Journal of the Korean Data & Information Science Society
2010, **21**(3), 589–595

한국데이터정보과학회지

Classification for intraclass correlation pattern by principal component analysis†

Hie Choon Chung[1] · Chien Pai Han[2]

[1]Department of Healthcare Management, Gwangju University
[2]Department of mathematics, University of Texas at Arlington

Received 2 April 2010, revised 18 May 2010, accepted 23 May 2010

Abstract

In discriminant analysis, we consider an intraclass correlation pattern by principal component analysis. We assume that the two populations are equally likely and the costs of misclassification are equal. In this situation, we consider two procedures, i.e., the test and proportion procedures, for selecting the principal components in classification. We compare the regular classification method and the proposed two procedures. We consider two methods for estimating error rate, i.e., the leave-one-out method and the bootstrap method.

Keywords: Bootstrap method, error rate, intraclass correlation pattern, leave-one-out method, principal component analysis, proportion procedure, test procedure.

1. Introduction

We consider the problem of classifying a px1 observation X of unknown origin to one of two multivariate normal populations π_i : $N(\mu^{(i)}, \Sigma\ )$, i=1, 2, where the covariance matrix has the intraclass correlation pattern, i.e., all variances are equal and covariances are equal. The covariance matrix can be written as $\Sigma = \sigma^2$ [(1-)I ρ + ρ J], where I is a pxp identity matrix, $J = ee^{'}$ and e is a px1 vector of one's. When the parameters are known, the linear discriminant function is

$$L = (\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}[X - \frac{1}{2}(\mu^{(1)} - \mu^{(2)})].$$

We assume that the costs of misclassification are equal and the prior probabilities are equal. Then the observation is classified into π_1 if L $\geq$ 0 and into π_2 if L<0. When the parameters are unknown, they need to be estimated. Suppose independent random samples of size n_i from π_i are available. We consider two cases.

† This study was conducted by research funds from Gwangju University in 2010.
[1] Professor, Department of Healthcare Management, Gwangju University, Gwangju 503-703, Korea. E-mail: hcc@gwangju.ac.kr
[2] Professor, Department of Mathematics, University of Texas at Arlington, TX 76019-0408, U.S.A.

The first case is that the components of the mean vector are equal and the second case is that the components are not equal. In these cases, Han and Chung (2001) found maximum likelihood estimates (MLE), and discussed the error rates and their estimations. Section 2 studies two procedures by principal component analysis for the above classification problem. Section 3 considers the comparison between the regular classification method and the two procedures, i.e., test procedure and proportion procedure, by principal component analysis. Section 4 considers the selection of significance level for test procedure.

2. Classification by principal component analysis

Principal component analysis has been suggested in the literature as a tool for the reduction of dimensionality. Among various optimal properties they possess, an important one is that the variables so constructed are ordered in their contribution to the total variability. However, as has been pointed out by Rao (1964) and Dempster (1969), this ordering often has little or no bearing in the ability of the new variables to 'discriminate' between two or more populations. In fact, criteria other than the magnitude of the variance have to be used to select appropriate principal components for the purpose of classification or discrimination among two (or more) populations. Such criteria are examined in this paper.

An observation X is to be classified into one of two populations, $\pi_1 : N(\mu^{(1)}, \Sigma)$ and $\pi_2 : N(\mu^{(2)}, \Sigma)$, where the covariance matrix has the intraclass correlation pattern. Let Y=HX where H is the Helmert matrix. Then the components of Y are the principal components. We have $\pi_1 : N(\mu_y^{(1)}, D)$ and $\pi_2 : N(\mu_y^{(2)}, D)$, where D is the diagonal covariance matrix of Y with

$$var(Y_1) = \sigma^2[1 + (p-1)\rho], \quad \text{and} \quad var(Y_i) = \sigma^2(1-\rho), i = 2, ..., p.$$

Suppose random samples of sizes n_1 and n_2 are taken from each of two populations. We have the MLE which are

$$\begin{aligned}
\widehat{\mu}_y^{(k)} &= \bar{Y}^{(k)} = (\bar{Y_1}^{(k)}, ..., \bar{Y_p}^{(k)})', k = 1, 2. \\
\widehat{\sigma}^2 &= \frac{A + (p-1)C}{p}, \\
\widehat{\rho} &= \frac{A - C}{A + (p-1)C},
\end{aligned}$$

where

$$\begin{aligned}
A &= \frac{1}{n_1 + n_2} \sum_{k=1}^{2} \sum_{j=1}^{n_k} (Y_{1j}^{(k)} - \bar{Y_1}^{(k)})^2, \\
C &= \frac{1}{(n_1 + n_2)(p-1)} \sum_{k=1}^{2} \sum_{i=2}^{p} \sum_{j=1}^{n_k} (Y_{1j}^{(k)} - \bar{Y_1}^{(k)})^2.
\end{aligned}$$

Now we consider two procedures for selecting the principal components.

2.1. Test procedure

We use the t test for testing the equality of the component means of the two populations. Let $\mu_{yj}^{(k)}$ be the jth component mean of the kth population. To test $H_{o1}: \mu_{y1}^{(1)} = \mu_{y1}^{(2)}$, the test statistic is

$$t_1 = \frac{\bar{Y_1}^{(1)} - \bar{Y_1}^{(2)}}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})S_1^2}}$$

where

$$S_1^2 = \frac{1}{n_1 + n_2 - 2} \sum_{k=1}^{2} \sum_{j=1}^{n_k} (Y_{1j}^{(k)} - \bar{Y_1}^{(k)})^2.$$

The null hypothesis H_{o1} is not rejected if $|t_1| < t_\alpha/2(n_1 + n_2 - 2)$.

To test $H_{oj}: \mu_{yj}^{(1)} = \mu_{yj}^{(2)}$, j=2,...,p, the test statistic is

$$t_j = \frac{\bar{Y_j}^{(1)} - \bar{Y_j}^{(2)}}{\sqrt{(\frac{1}{n}_1 + \frac{1}{n}_2)S_2^2}}$$

where

$$S_2^2 = \frac{1}{(n_1 + n_2 - 2)(p-1)} \sum_{k=1}^{2} \sum_{i=2}^{p} \sum_{j=1}^{n_k} (Y_{ij}^{(k)} - \bar{Y_i}^{(k)})^2.$$

The null hypothesis H_{oj} is not rejected if $|t_j| < t_\alpha/2((n_1 + n_2 - 2)(p - 1))$, j=2,...,p.

We discard all principal components that are not significant. Only the significant principal components are used in the discriminant function.

2.2. Proportion procedure

Let

$$D_1^{*2} = \frac{(\bar{Y_1}^{(1)} - \bar{Y_1}^{(2)})^2}{S_1^2} \quad \text{and} \quad D_j^{*2} = \frac{(\bar{Y_j}^{(1)} - \bar{Y_j}^{(2)})^2}{S_2^2} \quad , j = 2, ..., p.$$

Now arrange these in descending order and obtain $D_1^2 > D_2^2 > ... > D_p^2$. Now we define

$$prop = \frac{D_1^2 + D_2^2 + ... + D_k^2}{D_1^2 + D_2^2 + ... + D_p^2}.$$

The investigator determines a proportion of the total D_j^2 he is willing to accept. Then he will select the first k principal components to satisfy the selected proportion. For example, if the selected proportion is 0.7, we continue to select D_j^2 until prop reaches 0.7, or greater than 0.7. Then the corresponding principal components of the selected D_j^2 are used in the discriminant function.

3. Comparison of error rates

When the components of the mean vector are unequal, we first make an orthogonal transformation to diagonalize Σ. Let Y=HX, where H is the Helmert matrix. Then we have $\pi_1 : \ N(\mu_y^{(1)}, D)$, and $\pi_2 : \ N(\mu_y^{(2)}, D)$, where D is the diagonal covariance matrix. Substituting the MLE described in Section 2 into the discriminant function, we obtain

$$V = (\bar{Y}^{(1)} - \bar{Y}^{(2)})'\widehat{D}^{-}1[Y - \frac{1}{2}(\bar{Y}^{(1)} + \bar{Y}^{(2)})].$$

The observation Y is classified to π_1 if V ≥ 0 and to π_2 if V<0. Another classification statistic is the Anderson's classification statistic

$$W = (\bar{X}^{(1)} - \bar{X}^{(2)})'S^{-}1[X - \frac{1}{2}(\bar{X}^{(1)} + \bar{X}^{(2)})],$$

where $\bar{X}^{(1)}$, $\bar{X}^{(2)}$ and S are the usual sample means and covariance matrix respectively. The observation X is classified to π_1 if $W \geq 0$ and to π_2 if $W < 0$.

We consider the procedures by principal component analysis. Let Y=HX, where H is the Helmert matrix. Then we have $\pi_1 : \ N(\mu_y^{(1)}, D)$ and $\pi_2 : \ N(\mu_y^{(2)}, D)$, where D is the diagonal covariance matrix of Y with $var(Y_1) = \sigma^2[1 + (p-1)\rho]$, and $var(Y_i) = \sigma^2(1-\rho)$, i=2,...,p. When the components of the mean vector are unequal, the MLE $\widehat{\mu}_y^{(i)}$, $\widehat{D}$ of $\mu_y^{(i)}$ and D are described in Section 2. In the first case, i.e., when the components of the mean vector are equal, after we make the Helmert transformation, the population means of the second principal component to pth principal component are all zero for the two populations. Hence these principal components have no discriminant power. All discriminant power is with the first principal component. So we don't need to consider the procedures for selecting the principal components in this case. In the second case, we can select the principal components by the two procedures in Section 2.1 and 2.2. These selected principal components are used to construct the sample discriminant function. We obtain the sample discriminant function,

$$\widetilde{V} = (\bar{\widetilde{Y}}^{(1)} - \bar{\widetilde{Y}}^{(2)})'\widehat{\widetilde{D}^{-1}}[\widetilde{Y} - \frac{1}{2}(\bar{\widetilde{Y}}^{(1)} + \bar{\widetilde{Y}}^{(2)})].$$

The observation $\widetilde{Y}$ is classified to π_1 if $\widetilde{V} \geq 0$ and to π_2 if $\widetilde{V} < 0$.

To study the behavior of these sample discriminant functions, we compare the error rates. The error rate is $1/2[p(1|2) + p(2|1)]$, where $p(i|j)$ be the probability of misclassifying from π_j to π_i.

We compare the simulated error rates for (i) discriminant function L, which has the optimum error rate (OER), (ii) discriminant function V, which has the intraclass-correlation pattern error rate (IER), (iii) discriminant function W, which has Anderson-statistic error rate (AER), and (iv) discriminant function $\widetilde{V}$, which has the principal component error rate (PER). In the Monte Carlo study, we execute various values of μ, ρ, p, α, prop, and sample sizes. 500 runs are made for estimating the error rates. AP represents the average number of selected principal components which are used for discriminant function for each run. The number in the parenthesis represents the standard deviation. In Table 3.1 and Table 3.2, PER is smaller than IER in some cases. In genral, PER is about the same as IER or smaller than IER for small APs.

The error rate depends on the unknown parameters. Hence we need to estimate the error rate. We consider two methods for estimating error rate, i.e., the leave-one-out method and the bootstrap method. The estimators are denoted by LIER and BIER for estimating IER, by LAER and BAER for estimating AER, and by LPER and BPER for estimating PER. Tables give the estimators which are obtained by using 500 runs, and 300 bootstrap samples are generated for each run. From the tables, we can see that the leave-one-out method is almost the same as the bootstrap method.

Table 3.1 Error rate with intraclass correlation pattern (Unequal mean component)

$\pi_1 :\ N(\mu^{(1)}, \Sigma)$, $\mu^{(1)} = (0.0, 0.5, 1.0)$

p = 3 $\quad\pi_2 :\ N(\mu^{(2)}, \Sigma)$, $\mu^{(2)} = (1.5, 2.1, 2.7)$

n1,n2	ρ	OER	Intraclass	Anderson	Principal	α =.01	α =.1	Prop =.7
			IER.017(.004)	AER.023(.009)	PER	.016(.003)	.017(.003)	.016(.003)
	-.3	.014	LIER.017(.029)	LAER.023(.033)	LPER	.017(.028)	.017(.028)	.017(.028)
			BIER.016(.028)	BAER 018(.025)	BPER	.016(.028)	.016(.028)	.016(.028)
					AP	1.0	1.2	1.0
n1=9			IER 094(.012)	AER.106(.020)	PER	.089(.009)	.091(.011)	.088(.008)
	0	.083	LIER.094(.067)	LAER.106(.070)	LPER	.086(.063)	.084(.062)	.086(.063)
			BIER..090(.064)	BAER.095(.068)	BPER	.085(.062)	.083(.062)	.085(.061)
n2=12					AP	1.0	1.2	1.0
			IER.217(.022)	AER.227(.028)	PER	.217(.027)	.217(.026)	.216(.022)
	.9	.196	LIER.220(.103)	LAER.226(.101)	LPER	.205(.098)	.196(.090)	.197(.089)
			BIER.210(.096)	BAER.217(.100)	BPER	.205(.099)	.195(.092)	.196(.089)
					AP	0.9	1.4	1.2
			IER.015(.001)	AER.016(.002)	PER	.014(.001)	.015(.001)	.015(.001)
	-.3	.014	LIER.014(.014)	LAER.015(.015)	LPER	.014(.014)	.014(.014)	.014(.014)
n1=35			BIER.014(.014)	BAER.015(.014)	BPER	.014(.014)	.014(.014)	.014(.014)
					AP	1.0	1.3	1.0
	0	.083	IER.086(.003)	AER.089(.006)	PER	.085(.003)	.085(.003)	.084(.002)
			LIER.085(.033)	LAER.087(.033)	LPER	.082(.032)	.082(.032)	.082(.032)
			BIER.084(.033)	BAER.086(.034)	BPER	.082(.032)	.082(.032)	.082(.032)
n2=35					AP	1.0	1.3	1.0
			IER.203(.007)	AER.206(.009)	PER	.206(.005)	.205(.006)	.206(.004)
	.9	.196	LIER 202(.048)	LAER.206(.048)	LPER	.200(.048)	.198(.047)	.203(.048)
			BIER 202(.048)	BAER.204(0.47)	BPER	.200(.048)	.198(.046)	.203(.048)
					AP	1.2	1.7	1.0

Table 3.2 Error rate with intraclass correlation pattern (Unequal mean component)
π_1 : $N(\mu^{(1)}, \Sigma)$, $\mu^{(1)}$ = (0.1, 0.3, 0.5, 0.7, 0.9. 1.1)
p = 6 π_2 : $N(\mu^{(2)}, \Sigma)$, $\mu^{(2)}$ = (1.8, 2.0, 2.2, 2.4, 2.6, 2.8)

n1, n2	ρ	OER	Intraclass	Anderson	Principal	α =.01	α =.1	Prop =.7
			IER.0020(.0004)	AER.0042(.0023)	PER	.0018(.0003)	.0019(.0003)	.0018(.0003)
	-.1	.0016	LIER.0021(.0076)	LAER.0041(.0108)	LPER	.0018(.0071)	.0019(.0073)	.0018(.0070)
			BIER.0018(.0067)	BAER.0036(.0064)	BPER	.0018(.0069)	.0018(.0068)	.0018(.0067)
					AP	1.0	1.5	1.0
n1=17	.2	.070	IER.080(.008)	AER.097(.018)	PER	.074(.005)	.076(.007)	.073(.004)
			LIER.077(.045)	LAER.096(.050)	LPER	.070(.041)	.069(.041)	.071(.041)
			BIER.077(.046)	BAER.084(.046)	BPER	.070(.042)	.068(.041)	.070(.041)
n2=19					AP	1.0	1.5	1.0
			IER.211(.018)	AER.228(.026)	PER	.192(.011)	.199(.016)	.194(.014)
	.9	.187	LIER.211(.072)	LAER.230(.072)	LPER	.188(.064)	.185(.062)	.188(.062)
			BIER.203(.068)	BAER.209(.072)	BPER	.188(.064)	.183(.062)	.187(.063)
					AP	1.0	1.5	1.1
			IER.0018(.0002)	AER.0026(.0008)	PER	.0017(.0002)	.0018(.0002)	.0017(.0002)
	-.1	.0016	LIER.0017(.0049)	LAER.0022(.0059)	LPER	.0017(.0049)	.0016(.0048)	.0017(.0049)
			BIER.0015(.0047)	BAER.0021(.0047)	BPER	.0016(.0048)	.0015(.0046)	.0016(.0048)
n1=35					AP	1.0	1.5	1.0
			IER.075(.003)	AER.084(.008)	PER	.072(.003)	.073(.003)	.072(.002)
	.2	.070	LIER.074(.034)	LAER.082(.035)	LPER	.071(.032)	.070(.031)	.071(.032)
			BIER.073(.033)	BAER.078(.033)	BPER	.071(.031)	.070(.031)	.070(.032)
n2=35					AP	1.0	1.5	1.0
			IER.200(.008)	AER.209(.013)	PER	.190(.005)	.194(.008)	.189(.003)
	.9	.187	LIER.197(.050)	LAER.206(.052)	LPER	.185(.046)	.182(.046)	.186(.046)
			BIER.195(.050)	BAER.201(.061)	BPER	.185(.047)	.182(.047)	.186(.047)
					AP	1.0	1.5	1.0

4. Selection of significance level for test procedure

In Section 2.1, the test procedure is a preliminary test (Bancroft and Han, 1977; Han *et al.*, 1988). We consider the levels of the test, α =0.01 and 0.1 for various μ, ρ, p and sample sizes. In Table 3.1 and Table 3.2, we can see that PER is the smallest at α =0.1 in most cases. So we may recommend the significance level α =0.1 for test procedure.

5. Conclusion

We consider two procedures, i.e., the test and proportion procedures, for selecting the principal components in classification when the two populations are multivariate normal with intraclass correlation model. These procedures are useful for reducing the dimension and may give smaller error rate. For estimating the error rate, the leave-one-out method and the bootstrap method are about the same.

References

Bancroft, T. A. and Han, C. P. (1977). Inference based on conditional specification: A note and a bibliography. *International Statistical Review*, **45**, 117-127.

Bartlett, M. S. and Please, N. W. (1963). Discrimination in the case of zero mean differences. *Biometrika*, **50**, 17-21.

Dempster, A. P. (1969). *Elements of continuous multivariate analysis*, Addison Wesley, Reading, Mass.

Han, C. P. (1968). A note on discrimination in the case of unequal covariance matrices. *Biometrika*, **55**, 586-587.

Han, C. P. and Chung, H. C. (2001). Discriminant analysis with intraclass correlation pattern. *Proceedings of the Annual Meeting of the American Statistical Association*, **August 5-9**, 555-560.

Han, C. P., Rao, C. V. and Ravichandran, J. (1988). Inference based on conditional specification: A second bibliography. *Communications in Statistics-Theory and Methods*, **17**, 1945-1964.

Rao, C. R. (1964). Use and interpretation of principal components analysis in applied research. *Sankhya A*, **26**, 329-358.