CLASSIFICATION INTO MULTIVARIATE NORMAL POPULATIONS WHEN THE POPULATION MEANS ARE LINEARLY RESTRICTED

M. S. SRIVASTAVA

(Received Jan. 13, 1967)

Summary

This paper considers the problem of classifying a multivariate normal population into one of the k multivariate normal populations when the population means are linearly restricted and the common nonsingular covariance matrix is unknown. It is shown that the maximum likelihood rule is an admissible rule. An example is also given to explain the procedure.

1. Introduction

Let X_i, $i=0, 1, \cdots, k$, a p-dimensional vector, be independently normally distributed with mean vector μ_i, $i=0, 1, \cdots, k$ and common unknown nonsingular covariance matrix Δ. The μ_i's, $i=0, 1, \cdots, k$ are unknown but it is known that $\mu=\xi A$, i.e., we have the following model:

$$H=\begin{cases} \mathrm{E}\, X_0=\mu_0 \\ \mathrm{E} \underset{(p\times k)}{X} = \underset{(p\times k)}{\mu} = \underset{(p\times m)(m\times k)}{\xi A} \end{cases}$$

where $X=(X_1, \cdots, X_k)$, $\mu=(\mu_1, \cdots, \mu_k)$, ξ is a matrix of order $p\times m$ of unkown parameters, and A is a known, $m\times k$ $(m\leqq k)$ matrix of rank m. The case when A is not of full rank can be dealt with similarly. For experiments not involving regression, A is a matrix whose elements are ordinarily 0 or 1. For experiments which involve regression on the so-called "concomitant variables", A is a matrix, some of whose elements involve these "concomitant variates" or non stochastic observation, while the remaining elements are pure constants mostly 0 or 1.

Suppose it is known that $\mu_0=\mu_i$ for exactly one $i\in(1, 2, \cdots, k)$ i.e., we have the following model:

$$H_i = \begin{cases} \mu_0 = \mu_i & \text{for only one } i \ , \\ \mu = \xi A \ , & \end{cases}$$

$$i = 1, 2, \cdots, k.$$

The problem is to decide for which i this is true. Let H_i be the hypothesis that $\mu_0 = \mu_i$ and D_i be the decision of taking $\mu_0 = \mu_i$. The problem is thus: to find a statistical decision procedure for selecting one of the k decisions $(D_1, \cdots, D_k)$ which should be optimum in a certain sense. The next section is devoted to this end.

This problem has been considered by the author [4], [5] and admissible procedures were given under the invariance restriction. In this paper, we show that the maximum likelihood rule is an admissible procedure (in the whole class of procedures). For the known covariance case, this problem has been considered by Ellison [2] and Das Gupta [1]; both showed that the maximum likelihood rule is an admissible procedure.

2. Solution

For deriving the results of this section, we need the following lemma:

LEMMA. *Let S be the matrix of sums of products due to error under the model H and let S_i $(i=1, 2, \cdots, k)$ be the matrix of sums of products under the model H_i, then*

(i) $S = X[I_k - A'(AA')^{-1}A]X'$

(ii) $S_i = (X_0, X)\{I_{k+1} - (C_i, A)'[(C_i, A)(C_i, A)']^{-1}(C_i, A)\}(X_0, X)'$

and

(iii) $S_i = S + V_i V_i'$,

where

(1) $$A = (C_1, \cdots, C_k) \ ,$$

(2) $$Z^{(i)} = A'(AA')^{-1}C_i \ ,$$

(3) $$V_i = (1 + b_i)^{-1/2}(X_0 - XZ^{(i)})$$

and

(4) $$b_i = C_i'(AA')^{-1}C_i \ .$$

PROOF. The proof can easily be obtained from a general result due to Roy (1958). We need to express S_i in a different way. Let

$$\psi_\alpha^{(i)}=\begin{cases} X_\alpha & \text{for } \alpha\neq i, \quad \alpha=1, 2, \cdots, k, \\ (X_\alpha+X_0)/2 & \text{for } \alpha=i, \end{cases} \tag{5}$$

$$\psi^{(i)}=(\psi_1^{(i)}, \cdots, \psi_k^{(i)}), \tag{6}$$

$$U^{(i)}=\psi^{(i)}(D^{(i)})^{1/2}, \tag{7}$$

where $D^{(i)}$ is a diagonal matrix all of whose elements are equal to 1 except the ith element which is equal to 2. It is known (see, e.g., Roy (1958)) that A and $(AD^{(i)})^{1/2}$ can be written as

$$A=TL, \tag{8}$$

and

$$A(D^{(i)})^{1/2}=T^{(i)}L^{(i)}, \tag{9}$$

where T and $T^{(i)}$ are $(m\times m)$ nonsingular triangular matrices, and L and $L^{(i)}$ are $(m\times k)$ semi-orthogonal matrices. Complete $L^{(i)}$ with a matrix $M^{(i)}$ such that

$$\Gamma^{(i)}=\begin{pmatrix} L^{(i)} \\ M^{(i)} \end{pmatrix} \tag{10}$$

is an orthogonal matrix. (Similarly L can also be completed by a matrix M).

Then, under the model H_i, we have

$$\begin{aligned}\sum_0^k (X_\alpha-\mu_\alpha)(X_\alpha-\mu_\alpha)' &=(U^{(i)}L^{(i)\prime}-\xi T^{(i)})(U^{(i)}L^{(i)\prime}-\xi T^{(i)})' \\ &\quad + U^{(i)}M^{(i)\prime}M^{(i)}U^{(i)\prime}+\frac{1}{2}(X_i-X_0)(X_i-X_0)'.\end{aligned} \tag{11}$$

Hence,

$$S_i=U^{(i)}M^{(i)\prime}M^{(i)}U^{(i)\prime}+\frac{1}{2}(X_i-X_0)(X_i-X_0)'. \tag{12}$$

Let

$$\theta=(\theta_1, \theta_2, \cdots, \theta_m)=\xi T \tag{13}$$

and

$$\theta^{(i)}=(\theta_1^{(i)}, \theta_2^{(i)}, \cdots, \theta_m^{(i)})=\xi T^{(i)}. \tag{14}$$

We now proceed to derive the main result of this paper. The set of sufficient statistics for the unknown parameters (θ, μ_0, Δ) is (XL, X_0, S).

The joint probability density function (hereafter, pdf) of XL, X_0 and S, under the hypothesis (or model) H_i is

$$\text{(15)} \qquad \text{Const.}\,|\Delta^{-1}|^{(k+1)/2}\,\text{etr}-\frac{1}{2}\Delta^{-1}[S_i+(U^{(i)}L^{(i)\prime}-\theta^{(i)})(U^{(i)}L^{(i)\prime}-\theta^{(i)})'] \, ,$$

where the symbol etr stands for the exponential of the trace of a square matrix, and where the expression $(\det S)^{(e-p-1)/2}$ is included in the constant, $e=k-m$. We assume that $e\geqq p$.

We use the Bayes technique to prove the admissibility of the maximum likelihood rule. Hence, it will be shown that the maximum likelihood rule is an (a.e.) unique Bayes procedure. The construction of the prior distribution is similar to that of Kiefer and Schwartz [3].

We compute the Bayes procedure relative to the prior distribution $P=\sum_{i=1}^{k} a_i P_i$, $0\leqq a_i\leqq 1$, $\sum_{i=1}^{k} a_i=1$, on the parameter space $\Omega=\bigcup_{i=1}^{k} H_i$ with $\mathrm{P}(\Omega)<\infty$ and with P_i a finite measure on H_i, $i=1,\cdots,k$. In the present investigation, we consider *simple* loss function, i.e., the loss is assumed to be zero or one, according as a correct or incorrect decision was made. Let ϕ_i be the probability of accepting the ith decision, $\sum_{i=1}^{k}\phi_i=1$. Then, for the simple loss function, a decision rule is a *Bayes rule* relative to the apriori distribution P, if and only if, except on a set of Lebesgue measure zero, $\phi_i(T)=0$, whenever

$$\text{(16)} \qquad a_i\int f(T,\gamma)\,P_i(d\gamma)<\max_{j\neq i}\left\{\left(a_j\int f(T,\gamma)\,P_j(d\gamma)\right)\right\} \, ,$$

where $f(T,\gamma)$ is the density function with respect to Lebesgue measure of the distribution of $T=(XL, X_0, S)$, and $\gamma=(\theta,\mu_0,\Delta)\in\Omega$. Let us compute the Bayes procedure for the prior distribution P on Ω in which each H_i has probability $1/k$, $(a_i=1/k)$, and all measure is assigned to Δ's of the form $\Delta^{-1}=I_p+\eta\eta'$, where η is a $p\times q$ matrix, $q\leqq p$, and I_p is a $p\times q$ identity matrix. Let P_i assigns all its measure to $\theta^{(i)}$'s of the form $(I_p+\eta\eta')\theta^{(i)}=\eta\gamma^{(i)}$, where $\gamma^{(i)}=(\gamma_1^{(i)},\cdots,\gamma_m^{(i)})$ is a $(q\times m)$ matrix. Let the conditional prior density of each $\gamma^{(i)}$'s be independent and identically normally distributed with mean vector zero and covariance matrix $I_q+\eta'\eta$, and let the marginal density of η be given by

$$\text{(17)} \qquad \text{Const.}|I_p+\eta\eta'|^{-(e+1)/2} \, , \qquad e\leqq p+q-1 \, ,$$

the integrability of (17) follows from Kiefer and Schwartz [3].

Taking the expectation of (15) with respect to the prior measure of $\gamma_j^{(i)}$'s and then with respect to the prior measure of η, we find that the unconditional joint density of XL, X_0 and S under the hypothesis H_i is given by

$$\text{Const.}\int d\eta \int \Big[\operatorname{etr} - \frac{1}{2}(I_p + \eta\eta')(S_i + U^{(i)}L^{(i)\prime}L^{(i)}U^{(i)\prime}) - 2\eta\gamma^{(i)}L^{(i)}U^{(i)\prime} + \gamma^{(i)}\gamma^{(i)\prime} \Big] d\gamma^{(i)} , \tag{18}$$

by using the following two identities:

$$(I_q + W'W)^{-1} = I_q - W'(I_p + WW')^{-1}W \qquad \text{for } W: p\times q , \tag{19}$$

and

$$|I_q + W'W| = |I_p + WW'| . \tag{20}$$

Hence, the unconditional joint density of XL, X_0 and S under H_i is

$$\begin{aligned} &\text{Const.}\Big[\operatorname{etr} - \frac{1}{2}(S_i + U^{(i)}L^{(i)\prime}L^{(i)}U^{(i)\prime}) \Big] \int \Big[\operatorname{etr} - \frac{1}{2}\eta\eta' S_i \Big] d\eta \\ &= \text{Const.}\Big[\operatorname{etr} - \frac{1}{2}\Big(\sum_0^k X_\alpha X_\alpha' \Big) \Big] |S_i|^{-q/2} \\ &= \text{Const.}\Big[\operatorname{etr} - \frac{1}{2}\Big(\sum_0^k X_\alpha X_\alpha' \Big) \Big] |S|^{-q/2}[1 + V_i' S^{-1} V_i]^{-q/2} . \end{aligned} \tag{21}$$

Since the set of (XL, X_0, S) which yields ties for minimum among these statistics has Lebesgue measure zero, we get the following theorem:

THEOREM. *With simple loss, the maximum likelihood rule is an admissible classification procedure.*

Example. Consider the case when the number of factors affecting the outcome of an experiment is two.

Suppose that one observation is obtained at each of a number of levels of these factors, and denote by X_{ij}, $(i=1, 2, \cdots, a\,;\ j=1, 2, \cdots, b)$ the value observed when the first factor is at the ith level and the second at the jth level. It is assumed that the X_{ij} are independently distributed as a multivariate normal with mean vector μ_{ij} and common unknown nonsingular covariance matrix Δ, and also the two factors act independently so that

$$\mu_{ij} = \alpha + \beta_i + \delta_j$$

and

$$\sum_i \beta_i = 0 = \sum_j \delta_j .$$

Let X_0 be independently distributed as a multivariate normal with mean vector μ_0 and unknown non-singular covariance matrix Δ. The problem is to select one of the k decisions $D_1, \cdots, D_k$, where D_{ij}; $\mu_{ij} = \mu_0$, $k=ab$, $i=1, 2, \cdots, a$ and $j=1, 2, \cdots, b$.

For simplicity, let $a=2$ and $b=3$ so that $k=6$. We have then

$$X=(X_{11}, X_{12}, \cdots, X_{23})$$

and

$$\begin{aligned}\mu&=(\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23})\\ &=(\alpha, \beta_1, \delta_1, \delta_2)\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1\\ 1 & 1 & 1 & -1 & -1 & -1\\ 1 & 0 & -1 & 1 & 0 & -1\\ 0 & 1 & -1 & 0 & 1 & -1\end{pmatrix}\\ &=\xi A=\xi(C_1, \cdots, C_6)\ .\end{aligned}$$

It is easy to check that $AA'=6I$. Hence

$$b_1=\frac{1}{2}\,,\quad b_2=\frac{1}{2}\,,\quad b_3=\frac{1}{4}\,,\quad b_4=\frac{1}{2}\,,\quad b_5=\frac{1}{2}\,,\quad b_6=\frac{1}{4}\,,$$

$$XZ^{(1)}=\frac{1}{6}(3X_{11}+2X_{12}+X_{13}+X_{21}-X_{23})\,,$$

other $XZ^{(i)}$'s can be calculated similarly.

3. Acknowledgement

I wish to thank Professor Charles Stein for his valuable suggestions on the first draft of this paper.

UNIVERSITY OF TORONTO

REFERENCES

[1] S. D. Gupta, "Some special problems in classification," (Abstract), *Ann. Math. Statist.*, 33 (1962), 1504.

[2] B. E. Ellison, "Multivariate-normal classification with covariances known," *Ann. Math. Statist.*, 36 (1965), 1787-1793.

[3] J. Keifer and R. Schwartz, "Admissible Bayes character of T^2, R^2, and other fully invariant tests for classical multivariate normal problems," *Ann. Math. Statist.*, 36 (1965), 747-770.

[4] M. S. Srivastava, "Optimum procedures for classification and related problems," Technical Report No. 11, Stanford University, under NSF Grant GP-40, 1964.

[5] M. S. Srivastava, "Classification into multivariate normal populations when the population means are linearly restricted," (Abstract), *Ann. Math. Statist.*, 35 (1964), 933.