1 = \Delta_{12,\,13} / (\Delta_{12}\Delta_{13})$.

$$\text{PCC}(\pi_2) = P(U_{121} > U_{221},\ U_{123} > U_{223}) \quad (3.6)$$
$$\approx I_{w21}[p_{12}/2, p_{12}/2]\ I_{w23}[p_{12}/2, p_{12}/2]$$

where

$w_{21} = a_{21}/(a_{21} + d_{21})$, $w_{23} = a_{23}/(a_{23} + d_{23})$,

$p_{12} = p - \rho_2^2$

$d_{21} = b_{21}\{1 - \rho_2^2\}$, $d_{23} = b_{23}\{1 - \rho_2^2\}$ and

$\rho_2 = \Delta_{21,\,23} / (\Delta_{21}\Delta_{23})$.

$$\text{PCC}(\pi_3) = P(U_{131} > U_{231},\ U_{132} > U_{232})$$
$$\approx I_{w31}[p_{13}/2, p_{13}/2]\ I_{w32}[p_{13}/2, p_{13}/2] \quad (3.7)$$

where

$w_{31} = a_{31}/(a_{31} + d_{31})$, $w_{32} = a_{32}/(a_{32} + d_{32})$,

$p_{13} = p - \rho_3^2$

$d_{31} = b_{31}\{1 - \rho_3^2\}$, $d_{32} = b_{32}\{1 - \rho_3^2\}$ and

$\rho_3 = \Delta_{31,\,32} / (\Delta_{31}\Delta_{32})$.

Modified Johnson Approximation: A2

$$\text{PCC}(\pi_1) = P(U_{112} > U_{212},\ U_{113} > U_{213})$$
$$\approx I_{w'12}[p_{21}/2, p_{21}/2]\ I_{w'13}[p_{21}/2, p_{21}/2] \quad (3.8)$$

where

$w'_{12} = a_{12}/(a_{12} + d'_{12})$, $w'_{13} = a_{13}/(a_{13} + d'_{13})$,

$$p_{21} = \frac{p - 2\rho_1^2 p^{-1}}{1 - 2\rho_1^2 p^{-1}}$$

$d'_{12} = b_{12}\{1 - (2\rho_1^2/p)\}$ and

$d'_{13} = b_{13}\{1 - (2\rho_1^2/p)\}$

$$\text{PCC}(\pi_2) = P(U_{121} > U_{221},\ U_{123} > U_{223})$$
$$\approx I_{w'21}[p_{22}/2, p_{22}/2]\ I_{w'23}[p_{22}/2, p_{22}/2] \quad (3.9)$$

where

$w'_{21} = a_{21}/(a_{21} + d'_{21})$, $w'_{23} = a_{23}/(a_{23} + d'_{23})$,

$$p_{22} = \frac{p - 2\rho_2^2 p^{-1}}{1 - 2\rho_2^2 p^{-1}}$$

$d'_{21} = b_{21}\{1 - (2\rho_2^2/p)\}$ and

$d'_{23} = b_{23}\{1 - (2\rho_2^2/p)\}$.

$$\text{PCC}(\pi_3) = P(U_{131} > U_{231},\ U_{132} > U_{232})$$
$$\approx I_{w'31}[p_{13}/2, q_{13}/2]\ I_{w'32}[p_{23}/2, p_{23}/2] \quad (3.10)$$

where

$w'_{31} = a_{31}/(a_{31} + d'_{31})$, $w'_{32} = a_{32}/(a_{32} + d'_{32})$,

$$p_{23} = \frac{p - 2\rho_3^2 p^{-1}}{1 - 2\rho_3^2 p^{-1}}$$

$d'_{31} = b_{31}\{1 - (2\rho_3^2/p)\}$ and

$d'_{32} = b_{32}\{1 - (2\rho_3^2/p)\}$.

4. SIMULATION

Let X be $N_p(\mu, \mathbf{\Sigma})$ and Y be $N_p(0, I)$. The vector Y can be generated by p successive calls to a univariate normal generator. A popular technique for generating standard normal variate is due to Box and Muller (1958). If $\mathbf{\Sigma}$ is a non-singular matrix and $\mathbf{A}$ is $p \times p$ matrix such that $\mathbf{AA'} = \mathbf{\Sigma}$, then $X = AY + \mu$ is the appropriate linear transformation to Y to achieve $N_p(\mu, \mathbf{\Sigma})$. The choice of $\mathbf{A}$ is not unique. Perhaps the best choice for $\mathbf{A}$ is the Choleski factorization which is the lower triangular matrix $\mathbf{L}$ for which $\mathbf{LL'} = \mathbf{\Sigma}$.

Here, we generate $N_1 + N_2 + N_3 + 3$ observations from two 3-variate normal populations, $N_1 + 1$ from π_1, $N_2 + 1$ from π_2 and $N_3 + 1$ from π_3, with given values of parameters. First $N_1 + N_2 + N_3$ observations are used to obtain SDF. The remaining three observations, one from each population, were used to get numerical value for SDF using the leave-one out method for each group, separately. This process was repeated 1000 times to get one value of PMC associated with SDF for each group, separately, for one fixed set of parameters. The corresponding theoretical values are also computed from the expressions (3.5 – 3.10) in section 3 for comparison with simulated results. The numerical results presented in the Table are for the following parameter values of an example for three groups Discriminant function given in Anderson (1984, p 231);

$$m = 3,\ p = 4,\ \mu_1' = (164.51, 86.43, 25.49, 51.24),$$
$$\mu_2' = (160.53, 81.47, 23.84, 48.62) \text{ and}$$
$$\mu_3' = (158.17, 81.16, 21.44, 46.72),$$
$$\sigma_1 = 5.74,\ \sigma_2 = 3.20,\ \sigma_3 = 1.75 \text{ and } \sigma_4 = 3.50$$

The matrix of correlations for all three populations is given by

$$\begin{bmatrix} 1.0000 & 0.5849 & 0.1774 & 0.1974 \\ 0.5849 & 1.0000 & 0.2094 & 0.2170 \\ 0.1774 & 0.2094 & 1.0000 & 0.2910 \\ 0.1974 & 0.2170 & 0.2910 & 1.0000 \end{bmatrix}$$

The covariance matrix is calculated as

$$\mathbf{\Sigma} = \begin{bmatrix} 32.9476 & 10.7434 & 1.7820 & 3.9658 \\ 10.7434 & 10.2400 & 1.1726 & 2.4304 \\ 1.7820 & 1.1726 & 3.0625 & 1.7824 \\ 3.9658 & 2.4304 & 1.7824 & 12.2500 \end{bmatrix}$$

(N_1, N_2, N_3) = (20, 20, 20), (30, 20, 10), (40, 10, 10), (15, 15, 15), (20, 15, 10), (25, 10, 10) and (10, 10, 10)

$$\Delta_{12}^2 = 2.982,\ \Delta_{13}^2 = 6.974 \text{ and } \Delta_{23}^2 = 2.062$$

$$\rho_1 = 0.8658,\ \rho_2 = -0.3894 \text{ and } \rho_3 = 0.7983$$

5. NUMERICAL RESULTS

The actual values of PMC associated with 3 groups population LDF as given in Anderson (1984) are 0.21, 0.42 and 0.24, respectively. The numerical results in the Table reveal that the PMC (simulated) associated with SMDF are more than the PMC associated with population LDF. It is quite natural and has also been reported by Singh (2001) in case of two groups sample LDF. The PMC simulated (S) values for SLDF are closer to those provided by approximation $A2$. Hence this approximation may be used for practical applications in case of three groups Sample Linear Discriminant Analysis.

Table- Probability of misclassification

Sample Size	Population π_1			Population π_2			Population π_3		
(N_1, N_2, N_3)	S	A_1	A_2	S	A_1	A_2	S	A_1	A_2
(20, 20, 20)	0.251	0.183	0.253	0.424	0.491	0 .514	0.290	0.259	0.324
(30, 20, 10)	0.223	0.177	0.240	0.437	0.483	0.500	0.301	0.254	0.308
(40, 10, 20)	0.215	0.167	0.261	0.469	0.489	0.524	0.291	0.258	0.324
(15, 15, 15)	0.229	0.175	0.234	0.461	0.479	0.501	0.283	0.243	0.309
(20, 15, 10)	0.244	0.171	0.227	0.438	0.476	0.497	0.297	0.250	0.313
(25, 10, 10)	0.246	0.168	0.230	0.481	0.484	0.499	0.307	0.252	0.316
(10, 10, 10)	0.265	0.165	0.212	0.452	0.468	0.489	0.309	0.232	0.290

6. PRACTICAL EXAMPLE

The identification of economic meat animals at slaughter age on the basis of some variables (parameters) at earlier age preferably at birth is an important problem in animal sciences. Fifty six lambs were measured for 4 variables, namely body weight (X_1), height (X_2), length (X_3) and heart girth (X_4) at birth were measured. These animals were categorized into three (low, medium and high) groups based on nine-months (slaughter age) body weight. Three groups sample Discriminant function was used to classify these animals into 3 identified groups using SAS and two approximations. The PMC were calculated by using two approximations (A_1 and A_2) and by the leave-one out method (L) using SAS software. The sample means, the pooled covariance matrix (**S**) and of other details these animals are given below;

$$N_1 = N_2 = 17,\ N_3 = 22,\ m = 3,\ p = 4$$

$$\overline{X}_1 = (2.92, 33.35, 38.06, 33.29),$$

$$\overline{X}_2 = (3.55, 35.29, 40.65, 35.06) \text{ and}$$

$$\overline{X}_3 = (3.93, 37.41, 42.00, 36.18)$$

$$\mathbf{S} = \begin{pmatrix} 0.347 & 0.875 & 0.896 & 1.180 \\ 0.875 & 4.957 & 3.479 & 3.685 \\ 0.896 & 3.479 & 5.714 & 3.643 \\ 1.180 & 3.685 & 3.643 & 6.863 \end{pmatrix}$$

$$D_{12}^2 = 1.5279,\ D_{13}^2 = 0.9818 \text{ and } D_{23}^2 = 4.2767$$

$$\rho_1 = 0.9429,\ \rho_2 = -0.7223 \text{ and } \rho_3 = 0.9110$$

7. FINDINGS AND LIMITATIONS

The numerical results for PMC in section 6 though confirm the findings through simulation that approximation 2 provides results more close to the leave-one-out method the results using approximations (A_1, A_2) indicate relationship of PMC with the Mahalanobis distances. The group 1 is more distant from groups 2 and 3 in comparison to group 3 from groups 1 and 3 followed by group 2 from groups 1 and 3 and the PMC is in reverse order. This trend is not depicted by the values obtained through the software SAS. The PMC by two approaches is same for population 2 which has low distances from the two populations.

REFERENCES

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis* (second edition). John Wiley and Sons, New York.

Box, G.E.P. and Muller, M.E. (1958). A note on the generation of random normal deviates. *Ann. Maths. Stat.*, **29**, 610-611.

Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley and Sons.

Cox, D.R. and Wermuth, N. (1991). A simple approximation for bi-variate and tri-variate normal integrals. *Int. Statist. Rev.*, **59(2)**, 263-269.

Krzanowski, W.J. (1986). Multiple discriminant analysis in the presence of mixed continuous and categorical data. *Comp. Maths. Appl.*, **12A(2)**, 179-185.

Lachenbruch, P.A. and Mickey, M.R. (1968). Estimation of error rates in discriminant analysis. *Technometrics,* **10**, 1-11.

Moran, M.A. (1975). On the expectation of errors of allocation associated with linear discriminant function. *Biometrika,* **62**, 141-148.

Singh, B. (2001). On the performance of two sample linear discriminant function. *J. Ind. Soc. Agril. Statist.,* **54(2)**, 209-220.

Sitgreaves, R. (1961). Some results on the distribution of the W classification statistic. In : *Studies in Item Analysis and Prediction* (H. Solomon,Ed), Stanford University Press.

Teichroew, T. and Sitgreaves, R. (1961). Computation of an empirical sampling distribution for the W-classification statistic. In : *Studies in Item Analysis and Prediction* (H. Solomon, Ed), Stanford University Press.

Table- Probability of misclassification

Sample Size	Population π_1			Population π_2			Population π_3		
(N_1, N_2, N_3)	L	A_1	A_2	L	A_1	A_2	L	A_1	A_2
(17, 17, 22)	0.353	0.095	0.209	0.471	0.355	0.443	0.318	0.143	0.241

APPENDIX

Distribution of W_{ji}

We write W_{ij} in (3.1) as

$$W_{ji} = u_{0ji}'\, S^{-1}\, v_{0ji}, \tag{A.1}$$

where $u_{0ji} = (\bar{x}_j + \bar{x}_i)$, $v_{0ji} = [x - (1/2)(\bar{x}_j + \bar{x}_i)]$.

Now suppose that $X \in \pi_j$, then

$$u_{0ji} \sim N\,[\mu_j - m_i,\ (N_j^{-1} + N_i^{-1})\,\Sigma] \text{ and}$$
$$v_{0ji} \sim \mathrm{N}\,[(\mu_j - \mu_i)/2,\ (1+(4N_j)^{-1} + (4N_i)^{-1}))\,\Sigma]$$

Let

$$u_{1ji} = \sqrt{\,[\,N_jN_i\,/\,(N_j + N_i)]}\ u_{0ji} \text{ and}$$
$$v_{1ji} = \sqrt{\,[4N_jN_i\,/(N_j + N_i + 4N_jN_i)]}\ v_{0ji}$$

Then

$$u_{1ji} \sim N\,[(\mu_j - \mu_i)\,\sqrt{\{N_jN_i\,/(N_j + N_i)\}},\ \Sigma] \text{ and}$$
$$v_{1ji} \sim N\,[(\mu_j - \mu_i)\,\sqrt{\{N_jN_i\,/(N_j + N_i + 4N_jN_i)\}},\ \Sigma],$$
$$W_{ji} = k_{ji}\,[(u_{1ji} + v_{1ji})'S^{-1}(u_{1ji} + v_{1ji}) - (u_{1ji} - v_{1ji})'S^{-1}(u_{1ji} - \mathrm{v}_{1ji})], \tag{A.2}$$

where $k_{ji} = (1/8N_jN_i)\sqrt{\,[(N_j + N_i)\,(N_j + N_i + 4N_jN_i)]}$.

Note that $(u_{1ji} + v_{1ji})$ and $(u_{1ji} - v_{1ji})$ are independently normally distributed (see Moran 1975) as $(u_{1ji} + v_{1ji}) \sim N(\delta_{1ji},\ k_{1ji}\,\Sigma)$ and $(u_{1ji} - v_{1ji}) \sim N(\delta_{2\,ji},\ k_{2\,ji}\,\Sigma)$, where

$$\delta_{1ji} = [(\mu_j - \mu_i)\,\sqrt{N_jN_i}\ \{(N_j + N_i)^{-1/2} + (N_j + N_i + 4N_jN_i)^{-1/2}\}],$$
$$k_{1ji} = 2\,[1 + \{(N_j - N_i)\,/\{(N_j + N_i)\,(N_j + N_i + 4N_jN_i)\,\}^{1/2}\}],$$
$$\delta_{2ji} = [(\mu_j - \mu_i)\,\sqrt{N_jN_i}\ \{(N_j + N_i)^{-1/2} - (N_j + N_i + 4N_jN_i)^{-1/2}\,\}],$$
$$k_{2ji} = 2\,[1 - \{(N_j - N_i)\,/\{(N_j + N_i)\,(N_j + N_i + 4N_jN_i)\}^{1/2}\}]\,.$$

Let $t_{1ji} = (u_{1ji} + v_{1ji})\,(k_{1ji})^{-1/2}$ and

$$t_{2ji} = (u_{1ji} - v_{1ji})\,(k_{2ji})^{-1/2}$$

Then one writes that

$$W_{ji} = k_{ji}\,[k_{1ji}\,t_{1ji}'S^{-1}\,t_{1ji} - k_{2ji}\,t_{2ji}'S^{-1}\,t_{2ji}], \tag{A.3}$$

where $t_{1ji} \sim N\,[(\delta_{1ji}/\sqrt{k_{1ji}}),\ \Sigma]$ and $t_{2ji} \sim N\,[(d_{2ji}\,/\sqrt{k_{2ji}}),\ \Sigma]$ are independent.

Now, by using the theorem (5.2.2) of Anderson (1984, p163), we write the classification statistic W as

$$W_{ji} = k_{ji}\,k_{1ji}\,T_{1ji}^2 - k_{ji}\,k_{2ji}\,T_{2ji}^2, \tag{A.4}$$

where

$$T_{1ji}^2 \sim [np/(n - p + 1)]\ \mathrm{F}_{p,\,n - p + 1}(\Delta_{1ji}^2) \text{ and}$$
$$T_{2ji}^2 \sim [np/(n - p + 1)]\ \mathrm{F}_{p,\,n - p + 1}(\Delta_{2ji}^2)$$

with $\mathrm{F}_{a,b}\,(\Delta_{rji}^2)$ as non central F variates and $\Delta_{rji}^2 = (1/k_{rji})\ \delta_{rji}'\Sigma^{-1}\ \delta_{rji}$, $r = 1, 2$, that is,

$$\Delta_{1ji}^2 = (N_jN_i\,/k_{1ji})[\{(N_j + N_i)^{-1/2} + (N_j + N_i + 4N_jN_i)^{-1/2}]^2\ \Delta_{ji}^2 \text{ and}$$

$$\Delta_{2ji}^2 = (N_jN_i\,/k_{2ji})[\{(N_j + N_i)^{-1/2} - (N_j + N_i + 4N_jN_i)^{-1/2}]^2\ \Delta_{ji}^2\,.$$

The exact distribution of W_{ij} (A.4) is difficult to obtain since T_{1ji}^2 and T_{2ji}^2 are not independent. Their denominators are interrelated with identical distribution except when $p = 1$, in that case Hotelling T^2 variates are independent with same denominator.

Here, we assume the same denominator for all values of p and obtain an approximate expression for PMC associated with SLDF. The performance of this approximation in two-groups Discriminant function by Singh (2001) has been found good for practical applications.

With the assumption of same denominator we write W_{ji} as

$$W_{ji} = (U_{1ji} - U_{2ji})/V_{ji}, \tag{A.5}$$

where $U_{1ji} \sim g_{1ji}\chi_p^2(\Delta_{1ji}^2)$, $U_{2ji} \sim g_{2ji}c_p^2(\Delta_{2ji}^2)$ and $V_{ji} \sim \chi_{n-p+1}^2$ are independent chi-square variates. The constants $g_{rji} = nk_{ji}\,k_{rji}$, $r = 1, 2$ are defined as

$$g_{1ji} = (n/4N_jN_i)[N_j - N_i + \{(N_j + N_i)\,(N_j + N_i + 4N_jN_i\,)\}^{1/2}] \text{ and}$$

$$g_{2ji} = (n/4N_jN_i)[N_i - N_j + \{(N_j + N_i)\,(N_j + N_i + 4N_jN_i\,)\}^{1/2}]$$

$$E(U_{1ji}) = g_{1ji}(p + \Delta_{1ji}^2), \text{ and}$$

$$E(U_{2ji}) = g_{2ji}(p + \Delta_{2ji}^2) \tag{A.6}$$

The expressions for $X \in \pi_i$ can be obtained by interchanging δ_{1ji}^2 and δ_{2ji}^2.