Likelihood Ratio Tests in Multivariate Linear Model

Yasunori Fujikoshi

Department of Mathematics, Graduate School of Science,
Hiroshima University, 1-3-1 Kagamiyama, Higashi Hiroshima, Hiroshima 739-8626, Japan

Abstract

The aim of this paper is to review likelihood ratio test procedures in multivariate linear models, focusing on projection matrices. It is noted that the projection matrices to the spaces spanned by mean vectors in hypothesis and alternatives play an important role. Some basic properties are given for projection matrices. The models treated include multivariate regression model, discriminant analysis model and growth curve model. The hypotheses treated involve a generalized linear hypothesis and no additional information hypothesis, in addition to a usual liner hypothesis. The test statistics are expressed in terms of both projection matrices and sums of squares and products matrices.

AMS 2000 Subject Classification: primary 62H10; secondary 62E20

Key Words and Phrases: Algebraic Approach, Additional Information Hypothesis, Generalized Linear Hypothesis, Growth Curve Model, Multivariate Linear Model, Lambda Distribution, Likelihood Ratio Criterion (LRC), Projection Matrix. *Abbreviated title*: Tests in Multivariate Linear Model.

1. Introduction

In this paper we review statistical inference, especially LRC (Likelihood Ratio Criterion) in multivariate linear model, focusing on matrix theory. Cosider a multivariate linear model with p response variables $y_1, \ldots, y_p$ and k explanatory or dummy variables $x_1, \ldots, x_k$. Suppose that $\boldsymbol{y} = (y_1, \ldots, y_p)'$ and $\boldsymbol{x} = (x_1, \ldots, x_k)'$ are measured for n subjects, and let the observation of the ith subject be denoted by $\boldsymbol{y}_i$ and $\boldsymbol{x}_i$. Then, we have the observation matrices given by

$$\mathbf{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)', \quad \mathbf{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)'. \tag{1.1}$$

It is assumed that $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are independent and have the same covariance matrix $\boldsymbol{\Sigma}$. We express the mean of $\mathbf{Y}$ as follows:

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_p). \tag{1.2}$$

A multivariate linear model is defined by requiring that

$$\boldsymbol{\eta}_i \in \Omega, \quad \text{for all } i = 1, \ldots, p, \tag{1.3}$$

where Ω is a given subspace in the n dimensional Euclid space R^n. A typical Ω is given by

$$\boldsymbol{\Omega} = \mathcal{R}[\mathbf{X}] = \{\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\theta};\ \boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)', -\infty < \theta_i < \infty,\ i = 1, \ldots, k\}. \tag{1.4}$$

Here, $\mathcal{R}[\mathbf{X}]$ is the space spanned by the column vectors of $\mathbf{X}$. A general theory for statistical inference on the regression parameter $\boldsymbol{\Theta}$ can be seen in texts on multivariate analysis, e.g., see Anderson (2003), Arnold (1981), Fujikoshi et al. (2010), Muirhead (1982), Rencher (2002), Seber (2004), Seber (2008), Siotani et al. (1985), etc. In this chapter we discuss with algebraic approach in multivariate linear model.

In section 2 we consider a multivariate regression model in which x_i's are explanatory variables and $\Omega = \mathcal{R}[\mathbf{X}]$. The MLE (Maximum Likelihood Estimator) 's and LRC (Likelihood Ratio Criterion) for $\mathbf{\Theta}_2 = \mathbf{O}$ are derived by using projection matrices. Here, $\mathbf{\Theta} = (\mathbf{\Theta}_1\ \mathbf{\Theta}_2)$. The distribution of LRC is discussed by multivariate Cochran theorem. It is pointed out that projection matrices play an important role. In Section 3 we give a summary of projection matrices. In Section 4 we consider to test an additional information hypothesis of $\boldsymbol{y}_2$ in the prsence of $\boldsymbol{y}_1$, where $\boldsymbol{y}_1 = (y_1 . \ldots, y_q)'$ and $\boldsymbol{y}_2 = (y_{q+1} . \ldots, y_p)'$. In Section 5 we consider testing problems in discriminant analysis. Section 6 deals with a generalized multivariate linear model which is also called the growth curve model. Some related problems are discussed in Section 7.

2. Multivariate Regression Model

In this section we consider a multivariate regression model on p response variables and k explanatory variables denoted by $\boldsymbol{y} = (y_1, \ldots, y_p)'$ and $\boldsymbol{x} = (x_1, \ldots, x_k)'$, respectively. Suppose that we have the observatin matrices given by (1.1). A multivariate regression model is given by

$$\mathbf{Y} = \mathbf{X}\mathbf{\Theta} + \mathbf{E}, \tag{2.1}$$

where $\mathbf{\Theta}$ is a $k \times p$ unknown parameter matrix. It is assumed that the rows of the error matrix $\mathbf{E}$ are independently distributed as a p variate normal distribution with mean zero and unknown covariance matrix $\mathbf{\Sigma}$, i.e., $\mathrm{N}_p(\mathbf{0}, \mathbf{\Sigma})$.

Let $L(\mathbf{\Theta}, \mathbf{\Sigma})$ be the density function or the likelihood function. Then, we have

$$-2 \log L(\mathbf{\Theta}, \mathbf{\Sigma}) = n \log |\mathbf{\Sigma}| + \mathrm{tr}\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{\Theta})'(\mathbf{Y} - \mathbf{X}\mathbf{\Theta}) + np \log(2\pi).$$

The maximum likelihood estimators (MLE) $\hat{\mathbf{\Theta}}$ and $\hat{\mathbf{\Sigma}}$ of $\mathbf{\Theta}$ and $\mathbf{\Sigma}$ are defined by the maximizers of $L(\mathbf{\Theta}, \mathbf{\Sigma})$ or equivalently the minimizers of $-2 \log L(\mathbf{\Theta}, \mathbf{\Sigma})$.

Theorem 2.1. *Suppose that* $\mathbf{Y}$ *follows the multivariate regression model in* (2.1). *Then, the* MLE*'s of* $\mathbf{\Theta}$ *and* $\mathbf{\Sigma}$ *are given as*

$$\hat{\mathbf{\Theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$
$$\hat{\mathbf{\Sigma}} = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{\Theta}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{\Theta}}) = \frac{1}{n}\mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})\mathbf{Y},$$

where $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. *Further, it holds that*

$$-2\log L(\hat{\mathbf{\Theta}}, \hat{\mathbf{\Sigma}}) = n\log|\hat{\mathbf{\Sigma}}| + np\{\log(2\pi) + 1\}.$$

Theorem 2.1 can be shown by a linear algebraic method, which is discussed in the next section. Note that $\mathbf{P}_{\mathbf{X}}$ is the projection matrix on the range space $\mathbf{\Omega} = \mathcal{R}[\mathbf{X}]$. It is symmetric and idempotent, i.e.

$$\mathbf{P}_{\mathbf{X}}' = \mathbf{P}_{\mathbf{X}}, \quad \mathbf{P}_{\mathbf{X}}^2 = \mathbf{P}_{\mathbf{X}}.$$

Next, we consider to test the hypothesis

$$H:\ \mathrm{E}(\mathbf{Y}) = \mathbf{X}_1\mathbf{\Theta}_1 \quad \Leftrightarrow \quad \mathbf{\Theta}_2 = \mathbf{O}, \tag{2.2}$$

against $K;\ \mathbf{\Theta}_2 \neq \mathbf{O}$, where $\mathbf{X} = (\mathbf{X}_1\ \mathbf{X}_2),\ \mathbf{X}_1; n\times j$ and $\mathbf{\Theta} = (\mathbf{\Theta}_1'\ \mathbf{\Theta}_2')',\ \mathbf{\Theta}_1; j\times p$. The hypothesis means that the larst $k-j$ dimensional variate $\boldsymbol{x}_2 = (x_{j+1}, \ldots, x_k)'$ has no additional information in the presense of the first j variate $\boldsymbol{x}_1 = (x_1, \ldots, x_j)'$. In general, the likelihood ratio criterion (LRC) is defined by

$$\lambda = \frac{\max_H L(\mathbf{\Theta}, \mathbf{\Sigma})}{\max_K L(\mathbf{\Theta}, \mathbf{\Sigma})}. \tag{2.3}$$

Then we can express

$$\begin{aligned}
-2\log\lambda &= \min_H\{-2\log L(\mathbf{\Theta}, \mathbf{\Sigma})\} - \min_K\{-2\log L(\mathbf{\Theta}, \mathbf{\Sigma})\} \\
&= \min_H\{n\log|\mathbf{\Sigma}| + \mathrm{tr}(\mathbf{Y} - \mathbf{X\Theta})'(\mathbf{Y} - \mathbf{X\Theta})\} \\
&\quad - \min_K\{n\log|\mathbf{\Sigma}| + \mathrm{tr}(\mathbf{Y} - \mathbf{X\Theta})'(\mathbf{Y} - \mathbf{X\Theta})\}.
\end{aligned}$$

Using Theorem 2.1, we can expressed as

$$\lambda^{2/n} \equiv \mathbf{\Lambda} = \frac{|n\hat{\mathbf{\Sigma}}_\Omega|}{|n\hat{\mathbf{\Sigma}}_\omega|}.$$

Here, $\hat{\mathbf{\Sigma}}_\Omega$ and $\hat{\mathbf{\Sigma}}_\omega$ are the maximum likelihood estimators of $\mathbf{\Sigma}$ under the model (2.1) or K and H, respectively, which are given by

$$\begin{aligned} n\hat{\mathbf{\Sigma}}_\Omega &= (\mathbf{Y} - \mathbf{X}\hat{\mathbf{\Theta}}_\Omega)'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{\Theta}}_\Omega), \quad \hat{\mathbf{\Theta}}_\Omega = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_\Omega)\mathbf{Y} \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} n\hat{\mathbf{\Sigma}}_\omega &= (\mathbf{Y} - \mathbf{X}_1\hat{\mathbf{\Theta}}_{1\omega})'(\mathbf{Y} - \mathbf{X}_1\hat{\mathbf{\Theta}}_{1\omega}), \quad \hat{\mathbf{\Theta}}_{1\omega} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y} \\ &= \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_\omega)\mathbf{Y} \end{aligned} \tag{2.5}$$

Summarizing these results, we have the following theorem.

Theorem 2.2. *Let $\lambda = \Lambda^{n/2}$ be the* LRC *for testing H in* (2.2)*. Then, Λ is expressed as*

$$\Lambda = \frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_h|}, \tag{2.6}$$

where

$$\mathbf{S}_e = n\hat{\mathbf{\Sigma}}_\Omega, \quad \mathbf{S}_h = n\hat{\mathbf{\Sigma}}_\omega - n\hat{\mathbf{\Sigma}}_\Omega, \tag{2.7}$$

and $\mathbf{S}_\Omega$ and $\mathbf{S}_\omega$ are given by (2.4) *and* (2.5)*, respectively.*

The matrices $\mathbf{S}_\Omega$ and $\mathbf{S}_h$ in the testing problem are called the sums of squares and products (SSP) matrices due to the error and the hypothesis, respectively. We consider the distribution of Λ. If a $p \times p$ random matrix $\mathbf{W}$ is expressed as

$$\mathbf{W} = \sum_{j=1}^{n} \boldsymbol{z}_j \boldsymbol{z}_j',$$

where $\boldsymbol{z}_j \sim \mathrm{N}_p(\boldsymbol{\mu}_j, \mathbf{\Sigma})$ and $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ are independent, $\mathbf{W}$ is said to have a noncentral Wishart distribution with n degrees of freedom, covariance matrix $\mathbf{\Sigma}$, and noncentrality matrix $\mathbf{\Delta} = \boldsymbol{\mu}_1\boldsymbol{\mu}_1' + \cdots + \boldsymbol{\mu}_n\boldsymbol{\mu}_n'$. We write that $\mathbf{W} \sim \mathrm{W}_p(n, \mathbf{\Sigma}; \mathbf{\Delta})$. In the special case $\mathbf{\Delta} = \mathbf{O}$, $\mathbf{W}$ is said to have a Wishart distribution, denoted by $\mathbf{W} \sim \mathrm{W}_p(n, \mathbf{\Sigma})$.

Theorem 2.3. (multivariate Cochcran theorem) *Let* $\mathbf{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$, *where* $\boldsymbol{y}_i \sim \mathrm{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 1, \ldots, n$ *and* $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ *are independent. Let* $\mathbf{A}$, $\mathbf{A}_1$, *and* $\mathbf{A}_2$ *be* $n \times n$ *symmetric matrices. Then:*

(1) $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \mathrm{W}_p(k, \boldsymbol{\Sigma}; \boldsymbol{\Omega}) \Leftrightarrow \mathbf{A}^2 = \mathbf{A}, \mathrm{tr}\mathbf{A} = k, \boldsymbol{\Omega} = \mathrm{E}(\mathbf{Y})'\mathbf{A}\mathrm{E}(\mathbf{Y})$.

(2) $\mathbf{Y}'\mathbf{A}_1\mathbf{Y}$ *and* $\mathbf{Y}'\mathbf{A}_2\mathbf{Y}$ *are independent* $\Leftrightarrow \mathbf{A}_1\mathbf{A}_2 = \mathbf{O}$.

For a proof of multivariate Cochcran theorem, see, e.g. Seber (2004, 2008), Fujikoshi et al. (2010), Siotani et al. (1985), etc. Let $\mathbf{B}$ and $\mathbf{W}$ be independent random matrices following the Wishart distribution $\mathrm{W}_p(q, \boldsymbol{\Sigma})$ and $\mathrm{W}_p(n, \boldsymbol{\Sigma})$, respectively, with $n \geq p$. Then, the distribution of

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|}$$

is said to be the p-dimensional Lambda distribution with (q, n)-degrees of freedom, and is denoted by $\Lambda_p(q, n)$. For distributional results of $\Lambda_p(q, n)$, see Anderson (2003), Fujikoshi et al. (2010).

By using multivariate Cochran's Theorem, we have the following distributional results:

Theorem 2.4. *Let* $\mathbf{S}_e$ *and* $\mathbf{S}_h$ *be the random matrices in* (2.7). *Let* Λ *be the* Λ*-statistic defined by* (2.6). *Then,*

(1) $\mathbf{S}_e$ *and* $\mathbf{S}_h$ *are independently distributed as a Wishart distribution* $\mathrm{W}_p(n-k, \boldsymbol{\Sigma})$ *and a noncentral Wishart distribution* $\mathrm{W}_p(k - j, \boldsymbol{\Sigma}; \boldsymbol{\Delta})$, *respectively, where*
$$\boldsymbol{\Delta} = (\mathbf{X}\boldsymbol{\Theta})'(\mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}_1})\mathbf{X}\boldsymbol{\Theta}. \tag{2.8}$$

(2) *Under* H, *the statistic* Λ *is distributed as a lambda distribution* $\Lambda_p(k - j, n - k)$.

Proof. Note that $\mathbf{P}_\Omega = \mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{P}_\omega = \mathbf{P}_{\mathbf{X}_1} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}'$ and $\mathbf{P}_\Omega\mathbf{P}_\omega = \mathbf{P}_\omega\mathbf{P}_\Omega$. By multivariate Chochran's theorem the first result (1) follows by checking that

$$(\mathbf{I}_n-\mathbf{P}_\Omega)^2=(\mathbf{I}_n-\mathbf{P}_\Omega),\ (\mathbf{P}_\Omega-\mathbf{P}_\omega)^2=(\mathbf{P}_\Omega-\mathbf{P}_\omega),\ (\mathbf{I}_n-\mathbf{P}_\Omega)(\mathbf{P}_\Omega-\mathbf{P}_\omega)=\mathbf{O}.$$

The second result (2) follows by showing that $\boldsymbol{\Delta}_0=\mathbf{O}$, where $\boldsymbol{\Delta}_0$ is the $\boldsymbol{\Delta}$ under H. This is seen that

$$\boldsymbol{\Delta}_0=(\mathbf{X}_1\boldsymbol{\Theta}_1)'(\mathbf{P}_\Omega-\mathbf{P}_\omega)(\mathbf{X}_1\boldsymbol{\Theta}_1)=\mathbf{O},$$

since $\mathbf{P}_\Omega\mathbf{X}_1=\mathbf{P}_\omega\mathbf{X}_1=\mathbf{X}_1$. □

The matrices S_e and S_h in (2.7) are defined in terms of $n\times n$ matrices $\mathbf{P}_\Omega$ and $\mathbf{P}_\omega$. It is important to give expressions useful for their numerical computations. We have the following expressions:

$$\mathsf{S}_e=\mathsf{Y}'\mathsf{Y}-\mathsf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathsf{Y},\quad \mathsf{S}_h=\mathsf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathsf{Y}-\mathsf{Y}'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathsf{Y}.$$

Suppose that x_1 is 1 for all subjects, i.e., x_1 is an intercept term. Then, we can express these in terms of the SSP matrix of $(\boldsymbol{y}',\boldsymbol{x}')'$ defined by

$$\mathsf{S}=\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{y}_i-\bar{\boldsymbol{y}}\\ \boldsymbol{x}_i-\bar{\boldsymbol{x}}\end{pmatrix}\begin{pmatrix}\boldsymbol{y}_i-\bar{\boldsymbol{y}}\\ \boldsymbol{x}_i-\bar{\boldsymbol{x}}\end{pmatrix}'=\begin{pmatrix}\mathsf{S}_{yy} & \mathsf{S}_{yx}\\ \mathsf{S}_{xy} & \mathrm{S}_{xx}\end{pmatrix},\tag{2.9}$$

where $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{x}}$ are the sample mean vectors. Along the partition of $\boldsymbol{x}=(\boldsymbol{x}_1',\boldsymbol{x}_2')'$, we partition S as

$$\mathsf{S}=\begin{pmatrix}\mathsf{S}_{yy} & \mathsf{S}_{y1} & \mathsf{S}_{y2}\\ \mathsf{S}_{y1} & \mathsf{S}_{11} & \mathsf{S}_{12}\\ \mathsf{S}_{y2} & \mathsf{S}_{21} & \mathrm{S}_{22}\end{pmatrix}.\tag{2.10}$$

Then,

$$\mathsf{S}_e=\mathsf{S}_{yy\cdot x},\quad \mathsf{S}_h=\mathsf{S}_{y2\cdot 1}\mathsf{S}_{22\cdot 1}^{-1}\mathsf{S}_{2y\cdot 1}.\tag{2.11}$$

Here, we use the notation $\mathsf{S}_{yy\cdot x}=\mathsf{S}_{yy}-\mathsf{S}_{yx}\mathsf{S}_{xx}^{-1}\mathsf{S}_{xy}$, $\mathsf{S}_{y2\cdot 1}=\mathsf{S}_{y2}-\mathsf{S}_{y1}\mathsf{S}_{11}^{-1}\mathsf{S}_{1y}$, etc. These are derived in the next section by using projection matrices.

3. Idempotent Matrices and Max-Mini Problems

In the previous section we have seen that idempotent matrices play an important role on statistical inference inference in multivariate regression model. In fact, letting $\mathrm{E}(\mathbf{Y}) = \boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_p)$, consider a model satisfying

$$\boldsymbol{\eta}_i \in \Omega = \mathcal{R}[\mathbf{R}], \quad for\ all\ i = 1, \ldots, p, \tag{3.1}$$

Then the MLE of $\boldsymbol{\Theta}$ is $\hat{\boldsymbol{\Theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and hence the MLE of $\boldsymbol{\eta}$ is denoted by

$$\hat{\boldsymbol{\eta}}_\Omega = \mathbf{X}\hat{\boldsymbol{\Theta}} = \mathbf{P}_\Omega\mathbf{Y}.$$

Here, $\mathbf{P}_\Omega = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Further, the residual sums of squares and products (RSSP) matrix is expressed as

$$\mathbf{S}_\Omega = (\mathbf{Y} - \hat{\boldsymbol{\eta}}_\Omega)'(\mathbf{Y} - \hat{\boldsymbol{\eta}}_\Omega) = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_\Omega)\mathbf{Y}.$$

Under the hypothesis (2.2), the spaces $\boldsymbol{\eta}_i$'s belong are the same, and is given by $\omega = \mathcal{R}[\mathbf{X}]$. Similarly, we have

$$\begin{aligned} \hat{\boldsymbol{\eta}}_\omega &= \mathbf{X}\hat{\boldsymbol{\Theta}}_\omega = \mathbf{P}_\omega\mathbf{Y}. \\ \mathbf{S}_\omega &= (\mathbf{Y} - \hat{\boldsymbol{\eta}}_\omega)'(\mathbf{Y} - \hat{\boldsymbol{\eta}}_\omega) = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_\omega)\mathbf{Y}, \end{aligned}$$

where $\hat{\boldsymbol{\Theta}}_\omega = (\hat{\boldsymbol{\Theta}}'_{1\omega}\ \mathrm{O})'$ and $\hat{\boldsymbol{\Theta}}_{1\omega} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$. The LR criterion is based on the following decompsition of SSP matries;

$$\begin{aligned} \mathbf{S}_\omega = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_\omega)\mathbf{Y} &= \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_\Omega)\mathbf{Y} + \mathbf{Y}'(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{Y} \\ &= \mathbf{S}_e + \mathbf{S}_h. \end{aligned}$$

The degees of freedom in the Λ distribution $\Lambda_p(f_h, f_e)$ are given by

$$f_e = n - \dim[\Omega], \quad f_h = k - j = \dim[\Omega] - \dim[\omega].$$

In general, an $n \times n$ matrix $\mathbf{P}$ is called idemportent if $\mathbf{P}^2 = \mathbf{P}$. A symmetric and idempotent matrix is called projection matrix. Let $\mathbf{R}^n$ be the n dimensional Euclid space, and Ω be a subspace in $\mathbf{R}^n$. Then, any $n \times 1$ vector $\boldsymbol{y}$ can be uniquely decomposed into direct sum, i.e.,

$$\boldsymbol{y} = \boldsymbol{u} + \boldsymbol{v}, \quad \boldsymbol{u} \in \Omega, \quad \boldsymbol{v} \in \Omega^{\perp}, \tag{3.2}$$

where $\Omega^{\perp}$ is the orthocomplement space. Using decomposition (3.2), consider a mapping

$$\mathbf{P}_{\Omega} : \boldsymbol{y} \longrightarrow \boldsymbol{u}, \quad \text{i.e. } \mathbf{P}_{\Omega}\boldsymbol{y} = \boldsymbol{u}.$$

The mapping is linear, and hence it is expressed as a matrix. In this case, $\boldsymbol{u}$ is called the orthogonal projection of $\boldsymbol{y}$ into Ω, and $\mathbf{P}_{\Omega}$ is also called the orthogonal projection matrix to Ω. Then, we have the following basic properties:

(P1) $\mathbf{P}_{\Omega}$ is uniquely defined.

(P2) $\mathbf{I}_n - \mathbf{P}_{\Omega}$ is the projection matix to $\Omega^{\perp}$.

(P3) $\mathbf{P}_{\Omega}$ is a symmetric idempotent matrix.

(P4) $\mathcal{R}[\mathbf{P}_{\Omega}] = \Omega$, and $\dim[\Omega] = \mathrm{tr}\mathbf{P}_{\Omega}$.

Let ω be a subset of Ω. Then, we have the following properties:

(P5) $\mathbf{P}_{\Omega}\mathbf{P}_{\omega} = \mathbf{P}_{\omega}\mathbf{P}_{\Omega} = \mathbf{P}_{\omega}$.

(P6) $\mathbf{P}_{\Omega} - \mathbf{P}_{\omega} = \mathbf{P}_{\omega^{\perp}\cap\Omega}$, where $\omega^{\perp}$ is the orthcomplement space of ω.

(P7) Let $\mathbf{B}$ be a $q \times n$ matrix, and let $\mathrm{N}(\mathbf{B}) = \{\boldsymbol{y};\ \mathbf{B}\boldsymbol{y} = \mathbf{0}\}$. If $\omega = \mathrm{N}[\mathbf{B}] \cap \Omega$, then $\omega^{\perp} \cap \Omega = \mathcal{R}[\mathbf{P}_{\Omega}\mathbf{B}']$.

For more details, see, e.g. Rao (1973), Harville (1997), Seber (2008), Fujikoshi et al. (2010), etc.

The MLE's and LRC in multivariate regression model are derived by using the following Theorem.

Theorem 3.1.

(1) *Consider a function of* $f(\boldsymbol{\Sigma}) = \log|\boldsymbol{\Sigma}| + \mathrm{tr}\boldsymbol{\Sigma}^{-1}\mathbf{S}$ *of* $p \times p$ *positive definite matrix. Then,* $f(\boldsymbol{\Sigma})$ *takes uniquely the minimum at* $\boldsymbol{\Sigma} = \mathbf{S}$*, and the minimum value is given by*

$$\min_{\boldsymbol{\Sigma} > \mathbf{O}} f(\boldsymbol{\Sigma}) = f(\mathbf{S}) + p.$$

(2) *Let* $\mathbf{Y}$ *be an* $n \times p$ *known matrix and* $\mathbf{X}$ *an* $n \times k$ *known matrix of rank* k*. Consider a function of* $p \times p$ *positive definite matrix* $\boldsymbol{\Sigma}$ *and* $k \times p$ *matrix* $\boldsymbol{\Theta} = (\theta_{ij})$ *given by*

$$g(\boldsymbol{\Theta}, \boldsymbol{\Sigma}) = m \log|\Sigma| + \mathrm{tr}\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\Theta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\Theta}),$$

where $m > 0$*,* $-\infty < \theta_{ij} < \infty$*, for* $i = 1, \ldots, k;\ j = 1, \ldots, p$*. Then,* $g(\boldsymbol{\Theta}, \boldsymbol{\Sigma})$ *takes the minimum at*

$$\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad \boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}} = \frac{1}{m}\mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})\mathbf{Y},$$

and the minimum value is given by $m \log|\hat{\boldsymbol{\Sigma}}| + mp$*.*

Proof. Let $\ell_1, \ldots, \ell_p$ be the characteristic roots of $\boldsymbol{\Sigma}^{-1}\mathbf{S}$. Note that the characteristic roots of $\Sigma^{-1}\mathbf{S}$ and $\boldsymbol{\Sigma}^{-1/2}\mathbf{S}\boldsymbol{\Sigma}^{-1/2}$ are the same. The latter matrix is positive definite, and hence we may assume $\ell_1 \geq \cdots \geq \ell_p > 0$. Then

$$\begin{aligned} f(\boldsymbol{\Sigma}) - f(\mathbf{S}) &= \log|\boldsymbol{\Sigma}\mathbf{S}^{-1}| + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - p \\ &= -\log|\boldsymbol{\Sigma}^{-1}\mathbf{S}| + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - p \\ &= \sum_{i=1}^{p}(-\log \ell_i + \ell_i - 1) \geq 0. \end{aligned}$$

The last inequality follows from $x - 1 \geq \log x$ $(x > 0)$. The equality holds if and only if $\ell_1 = \cdots = \ell_p = 1 \Leftrightarrow \boldsymbol{\Sigma} = \mathbf{S}$.

Next we prove (2). For any positive definite matrix $\boldsymbol{\Sigma}$, we have

$$\begin{aligned} &\mathrm{tr}\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\Theta})'(\mathbf{Y} - \mathbf{X}\Theta) \\ &\quad = \mathrm{tr}\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\Theta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\Theta}}) + \mathrm{tr}\boldsymbol{\Sigma}^{-1}\{\mathbf{X}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})\}'\mathbf{X}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) \\ &\quad \geq \mathrm{tr}\boldsymbol{\Sigma}^{-1}\mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})\mathbf{Y}. \end{aligned}$$

The first equality follow from that $\mathbf{Y} - \mathbf{X}\boldsymbol{\Theta} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\Theta}} + \mathbf{X}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})$ and $(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\Theta}})'\mathbf{X}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) = \mathbf{O}$. In the last step, the equality holds when $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}$. The required result is obtained by noting that $\hat{\boldsymbol{\Theta}}$ does not depend on $\boldsymbol{\Sigma}$ and combining this result with the first result (1). □

Theorem 3.2. *Let* $\mathbf{X}$ *be an* $n \times k$ *matrix of rank* k, *and let* $\Omega = \mathcal{R}[\mathbf{X}]$ *which is defined also by the set* $\{\boldsymbol{y} : \ \boldsymbol{y} = \mathbf{X}\boldsymbol{\theta}\}$, *where* $\boldsymbol{\theta}$ *is a* $k \times 1$ *unknown parameter vector. Let* $\mathbf{C}$ *be a* $c \times k$ *matrix of rank* c, *and define* ω *by the set* $\{\boldsymbol{y} : \ \boldsymbol{y} = X\boldsymbol{\theta}, \ \mathbf{C}\boldsymbol{\theta} = \mathbf{0}\}$. *Then,*

(1) $\mathbf{P}_\Omega = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

(2) $\mathbf{P}_\Omega - \mathbf{P}_\omega = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Proof. (1) Let $\hat{\boldsymbol{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and consider a decomposition $\boldsymbol{y} = \hat{\boldsymbol{y}} + (\boldsymbol{y} - \hat{\boldsymbol{y}})$. Then, $\hat{\boldsymbol{y}}'(\boldsymbol{y} - \hat{\boldsymbol{y}}) = 0$. Therefore, $\mathbf{P}_\Omega \boldsymbol{y} = \hat{\boldsymbol{y}}$ and hence $\mathbf{P}_\Omega = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

(2) Since $\mathbf{C}\boldsymbol{\theta} = \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \cdot \mathbf{X}\boldsymbol{\theta}$, we can write $\omega = \mathrm{N}[\mathbf{B}] \cap \Omega$, where $\mathbf{B} = \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Using (P7),

$$\omega^\perp \cap \Omega = \mathcal{R}[\mathbf{P}_\Omega \mathbf{B}'] = \mathcal{R}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'].$$

The final result is obtained by using (1) and (P7). □

Consider a special case $\mathbf{C} = (\mathbf{O} \ \mathbf{I}_{k-q})$. Then $\omega = \mathcal{R}[\mathbf{X}_1]$, where $\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2)$, $\mathbf{X}_1 : n \times q$. We have the following results:

$$\omega^\perp \cap \Omega = \mathcal{R}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1})\mathbf{X}_2],$$
$$\mathbf{P}_{\omega^\perp \cap \Omega} = (\mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1})\mathbf{X}_2\{\mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1})\mathbf{X}_2\}^{-1}\mathbf{X}_2'(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1}).$$

The expressions (2.11) for $\mathbf{S}_e$ and $\mathbf{S}_h$ in terms of $\mathbf{S}$ can be obtained from projection matrices based on

$$\Omega = \mathcal{R}[\mathbf{X}] = \mathcal{R}[\mathbf{1}_n] + \mathcal{R}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{X}],$$
$$\omega^\perp \cap \Omega = \mathcal{R}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n} - \mathbf{P}_{(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{X}_1})\mathbf{X}_2].$$

4. General Linear Hypothesis

In this section we consider to test a general linear hypothesis

$$H_g: \ \mathbf{C}\Theta\mathbf{D} = \mathbf{O}, \tag{4.1}$$

against alternatives $K_g: \ \mathbf{C}\Theta\mathbf{D} \neq \mathbf{O}$ under a multivariate linear model given by (2.1), where where $\mathbf{C}$ is a $c \times k$ given matrix with rank c and $\mathbf{D}$ is a $p \times d$ given matrix with rank d. When $\mathbf{C} = (\mathbf{O}\ \mathbf{I}_{k-j})$ and $\mathbf{D} = \mathbf{I}_p$, the hypothesis H_g becomes $H : \Theta_2 = \mathbf{O}$.

For the derivation of LR test of (4.1), we can use the following conventional approach: If $\mathsf{U} = \mathsf{Y}\mathbf{D}$, then the rows of U are independent and normally distributed with the identical covariance matrix $\mathbf{D}'\boldsymbol{\Sigma}\mathbf{D}$, and

$$\mathrm{E}(\mathsf{U}) = \mathbf{X}\boldsymbol{\Xi}, \tag{4.2}$$

where $\Xi = \boldsymbol{\Theta}\mathbf{D}$. The hypothesis (4.1) is expressed as

$$H_g : \mathbf{C}\Xi = \mathbf{O}. \tag{4.3}$$

Applying a general theory for testing H_g in (2.1), we have the LRC λ:

$$\lambda^{2/n} = \Lambda = \frac{|\mathsf{S}_e|}{|\mathsf{S}_e + \mathsf{S}_h|}, \tag{4.4}$$

where

$$\begin{aligned}\mathsf{S}_e &= \mathsf{U}'(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathsf{U} \\ &= \mathbf{D}'\mathsf{Y}'(\mathbf{I}_n - \mathbf{P}_\mathbf{A})\mathsf{Y}\mathbf{D},\end{aligned}$$

and

$$\begin{aligned}\mathsf{S}_h &= (\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathsf{U})'\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathsf{U}, \\ &= (\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathsf{Y}\mathbf{D})'\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathsf{Y}D.\end{aligned}$$

Theorem 4.1. *The statistic* Λ *in* (4.4) *is an LR statistic for testing* (4.1) *under* (2.1). *Further, under* H_g, $\Lambda \sim \Lambda_d(c, n-k)$.

Proof. Let $\mathbf{G} = (\mathbf{G}_1\ \mathbf{G}_2)$ be a $p \times p$ matrix such that $\mathbf{G}_1 = \mathbf{D}$, $\mathbf{G}_1'\mathbf{G}_2 = \mathbf{O}$, and $|\mathbf{G}| \neq 0$. Consider a transformation from $\mathbf{Y}$ to

$$(\mathbf{U}\ \mathbf{V}) = \mathbf{Y}(\mathbf{G}_1\ \mathbf{G}_2).$$

Then the rows of $(\mathbf{U}\ \mathbf{V})$ are independently normal with the same covariance matrix

$$\Psi = \mathbf{G}'\Sigma\mathbf{G} = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix}, \quad \Psi_{12} : d \times (p-d),$$

and

$$\begin{aligned} \mathrm{E}[(\mathbf{U}\ \mathbf{V})] &= \mathbf{X}\Theta(\mathbf{G}_1\ \mathbf{G}_2) \\ &= \mathbf{X}(\Xi\ \Delta), \quad \Xi = \Theta\mathbf{G}_1,\ \Delta = \Theta\mathbf{G}_2. \end{aligned}$$

The conditional of $\mathbf{V}$ given $\mathbf{U}$ is normal. The rows of $\mathbf{V}$ given $\mathbf{U}$ are independently normal with the same covariance matrix $\boldsymbol{\Psi}_{11\cdot 2}$, and

$$\begin{aligned} \mathrm{E}(\mathbf{V}|\mathbf{U}) &= \mathbf{X}\boldsymbol{\Delta} + (\mathbf{U} - \mathbf{X}\boldsymbol{\Xi})\boldsymbol{\Gamma} \\ &= \mathbf{X}\boldsymbol{\Delta}^* + \mathbf{U}\Gamma, \end{aligned}$$

where $\boldsymbol{\Delta}^* = \boldsymbol{\Delta} - \Xi\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma} = \boldsymbol{\Psi}_{11}^{-1}\boldsymbol{\Psi}_{12}$. We see that the maximum likelihood of $\mathbf{V}$ given $\mathbf{U}$ does not depend on the hypothesis. Therefore, an LR statistic is obtained from the marginal distribution of $\mathbf{U}$, which implies the results required. □

5. Additional Information Tests for Reseponse Variables

We consider a mutivariate regression model with an intercept term x_0 and k explanatory variables $x_1, \ldots, x_k$ as follows.

$$\mathbf{Y} = \mathbf{1}\boldsymbol{\theta}' + \mathbf{X}\Theta + \mathbf{E}, \tag{5.1}$$

where Y and $\mathbf{X}$ are the observation matrices on $\boldsymbol{y} = (y_1, \ldots, y_p)'$ and $\boldsymbol{x} = (x_1, \ldots, x_k)'$. We assume that the error matrix E has the same property as in (2.1), and rank$(\mathbf{1}_n\ \mathbf{X}) = k+1$. Our interest is to test a hypothesis $H_{2\cdot 1}$ on no additional information of $\boldsymbol{y}_2 = (y_{q+1}, \ldots, y_p)'$ in presense of $\boldsymbol{y}_1 = (y_1, \ldots, y_q)'$.

Along the partition of $\boldsymbol{y}$ into $(\boldsymbol{y}_1', \boldsymbol{y}_2')$, let Y, $\boldsymbol{\theta}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Sigma}$ partition as

$$\mathsf{Y} = (\mathsf{Y}_1\ \mathsf{Y}_2), \quad \boldsymbol{\Theta} = (\boldsymbol{\Theta}_1\ \boldsymbol{\Theta}_2),$$
$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

The conditional distribution of Y_2 given Y_1 is normal with mean

$$\begin{aligned} \mathrm{E}(\mathsf{Y}_2|\mathsf{Y}_1) &= \mathbf{1}\boldsymbol{\theta}_2' + \mathbf{X}\Theta_2 + (\mathsf{Y}_1 - \mathbf{1}_n\boldsymbol{\theta}_1' - \mathsf{X}\boldsymbol{\Theta}_1)\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ &= \mathbf{1}_n\tilde{\boldsymbol{\theta}}_{02}' + \mathbf{X}\tilde{\boldsymbol{\Theta}}_2 + \mathsf{Y}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}, \end{aligned} \tag{5.2}$$

and the conditional covariance matrix is expressed as

$$\mathrm{Var}[\mathrm{vec}\,(\mathsf{Y}_2|\mathsf{Y}_1)] = \boldsymbol{\Sigma}_{22\cdot 1} \otimes \mathbf{I}_n, \tag{5.3}$$

where $\boldsymbol{\Sigma}_{22\cdot 1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, and

$$\tilde{\boldsymbol{\theta}}_2' = \boldsymbol{\theta}_2' - \boldsymbol{\theta}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}, \quad \tilde{\boldsymbol{\Theta}}_2 = \boldsymbol{\Theta}_2 - \boldsymbol{\Theta}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}.$$

Here, for a $n \times p$ matrix $\mathsf{Y} = (\boldsymbol{y}_{(1)}, \ldots, \boldsymbol{y}_{(p)})$, vec$(\boldsymbol{Y})$ means an np-vector $(\boldsymbol{y}_{(1)}', \ldots, \boldsymbol{y}_{(p)}')'$. Now we define the hypothesis $H_{2\cdot 1}$ as

$$H_{2\cdot 1}\colon \boldsymbol{\Theta}_2 = \boldsymbol{\Theta}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \Leftrightarrow \tilde{\boldsymbol{\Theta}}_2 = \mathsf{O}. \tag{5.4}$$

The hypothesis $H_{2\cdot 1}$ means that $\boldsymbol{y}_2$ after removing the effects of $\boldsymbol{y}_1$ does not depend on $\boldsymbol{x}$. In other words, the relationship between $\boldsymbol{y}_2$ and $\boldsymbol{x}$ can be described by the relationship between $\boldsymbol{y}_1$ and $\boldsymbol{x}$. In this sense $\boldsymbol{y}_2$ is redundant in the relationship between $\boldsymbol{y}$ and $\boldsymbol{x}$.

The LR criterion for testing the hypothesis $H_{2\cdot 1}$ against alternatives $K_{2\cdot 1}$: $\tilde{\boldsymbol{\Theta}}_{2\cdot 1} \neq \mathsf{O}$ can be obtained through the following steps.

(D1) The density function of $\mathsf{Y} = (\mathsf{Y}_1\ \mathsf{Y}_2)$ can be expressed as the product of the marginal density function of Y_1 and the conditional density function of Y_2 given Y_1. Note that the density functions of Y_1 under $H_{2\cdot 1}$ and $K_{2\cdot 1}$ are the same.

(D2) The spaces spanned by each columns of $\mathrm{E}(\mathbf{Y}_2|\mathbf{Y}_1)$ are the same, and let the spaces under $K_{2\cdot1}$ and $H_{2\cdot1}$ denote by Ω and ω, respectively. Then

$$\Omega = \mathcal{R}[(\mathbf{1}_n\ \mathbf{Y}_1\ \mathbf{X})], \quad \omega = \mathcal{R}[(\mathbf{1}_n\ \mathbf{Y}_1)],$$

and $\dim(\Omega) = q + k + 1$, $\dim(\omega) = k + 1$.

(D3) The likelihood ratio criterion λ is expressed as

$$\lambda^{2/n} = \Lambda = \frac{|\mathbf{S}_\Omega|}{|\mathbf{S}_\omega|} = \frac{|\mathbf{S}_\Omega|}{|\mathbf{S}_\Omega + (\mathbf{S}_\omega - \mathbf{S}_\Omega)|}.$$

where $\mathbf{S}_\Omega = \mathbf{Y}_2'(\mathbf{I}_n - \mathbf{P}_\Omega)\mathbf{Y}_2$, and $\mathbf{S}_\omega = \mathbf{Y}_2'(\mathbf{I}_n - \mathbf{P}_\omega)\mathbf{Y}_2$.

(D4) Note that $\mathrm{E}(\mathbf{Y}_2|\mathbf{Y}_1)'(\mathbf{P}_\omega - \mathbf{P}_\omega)\mathrm{E}(\mathbf{Y}_2|\mathbf{Y}_1) = \mathbf{O}$ under $\mathrm{H}_{2\cdot1}$. The conditional distribution of Λ under $H_{2\cdot1}$ is $\Lambda_{p-q}(k, n-q-k-1)$, and hence the distribution of Λ under $H_{2\cdot1}$ is $\Lambda_{p-q}(k, n-q-k-1)$.

Note that the Λ statistic is defined through $\mathbf{Y}_2'(\mathbf{I}_n - \mathbf{P}_\Omega)\mathbf{Y}_2$, and $\mathbf{Y}_2'(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{Y}_2$, which involve $n \times n$ matrices. We try to write these statistics in termes of the SSP matrix of $(\boldsymbol{y}', \boldsymbol{x}')'$ defined by

$$\begin{aligned}\mathbf{S} &= \sum_{i=1}^{n} \begin{pmatrix} \boldsymbol{y}_i - \bar{\boldsymbol{y}} \\ \boldsymbol{x}_i - \bar{\boldsymbol{x}} \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_i - \bar{\boldsymbol{y}} \\ \boldsymbol{x}_i - \bar{\boldsymbol{x}} \end{pmatrix}' \\ &= \begin{pmatrix} \mathbf{S}_{yy} & \mathbf{S}_{yx} \\ \mathbf{S}_{xy} & \mathrm{S}_{xx} \end{pmatrix},\end{aligned}$$

where $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{x}}$ are the sample mean vectors. Along the partition of $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2')'$, we partition $\mathbf{S}$ as

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \mathbf{S}_{1x} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \mathbf{S}_{2x} \\ \mathbf{S}_{x1} & \mathbf{S}_{x2} & \mathrm{S}_{xx} \end{pmatrix}.$$

We can show that

$$\begin{aligned}\mathbf{S}_\omega &= \mathbf{S}_{22\cdot1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}, \\ \mathbf{S}_\Omega &= \mathbf{S}_{22\cdot1x} = \mathbf{S}_{22\cdot x} - \mathbf{S}_{21\cdot x}\mathbf{S}_{11\cdot x}^{-1}\mathbf{S}_{12\cdot x}.\end{aligned}$$

The first result is obtained by using

$$\omega = \mathcal{R}[\mathbf{1}_n] + \mathcal{R}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y}_1].$$

The second result is obtained by using

$$\begin{aligned}\Omega &= \mathcal{R}[\mathbf{1}_n] + \mathcal{R}[(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{X}})] \\ &= \mathcal{R}[\mathbf{1}_n] + \mathcal{R}[(\mathbf{I}_n - \mathbf{P}_0)\mathbf{X}] + \mathcal{R}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})(\mathbf{I}_n - \mathbf{P}_0)\mathbf{Y}_1],\end{aligned}$$

where $\tilde{\mathbf{Y}}_1 = (\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y}_1$ and $\tilde{\mathbf{X}} = (\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{X}$.

Summarizing the above results, we have the following theorem.

Theorem 5.1. *In the multivariate regression model* (5.1)*, consider to test the hypothesis* $H_{2\cdot 1}$ *in* (5.4) *against* $K_{2\cdot 1}$*. Then the LR criterion* λ *is given by*

$$\lambda^{2/n} = \Lambda = \frac{|\mathbf{S}_{22\cdot 1x}|}{|\mathbf{S}_{22\cdot 1}|},$$

whose null distribution is $\Lambda_{p-q}(k, n-q-k-1)$*.*

Note that $\mathbf{S}_{22\cdot 1}$ can be decomposed as

$$\mathbf{S}_{22\cdot 1} = \mathbf{S}_{22\cdot 1x} + \mathbf{S}_{2x\cdot 1}\mathbf{S}_{xx\cdot 1}^{-1}\mathbf{S}_{x2\cdot 1}.$$

This decomposition is obtained by expressing $\mathbf{S}_{221x}$ in terms of $\mathbf{S}_{22\cdot 1}$, $\mathbf{S}_{2x\cdot 1}$, $\mathbf{S}_{xx\cdot 1}$ and $\mathbf{S}_{x2\cdot 1}$ by using an inverse formula

$$\begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{H}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} + \begin{pmatrix} -\mathbf{H}_{11}^{-1}\mathbf{H}_{12} \\ \mathbf{I} \end{pmatrix} \mathbf{H}_{22\cdot 1}^{-1} \begin{pmatrix} -\mathbf{H}_{21}\mathbf{H}_{11}^{-1} & \mathbf{I} \end{pmatrix}. \tag{5.5}$$

The decomposition is expressed as

$$\mathbf{S}_{22\cdot 1} - \mathbf{S}_{22\cdot 1x} = \mathbf{S}_{2x\cdot 1}\mathbf{S}_{xx\cdot 1}^{-1}\mathbf{S}_{x2\cdot 1}. \tag{5.6}$$

The result may be also obtained by the following algebraic method. We have

$$\begin{aligned}\mathbf{S}_{22\cdot 1} - \mathbf{S}_{22\cdot 1x} &= \mathbf{Y}_2'(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{Y}_2 \\ &= \mathbf{Y}_2'(\mathbf{P}_{\omega^\perp \cap \Omega})\mathbf{Y}_2,\end{aligned}$$

and

$$\Omega = \mathcal{R}[\mathbf{1}_n] + \mathcal{R}[(\tilde{\mathbf{Y}}_1 \ \tilde{\mathbf{X}})], \quad \omega = \mathcal{R}[\mathbf{1}_n] + \mathcal{R}[\tilde{\mathbf{Y}}_1].$$

Therefore,

$$\begin{aligned} \omega^{\perp} \cap \Omega &= \mathcal{R}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_1} - \mathbf{P}_{\tilde{\mathbf{Y}}_1})(\tilde{\mathbf{Y}}_1 \ \tilde{\mathbf{X}})] \\ &= \mathcal{R}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_1} - \mathbf{P}_{\tilde{\mathbf{Y}}_1})\tilde{\mathbf{X}}], \end{aligned}$$

which gives an expression for $\mathbf{P}_{\omega^{\perp}\cap\Omega}$ by using Theoren 3.1 (1). This leads to (5.6).

6. Tests in Discriminant Analysis

We consider q p-variate normal populations with common covariance matrix $\boldsymbol{\Sigma}$ and the ith population having mean vector $\boldsymbol{\theta}_i$. Suppose that a sample of size n_i is available from the ith population, and let $\boldsymbol{y}_{ij}$ be the jth observation from the ith population. The observation matrix for all the observations is expressed as

$$\mathbf{Y} = (\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{1n_1}, \boldsymbol{y}_{21}, \ldots, \boldsymbol{y}_{q1}, \ldots, \boldsymbol{y}_{qn_q})'. \tag{6.1}$$

It is assumed that $\boldsymbol{y}_{ij}$ are independent, and

$$\boldsymbol{y}_{ij} \sim \mathrm{N}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}),\ j = 1, \ldots, n_i;\ i = 1, \ldots, q, \tag{6.2}$$

The model is expressed as

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\Theta} + \mathbf{E}, \tag{6.3}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_q} \end{pmatrix}, \quad \boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\theta}_1' \\ \boldsymbol{\theta}_2' \\ \vdots \\ \boldsymbol{\theta}_q' \end{pmatrix}.$$

Here, the error matrix $\mathbf{E}$ has the same property as in (2.1).

First we consider to test

$$H: \boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_q (= \boldsymbol{\theta}), \tag{6.4}$$

against alternatives $K: \boldsymbol{\theta}_i \neq \boldsymbol{\theta}_j$ for some i, j. The hypothesis can be expressed as

$$H: \mathbf{C}\boldsymbol{\Theta} = \mathbf{O}, \quad \mathbf{C} = (\mathbf{I}_{q-1},\ -\mathbf{1}_{q-1}). \tag{6.5}$$

The tests including LRC are based on three basic statistics, the within-group SSP matrix $\mathbf{W}$, the between-group SSP matrix $\mathbf{B}$ and the total SSP matrix $\mathbf{T}$ given by

$$\mathbf{W} = \sum_{i=1}^{q} (n_i - 1)\mathbf{S}_i, \quad \mathbf{B} = \sum_{i=1}^{q} n_i (\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})',$$
$$\mathbf{T} = \mathbf{B} + \mathbf{W} = \sum_{i=1}^{q} \sum_{j=1}^{n_i} (\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}})(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}})', \tag{6.6}$$

where $\bar{\boldsymbol{y}}_i$ and $\mathbf{S}_i$ are the mean vector and sample covariance matrix of the ith population, and $\bar{\boldsymbol{y}}$ is the total mean vector defined by $(1/n)\sum_{i=1}^{q} n_i \bar{\boldsymbol{y}}_i$, and $n = \sum_{i=1}^{q} n_i$. In general, $\mathbf{W}$ and $\mathbf{B}$ are independently distributed as a Wishart distribution $\mathrm{W}_p(n-q, \boldsymbol{\Sigma})$ and a noncentral Wishart distribution $\mathrm{W}_p(q-1, \boldsymbol{\Sigma}; \boldsymbol{\Delta})$, respectively where

$$\boldsymbol{\Delta} = \sum_{i=1}^{q} n_i (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})',$$

where $\bar{\boldsymbol{\theta}} = (1/n)\sum_{i=1}^{q+1} n_i \boldsymbol{\theta}_i$. Then, the following theorem is well known.

Theorem 6.1. *Let $\lambda = \Lambda^{n/2}$ be the* LRC *for testing H in* (6.4). *Then, Λ is expressed as*

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|} = \frac{|\mathbf{W}|}{|\mathbf{T}|}, \tag{6.7}$$

where $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{T}$ are given in (6.6). *Further, Under H, the statistic Λ is ditributed as a lambda distribution $\Lambda_p(q-1, n-q)$.*

Now we shall show Theorem 6.1 by an algebraic method. It is easy to see that

$$\Omega = \mathcal{R}[\mathbf{A}], \quad \omega = \mathrm{N}[\mathbf{C}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'] \cap \Omega = \mathcal{R}[\mathbf{1}_n].$$

The last equality is also checked from that under H

$$\mathrm{E}[\mathsf{Y}] = \mathbf{A}\mathbf{1}_q\boldsymbol{\theta}' = \mathbf{1}_n\boldsymbol{\theta}'.$$

We have

$$\begin{aligned}\mathsf{T} &= \mathsf{Y}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathsf{Y} \\ &= \mathsf{Y}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathsf{Y} + \mathsf{Y}'(\mathbf{P}_{\mathbf{A}} - \mathbf{P}_{\mathbf{1}_n})\mathsf{Y} \\ &= \mathsf{W} + \mathsf{B}.\end{aligned}$$

Further, it is easily checked that

(1) $(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})^2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{A}}, \quad (\mathbf{P}_{\mathbf{A}} - \mathbf{P}_{\mathbf{1}_n})^2 = \mathbf{P}_{\mathbf{A}} - \mathbf{P}_{\mathbf{1}_n}.$

(2) $(\mathbf{P}_{\mathbf{A}} - \mathbf{P}_{\mathbf{1}_n})(\mathbf{P}_{\mathbf{A}} - \mathbf{P}_{\mathbf{1}_n}) = \mathbf{O}.$

(3) $f_e = \dim[\mathcal{R}[\mathbf{A}]^{\perp}] = \mathrm{tr}(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}}) = n - q,$
$f_h = \dim[\mathcal{R}[\mathbf{1}_n]^{\perp} \cap \mathcal{R}[\mathbf{A}]] = \mathrm{tr}(\mathbf{P}_{\mathbf{A}} - \mathbf{P}_{\mathbf{1}_n}) = q - 1.$

Related to the test of H, we are interested in whether a subset of variables $y_1, \ldots, y_p$ is sufficient for discriminant analysis, or the set of remainder variables has no additional information or is redundant. Without loss of generality we consider the sufficiency of a subvector $\boldsymbol{y}_1 = (y_1, \ldots, y_k)'$ of $\boldsymbol{y}$, or redundancy of the remainder vector $\boldsymbol{y}_2 = (y_{k+1}, \ldots, y_p)'$. Consider to test

$$H_{2\cdot 1} : \boldsymbol{\theta}_{1;2\cdot 1} = \cdots = \boldsymbol{\theta}_{q;2\cdot 1}(= \boldsymbol{\theta}_{2\cdot 1}), \tag{6.8}$$

where

$$\boldsymbol{\theta}_i = \begin{pmatrix} \boldsymbol{\theta}_{i;1} \\ \boldsymbol{\theta}_{i;2} \end{pmatrix}, \quad \boldsymbol{\theta}_{i;1};\ k \times 1, \quad i = 1, \ldots, q,$$

and

$$\boldsymbol{\theta}_{i;2\cdot 1} = \boldsymbol{\theta}_{i;2} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\theta}_{i;1}, \quad i = 1, \ldots, q.$$

The testing problem was considered by Rao (1948). The hypothsis can be formulated in terms of Maharanobis distance and discriminant functions. For its details, see Rao (1973) and Fujikoshi (1987). To obtain a likelihood ratio for $H_{2\cdot 1}$, we partition the observation matrix as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 & \mathbf{Y}_2 \end{pmatrix}, \quad \mathbf{Y}_1 : n \times k.$$

Then the conditional distribution of $\mathbf{Y}_2$ given $\mathbf{Y}_1$ is normal such that the rows of $\mathbf{Y}_2$ are independently distributed with covariance matrix $\boldsymbol{\Sigma}_{22\cdot 1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, and the conditional mean is given by

$$\mathrm{E}(\mathbf{Y}_2|\mathbf{Y}_1) = \mathbf{A}\boldsymbol{\Theta}_{2\cdot 1} + \mathbf{Y}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}, \tag{6.9}$$

where $\boldsymbol{\Theta}_{2\dot{1}} = (\boldsymbol{\theta}_{1;2\cdot 1}, \ldots, \boldsymbol{\theta}_{q;2\cdot 1})'$. The LRC for $H_{2\cdot 1}$ can be obtained by use of the conditional distribution, and following the steps (D1)-(D4) in Section 5. In fact, The spaces spanned by each columns of $\mathrm{E}(\mathbf{Y}_2|\mathbf{Y}_1)$ are the same, and let the spaces under $K_{2\cdot 1}$ and $H_{2\cdot 1}$ denote by Ω and ω, respectively. Then

$$\Omega = \mathcal{R}[(\mathbf{A}\ \mathbf{Y}_1)], \quad \omega = \mathcal{R}[(\mathbf{1}_n\ \mathbf{Y}_1)],$$

$\dim(\Omega) = q + k$, and $\dim(\omega) = q + 1$. The likelihood ratio criterion λ can be expressed as

$$\lambda^{2/n} = \Lambda = \frac{|\mathbf{S}_\Omega|}{|\mathbf{S}_\omega|} = \frac{|\mathbf{S}_\Omega|}{|\mathbf{S}_\Omega + (\mathbf{S}_\omega - \mathbf{S}_\Omega)|}.$$

where $\mathbf{S}_\Omega = \mathbf{Y}_2'(\mathbf{I}_n - \mathbf{P}_\Omega)\mathbf{Y}_2$, and $\mathbf{S}_\omega = \mathbf{Y}_2'(\mathbf{I}_n - \mathbf{P}_\omega)\mathbf{Y}_2$. We exprese the LRC in terms of $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{T}$. Let us partition $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{T}$ as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{pmatrix}, \tag{6.10}$$

where $\mathbf{W}_{12} : q \times (p - q)$, $\mathbf{B}_{12} : q \times (p - q)$ and $\mathbf{T}_{12}$: $q \times (p - q)$. Notig that $\mathbf{P}_\Omega = \mathbf{P}_\mathbf{A} + \mathbf{P}_{(\mathbf{I}_n - \mathbf{P}_\mathbf{A})\mathbf{Y}_1}$, we have

$$\begin{aligned} \mathbf{S}_\Omega &= \mathbf{Y}_2' \left\{ \mathbf{I}_n - \mathbf{P}_\mathbf{A} - (\mathbf{I}_n - \mathbf{P}_\mathbf{A})\mathbf{Y}_1\{\mathbf{Y}_1'(\mathbf{I}_n - \mathbf{P}_\mathbf{A})\mathbf{Y}_1\}^{-1}\mathbf{Y}_1'(\mathbf{I}_n - \mathbf{P}_\mathbf{A}) \right\} \mathbf{Y}_2 \\ &= \mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12} = \mathbf{W}_{22\cdot 1}. \end{aligned}$$

Similarly, notig that $\mathbf{P}_\omega = \mathbf{P}_{\mathbf{1}_n} + \mathbf{P}_{(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y}_1}$, we have

$$\begin{aligned}\mathbf{S}_\omega &= \mathbf{Y}_2' \left\{ \mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n} - (\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y}_1\{\mathbf{Y}_1'(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y}_1\}^{-1}\mathbf{Y}_1'(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n}) \right\} \mathbf{Y}_2 \\ &= \mathbf{T}_{22} - \mathbf{T}_{21}\mathbf{T}_{11}^{-1}\mathbf{T}_{12} = \mathbf{T}_{22\cdot 1}.\end{aligned}$$

Theorem 6.2. *Suppose that the observation matrix* $\mathbf{Y}$ *in* (6.1) *is a set of samples from* $\mathrm{N}_p(\boldsymbol{\theta}_i, \boldsymbol{\Sigma})$, $i = 1, \ldots, q$. *Then the likelihood ratio criterion* λ *for the hypothesis* $H_{2\cdot 1}$ *in* (6.8) *is given by*

$$\lambda = \left(\frac{|\mathbf{W}_{22\cdot 1}|}{|\mathbf{T}_{22\cdot 1}|} \right)^{n/2},$$

where $\mathbf{W}$ *and* $\mathbf{T}$ *are given by* (6.6). *Further, under* $H_{2\cdot 1}$,

$$\frac{|\mathbf{W}_{22\cdot 1}|}{|\mathbf{T}_{22\cdot 1}|} \sim \Lambda_{p-k}(q-1, n-q-k).$$

Proof. We consider the conditional distributions of $\mathbf{W}_{22\cdot 1}$ and $\mathbf{T}_{22\cdot 1}$ given $\mathbf{Y}_1$ by using Theorem 2.3, and see also that they does not depend on $\mathbf{Y}_1$. We have seen that

$$\mathbf{W}_{22\cdot 1} = \mathbf{Y}_2'\mathbf{Q}_1\mathbf{Y}_2, \quad \mathbf{Q}_1 = \mathbf{I}_n - \mathbf{P}_{\mathbf{A}} - \mathbf{P}_{(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{Y}_1}.$$

It is easy to see that $\mathbf{Q}_1^2 = \mathbf{Q}_1$, $\mathrm{rank}(\mathbf{Q}_1) = \mathrm{tr}\mathbf{Q}_1 = n-q-k$, $\mathbf{Q}_1\mathbf{A} = \mathbf{O}$, $\mathbf{Q}_1\mathbf{X}_1 = \mathbf{O}$, and

$$\mathrm{E}(\mathbf{Y}_2|\mathbf{Y}_1)'\mathbf{Q}_1\mathrm{E}(\mathbf{Y}_2|\mathbf{Y}_1) = \mathbf{O}.$$

This implies that $\mathbf{W}_{22\cdot 1}|\mathbf{Y}_1 \sim \mathrm{W}_{p-k}(n-q-k, \boldsymbol{\Sigma}_{22\cdot 1})$ and hence $\mathbf{W}_{22\cdot 1} \sim \mathrm{W}_{p-k}(n-q-k, \boldsymbol{\Sigma}_{22\cdot 1})$. For $\mathbf{T}_{22\cdot 1}$, we have

$$\mathbf{T}_{22\cdot 1} = \mathbf{Y}_2'\mathbf{Q}_2\mathbf{Y}_2, \quad \mathbf{Q}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n} - \mathbf{P}_{(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y}_1},$$

and hence

$$\mathbf{T}_{22\cdot 1} - \mathbf{W}_{22\cdot 1} = \mathbf{Y}_2'(\mathbf{Q}_2 - \mathbf{Q}_1)\mathbf{Y}_2.$$

Similarly $\mathbf{Q}_2$ is idempotent. Using $\mathbf{P}_{\mathbf{1}_n}\mathbf{P}_{\mathbf{A}} = \mathbf{P}_{\mathbf{A}}\mathbf{P}_{\mathbf{1}_n} = \mathbf{P}_{\mathbf{1}_n}$, we have $\mathbf{Q}_1\mathbf{Q}_2 = \mathbf{Q}_2\mathbf{Q}_1 = \mathbf{Q}_1$, and hence

$$(\mathbf{Q}_2 - \mathbf{Q}_1)^2 = \mathbf{Q}_2 - \mathbf{Q}_1, \qquad \mathbf{Q}_1 \cdot (\mathbf{Q}_2 - \mathbf{Q}_1) = \mathbf{O}.$$

Further, under $H_{2\cdot 1}$,

$$\mathrm{E}(\mathbf{X}_2|\mathbf{X}_1)'(\mathbf{Q}_2 - \mathbf{Q}_1)\mathrm{E}(\mathbf{X}_2|\mathbf{X}_1) = \mathbf{O}.$$

$\square$

7. General Multivariate Linear Model

In this section we consider a general multivariate linear model as follows. Let $\mathbf{Y}$ be an $n \times p$ observation matrix whose rows are independently distributed as p-variate normal distribution with a common covariance matrix $\boldsymbol{\Sigma}$. Suppose that the mean of $\mathbf{Y}$ is given as

$$\mathrm{E}(\mathbf{Y}) = \mathbf{A}\boldsymbol{\Theta}\mathbf{X}', \tag{7.1}$$

where $\mathbf{A}$ is an $n \times k$ given matrix with rank k, $\mathbf{X}$ is a $p \times q$ matrix with rank q, and $\boldsymbol{\Theta}$ is a $k \times q$ unknown parameter matrix. For a motivation of (7.1), consider the case when a single variable y is measured at p time points $t_1, \ldots, t_p$ (or different conditions) on n subjects chosen at random from a group. Suppose that we denote the variable y at time point t_j by y_j. Let the observations $y_{i1}, \ldots, y_{ip}$ of the ith subject be denoted by

$$\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ip})', \quad i = 1, \ldots, n.$$

If we consider a polynomial regression of degree $q-1$ of y on the time variable t, then

$$\mathrm{E}(\boldsymbol{y}_i) = \mathbf{X}\boldsymbol{\theta},$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & t_1 & \cdots & t_1^{q-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & t_p & \cdots & t_p^{q-1} \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_q \end{pmatrix}.$$

If there are k different groups and each goups have a polynomial regression of degree $q-1$ of y, we have a model given by (7.1). From such motivation the model (7.1) is also called a growth curve model. For its detail, see Pothoff and Roy (1964).

Now, let us consider to derive LRC for a general linear hypothesis

$$H_g : \mathbf{C\Theta D} = \mathbf{O}, \tag{7.2}$$

against alternatives $K_g : \mathbf{C\Theta D} \neq \mathbf{O}$. Here, $\mathbf{C}$ is a $c \times k$ given matrix with rank c, and $\mathbf{D}$ is a $q \times d$ given matrix with rank d. This problem was discussed by Khatri (1966), Gleser and Olkin (1970), etc. Here, we obtain LRC by reducing it to the problem of obtaining LRC for a general linear hypothesis in a multivariate linear model. In order to relate the model (7.1) to a multivariate linear model, consider the transformation from Y to $(\mathsf{U}\ \mathsf{V})$:

$$(\mathsf{U}\ \mathsf{V}) = \mathsf{Y}\mathbf{G}, \quad \mathbf{G} = (\mathbf{G}_1\ \mathbf{G}_2), \tag{7.3}$$

where $\mathbf{G}_1 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, $\mathbf{G}_2 = \tilde{\mathbf{X}}$, and $\tilde{\mathbf{X}}$ is a $p \times (p-q)$ matrix satisfying $\tilde{\mathbf{X}}'\mathbf{X} = \mathsf{O}$ and $\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \mathbf{I}_{p-q}$. Then, the rows of $(\mathsf{U}\ \mathsf{V})$ are independently distributed as p-variate normal distributions with means

$$\mathrm{E}[(\mathsf{U}\ \mathsf{V})] = (\mathbf{A}\Theta\ \mathbf{O}),$$

and the common covariance matrix

$$\Psi = \mathbf{G}'\Sigma\mathbf{G} = \begin{pmatrix} \mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1 & \mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_2 \\ \mathbf{G}_2'\mathbf{\Sigma}\mathbf{G}_1 & \mathbf{G}_2'\mathbf{\Sigma}\mathbf{G}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{\Psi}_{11} & \mathbf{\Psi}_{12} \\ \mathbf{\Psi}_{21} & \mathbf{\Psi}_{22} \end{pmatrix}.$$

This transformation can be regarded as one from $\boldsymbol{y} = (y_1, \ldots, y_p)'$ to a q-variate main variable $\boldsymbol{u} = (u_1, \ldots, u_q)'$ and a $(p-q)$-variate auxiliary variable $\boldsymbol{v} = (v_1, \ldots, v_{p-q})'$. The model (7.1) is equivalent to the following joint model of two componets:

(1) The conditional distribution of U given V is

$$\mathsf{U} \mid \mathsf{V} \sim \mathrm{N}_{n\times q}(\mathsf{A}^*\mathbf{\Xi},\ \mathbf{\Psi}_{11\cdot 2}). \tag{7.4}$$

(2) The marginal distribution of $\mathbf{V}$ is

$$\mathbf{V} \sim \mathrm{N}_{n\times(p-q)}(\mathbf{O},\ \mathbf{\Psi}_{22}), \tag{7.5}$$

where

$$\mathbf{A}^* = (\mathbf{A}\ \mathbf{V}), \quad \mathbf{\Xi} = \begin{pmatrix} \mathbf{\Theta} \\ \mathbf{\Gamma} \end{pmatrix},$$
$$\mathbf{\Gamma} = \mathbf{\Psi}_{22}^{-1}\mathbf{\Psi}_{21},\ \ \mathbf{\Psi}_{11\cdot 2} = \mathbf{\Psi}_{11} - \mathbf{\Psi}_{12}\mathbf{\Psi}_{22}^{-1}\mathbf{\Psi}_{21}.$$

Befor we obtain LRC, first we consider the MLE's in (7.1). Applying a general theory of multivariate linear model to (7.4) and (7.5), the MLEs of $\mathbf{\Xi}$, $\mathbf{\Psi}_{11\cdot 2}$, and $\mathbf{\Psi}_{22}$ are given by

$$\hat{\Xi} = (\mathbf{A}^{*\prime}\mathbf{A}^*)^{-1}\mathbf{A}^{*\prime}\mathbf{U}, \quad n\hat{\mathbf{\Psi}}_{11\cdot 2} = \mathbf{U}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}^*})\mathbf{U}, \quad n\hat{\mathbf{\Psi}}_{22} = \mathbf{V}'\mathbf{V}. \tag{7.6}$$

Let

$$\mathbf{S} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{Y}, \quad \mathbf{W} = \mathbf{G}'\mathbf{S}\mathbf{G} = (\mathbf{U}\ \mathbf{V})'(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})(\mathbf{U}\ \mathbf{V}),$$

and partition $\mathbf{W}$ as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}, \quad \mathbf{W}_{12} : q \times (p-q).$$

Theorem 7.1. *For an $n \times p$ observation matrix $\mathbf{Y}$, assume a general multivariate linear model given by* (7.1)*. Then:*

(1) *The MLE $\hat{\mathbf{\Theta}}$ of $\mathbf{\Theta}$ is given by*

$$\hat{\mathbf{\Theta}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{S}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}.$$

(2) *The MLE $\hat{\mathbf{\Psi}}_{11\cdot 2}$ of $\mathbf{\Psi}_{11\cdot 2}$ is given by*

$$n\hat{\mathbf{\Psi}}_{11\cdot 2} = \mathbf{W}_{11\cdot 2} = (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}.$$

Proof. The MLE of $\boldsymbol{\Xi}$ is $\hat{\boldsymbol{\Xi}} = (\mathbf{A}^{*'}\mathbf{A}^*)^{-1}\mathbf{A}^{*'}\mathbf{U}$. The inverse formula (see (5.5)) gives

$$\begin{aligned}\mathbf{Q} = (\mathbf{A}^{*'}\mathbf{A}^*)^{-1} &= \begin{pmatrix} (\mathbf{A}'\mathbf{A})^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \\ &+ \begin{pmatrix} -(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{V} \\ \mathbf{I}_{p-q} \end{pmatrix} [\mathbf{V}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{V}]^{-1} \begin{pmatrix} -(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{V} \\ \mathbf{I}_{p-q} \end{pmatrix}' \\ &= \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\hat{\boldsymbol{\Theta}} &= (\mathbf{Q}_{11}\mathbf{A}' + \mathbf{Q}_{12}\mathbf{V}')\mathbf{U} \\ &= (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{G}_1 - (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{G}_2(\mathbf{G}_2'\mathbf{S}\mathbf{G}_2)^{-1}\mathbf{G}_2'SG_1.\end{aligned}$$

Using

$$\mathbf{G}_2(\mathbf{G}_2'\mathbf{S}\mathbf{G}_2)^{-1}\mathbf{G}_2' = \mathbf{S}^{-1} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{S}^{-1}\mathbf{G}_2)^{-1}\mathbf{G}_1'\mathbf{S}^{-1},$$

we obtain (1). For a derivation of (2), let $\mathbf{B} = (\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{V}$. Then, using $\mathbf{P}_{\mathbf{A}^*} = \mathbf{P}_{\mathbf{A}} + \mathbf{P}_{\mathbf{B}}$, the first expression of (1) is obtained. Similarly, the second expression of (2) is obtained. □

Theorem 7.2. *Let* $\lambda = \Lambda^{n/2}$ *be the LRC for testing the hypothesis* (7.2) *in the generalized multivariate linear model* (7.1). *Then,*

$$\Lambda = |\mathbf{S}_e|/|\mathbf{S}_e + \mathbf{S}_h|,$$

where

$$\mathbf{S}_e = \mathbf{D}'(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{D}, \quad \mathbf{S}_h = (\mathbf{C}\hat{\Theta}\mathbf{D})(\mathbf{C}\mathbf{R}\mathbf{C}')^{-1}\mathbf{C}\hat{\Theta}\mathbf{D}$$

and

$$\begin{aligned}\mathbf{R} = (\mathbf{A}'\mathbf{A})^{-1} &+ (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{S}^{-1}\{\mathbf{S} - \mathbf{X}'(\mathbf{X}\mathbf{S}^{-1}\mathbf{X}')^{-1}\mathbf{X}\} \\ &\times \mathbf{S}^{-1}\mathbf{Y}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}.\end{aligned}$$

Here $\hat{\boldsymbol{\Theta}}$ *is given in Theorem* 7.1 (1). *Further, the null distribution is* $\Lambda_d(c, n-k-(p-q))$.

Proof. The test of H_g in (7.2) agaist alternatives K_g is equivalent to testing

$$H_g : \mathbf{C}^* \Xi \mathbf{D} = \mathbf{O} \tag{7.7}$$

under the conditional model (7.4), where $\mathbf{C}^* = (\mathbf{C}\ \mathbf{O})$. Since the distribution of $\mathbf{V}$ does not depend on H_g, the LR test under the conditional model is the LR test under the unconditional model. Using a general result for a general linear hypothesis given in Theorem 4.1, we obtain

$$\Lambda = |\tilde{\mathbf{S}}_e| / |\tilde{\mathbf{S}}_e + \tilde{\mathbf{S}}_h|,$$

where

$$\tilde{\mathbf{S}}_e = \mathbf{D}'\mathbf{U}(\mathbf{I}_n - \mathbf{A}^*(\mathbf{A}^{*'}\mathbf{A}^*)^{-1}\mathbf{A}^{*'})\mathbf{U}\mathbf{D},$$
$$\tilde{\mathbf{S}}_h = (\mathbf{C}\hat{\Xi}\mathbf{D})(\mathbf{C}^*(\mathbf{A}^{*'}\mathbf{A}^*)^{-1}\mathbf{C}^{*'})^{-1}\mathbf{C}\hat{\Xi}D.$$

By reduction similar to those of MLEs, it is seen that $\tilde{\mathbf{S}}_e = \mathbf{S}_e$ and $\tilde{\mathbf{S}}_h = \mathbf{S}_h$. This completes the proof. □

8. Concluding Remarks

In this paper we discuss LRC in multivariate linear model, focussing on the role of projection matrices. Testing problems considered involve the hypotheses on selection of variables or no additional information of a set of variables, in addition to a typical linear hypothesis. It may be noted that various LRC's and their distributions are obtained by aljebraic methods.

We have not discussed with LRC's for the hypothsis of selection of variables in canonical correlation analysis, and for dimensionality in multivariate linear model. Some results for these problems can be founded in Fujikoshi (1982), Fujikoshi et al. (2010).

In multivariate analysis, there are some other test crteria such as Lawley-Hotelling trace criterion and Bartlett-Nanda-Pillai trace criterion. For the testing problems treated in this chapter, it is possible to propose such criteria as in Fujikoshi (1989).

The LRC's for tests of no additional information of a set of variables will be useful in selection of variables. For example, it is possible to propose model selection criteria such as AIC (see Akakike (1973)).

Acknowledgements

The author wishes to thank Dr.Tetsuro Sakurai for his variable comments for the first draft. The author's research is partially supported by the Ministry of Education, Science, Sports, and Culture, a Grant-in-Aid for Scientific Research (C), #25330038, 2013-2015.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (B.N. Petrov and F. Csáki, eds.). Akadémia Kiado, Budapest, Hungary, 267–281.

[2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, N.J.

[3] Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*, Wiley, Hoboken, N.J.

[4] Fujikoshi, Y. (1982). A test for additional information in canonical correlation analysis, *Ann. Inst. Statist. Math.*, **34**, 137–144.

[5] Fujikoshi, Y. (1989). Tests for redundancy of some variables in multivariate analysis. *Recent Developments in Statistical Data Analysis and Inference* (Y. Dodge, Ed.), 141-163, Elesevier Science Publishers B.V., Amsterdam.

[6] Fujikoshi, Y., Ulyanov, V. V. and Shimizu, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley, Hobeken, N.J.

[7] Gleser, L. J. and Olkin, I. (1970). Linear models in multivariate analysis, In *Essays in Probability and Statistics*, (R.C. Bose, ed.), University of North Carolina Press, Chapel Hill, NC, 267–292.

[8] Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.

[9] Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curve. *Ann. Inst. Statist. Math.*, **18**, 75–86.

[10] Kshirsagar, A. M. (1995). *Growth Curves*. Marcel Dekker, Inc.

[11] Muirhead, R. J. (1982). *Aspect of Multivariate Statistical Theory*, Wiley, New York.

[12] Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika,*, **51**, 313–326.

[13] Rao, C. R. (1948). Tests of significance in multivariate analysis. *Biometrika*, **35**, 58–79.

[14] Rao, C. R. (1970). Inference on discriminant function coefficients. In *Essays in Prob. and Statist.* (R. C. Bose, ed.), 587–602.

[15] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

[16] Rencher, A. C. (2002). *Methods of Multivariate Analysis*, 2nd ed. Wiley, New York.

[17] Seber, G. A. F. (1984). *Multivariate Observations*. Wiley, New York.

[18] Seber, G. A. F. (2008). 2nd. *A Matrix Handbook for Statisticians*. Wiley, Hobeken, N.J.

[19] Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*, American Sciences Press, Columbus, OH.