CHAPTER 9

Inference for Means in Multivariate Linear Models

Essentially, this chapter consists of a number of examples of estimation and testing problems for means when an observation vector has a normal distribution. Invariance is used throughout to describe the structure of the models considered and to suggest possible testing procedures. Because of space limitations, maximum likelihood estimators are the only type of estimators discussed. Further, likelihood ratio tests are calculated for most of the examples considered.

Before turning to the concrete examples, it is useful to have a general model within which we can view the results of this chapter. Consider an n-dimensional inner product space $(V, (\cdot,\cdot))$ and suppose that X is a random vector in V. To describe the type of parametric models we consider for X, let f be a decreasing function on $[0, \infty)$ to $[0, \infty)$ such that $f[(x, x)]$ is a density with respect to Lebesgue measure on $(V, (\cdot,\cdot))$. For convenience, it is assumed that f has been normalized so that, if $Z \in V$ has density f, then $\text{Cov}(Z) = I$. Obviously, such a Z has mean zero. Now, let M be a subspace of V and let γ be a set of positive definite linear transformations on V to V such that $I \in \gamma$. The pair (M, γ) serves as the parameter space for a model for X. For $\mu \in M$ and $\Sigma \in \gamma$,

$$p(x|\mu, \Sigma) \equiv |\Sigma|^{-n/2} f\left[(x - \mu, \Sigma^{-1}(x - \mu))\right]$$

is a density on V. The family

$$\{p(\cdot|\mu, \Sigma) | \mu \in M, \Sigma \in \gamma\}$$

determines a parametric model for X. It is clear that if $p(\cdot|\mu, \Sigma)$ is the density of X, then $\mathcal{E}X = \mu$ and $\text{Cov}(X) = \Sigma$. In particular, when

$$f(u) = (\sqrt{2\pi})^{-n}\exp[-\tfrac{1}{2}u], \qquad u \geq 0,$$

then X has a normal distribution with mean $\mu \in M$ and covariance $\Sigma \in \gamma$. The parametric model for X is in fact a linear model for X with parameter set (M, γ). Now, assume that $\Sigma(M) = M$ for all $\Sigma \in \gamma$. Since $I \in \gamma$, the least-squares and Gauss–Markov estimator of μ are equal to PX where P is the orthogonal projection onto M. Further, $\hat{\mu} \equiv PX$ is also the maximum likelihood estimator of μ. To see this, first note that $P\Sigma = \Sigma P$ for $\Sigma \in \gamma$ since M is invariant under $\Sigma \in \gamma$. With $Q = I - P$, we have

$$\begin{aligned}(x - \mu, \Sigma^{-1}(x - \mu)) &= (P(x - \mu) + Qx, \Sigma^{-1}(P(x - \mu) + Qx)) \\ &= (Px - \mu, \Sigma^{-1}(Px - \mu)) + (Qx, \Sigma^{-1}Qx).\end{aligned}$$

The last equality is a consequence of

$$(Qx, \Sigma^{-1}P(x - \mu)) = (x, Q\Sigma^{-1}P(x - \mu)) = (x, QP\Sigma^{-1}(x - \mu)) = 0$$

as $QP = 0$ and $\Sigma^{-1}P = P\Sigma^{-1}$. Therefore, for each $\Sigma \in \gamma$,

$$(x - \mu, \Sigma^{-1}(x - \mu)) \geq (Qx, \Sigma^{-1}Qx)$$

with equality iff $\mu = Px$. Since the function f was assumed to be decreasing, it follows that $\hat{\mu} = PX$ is the maximum likelihood estimator of μ, and $\hat{\mu}$ is unique if f is strictly decreasing. Thus under the assumptions made so far, $\hat{\mu} = PX$ is the maximum likelihood estimator for μ. These assumptions hold for most of the examples considered in this chapter. To find the maximum likelihood estimator of Σ, it is necessary to compute

$$\sup_{\Sigma \in \gamma} |\Sigma|^{-n/2} f[(Qx, \Sigma^{-1}Qx)]$$

and find the point $\hat{\Sigma} \in \gamma$ where the supremum is achieved, assuming it exists. The solution to this problem depends crucially on the set γ and this is what generates the infinite variety of possible models, even with the assumption that $\Sigma M = M$ for $\Sigma \in \gamma$. The examples of this chapter are generated by simply choosing some γ's for which $\hat{\Sigma}$ can be calculated explicitly.

We end this rather lengthy introduction with a few general comments about testing problems. In the notation of the previous paragraph, consider a parameter set (M, γ) with $I \in \gamma$ and assume $\Sigma M = M$ for $\Sigma \in \gamma$. Also, let

$M_0 \subset M$ be a subspace of V and assume that $\Sigma M_0 = M_0$ for $\Sigma \in \gamma$. Consider the problem of testing the null hypothesis that $\mu \in M_0$ versus the alternative that $\mu \in (M - M_0)$. Under the null hypothesis, the maximum likelihood estimator for μ is $\hat{\mu}_0 = P_0 X$ where P_0 is the orthogonal projection onto M_0. Thus the likelihood ratio test rejects the null hypothesis for small values of

$$\Lambda(x) = \frac{\sup_{\Sigma \in \gamma} |\Sigma|^{-n/2} f[(Q_0 x, \Sigma^{-1} Q_0 x)]}{\sup_{\Sigma \in \gamma} |\Sigma|^{-n/2} f[(Qx, \Sigma^{-1} Qx)]}$$

where $Q_0 = I - P_0$. Again, the set γ is the major determinant with regard to the distribution, invariance, and other properties of $\Lambda(x)$. The examples in this chapter illustrate some of the properties of γ that lead to tractable solutions to the estimation problem for Σ and the testing problem described above.

9.1. THE MANOVA MODEL

The multivariate general linear model introduced in Example 4.4, also known as the multivariate analysis of variance model (the MANOVA model), is the subject of this section. The vector space under consideration is $\mathcal{L}_{p,n}$ with the usual inner product $\langle \cdot, \cdot \rangle$ and the subspace M of $\mathcal{L}_{p,n}$ is

$$M = \{x | x = Z\beta,\ \beta \in \mathcal{L}_{p,k}\}$$

where Z is a fixed $n \times k$ matrix of rank k. Consider an observation vector $X \in \mathcal{L}_{p,n}$ and assume that

$$\mathcal{L}(X) = N(\mu, I_n \otimes \Sigma)$$

where $\mu \in M$ and Σ is an unknown $p \times p$ positive definite matrix. Thus the set of covariances for X is

$$\gamma = \{I_n \otimes \Sigma | \Sigma \in \mathcal{S}_p^+\}$$

and (M, γ) is the parameter set of the linear model for X. It was verified in Example 4.4 that M is invariant under each element of γ. Also, the orthogonal projection onto M is $P = P_z \otimes I_p$ where

$$P_z = Z(Z'Z)^{-1}Z'.$$

Further, $Q = I - P = Q_z \otimes I_p$ is the orthogonal projection onto $M^{\perp}$ where

$Q_z = I_n - P_z$. Thus

$$\hat{\mu} = PX = (P_z \otimes I_p)X = P_z X$$

is the maximum likelihood estimator of $\mu \in M$ and, from Example 7.10,

$$\hat{\Sigma} = \frac{1}{n} X' Q_z X$$

is the maximum likelihood estimator of Σ when $n - k \geqslant p$, which we assume throughout this discussion. Thus for the MANOVA model, the maximum likelihood estimators have been derived. The reader should check that the MANOVA model is a special case of the linear model described at the beginning of this chapter.

We now turn to a discussion of the classical MANOVA testing problem. Let K be a fixed $r \times k$ matrix of rank r and consider the problem of testing

$$H_0 : K\beta = 0 \quad \text{versus} \quad H_1 : K\beta \neq 0$$

where $\mu = Z\beta$ is the mean of X. It is not obvious that this testing problem is of the general type described in the introduction to this chapter. However, before proceeding further, it is convenient to transform this problem into what is called the canonical form of the MANOVA testing problem. The essence of the argument below is that it suffices to take

$$Z = Z_0 \equiv \begin{pmatrix} I_k \\ 0 \end{pmatrix}, \qquad K = K_0 \equiv (I_r \;\; 0)$$

in the above problem. In other words, a transformation of the original problem results in a problem where $Z = Z_0$ and $K = K_0$. We now proceed with the details. The parametric model for $X \in \mathcal{L}_{p,n}$ is

$$\mathcal{L}(X) = N(Z\beta, I_n \otimes \Sigma)$$

and the statistical problem is to test $H_0 : K\beta = 0$ versus $H_1 : K\beta \neq 0$. Since Z has rank k, $Z = \Psi U$ for some linear isometry $\Psi : n \times k$ and some $k \times k$ matrix $U \in G_U^+$. The k columns of Ψ are the first k columns of some $\Gamma \in \mathcal{O}_n$ so

$$\Psi = \Gamma \begin{pmatrix} I_k \\ 0 \end{pmatrix} = \Gamma Z_0.$$

Setting $\tilde{X} = \Gamma' X$, $\tilde{\beta} = U\beta$, and $\tilde{K} = KU^{-1}$, we have

$$\mathcal{L}(\tilde{X}) = N(Z_0 \tilde{\beta}, I_n \otimes \Sigma)$$

and the testing problem is $H_0: \tilde{K}\tilde{\beta} = 0$ versus $H_1: \tilde{K}\tilde{\beta} \neq 0$. Applying the same argument to $\tilde{K}'$ as we did to Z,

$$\tilde{K}' = \Delta\begin{pmatrix} I_r \\ 0 \end{pmatrix}U_1$$

for some $\Delta \in \mathcal{O}_k$ and some $r \times r$ matrix U_1 in G_U^+. Let

$$\Gamma_1 = \begin{pmatrix} \Delta' & 0 \\ 0 & I_{n-k} \end{pmatrix} \in \mathcal{O}_n$$

and set $Y = \Gamma_1 X$, $B = \Delta'\tilde{\beta}$. Since

$$\Gamma_1 Z_0 \tilde{\beta} = \begin{pmatrix} \Delta' & 0 \\ 0 & I_{n-k} \end{pmatrix}\begin{pmatrix} \tilde{\beta} \\ 0 \end{pmatrix} = \begin{pmatrix} B \\ 0 \end{pmatrix} = Z_0 B,$$

it follows that

$$\mathcal{L}(Y) = N(Z_0 B, I_n \otimes \Sigma)$$

and the testing problem is $H_0: K_0 B = 0$ versus $H_1: K_0 B \neq 0$. Thus after two transformations, the original problem has been transformed into a problem with $Z = Z_0$ and $K = K_0$. Since $K_0 = (I_r\ 0)$, the null hypothesis is that the first r rows of B are zero. Partition B into

$$B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}; \qquad B_1 : r \times p, \qquad B_2 : (k - r) \times p$$

and partition Y into

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}; \qquad Y_1 : r \times p, \qquad Y_2 : (k - r) \times p, \qquad Y_3 : (n - k) \times p.$$

Because $\text{Cov}(Y) = I_n \otimes \Sigma$, Y_1, Y_2, and Y_3 are mutually independent and it is clear that

$$\mathcal{L}(Y_1) = N(B_1, I_r \otimes \Sigma)$$

$$\mathcal{L}(Y_2) = N(B_2, I_{(k-r)} \otimes \Sigma)$$

$$\mathcal{L}(Y_3) = N(0, I_{(n-k)} \otimes \Sigma).$$

Also, the testing problem is $H_0 : B_1 = 0$ versus $H_1 : B_1 \neq 0$. It is this form of the problem that is called the canonical MANOVA testing problem. The only reason for transforming from the original problem to the canonical problem is that certain expressions become simpler and the invariance of the MANOVA testing problem is more easily articulated when the problem is expressed in canonical form.

We now proceed to analyze the canonical MANOVA testing problem. To simplify some later formulas, the notation is changed a bit. Let Y_1, Y_2, and Y_3 be independent random matrices that satisfy

$$\mathcal{L}(Y_1) = N(B_1, I_r \otimes \Sigma)$$

$$\mathcal{L}(Y_2) = N(B_2, I_s \otimes \Sigma)$$

$$\mathcal{L}(Y_3) = N(0, I_m \otimes \Sigma)$$

so B_1 is $r \times p$ and B_2 is $s \times p$. As usual Σ is a $p \times p$ unknown positive definite matrix. To insure the existence of a maximum likelihood estimator for Σ, it is assumed that $m \geqslant p$ and the sample space for Y_3 is taken to be the set of all $m \times p$ real matrices of rank p. A set of Lebesgue measure zero has been deleted from the natural sample space $\mathcal{L}_{p,m}$ of Y_3. The testing problem is

$$H_0 : B_1 = 0 \quad \text{versus} \quad H_1 : B_1 \neq 0.$$

Setting $n = r + s + m$ and

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \in \mathcal{L}_{p,n},$$

$\mathcal{L}(Y) = N(\mu, I_n \otimes \Sigma)$ where μ is an element of the subspace

$$M = \left\{ \mu | \mu = \begin{pmatrix} B_1 \\ B_2 \\ 0 \end{pmatrix}; B_1 \in \mathcal{L}_{p,r}, B_2 \in \mathcal{L}_{p,s} \right\}.$$

In this notation, the null hypothesis is that $\mu \in M_0 \subset M$ where

$$M_0 = \left\{ \mu | \mu = \begin{pmatrix} 0 \\ B_2 \\ 0 \end{pmatrix}; B_2 \in \mathcal{L}_{p,s} \right\}.$$

Since M and M_0 are both invariant under $I_n \otimes \Sigma$ for all $\Sigma > 0$, the testing problem under consideration is of the type described in general terms earlier, and

$$\gamma = \{I_n \otimes \Sigma | \Sigma > 0\}.$$

When the model for Y is (M, γ), the density of Y is

$$\begin{aligned} p(Y|B_1, B_2, \Sigma) = {} & \left(\sqrt{2\pi}\right)^{-n} |\Sigma|^{-n/2} \\ & \times \exp\left[-\tfrac{1}{2}\operatorname{tr}(Y_1 - B_1)\Sigma^{-1}(Y_1 - B_1)' \right. \\ & \left. -\tfrac{1}{2}\operatorname{tr}(Y_2 - B_2)\Sigma^{-1}(Y_2 - B_2)' - \tfrac{1}{2}\operatorname{tr} Y_3\Sigma^{-1}Y_3'\right]. \end{aligned}$$

In this case, the maximum likelihood estimators of B_1, B_2, and Σ are easily seen to be

$$\hat{B}_1 = Y_1, \qquad \hat{B}_2 = Y_2, \qquad \hat{\Sigma} = \frac{1}{n} Y_3'Y_3.$$

When the model for Y is (M_0, γ), the density of Y is $p(Y|0, B_2, \Sigma)$ and the maximum likelihood estimators of B_2 and Σ are

$$\tilde{B}_2 = Y_2, \qquad \tilde{\Sigma} = \frac{1}{n}(Y_3'Y_3 + Y_1'Y_1).$$

Therefore, the likelihood ratio test rejects for small values of

$$\Lambda(Y) = \frac{p\left(Y|0, \tilde{B}_2, \tilde{\Sigma}\right)}{p\left(Y|\hat{B}_1, \hat{B}_2, \hat{\Sigma}\right)} = \frac{|\tilde{\Sigma}|^{-n/2}}{|\hat{\Sigma}|^{-n/2}} = \frac{|Y_3'Y_3|^{n/2}}{|Y_3'Y_3 + Y_1'Y_1|^{n/2}}.$$

Summarizing this, we have the following result.

Proposition 9.1. For the canonical MANOVA testing problem, the likelihood ratio test rejects the null hypothesis for small values of the statistic

$$U = \frac{|Y_3'Y_3|}{|Y_3'Y_3 + Y_1'Y_1|}.$$

Under H_0, $\mathcal{L}(U) = U(m, r, p)$ where the distribution $U(m, r, p)$ is given in Proposition 8.15.

Proof. The first assertion is clear. Under H_0, $\mathcal{L}(Y_1) = N(0, I_r \otimes \Sigma)$ and $\mathcal{L}(Y_3) = N(0, I_m \otimes \Sigma)$. Therefore, $\mathcal{L}(Y_1'Y_1) = W(\Sigma, p, r)$ and $\mathcal{L}(Y_3'Y_3) = W(\Sigma, p, m)$. Since $m \geqslant p$, Proposition 8.18 implies the result. □

Before attempting to interpret the likelihood ratio test, it is useful to see first what implications can be obtained from invariance considerations in the canonical MANOVA problem. In the notation of the previous paragraph, (M, γ) is the parameter set for the model for Y and under the null hypothesis, (M_0, γ) is the parameter set for Y. In order that the testing problem be invariant under a group of transformations, both of the parameter sets (M, γ) and (M_0, γ) must be invariant. With this in mind, consider the group G defined by

$$G = \{g | g = (\Gamma_1, \Gamma_2, \Gamma_3, \xi, A); \Gamma_1 \in \mathcal{O}_r, \Gamma_2 \in \mathcal{O}_s,$$

$$\Gamma_3 \in \mathcal{O}_m, \xi \in \mathcal{L}_{p,s}, A \in Gl_p\}$$

where the group action on the sample space is given by

$$(\Gamma_1, \Gamma_2, \Gamma_3, \xi, A)\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \Gamma_1 Y_1 A' \\ \Gamma_2 Y_2 A' + \xi \\ \Gamma_3 Y_3 A' \end{pmatrix}.$$

The group composition, defined so that the above action is a left action on the sample space, is

$$(\Gamma_1, \Gamma_2, \Gamma_3, \xi, A)(\Delta_1, \Delta_2, \Delta_3, \eta, C) = (\Gamma_1\Delta_1, \Gamma_2\Delta_2, \Gamma_3\Delta_3, \Gamma_2\eta A' + \xi, AC).$$

Further, the induced group action on the parameter set (M, γ) is

$$(\Gamma_1, \Gamma_2, \Gamma_3, \xi, A)(B_1, B_2, \Sigma) = (\Gamma_1 B_1 A', \Gamma_2 B_2 A' + \xi, A\Sigma A'),$$

where the point

$$\begin{pmatrix} B_1 \\ B_2 \\ 0 \end{pmatrix} \in M, \qquad (I_n \otimes \Sigma) \in \gamma$$

has been represented simply by (B_1, B_2, Σ). Now it is routine to check that when Y has a normal distribution with $\mathcal{E}Y \in M(\mathcal{E}Y \in M_0)$ and $\text{Cov}(Y) \in \gamma$, then $\mathcal{E}gY \in M(\mathcal{E}gY \in M_0)$ and $\text{Cov}(gY) \in \gamma$, for $g \in G$. Thus the hypothesis testing problem is G-invariant and the likelihood ratio test is a G-invariant function of Y. To describe the invariant tests, a maximal invariant under the action of G on the sample space needs to be computed. The following result provides one form of a maximal invariant.

Proposition 9.2. Let $t = \min\{r, p\}$, and define $h(Y_1, Y_2, Y_3)$ to be the t-dimensional vector $(\lambda_1, \ldots, \lambda_t)'$ where $\lambda_1 \geqslant \cdots \geqslant \lambda_t$ are the t largest eigenvalues of $Y_1'Y_1(Y_3'Y_3)^{-1}$. Then h is a maximal invariant under the action of G on the sample space of Y.

Proof. Note that $Y_1'Y_1(Y_3'Y_3)^{-1}$ has at most t nonzero eigenvalues, and these t eigenvalues are nonnegative. First, consider the case when $r \leqslant p$ so $t = r$. By Proposition 1.39, the nonzero eigenvalues of $Y_1'Y_1(Y_3'Y_3)^{-1}$ are the same as the nonzero eigenvalues of $Y_1(Y_3'Y_3)^{-1}Y_1'$, and these eigenvalues are obviously invariant under the action of g on Y. To show that h is maximal invariant for this case, a reduction argument similar to that in Example 7.4 is used. Given

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix},$$

we claim that there exists a $g_0 \in G$ such that

$$g_0(Y) = \begin{pmatrix} (D0) \\ 0 \\ \begin{pmatrix} I_p \\ 0 \end{pmatrix} \end{pmatrix} \in \begin{pmatrix} \mathcal{L}_{p,r} \\ \mathcal{L}_{p,s} \\ \mathcal{L}_{p,m} \end{pmatrix}$$

where D is $r \times r$ and diagonal and has diagonal elements $\sqrt{\lambda}_1, \ldots, \sqrt{\lambda}_r$. For $g = (\Gamma_1, \Gamma_2, \Gamma_3, \xi, A)$,

$$gY = \begin{pmatrix} \Gamma_1 Y_1 A' \\ \Gamma_2 Y_2 A' + \xi \\ \Gamma_3 Y_3 A' \end{pmatrix}.$$

By Proposition 5.2, $Y_3 = \Psi_3 U_3$ where $\Psi_3 \in \mathcal{F}_{p,m}$ and $U_3 \in G_U^+$ is $p \times p$. Choose $A' = U_3^{-1}\Delta$ where $\Delta \in \mathcal{O}_p$ is, as yet, unspecified. Then

$$\Gamma_1 Y_1 A' = \Gamma_1 Y_1 U_3^{-1}\Delta$$

and, by the singular value decomposition theorem for matrices, there exists a $\Gamma_1 \in \mathcal{O}_r$ and a $\Delta \in \mathcal{O}_p$ such that

$$\Gamma_1 Y_1 U_3^{-1} \Delta = (D0)$$

where D is an $r \times r$ diagonal matrix whose diagonal elements are the square roots of the eigenvalues of $Y_1(U_3U_3')^{-1}Y_1' = Y_1(Y_3Y_3')^{-1}Y_1'$. With this choice for $\Delta \in \mathcal{O}_r$ it is clear that $Y_3A' = Y_3U_3^{-1}\Delta \in \mathcal{F}_{p,m}$ so there exists a $\Gamma_3 \in \mathcal{O}_m$ such that

$$\Gamma_3 Y_3 U_3^{-1} \Delta = \begin{pmatrix} I_p \\ 0 \end{pmatrix}.$$

Choosing $\Gamma_2 = I_s$, $\xi = -Y_2A'$, and setting

$$g_0 = \left(\Gamma_1, I_s, \Gamma_3, -Y_2U_3^{-1}\Delta, (U_3^{-1}\Delta)'\right),$$

g_0Y has the representation claimed. To show h is maximal invariant, suppose $h(Y_1, Y_2, Y_3) = h(Z_1, Z_2, Z_3)$. Let D be the $r \times r$ diagonal matrix, the squares of whose diagonal elements are the eigenvalues of $Y_1(Y_3Y_3')^{-1}Y_1'$ and $Z_1(Z_3Z_3')^{-1}Z_1'$. Then there exist g_0 and $g_1 \in G$ such that

$$g_0Y = \begin{pmatrix} (D0) \\ 0 \\ \begin{pmatrix} I_p \\ 0 \end{pmatrix} \end{pmatrix} = g_1Z$$

so $Y = g_0^{-1}g_1Z$. Thus Y and Z are in the same orbit and h is a maximal invariant function.

When $r > p$, basically the same argument establishes that h is a maximal invariant. To show h is invariant, if $g = (\Gamma_1, \Gamma_2, \Gamma_3, \xi, A)$, then the matrix $Y_1'Y_1(Y_3'Y_3)^{-1}$ gets transformed into $AY_1'Y_1(Y_3'Y_3)^{-1}A^{-1}$ when Y is transformed to gY. By Proposition 1.39, the eigenvalues of $AY_1'Y_1(Y_3'Y_3)^{-1}A^{-1}$ are the same as the eigenvalues of $Y_1'Y_1(Y_3'Y_3)^{-1}$, so h is invariant. To show h is maximal invariant, first show that, for each Y, there exists a $g_0 \in G$ such that

$$g_0Y = \begin{pmatrix} \begin{pmatrix} D \\ 0 \end{pmatrix} \\ 0 \\ \begin{pmatrix} I_p \\ 0 \end{pmatrix} \end{pmatrix} \in \begin{pmatrix} \mathcal{L}_{p,r} \\ \mathcal{L}_{p,s} \\ \mathcal{L}_{p,m} \end{pmatrix}$$

where D is the $p \times p$ diagonal matrix of square roots of eigenvalues

$(Y_1'Y_1)(Y_3'Y_3)^{-1}$. The argument for this is similar to that given previously and is left to the reader. Now, by mimicking the proof for the case $r \leqslant p$, it follows that h is maximal invariant. □

Proposition 9.3. The distribution of the maximal invariant $h(Y_1, Y_2, Y_3)$ depends on the parameters (B_1, B_2, Σ) only through the vector of the t largest eigenvalues of $B_1'B_1\Sigma^{-1}$.

Proof. Since h is a G-invariant function, the distribution of h depends on (B_1, B_2, Σ) only through a maximal invariant parameter under the induced action of G on the parameter space. This action, given earlier, is

$$(\Gamma_1, \Gamma_2, \Gamma_3, \xi, A)(B_1, B_2, \Sigma) = (\Gamma_1 B_1 A', \Gamma_2 B_2 A' + \xi, A\Sigma A').$$

However, an argument similar to that used to prove Proposition 9.2 shows that the vector of the t largest eigenvalues of $B_1'B_1\Sigma^{-1}$ is maximal invariant in the parameter space. □

An alternative form of the maximal invariant is sometimes useful.

Proposition 9.4. Let $t = \min\{r, p\}$ and define $h_1(Y_1, Y_2, Y_3)$ to be the t-dimensional vector $(\theta_1, \dots, \theta_t)'$ where $\theta_1 \leqslant \cdots \leqslant \theta_t$ are the t smallest eigenvalues of $Y_3'Y_3(Y_3'Y_3 + Y_1'Y_1)^{-1}$. Then $\theta_i = 1/(1 + \lambda_i)$, $i = 1, \dots, t$, where λ_i's are defined in Proposition 9.2. Further, $h_1(Y_1, Y_2, Y_3)$ is a maximal invariant.

Proof. For $\lambda \in [0, \infty)$, let $\theta = 1/(1 + \lambda)$. If λ satisfies the equation

$$\left|Y_1'Y_1(Y_3'Y_3)^{-1} - \lambda I_p\right| = 0,$$

then a bit of algebra shows that θ satisfies the equation

$$\left|Y_3'Y_3(Y_3'Y_3 + Y_1'Y_1)^{-1} - \theta I_p\right| = 0,$$

and conversely. Thus $\theta_i = 1/(1 + \lambda_i)$, $i = 1, \dots, t$, are the t smallest eigenvalues of $Y_3'Y_3(Y_3'Y_3 + Y_1'Y_1)^{-1}$. Since $h_1(Y_1, Y_2, Y_3)$ is a one-to-one function of $h(Y_1, Y_2, Y_3)$, it is clear that $h_1(Y_1, Y_2, Y_3)$ is a maximal invariant. □

Since every G-invariant test is a function of a maximal invariant, the problem of choosing a reasonable invariant test boils down to studying tests based on a maximal invariant. When $t \equiv \min\{p, r\} = 1$, the following result shows that there is only one sensible choice for an invariant test.

Proposition 9.5. If $t = 1$ in the MANOVA problem, then the test that rejects for large values of λ_1 is uniformly most powerful within the class of G-invariant tests. Further, this test is equivalent to the likelihood ratio test.

Proof. First consider the case when $p = 1$. Then $Y_1'Y_1(Y_3'Y_3)^{-1}$ is a nonnegative scalar and

$$\lambda_1 = \frac{Y_1'Y_1}{Y_3'Y_3}.$$

Also, $\mathcal{L}(Y_1) = N(B_1, \sigma^2 I_r)$ and $\mathcal{L}(Y_3) = N(0, \sigma^2 I_m)$ where Σ has been set equal to σ^2 to conform to classical notation when $p = 1$. By Proposition 8.14,

$$\mathcal{L}(\lambda_1) = F(r, m; \delta)$$

where $\delta = B_1'B_1/\sigma^2$ and the null hypothesis is that $\delta = 0$. Since the noncentral F distribution has a monotone likelihood ratio, it follows that the test that rejects for large values of λ_1 is uniformly most powerful for testing $\delta = 0$ versus $\delta > 0$. As every invariant test is a function of λ_1, the case for $p = 1$ follows.

Now, suppose $r = 1$. Then the only nonzero eigenvalue of $Y_1'Y_1(Y_3'Y_3)^{-1}$ is $Y_1(Y_3'Y_3)^{-1}Y_1'$ by Proposition 1.39. Thus

$$\lambda_1 = Y_1(Y_3'Y_3)^{-1}Y_1'$$

and, by Proposition 8.14,

$$\mathcal{L}(\lambda_1) = F(p, m - p + 1; \delta)$$

where $\delta = B_1\Sigma^{-1}B_1' \geqslant 0$. The problem is to test $\delta = 0$ versus $\delta > 0$. Again, the noncentral F distribution has a monotone likelihood ratio and the test that rejects for large values of λ_1 is uniformly most powerful among tests based on λ_1.

The likelihood ratio test rejects H_0 for small values of

$$\Lambda = \frac{|Y_3'Y_3|}{|Y_3'Y_3 + Y_1'Y_1|} = \frac{1}{|I_p + Y_1'Y_1(Y_3'Y_3)^{-1}|}.$$

If $p = 1$, then $\Lambda = (1 + \lambda_1)^{-1}$ and rejecting for small values of Λ is equivalent to rejecting for large values of λ_1. When $r = 1$, then

$$|I_p + Y_1'Y_1(Y_3'Y_3)^{-1}| = 1 + Y_1(Y_3'Y_3)^{-1}Y_1' = 1 + \lambda_1$$

so again $\Lambda = (1 + \lambda_1)^{-1}$. □

When $t > 1$, the situation is not so simple. In terms of the eigenvalues $\lambda_1, \ldots, \lambda_t$, the likelihood ratio criterion rejects H_0 for small values of

$$\Lambda = \frac{|Y_3'Y_3|}{|Y_3'Y_3 + Y_1'Y_1|} = \frac{1}{|I_p + Y_1'Y_1(Y_3'Y_3)^{-1}|} = \prod_{i=1}^{t} \frac{1}{1 + \lambda_i}.$$

However, there are no compelling reasons to believe that other tests based on $\lambda_1, \ldots, \lambda_t$ would not be reasonable. Before discussing possible alternatives to the likelihood ratio test, it is helpful to write the maximal invariant statistic in terms of the original variables that led to the canonical MANOVA problem. In the original MANOVA problem, we had an observation vector $X \in \mathcal{L}_{p,n}$ such that

$$\mathcal{L}(X) = N(Z\beta, I_n \otimes \Sigma)$$

and the problem was to test $K\beta = 0$. We know that

$$\hat{\beta} = (Z'Z)^{-1}ZX$$

and

$$\hat{\Sigma} = \frac{1}{n} X'Q_z X \equiv \frac{1}{n} S$$

are the maximum likelihood estimators of β and Σ.

Proposition 9.6. Let $t = \min\{p, r\}$. Suppose the original MANOVA problem is reduced to a canonical MANOVA problem. Then a maximal invariant in the canonical problem expressed in terms of the original variables is the vector $(\lambda_1, \ldots, \lambda_t)'$, $\lambda_1 \geqslant \cdots \geqslant \lambda_t$, of the t largest eigenvalues of

$$V \equiv \left[(K\hat{\beta})'\left(K(Z'Z)^{-1}K'\right)^{-1}K\hat{\beta}\right]S^{-1}.$$

Proof. The transformations that reduced the original problem to canonical form led to the three matrices Y_1, Y_2, and Y_3 where Y_1 is $r \times p$, Y_2 is $(k - r) \times p$, and Y_3 is $(n - k) \times p$. Expressing Y_1 and Y_3 in terms of X, Z, and K, it is not too difficult (but certainly tedious) to show that

$$Y_1'Y_1(Y_3'Y_3)^{-1} = V.$$

By Proposition 9.2, the vector $(\lambda_1, \ldots, \lambda_t)'$ of the t largest eigenvalues of

$Y_1'Y_1(Y_3'Y_3)^{-1}$ is a maximal invariant. Thus the vector of the t largest eigenvalues of V is maximal invariant. □

In terms of X, Z, and K, the likelihood ratio test rejects the null hypothesis if

$$\Lambda = \frac{|S|}{|\hat{\beta}'K'\left(K(Z'Z)^{-1}K'\right)^{-1}K\hat{\beta} + S|}$$

is too small. Also, the distribution of Λ under H_0 is given in Proposition 9.1 as $U(n-k, r, p)$. The distribution of Λ when $K\beta \neq 0$ is quite complicated when $t > 1$ except in the case when β has rank one. In this case, the distribution of Λ is given in Proposition 8.16.

We now turn to the question of possible alternatives to the likelihood ratio test. For notational convenience, the canonical form of the MANOVA problem is treated. However, the reader can express statistics in terms of the original variables by applying Proposition 9.6. Since our interest is in invariant tests, consider Y_1 and Y_3, which are independent, and satisfy

$$\mathcal{L}(Y_1) = N(B_1, I_n \otimes \Sigma)$$

$$\mathcal{L}(Y_3) = N(0, I_m \otimes \Sigma).$$

The random vector Y_2 need not be considered as invariant tests do not involve Y_2. Intuitively, the null hypothesis $H_0: B_1 = 0$ should be rejected, on the basis of an invariant test, if the nonzero eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_t$ of $Y_1'Y_1(Y_3'Y_3)^{-1}$ are too large in some sense. Since $\mathcal{L}(Y_1) = N(B_1, I_r \otimes \Sigma)$,

$$\mathcal{E}Y_1'Y_1 = B_1'B_1 + r\Sigma.$$

Also, it is not difficult to verify that (see the problems at the end of this chapter)

$$\mathcal{E}(Y_3'Y_3)^{-1} = \frac{1}{m-p-1}\Sigma^{-1}$$

when $m - p - 1 > 0$. Since Y_1 and Y_3 are independent,

$$\mathcal{E}Y_1'Y_1(Y_3'Y_3)^{-1} = \frac{r}{m-p-1}I_p + \frac{1}{m-p-1}B_1'B_1\Sigma^{-1}.$$

Therefore, the further B_1 is away from zero, the larger we expect the eigenvalues of $B_1'B_1\Sigma^{-1}$ to be, and hence the larger we expect the eigenvalues of $Y_1'Y_1(Y_3'Y_3)^{-1}$ to be. In particular,

$$\mathcal{E}\,\mathrm{tr}\, Y_1'Y_1(Y_3'Y_3)^{-1} = \frac{rp}{m-p-1} + \frac{1}{m-p-1}\,\mathrm{tr}\, B_1B_1'\Sigma^{-1}$$

and $\mathrm{tr}\, B_1'B_1\Sigma^{-1}$ is just the sum of the eigenvalues of $B_1'B_1\Sigma^{-1}$. The test that rejects for large values of the statistic

$$\sum_1^t \lambda_i = \mathrm{tr}\, Y_1'Y_1(Y_3'Y_3)^{-1}$$

is called the Lawley–Hotelling trace test and is one possible alternative to the likelihood ratio test. Also, the test that rejects for large values of

$$\sum_1^t \frac{\lambda_i}{1+\lambda_i} = \mathrm{tr}\, Y_1'Y_1(Y_3'Y_3 + Y_1'Y_1)^{-1}$$

was introduced by Pillai as a competitor to the likelihood ratio test. A third competitor is based on the following considerations. The null hypothesis $H_0: B_1 = 0$ is equivalent to the intersection over $u \in R^r$, $\|u\| = 1$, of the null hypotheses $H_u: u'B_1 = 0$. Combining Propositions 9.5 and 9.6, it follows that the test that accepts H_u iff

$$u'Y_1(Y_3'Y_3)^{-1}Y_1'u \leqslant c$$

is a uniformly most powerful test within the class of tests that are invariant under the group of transformations preserving H_u. Here, c is a constant. Under H_u,

$$\mathcal{L}\left(u'Y_1(Y_3'Y_3)^{-1}Y_1'u\right) = F_{p,\,m-p+1}$$

so it seems reasonable to require that c not depend on u. Since H_0 is equivalent to $\cap\{H_u \mid \|u\| = 1, u \in R^r\}$, H_0 should be accepted iff all the H_u are accepted—that is, H_0 should be accepted iff

$$\sup_{\|u\|=1} u'Y_1(Y_3'Y_3)^{-1}Y_1'u \leqslant c.$$

However, this supremum is just the largest eigenvalue of $Y_1(Y_3'Y_3)^{-1}Y_1'$, which is λ_1. Thus the proposed test is to accept H_0 iff $\lambda_1 \leqslant c$ or equivalently, to reject H_0 for large values of λ_1. This test is called Roy's maximum root test.

Unfortunately, there is very little known about the comparative behavior of the tests described above. A few numerical studies have been done for small values of r, m, and p but no single test stands out as dominating the others over a substantial portion of the set of alternatives. Since very accurate approximations are available for the null distribution of the likelihood ratio test, this test is easier to apply than the above competitors. Further, there is an interesting decomposition of the test statistic

$$\Lambda = \frac{|Y_3'Y_3|}{|Y_3'Y_3 + Y_1'Y_1|},$$

which has some applications in practice. Let $S = Y_3'Y_3$ so $\mathcal{L}(S) = W(\Sigma, p, m)$ and let $X_1', \dots, X_r'$ denote the rows of Y_1. Under $H_0: B_1 = 0$, $X_1, \dots, X_r$ are independent and $\mathcal{L}(X_i) = N(0, \Sigma)$. Further,

$$\Lambda = \frac{|S|}{\left|S + \sum_1^r X_i X_i'\right|} = \prod_{i=1}^r \Lambda_i$$

where

$$\Lambda_1 = \frac{|S|}{|S + X_1 X_1'|}$$

and

$$\Lambda_i = \frac{\left|S + \sum_1^{i-1} X_i X_i'\right|}{\left|S + \sum_1^i X_i X_i'\right|}, \qquad i = 2, \dots, r.$$

Proposition 8.15 gives the distribution of Λ_i under H_0 and shows that $\Lambda_1, \dots, \Lambda_r$ are independent under H_0. Let $\beta_1', \dots, \beta_r'$ denote the rows of B_1 and consider the r testing problems given by the null hypotheses

$$H_i: \{(\beta_1, \dots, \beta_r) | \beta_1 = \beta_2 = \cdots = \beta_i = 0\}$$

and the alternatives

$$\overline{H}_i: \{(\beta_1, \dots, \beta_r) | \beta_1 = \beta_2 = \cdots = \beta_{i-1} = 0\}$$

for $i = 1, \dots, r$. Obviously, $H_0 = \cap_1^r H_i$ and the likelihood ratio test for testing H_i against $\overline{H}_i$ rejects H_i iff Λ_i is too small. Thus the likelihood ratio test for H_0 can be thought of as one possible way of combining the r independent test statistics into an overall test of $\cap_1^r H_i$.

9.2. MANOVA PROBLEMS WITH BLOCK DIAGONAL COVARIANCE STRUCTURE

The parameter set of the MANOVA model considered in the previous section consisted of a subspace $M = \{\mu | \mu = ZB, B \in \mathcal{L}_{p,k}\} \subseteq \mathcal{L}_{p,n}$ and a set of covariance matrices

$$\gamma = \{I_n \otimes \Sigma | \Sigma \in \mathcal{S}_p^+\}.$$

It was assumed that the matrix Σ was completely unknown. In this section, we consider estimation and testing problems when certain things are known about Σ. For example, if $\Sigma = \sigma^2 I_p$ with σ^2 unknown and positive, then we have the linear model discussed in Section 3.1. In this case, the linear model with parameter set $\{M, \gamma\}$ is just a univariate linear model in the sense that $I_n \otimes \Sigma = \sigma^2 I_n \otimes I_p$ and $I_n \otimes I_p$ is the identity linear transformation on the vector space $\mathcal{L}_{p,n}$. This model is just the linear model of Section 9.1 when $p = 1$ and np plays the role of n. Of course, when $\Sigma = \sigma^2 I_p$, the subspace M need not have the structure above in order for Proposition 4.6 to hold.

In what follows, we consider another assumption concerning Σ and treat certain estimation and testing problems. For the models treated, it is shown that these models are actually "products" of the MANOVA models discussed in Section 9.1.

Suppose $Y \in \mathcal{L}_{p,n}$ is a random vector with $\mathcal{E}Y \in M$ where

$$M = \{\mu | \mu = ZB, B \in \mathcal{L}_{p,k}\}$$

and Z is a known $n \times k$ matrix of rank k. Write $p = p_1 + p_2$, $p_i \geqslant 1$, for $i = 1, 2$. The covariance of Y is assumed to be an element of

$$\gamma_0 = \left\{ I_n \otimes \Sigma | \Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}, \Sigma_{ii} \in \mathcal{S}_{p_i}^+, i = 1, 2 \right\}.$$

Thus the rows of Y, say $Y_1', \ldots, Y_n'$, are uncorrelated. Further, if Y_i is partitioned into $X_i \in R^{p_1}$ and $W_i \in R^{p_2}$, $Y_i' = (X_i', W_i')$, then X_i and W_i are also uncorrelated, since

$$\operatorname{Cov}(Y_i) = \operatorname{Cov}\left\{ \begin{pmatrix} X_i \\ W_i \end{pmatrix} \right\} = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}.$$

Thus the interpretation of the assumed structure of γ_0 is that the rows of Y are uncorrelated and within each row, the first p_1 coordinates are uncorrelated with the last p_2 coordinates. This suggests that we decompose Y into $X \in \mathcal{L}_{p_1, n}$ and $W \in \mathcal{L}_{p_2, n}$ where

$$Y = (X, W) \in \mathcal{L}_{p, n}.$$

Obviously, the rows of $X(W)$ are $X_1', \ldots, X_n'(W_1', \ldots, W_n')$. Also, partition $B \in \mathcal{L}_{p,k}$ into $B_1 \in \mathcal{L}_{p_1,k}$ and $B_2 \in \mathcal{L}_{p_2,k}$. It is clear that

$$\mathcal{E}X \in M_1 \equiv \{\mu_1 | \mu_1 = ZB_1, B_1 \in \mathcal{L}_{p_1,k}\}$$

and

$$\mathcal{E}W \in M_2 \equiv \{\mu_2 | \mu_2 = ZB_2, B_2 \in \mathcal{L}_{p_2,k}\}.$$

Further,

$$\operatorname{Cov}(X) \in \gamma_1 \equiv \{I_n \otimes \Sigma_{11} | \Sigma_{11} \in \mathcal{S}_{p_1}^+\}$$

and

$$\operatorname{Cov}(W) \in \gamma_2 \equiv \{I_n \otimes \Sigma_{22} | \Sigma_{22} \in \mathcal{S}_{p_2}^+\}.$$

Since X and W are uncorrelated, if Y has a normal distribution, then X and W are independent and normal and we have a MANOVA model of Section 9.1 for both X and W (with parameter sets (M_1, γ_1) and (M_2, γ_2)). In summary, when Y has a normal distribution, Y can be partitioned into X and W, which are independent. Therefore, the density of Y is

$$f(Y|\mu, \Sigma) = f_1(X|\mu_1, \Sigma_{11}) f_2(W|\mu_2, \Sigma_{22})$$

where f, f_1, and f_2 are normal densities on the appropriate spaces. Since we have MANOVA models for both X and W, the maximum likelihood estimators of μ_1, μ_2, Σ_{11}, and Σ_{22} follow from the result of the first section. For testing the null hypothesis $H_0: KB = 0$, $K: r \times k$ of rank r, a similar decomposition occurs. As $B = (B_1 B_2)$, $H_0: KB = 0$ is equivalent to the two null hypotheses $H_0^1: KB_1 = 0$ and $H_0^2: KB_2 = 0$.

Proposition 9.7. Assume that $n - k \geqslant \max\{p_1, p_2\}$. For testing $H_0: KB = 0$, the likelihood ratio test rejects for small values of $\Lambda = \Lambda_1 \Lambda_2$ where

$$\Lambda_1 = \frac{|X'Q_z X|}{\left|X'Q_z X + (K\hat{B}_1)'\left(K(Z'Z)^{-1}K'\right)^{-1} K\hat{B}_1\right|}$$

and

$$\Lambda_2 = \frac{|W'Q_zW|}{|W'Q_zW + (K\hat{B}_2)'(K(Z'Z)^{-1}K')^{-1}K\hat{B}_2|}.$$

Here, $Q_z = I - P_z$ where $P_z = Z(Z'Z)^{-1}Z'$ and

$$\hat{B}_1 = (Z'Z)^{-1}Z'X, \qquad \hat{B}_2 = (Z'Z)^{-1}Z'W.$$

Proof. We need to calculate

$$\Psi(Y) \equiv \frac{\sup\limits_{(\mu, \Sigma) \in H_0} f(Y|\mu, \Sigma)}{\sup\limits_{(\mu, \Sigma) \in \mathfrak{M}} f(Y|\mu, \Sigma)}$$

where $\mathfrak{M}$ is the set of (μ, Σ) such that $\mu \in M$ and $I_n \otimes \Sigma \in \gamma_0$. As noted earlier,

$$f(Y|\mu, \Sigma) = f_1(X|\mu_1, \Sigma_{11})f_2(W|\mu_2, \Sigma_{22}).$$

Also, $(\mu, \Sigma) \in H_0$ iff $(\mu_1, \Sigma_{11}) \in H_0^1$ and $(\mu_2, \Sigma_{22}) \in H_0^2$. Further, $(\mu, \Sigma) \in \mathfrak{M}$ iff $(\mu_i, \Sigma_{ii}) \in \mathfrak{M}_i$ where $\mathfrak{M}_i$ is the set of (μ_i, Σ_{ii}) such that $\mu_i \in M_i$ and $I_n \otimes \Sigma_{ii} \in \gamma_i$ for $i = 1, 2$. From these remarks, it follows that

$$\Psi(Y) = \Psi_1(X)\Psi_2(W)$$

where

$$\Psi_1(X) = \frac{\sup\limits_{(\mu_1, \Sigma_{11}) \in H_0^1} f_1(X|\mu_1, \Sigma_{11})}{\sup\limits_{(\mu_1, \Sigma_{11}) \in \mathfrak{M}_1} f_1(X|\mu_1, \Sigma_{11})}$$

and

$$\Psi_2(W) = \frac{\sup\limits_{(\mu_2, \Sigma_{22}) \in H_0^2} f_2(W|\mu_2, \Sigma_{22})}{\sup\limits_{(\mu_2, \Sigma_{22}) \in \mathfrak{M}_2} f_2(W|\mu_2, \Sigma_{22})}.$$

However, $\Psi_1(X)$ is simply the likelihood ratio statistic for testing H_0^1 in the

MANOVA model for X. The results of Propositions 9.6 and 9.1 show that $\Psi_1(X) = (\Lambda_1)^{n/2}$. Similarly, $\Psi_2(W) = (\Lambda_2)^{n/2}$. Thus $\Psi(Y) = (\Lambda_1\Lambda_2)^{n/2}$ so the test that rejects for small values of $\Lambda = \Lambda_1\Lambda_2$ is equivalent to the likelihood ratio test. □

Since X and W are independent, the statistics Λ_1 and Λ_2 are independent. The distribution of Λ_i when H_0^i is true is $U(n - p_i, r, p_i)$ for $i = 1, 2$. Therefore, when H_0 is true, $\Lambda_1\Lambda_2$ is distributed as a product of independent beta random variables and the results in Anderson (1958) provide an approximation to the null distribution of $\Lambda_1\Lambda_2$.

We now turn to a discussion of the invariance aspects of testing $H_0: KB = 0$ on the basis of the observation vector Y. The argument used to reduce the MANOVA model of Section 9.1 to canonical form is valid here, and this leads to a group of transformations G_1, which preserve the testing problem H_0^1 for the MANOVA model for X. Similarly, there is a group G_2 that preserves the testing problem H_0^2 for the MANOVA model for W. Since $Y = (X, W)$, we can define the product group $G_1 \times G_2$ acting on Y by

$$(g_1, g_2)Y \equiv (g_1X, g_2W)$$

and the testing problem H_0 is clearly invariant under this group action. A maximal invariant is derived as follows. Let $t_i = \min\{r, p_i\}$ for $i = 1, 2$, and in the notation of Proposition 9.7, let

$$V_1 = \left[(K\hat{B}_1)'\left(K(Z'Z)^{-1}K'\right)^{-1}K\hat{B}_1\right](X'Q_zX)^{-1}$$

and

$$V_2 = \left[(K\hat{B}_2)'\left(K(Z'Z)^{-1}K'\right)^{-1}K\hat{B}_2\right](W'Q_zW)^{-1}.$$

Let $\eta_1 \geqslant \cdots \geqslant \eta_{t_1}$ be the t_1 largest eigenvalues of V_1 and $\theta_1 \geqslant \cdots \geqslant \theta_{t_2}$ be the t_2 largest eigenvalues of V_2.

Proposition 9.8. A maximal invariant under the action of $G_1 \times G_2$ on Y is the $(t_1 + t_2)$-dimensional vector $(\eta_1, \ldots, \eta_{t_1}; \theta_1, \ldots, \theta_{t_2}) = h(Y) \equiv (h_1(X); h_2(W))$. Here, $h_1(X) = (\eta_1, \ldots, \eta_{t_1})$ and $h_2(W) = (\theta_1, \ldots, \theta_{t_2})$.

Proof. By Proposition 9.6, $h_1(X)(h_2(W))$ is maximal invariant under the action of $G_1(G_2)$ on $X(W)$. Thus h is G-invariant. If $h(Y_1) = h(Y_2)$ where $Y_1 = (X_1, W_1)$ and $Y_2 = (X_2, W_2)$, then $h_1(X_1) = h_1(X_2)$ and $h_2(W_1) = h_2(W_2)$. Thus there exists $g_1 \in G_1(g_2 \in G_2)$ such that $g_1X_1 = X_2(g_2W_1 =$

W_2). Therefore,

$$(g_1, g_2)Y_1 = (g_1X_1, g_2W_1) = (X_2, W_2) = Y_2$$

so h is maximal invariant. □

As a function of $h(Y)$, the likelihood ratio test rejects H_0 if

$$\Lambda = \Lambda_1\Lambda_2 = \prod_1^{t_1}\left(\frac{1}{1+\eta_i}\right)\prod_1^{t_2}\left(\frac{1}{1+\theta_i}\right)$$

is too small. Since $t_1 + t_2 > 1$, the maximal invariant $h(Y)$ is always of dimension greater than one. Thus the situation described in Proposition 9.5 cannot arise in the present context. In no case will there exist a uniformly most powerful invariant test of $H_0: KB = 0$ even if K has rank 1. This completes our discussion of the present linear model.

It should be clear by now that the results described above can be easily extended to the case when Σ has the form

$$\Sigma = \begin{pmatrix} \Sigma_{11} & & & \\ & \Sigma_{22} & & \\ & & \ddots & \\ & & & \Sigma_{ss} \end{pmatrix}$$

where the off-diagonal blocks of Σ are zero. Here $\Sigma \in \mathcal{S}_p^+$ and $\Sigma_{ii} \in \mathcal{S}_{p_i}^+$, $\Sigma_1^s p_i = p$. In this case, the set of covariances for $Y \in \mathcal{L}_{p,n}$ is the set γ_0, which consists of all $I_n \otimes \Sigma$ where Σ has the above form and each Σ_{ii} is unknown. The mean space for Y is M as before. For this model, Y can be decomposed into s independent pieces and we have a MANOVA model in $\mathcal{L}_{p_i,n}$ for each piece. Also, the matrix $B(\mathcal{E}Y = ZB)$ decomposes into $B_1,\ldots,$ B_s, $B_i \in \mathcal{L}_{p_i,k}$ and a null hypothesis $H_0: KB = 0$ is equivalent to the intersection of the s null hypotheses $H_0^i: KB_i = 0$, $i = 1,\ldots, s$. The likelihood ratio test of H_0 is now based on a product of s independent statistics, say $\Lambda \equiv \prod_1^s \Lambda_i$, where $\mathcal{L}(\Lambda_i) = U(n - p_i, r, p_i)$ and thus Λ is distributed as a product of independent beta random variables when H_0 is true. Further, invariance considerations lead to an s-fold product group that preserves the testing problem and a maximal invariant is of dimension $t_1 + \cdots + t_s$ where $t_i = \min\{r, p_i\}$, $i = 1,\ldots, s$. The details of all this, which are mainly notational, are left to the reader.

In this section, it has been shown that the linear model with a block diagonal covariance matrix can be decomposed into independent component models, each of which is a MANOVA model of the type treated in Section 9.1. This decomposition technique also appears in the next two sections in which we treat linear models with different types of covariance structure.

9.3. INTRACLASS COVARIANCE STRUCTURE

In some instances, it is natural to assume that the covariance matrix of a random vector possesses certain symmetry properties that are suggested by the sampling situation. For example, if n measurements are taken under the same experimental conditions, it may be reasonable to suppose that the order in which the observations are taken is immaterial. In other words, if $X_1, \ldots, X_p$ denote the observations and $X' = (X_1, \ldots, X_p)$ is the observation vector, then X and any permutation of X have the same distribution. Symbolically, this means that $\mathcal{L}(X) = \mathcal{L}(gX)$ where g is a permutation matrix. If $\Sigma \equiv \operatorname{Cov}(X)$ exists, this implies that $\Sigma = g\Sigma g'$ for $g \in \mathcal{P}_p$ where $\mathcal{P}_p$ denotes the group of $p \times p$ permutation matrices. Our first task is to characterize those covariance matrices that are invariant under $\mathcal{P}_p$—that is, those covariance matrices that satisfy $\Sigma = g\Sigma g'$ for all $g \in \mathcal{P}_p$. Let $e \in R^p$ be the vector of ones and set $P_e = (1/p)ee'$ so P_e is the orthogonal projection onto span$\{e\}$. Also, let $Q_e = I_p - P_e$.

Proposition 9.9. Let Σ be a positive definite $p \times p$ matrix. The following are equivalent:

(i) $\Sigma = g\Sigma g'$ for $g \in \mathcal{P}_p$.
(ii) $\Sigma = \alpha P_e + \beta Q_e$ for $\alpha > 0$ and $\beta > 0$.
(iii) $\Sigma = \sigma^2 A(\rho)$ where $\sigma^2 > 0$, $-1/(p-1) < \rho < 1$, and $A(\rho)$ is a $p \times p$ matrix with elements $a_{ii} = 1$, $i = 1, \ldots, p$, and $a_{ij}(\rho) = \rho$ for $i \neq j$.

Proof. Since

$$A(\rho) = (1-\rho)I_p + \rho ee' = (1-\rho)I_p + p\rho P_e$$

$$= (1-\rho)Q_e + (1 + (p-1)\rho)P_{e'}$$

the equivalence of (ii) and (iii) follows by taking $\alpha = \sigma^2(1 + (p-1)\rho)$ and $\beta = \sigma^2(1-\rho)$. Since $ge = e$ for $g \in \mathcal{P}_p$, $gP_e = P_e g$. Thus if (ii) holds, then

$$g\Sigma g' = \alpha g P_e g' + \beta g Q_e g' = \alpha P_e + \beta Q_e = \Sigma$$

so (i) holds. To show (i) implies (ii), let $X \in R^p$ be a random vector with $\text{Cov}(X) = \Sigma$. Then (i) implies that $\text{Cov}(X) = \text{Cov}(gX)$ for $g \in \mathcal{P}_p$. Therefore,

$$\text{var}(X_i) = \text{var}(X_j), \qquad i, j = 1, \dots, p$$

and

$$\text{cov}(X_i, X_j) = \text{cov}(X_{i'}, X_{j'}); \qquad i \neq j, i' \neq j'.$$

Let $\gamma = \text{var}(X_1)$ and $\delta = \text{cov}(X_1, X_2)$. Then

$$\Sigma = \delta ee' + (\gamma - \delta) I_p = p\delta P_e + (\gamma - \delta)(P_e + Q_e)$$

$$= (\gamma + (p-1)\delta) P_e + (\gamma - \delta) Q_e = \alpha P_e + \beta Q_e$$

where $\alpha = \gamma + (p-1)\delta$ and $\beta = \gamma - \delta$. The positivity of α and β follows from the assumption that Σ is positive definite. □

A covariance matrix Σ that satisfies one of the conditions of Proposition 9.9 is called an *intraclass covariance matrix* and is said to have intraclass covariance structure. Now that intraclass covariance matrices have been described, suppose that $X \in \mathcal{L}_{p,n}$ has a normal distribution with $\mu \equiv \mathcal{E}X \in M$ and $\text{Cov}(X) \in \gamma$ where M is a linear subspace of $\mathcal{L}_{p,n}$ and

$$\gamma = \{I_n \otimes \Sigma | \Sigma \in \mathcal{S}_p^+, \Sigma = \alpha P_e + \beta Q_e, \alpha > 0, \beta > 0\}.$$

The covariance structure assumed for X means that the rows of X are independent and each row of X has the same intraclass covariance structure. In terms of invariance, if $\Gamma \otimes g \in \mathcal{O}_n \otimes \mathcal{P}_p$, and $I_n \otimes \Sigma \in \gamma$, it is clear that

$$\text{Cov}((\Gamma \otimes g) X) = \text{Cov}(X)$$

since

$$(\Gamma \otimes g)(I_n \otimes \Sigma)(\Gamma \otimes g)' = (\Gamma I_n \Gamma') \otimes (g \Sigma g') = I_n \otimes \Sigma.$$

Conversely, if T is a positive definite linear transformation on $\mathcal{L}_{p,n}$ that satisfies

$$(\Gamma \otimes g) T (\Gamma \otimes g)' = T \quad \text{for } \Gamma \otimes g \in \mathcal{O}_n \otimes \mathcal{P}_p,$$

it is not difficult to show that $T \in \gamma$. The proof of this is left to the reader.

Since the identity linear transformation is an element of γ, in order that the least-squares estimator of $\mu \in M$ be the maximum likelihood estimator, it is sufficient that

$$(I_n \otimes \Sigma)M \subseteq M \quad \text{for } I_n \otimes \Sigma \in \gamma.$$

Our next task is to describe a class of linear subspaces M that satisfy the above condition.

Proposition 9.10. Let C be an $r \times p$ real matrix of rank r with rows $c_1', \dots, c_r'$. If $u_1, \dots, u_r$ is any basis for $N \equiv \text{span}\{c_1, \dots, c_r\}$ and U is an $r \times p$ matrix with rows $u_1', \dots, u_r'$, then there exists an $r \times r$ nonsingular matrix A such that $AU = C$.

Proof. Since $u_1, \dots, u_r$ is a basis for N,

$$c_i = \sum_{k=1}^{r} a_{ik} u_k, \qquad i = 1, \dots, r$$

for some real numbers a_{ik}. Setting $A = \{a_{ik}\}$, $AU = C$ follows. As the basis $\{u_1, \dots, u_r\}$ is mapped onto the basis $\{c_1, \dots, c_r\}$ by the linear transformation defined by A, the matrix A is nonsingular. □

Given positive integers n and p, let k and r be positive integers that satisfy $k < n$ and $r \leqslant p$. Define a subspace $M \subseteq \mathcal{L}_{p,n}$ by

$$M = \{\mu | \mu = Z_1 B Z_2;\ B \in \mathcal{L}_{r,k}\}$$

where Z_1 is $n \times k$ of rank k, Z_2 is $r \times p$ of rank r, and assume that $e \in R^p$ is an element of the subspace spanned by rows of Z_2, say $e \in N = \text{span}\{z_1, \dots, z_r\}$ and the rows of Z_2 are $z_1', \dots, z_r'$. At this point, it is convenient to relabel things a bit. Let $u_1 = e/\sqrt{p}$, $u_2, \dots, u_r$, be an orthonormal basis for N and let $U: r \times p$ have rows $u_1', \dots, u_r'$. By Proposition 9.10, $Z_2 = AU$ for some $r \times r$ nonsingular matrix A so

$$M = \{\mu | \mu = Z_1 B U, B \in \mathcal{L}_{r,k}\}.$$

Summarizing, $X \in \mathcal{L}_{p,n}$ is assumed to have a normal distribution with $\mathcal{E}X \in M$ and $\text{Cov}(X) \in \gamma$ where M and γ are given above. To decompose this model for X into the product of two simple univariate linear models, let $\Gamma \in \mathcal{O}_p$ have $u_1', \dots, u_r'$ as its first r rows. With $Y = (I_n \otimes \Gamma)X$,

$$\mathcal{E}Y = \mathcal{E}X\Gamma' = Z_1 B U \Gamma'$$

and

$$\begin{aligned}\operatorname{Cov}(Y) &= (I_n \otimes \Gamma)\operatorname{Cov}(X)(I_n \otimes \Gamma)' \\ &= (I_n \otimes \Gamma)(I_n \otimes (\alpha P_e + \beta Q_e))(I_n \otimes \Gamma)' \\ &= I_n \otimes (\alpha \Gamma P_e \Gamma' + \beta \Gamma Q_e \Gamma').\end{aligned}$$

However,

$$U\Gamma' = (I_r 0) \in \mathcal{L}_{p,r},$$

$$\Gamma P_e \Gamma' = \varepsilon_1 \varepsilon_1'$$

and

$$\Gamma Q_e \Gamma' = I_p - \varepsilon_1 \varepsilon_1'$$

where $\varepsilon_1' = (1, 0, \ldots, 0)$. Therefore, the matrix $D \equiv \alpha\Gamma P_e \Gamma' + \beta \Gamma Q_e \Gamma'$ is diagonal with diagonal elements $d_1, \ldots, d_p$ given by $d_1 = \alpha$ and $d_2 = \cdots = d_p = \beta$. Let $Y_1, \ldots, Y_p$ be the columns of Y and let $b_1, \ldots, b_r$ be the columns of B. Then it is clear that $Y_1, \ldots, Y_p$ are independent,

$$\mathcal{L}(Y_1) = N(Z_1 b_1, \alpha I_n)$$

$$\mathcal{L}(Y_i) = N(Z_1 b_i, \beta I_n), \qquad i = 2, \ldots, r,$$

and

$$\mathcal{L}(Y_i) = N(0, \beta I_n), \qquad i = r+1, \ldots, p.$$

To piece things back together, set $m = n(p-1)$ and let $V \in R^m$ be given by $V' = (Y_2', Y_3', \ldots, Y_p')$. Then

$$\mathcal{L}(V) = N(\tilde{Z}\delta, \beta I_m)$$

where $\delta \in R^{(r-1)p}$, $\delta' = (b_2', \ldots, b_r')$, and

$$\tilde{Z} = \begin{pmatrix} Z_1 & & & 0 \\ & Z_1 & & \\ & & \ddots & \\ 0 & & & Z_1 \\ \hline & & 0 & \end{pmatrix} : m \times ((r-1)p).$$

Thus X has been decomposed into the two independent random vectors Y_1 and V and the linear models for Y_1 and V are given by the parameter sets (M_1, γ_1) and (M_2, γ_2) where

$$M_1 = \{\mu_1 | \mu_1 = Z_1 b_1;\ b_1 \in R^k\}$$

$$\gamma_1 = \{\alpha I_n | \alpha > 0\}$$

$$M_2 = \{\mu_2 | \mu_2 = \tilde{Z}\delta,\ \delta \in R^{(r-1)p}\}$$

and

$$\gamma_2 = \{\beta I_m | \beta > 0\}.$$

Both of these linear models are univariate in the sense that γ_1 and γ_2 consist of a constant times an identity matrix.

It is obvious that the general theory developed in Section 9.1 for the MANOVA model applies directly to the above two linear models individually. In particular, the maximum likelihood estimators of b_1, α, δ, and β can simply be written down. Also, linear hypotheses about b_1 or δ can be tested separately, and uniformly most powerful invariant tests will exist for such testing problems when the two linear models are treated separately. However, an interesting phenomenon occurs when we test a joint hypothesis about b_1 and δ. For example, suppose the null hypothesis H_0 is that $b_1 = 0$ and $\delta = 0$ and the alternative is that $b_1 \neq 0$ or $\delta \neq 0$. This null hypothesis is equivalent to the hypothesis that $B = 0$ in the original model for X. By simply writing down the densities of Y_1 and V and substituting in the maximum likelihood estimators of the parameters, the likelihood ratio test for H_0 rejects if

$$\Lambda \equiv \left(\frac{\|Y_1 - Z_1\hat{b}_1\|^2}{\|Y_1\|^2}\right)^{n/2}\left(\frac{\|V - \tilde{Z}\hat{\delta}\|^2}{\|V\|^2}\right)^{m/2}$$

is too small. Here, $\|\cdot\|$ denotes the standard norm on the coordinate Euclidean space under consideration. Let

$$W_1 = \frac{\|Y_1 - Z_1\hat{b}_1\|^2}{\|Y_1\|^2}$$

and

$$W_2 = \frac{\|V - \tilde{Z}\hat{\delta}\|^2}{\|V\|^2}$$

so W_1 and W_2 are independent and each has a beta distribution. When $p \geqslant 3$, then $m = n(p-1) > n$ and it follows that $\Lambda^{2/n} = W_1 W_2^{m/n}$ is not in general distributed as a product of independent beta random variables. This is in contrast to the situation treated in Section 9.2 of this chapter.

We end this section with a brief description of what might be called multivariate intraclass covariance matrices. If $X \in R^p$ and $\text{Cov}(X) = \Sigma$, then Σ is an intraclass covariance matrix iff $\text{Cov}(gX) = \text{Cov}(X)$ for all $g \in \mathcal{P}_p$. When the random vector X is replaced by the random matrix $Y: p \times q$, then the expression $gY = (g \otimes I_q)Y$ still makes sense for $g \in \mathcal{P}_p$, and it is natural to seek a characterization of $\text{Cov}(Y)$ when $\text{Cov}(Y) = \text{Cov}((g \otimes I_q)Y)$ for all $g \in \mathcal{P}_p$. For $g \in \mathcal{P}_p$, the linear transformation $g \otimes I_q$ just permutes the rows of Y and, to characterize $T = \text{Cov}(Y)$, we must describe how permutations of the rows of Y affect T. The condition that $\text{Cov}(Y) = \text{Cov}((g \otimes I_q)Y)$ is equivalent to the condition

$$T = (g \otimes I_q)T(g \otimes I_q)', \qquad g \in \mathcal{P}_p.$$

For A and B in $\mathbb{S}_q^+$, consider

$$T_0 \equiv P_e \otimes A + Q_e \otimes B.$$

Then T_0 is a self-adjoint and positive definite linear transformation on $\mathcal{L}_{q,p}$ to $\mathcal{L}_{q,p}$. It is readily verified that

$$T_0 = (g \otimes I_q)T_0(g \otimes I_q)', \qquad g \in \mathcal{P}_p.$$

That T_0 is a possible generalization of an intraclass covariance matrix is fairly clear—the positive scalars α and β of Proposition 9.9 have become the positive definite matrices A and B. The following result shows that if T is $(\mathcal{P}_p \otimes I_q)$-invariant—that is, if T satisfies $T = (g \otimes I_q)T(g \otimes I_q)'$—then T must be a T_0 for some positive definite A and B.

Proposition 9.11. If T is positive definite and $(\mathcal{P}_p \otimes I_q)$-invariant, then there exist $q \times q$ positive definite matrices A and B such that

$$T = P_e \otimes A + Q_e \otimes B.$$

Proof. The proof of this is left to the reader. □

Unfortunately, space limitations prevent a detailed description of linear models that have covariances of the form $I_n \otimes T$ where T is given in

Proposition 9.11. However, the analysis of these models proceeds along the lines of that given for intraclass covariance models and, as usual, these models can be decomposed into independent pieces, each of which is a MANOVA model.

9.4. SYMMETRY MODELS: AN EXAMPLE

The covariance structures studied thus far in this chapter are special cases of a class of covariance models called symmetry models. To describe these, let $(V, (\cdot,\cdot))$ be an inner product space and let G be a compact subgroup of $\mathcal{O}(V)$. Define the class of positive definite transformations γ_G by

$$\gamma_G = \{\Sigma | \Sigma \in \mathcal{L}(V, V), \Sigma > 0, g\Sigma g' = \Sigma \quad \text{for all } g \in G\}.$$

Thus γ_G is the set of positive definite covariances that are invariant under G in the sense that $\Sigma = g\Sigma g'$ for $g \in G$. To justify the term symmetry model for γ_G, first observe that the notion of "symmetry" is most often expressed in terms of a group acting on a set. Further, if X is a random vector in V with $\text{Cov}(X) = \Sigma$, then $\text{Cov}(gX) = g\Sigma g'$. Thus the condition that $\Sigma = g\Sigma g'$ is precisely the condition that X and gX have the same covariance—hence, the term symmetry model.

Most of the covariance sets considered in this book have been symmetry models for a particular choice of $(V, (\cdot,\cdot))$ and G. For example, if $G = \mathcal{O}(V)$, then

$$\gamma_G = \{\Sigma | \Sigma = \sigma^2 I, \sigma^2 > 0\},$$

as Proposition 2.13 shows. Hence $\mathcal{O}(V)$ generates the weakly spherical symmetry model. The result of Proposition 2.19 establishes that when $(V, (\cdot,\cdot)) = (\mathcal{L}_{p,n}, \langle \cdot,\cdot \rangle)$ and

$$G = \{g | g = \Gamma \otimes I_p, \Gamma \in \mathcal{O}_n\},$$

then

$$\gamma_G = \{\Sigma | \Sigma = I_n \otimes A, A \in \mathcal{S}_p^+\}.$$

Of course, this symmetry model has occurred throughout this book. Using techniques similar to that in Proposition 2.19, the covariance models considered in Section 9.2 are easily shown to be symmetry models for an appropriate group. Moreover, Propositions 9.9 and 9.11 describe sets of covariances (the intraclass covariances and their multivariate extensions) in exactly the manner in which the set γ_G was defined. Thus symmetry models are not unfamiliar objects.

Now, given a closed group $G \subseteq \mathcal{O}(V)$, how can we explicitly describe the model γ_G? Unfortunately, there is no one method or approach that is appropriate for all groups G. For example, the results of Proposition 2.19 and Proposition 9.9 were proved by quite different means. However, there is a general structure theory known for the models γ_G (see Andersson, 1975), but we do not discuss that here. The general theory tells us what γ_G should look like, but does not tell us how to derive the particular form of γ_G.

The remainder of this section is devoted to an example where the methods are a bit different from those encountered thus far. To motivate the considerations below, consider observations $X_1, \ldots, X_p$, which are taken at p equally spaced points on a circle and are numbered sequentially around the circle. For example, the observations might be temperatures at a fixed cross section on a cylindrical rod when a heat source is present at the center of the rod. Impurities in the rod and the interaction of adjacent measuring devices may make an exchangeability assumption concerning the joint distribution of $X_1, \ldots, X_p$ unreasonable. However, it may be quite reasonable to assume that the covariance between X_j and X_k depends only on how far apart X_j and X_k are on the circle—that is, $\text{cov}(X_j, X_{j+1})$ does not depend on j, $j = 1, \ldots, p$, where $X_{p+1} \equiv X_1$; $\text{cov}(X_j, X_{j+2})$ does not depend on j, $j = 1, \ldots, p$, where $X_{p+2} \equiv X_2$, and so on. Assuming that $\text{cov}(X_j, X_j)$ does not depend on j, these assumptions can be succinctly expressed as follows. Let $X \in R^p$ have coordinates $X_1, \ldots, X_p$ and let C be a $p \times p$ matrix with

$$c_{p1} = c_{j(j+1)} = 1, \qquad j = 1, \ldots, p-1$$

and the remaining elements of C zero. A bit of reflection will convince the reader that the conditions assumed on the covariances is equivalent to the condition that $\text{Cov}(CX) = \text{Cov}(X)$. The matrix C is called a cyclic permutation matrix since, if $x \in R^p$ has coordinates $x_1, \ldots, x_p$, then Cx has coordinates $x_2, x_3, \ldots, x_p, x_1$. In the case that $p = 5$, a direct calculation shows that

$$\Sigma = \text{Cov}(X) = \text{Cov}(CX) = C\Sigma C'$$

iff Σ has the form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_2 & \rho_1 \\ & 1 & \rho_1 & \rho_2 & \rho_2 \\ & & 1 & \rho_1 & \rho_2 \\ & & & 1 & \rho_1 \\ & & & & 1 \end{pmatrix}$$

where $\sigma^2 > 0$. The conditions on ρ_1 and ρ_2 so that Σ is positive definite are given later. Covariances that satisfy the condition $\Sigma = C\Sigma C'$ are called *cyclic covariances*. Some further motivation for the study of cyclic covariances can be found in Olkin and Press (1969).

To begin the formal treatment of cyclic covariances, first observe that $C^p = I_p$ so the group generated by C is

$$G_0 = \{I_p, C, C^2, \dots, C^{p-1}\}.$$

Since C generates G_0, it is clear that $C\Sigma C' = \Sigma$ iff $g\Sigma g' = \Sigma$ for all $g \in G_0$. In what follows, only the case of $p = 2q + 1$, $q \geqslant 1$, is treated. When p is even, slightly different expressions are obtained but the analyses are similar. Rather than characterize the covariance set γ_{G_0} directly, it is useful and instructive to first describe the set

$$\mathcal{Q}_{G_0} = \{B | BC = CB, B \in \mathcal{C}_p\}.$$

Recall that $\mathbb{C}^p$ is the complex vector space of p-dimensional coordinate complex vectors and $\mathcal{C}_p$ is the set of all $p \times p$ complex matrices. Consider the complex number $r \equiv \exp[2\pi i/p]$ and define complex column vectors $w_k \in \mathbb{C}^p$ with jth coordinate given by

$$w_{kj} = p^{-1/2}\exp\left[\frac{2\pi i(j-1)(k-1)}{p}\right]; \quad j = 1, \dots, p$$

for $k = 1, \dots, p$. A direct calculation shows that

$$w_k^* w_l = \delta_{kl}, \qquad k, l = 1, \dots, p$$

so $w_1, \dots, w_p$ is an orthonormal basis for $\mathbb{C}^p$. For future reference note that

$$w_1 = p^{-1/2}e, \qquad \bar{w}_k = w_{p-k+2}, \qquad k = 2, \dots, q+1$$

where $p = 2q + 1$, $q \geqslant 1$. Here, the bar over w_k denotes complex conjugate, and e is the vector of ones in $\mathbb{C}^p$. The basic relation

$$Cw_k = r^{k-1}w_k, \qquad k = 1, \dots, p$$

shows that

$$C = \sum_{k=1}^{p} r^{k-1} w_k w_k^*. \tag{9.1}$$

As usual, * denotes conjugate transpose. Obviously, $1, r, \dots, r^{p-1}$ are eigenvalues of C with corresponding eigenvectors $w_1, \dots, w_p$. Let $D_0 \in \mathcal{C}_p$ be diagonal with $d_{kk} = r^{k-1}$, $k = 1, \dots, p$ and let $U \in \mathcal{C}_p$ have columns $w_1, \dots, w_k$. The relation (9.1) can be written $C = UD_0U^*$. Since $UU^* = I_p$, U is a unitary complex matrix.

Proposition 9.12. The set $\mathcal{Q}_{G_0}$ consists of those $B \in \mathcal{C}_p$ that have the form

$$B = \sum_1^p \beta_k w_k w_k^*. \tag{9.2}$$

where $\beta_1, \dots, \beta_p$ are arbitrary complex numbers.

Proof. If B has the form (9.2), the identity $BC = CB$ follows easily from (9.1). Conversely, suppose $BC = CB$. Then

$$BUD_0U^* = UD_0U^*B$$

so

$$U^*BUD_0 = D_0U^*BU$$

since $U^*U = I_p$. In other words, U^*BU commutes with D_0. But D_0 is a diagonal matrix with distinct nonzero diagonal elements. This implies that U^*BU must be diagonal, say D, with diagonal elements $\beta_1, \dots, \beta_p$. Thus $U^*BU = D$ so $B = UDU^*$. Then B has the form (9.2). □

The next step is to identify those elements of $\mathcal{Q}_{G_0}$ that are real and symmetric. Consider $B \in \mathcal{Q}_{G_0}$ so

$$B = \sum_1^p \beta_k w_k w_k^*.$$

Now, suppose that B is real and symmetric. Then the eigenvalues of B, namely $\beta_1, \dots, \beta_p$, are real. Since $\beta_1, \dots, \beta_p$ are real and B is real, we have

$$\sum_1^p \beta_k w_k w_k^* = B = \overline{B} = \sum_1^p \beta_k \overline{w}_k \overline{w}_k^*.$$

The relationship $\overline{w}_k = w_{p-k+2}$, $k = 2, \dots, q+1$, implies that $\beta_k = \beta_{p-k+2}$,

$k = 2, \ldots, q + 1$, so

$$B = \beta_1 w_1 w_1^* + \sum_{k=2}^{q+1} \beta_k (w_k w_k^* + \bar{w}_k \bar{w}_k^*). \tag{9.3}$$

But any B given by (9.3) is real, symmetric, and commutes with C and conversely. We now show that (9.3) yields a spectral form for the real symmetric elements of $\mathcal{Q}_{G_0}$. Write $w_k = x_k + iy_k$ with $x_k, y_k \in R^p$, and define $u_k \in R^p$ by

$$u_k = x_k + y_k, \qquad k = 1, \ldots, p.$$

The two identities

$$w_k^* w_l = \delta_{kl}, \qquad k, l = 1, \ldots, p$$

$$\bar{w}_k = w_{p-k+2}, \qquad k = 2, \ldots, p$$

and the reality of w_1 yield the identities

$$u_k' u_l = \delta_{kl}, \qquad k, l = 1, \ldots, p$$

$$w_k w_k^* + \bar{w}_k \bar{w}_k^* = u_k u_k' + u_{p-k+2} u_{p-k+2}', \qquad k = 2, \ldots, p.$$

Thus $u_1, \ldots, u_p$ is an orthonormal basis for R^p. Hence any B of the form (9.3) can be written

$$B = \beta_1 u_1 u_1' + \sum_{2}^{q+1} \beta_k (u_k u_k' + u_{p-k+2} u_{p-k+2}')$$

and this is a spectral form for B. Such a B is positive definite iff $\beta_k > 0$ for $k = 1, \ldots, q + 1$. This discussion yields the following.

Proposition 9.13. The symmetry model γ_{G_0} consists of those covariances Σ that have the form

$$\Sigma = \alpha_1 u_1 u_1' + \sum_{k=2}^{q+1} \alpha_k (u_k u_k' + u_{p-k+2} u_{p-k+2}') \tag{9.4}$$

where $\alpha_k > 0$ for $k = 1, \ldots, q + 1$.

Let Γ have rows $u_1', \ldots, u_p'$. Then Γ is a $p \times p$ symmetric orthogonal matrix with elements

$$\gamma_{jk} = \cos\left[\frac{2\pi}{p}(j-1)(k-1)\right] + \sin\left[\frac{2\pi}{p}(j-1)(k-1)\right]$$

for $j, k = 1, \ldots, p$. Further, any Σ given by (9.4) will be diagonalized by Γ—that is, $\Gamma\Sigma\Gamma$ is diagonal, say D, with diagonal elements

$$d_k = \alpha_k, \qquad k = 1, \ldots, q+1; \qquad d_{p-k+2} = \alpha_k, \qquad k = 2, \ldots, q+1.$$

Since Γ simultaneously diagonalizes all the elements of γ_{G_0}, Γ can sometimes be used to simplify the analysis of certain models with covariances in γ_{G_0}. This is done in the following example.

As an application of the foregoing analysis, suppose $Y_1, \ldots, Y_n$ are independent with $Y_j \in R^p$, $p = 2q+1$, and $\mathcal{L}(Y_j) = N(\mu, \Sigma)$, $j = 1, \ldots, n$. It is assumed that Σ is a cyclic covariance so $\Sigma \in \gamma_{G_0}$. In what follows, we derive the likelihood ratio test for testing H_0, the null hypothesis that the coordinates of μ are all equal, versus H_1, the alternative that μ is completely unknown. As usual, form the matrix $Y: n \times p$ with rows Y_j', $j = 1, \ldots, n$, so

$$\mathcal{L}(Y) = N(e\mu', I_n \otimes \Sigma)$$

where $\mu \in R^p$ and $\Sigma \in \gamma_{G_0}$. Consider the new random vector $Z = (I_n \otimes \Gamma)Y$ where Γ is defined in the previous paragraph. Setting $\nu = \Gamma\mu$, we have

$$\mathcal{L}(Z) = N(e\nu', I_n \otimes D)$$

where $D = \Gamma\Sigma\Gamma$. As noted earlier, D is diagonal with diagonal elements

$$d_k = \alpha_k, \qquad k = 1, \ldots, q+1; \qquad d_{p-k+2} = \alpha_k, \qquad k = 2, \ldots, q+1.$$

Since Σ was assumed to be a completely unknown element of γ_{G_0}, the diagonal elements of D are unknown parameters subject only to the restriction that $\alpha_j > 0$, $j = 1, \ldots, q+1$. In terms of $\nu = \Gamma\mu$, the null hypothesis is $H_0: \nu_2 = \cdots = \nu_p = 0$. Because of the structure of D, it is convenient to relabel things once more. Denote the columns of Z by $Z_1, \ldots, Z_p$ and consider $W_1, \ldots, W_{q+1}$ defined by

$$W_1 = Z_1, \qquad W_j = (Z_j Z_{p-j+2}), \qquad j = 2, \ldots, q+1.$$

Thus $W_1 \in R^n$ and $W_j \in \mathcal{L}_{2,n}$ for $j = 2, \ldots, q+1$. Define vectors $\xi_j \in R^2$

by

$$\xi_j = \begin{pmatrix} \nu_j \\ \nu_{p-j+2} \end{pmatrix}, \qquad j = 2,\dots, q+1.$$

Now, it is clear that $W_1,\dots, W_{q+1}$ are independent and

$$\mathcal{L}(W_1) = N(\nu_1 e, \alpha_1 I_n), \qquad \mathcal{L}(W_j) = N(e\xi_j', \alpha_j I_n \otimes I_2),$$
$$j = 2,\dots, q+1.$$

Further, the null hypothesis is $H_0: \xi_j = 0, j = 2,\dots, q+1$, and the alternative is that $\xi_j \neq 0$ for some $j = 2,\dots, q+1$. With the model written in this form, a derivation of the likelihood ratio test is routine. Let $P_e = ee'/n$ and let $\|\cdot\|$ denote the usual norm on $\mathcal{L}_{2,n}$. Then the likelihood ratio test rejects H_0 for small values of

$$\Lambda \equiv \prod_{j=2}^{q+1} \frac{\|W_j - P_e W_j\|^2}{\|W_j\|^2}.$$

Of course, the likelihood ratio test of $H_0^{(j)}: \xi_j = 0$ versus $H_1^{(j)}: \xi_j \neq 0$ rejects for small values of

$$\Lambda_j = \frac{\|W_j - P_e W_j\|^2}{\|W_j\|^2}, \qquad j = 2,\dots, q+1.$$

The random variables $\Lambda_2,\dots, \Lambda_{q+1}$ are independent, and under $H_0^{(j)}$,

$$\mathcal{L}(\Lambda_j) = \mathcal{B}(n-1, 1).$$

Thus under H_0, Λ is distributed as a product of the independent beta random variables, each with parameters $n-1$ and 1.

We end this section with a discussion that leads to a new type of structured covariance—namely, the complex covariance structure that is discussed more fully in the next section. This covariance structure arises when we search for an analog of Proposition 9.11 for the cyclic group G_0. To keep things simple, assume $p = 3$ (i.e., $q = 1$) so G_0 has three elements and is a subgroup of the permutation group $\mathcal{P}_3$, which has six elements. Since $p = 3$, Propositions 9.9 and 9.13 yield that $\gamma_{\mathcal{P}_3} = \gamma_{G_0}$ and these symmetry models consist of those covariances of the form

$$\Sigma = \alpha P_e + \beta Q_e, \qquad \alpha > 0, \beta > 0$$

where $P_e = \frac{1}{3}ee'$ and $Q_e = I_3 - P_e$.

Now, consider the two groups $\mathcal{P}_3 \otimes I_r$ and $G_0 \otimes I_r$ acting on $\mathcal{L}_{r,3}$ by

$$(g \otimes I_r)(x) = gx, \qquad g \in \mathcal{P}_3, \qquad x \in \mathcal{L}_{r,3}.$$

Proposition 9.11 states that a covariance T on $\mathcal{L}_{r,3}$ is $\mathcal{P}_3 \otimes I_r$ invariant iff

$$T = P_e \otimes A + Q_e \otimes B \tag{9.5}$$

for some $r \times r$ positive definite A and B. We now claim that for $r > 1$, there are covariances on $\mathcal{L}_{r,3}$ that cannot be written in the form (9.5), but that are $G_0 \otimes I_r$ invariant.

To establish the above claim, recall that the vectors u_1, u_2, and u_3 defined earlier are an orthonormal basis for R^3 and

$$P_e = u_1u_1', \qquad Q_e = u_2u_2' + u_3u_3'.$$

These vectors were defined from the vectors $w_k = x_k + iy_k$, $k = 1,2,3$, by $u_k = x_k + y_k$, $k = 1,2,3$. Define the matrix J by

$$J = i[w_2w_2^* - w_3w_3^*].$$

By Proposition 9.12, J commutes with C. Consider vectors v_2 and v_3 given by

$$v_2 = \frac{1}{\sqrt{2}}(u_2 + u_3), \qquad v_3 = \frac{1}{\sqrt{2}}(u_2 - u_3)$$

so $\{v_2, v_3\}$ is an orthonormal basis for span $\{u_2, u_3\}$. Since $w_3 = \bar{w}_2$, we have $u_3 = x_2 - y_2$, which implies that $v_2 = \sqrt{2}\,x_2$ and $v_3 = \sqrt{2}\,y_2$. This readily implies that

$$J = v_2v_3' - v_3v_2'$$

so J is skew-symmetric, nonzero, and $Ju_1 = 0$. Now, consider the linear transformation T_0 on $\mathcal{L}_{r,3}$ to $\mathcal{L}_{r,3}$ given by

$$T_0 = P_e \otimes A + Q_e \otimes B + J \otimes F$$

where A and B are $r \times r$ and positive definite and F is skew-symmetric. It is now a routine matter to show that $(C \otimes I_r)T_0 = T_0(C \otimes I_r)$ since $CP_e = P_eC$, $CQ_e = Q_eC$, and $JC = CJ$. Thus T_0 commutes with each element of $G_0 \otimes I_r$ and T_0 is symmetric as both J and F are skew-symmetric. We now make two claims: first, that a nonzero F exists such that T_0 is positive definite, and second, that such a T_0 cannot be written in the form (9.5). Since $P_e \otimes A + Q_e \otimes B$ is positive definite, it follows that for all skew-symmetric F's that are sufficiently small,

$$P_e \otimes A + Q_e \otimes B + J \otimes F$$

is positive definite. Thus there exists a nonzero skew-symmetric F so that T_0 is positive definite. To establish the second claim, we have the following.

Proposition 9.14. Suppose that

$$P_e \otimes A_1 + Q_e \otimes B_1 + J \otimes F_1 = P_e \otimes A_2 + Q_e \otimes B_2 + J \otimes F_2$$

where A_j and $B_j, j = 1, 2$, are symmetric and $F_j, j = 1, 2$, is skew-symmetric. This implies that $A_1 = A_2$, $B_1 = B_2$, and $F_1 = F_2$.

Proof. Recall that $\{u_1, v_2, v_3\}$ is an orthonormal basis for R^3. The relation $Q_e u_1 = J u_1 = 0$ implies that for $x \in R^r$

$$(P_e \otimes A_j + Q_e \otimes B_j + J \otimes F_j)(u_1 \Box x) = u_1 \Box (A_j x)$$

for $j = 1, 2$ so $u_1 \Box (A_1 x) = u_1 \Box (A_2 x)$. With $\langle \cdot , \cdot \rangle$ denoting the natural inner product on $\mathfrak{L}_{r,3}$, we have

$$x' A_1 x = \langle u_1 \Box x, u_1 \Box (A_1 x) \rangle = \langle u_1 \Box x, u_1 \Box (A_2 x) \rangle = x' A_2 x$$

for all $x \in R^r$. The symmetry of A_1 and A_2 yield $A_1 = A_2$. Since $P_e v_2 = 0$, $Q_e v_2 = v_2$, and $J v_2 = -v_3$, we have

$$\begin{aligned}(P_e \otimes A_1 + Q_e \otimes B_1 + J \otimes F_1)(v_2 \Box x) &= v_2 \Box (B_1 x) - v_3 \Box (F_1 x) \\ &= v_2 \Box (B_2 x) - v_3 \Box (F_2 x)\end{aligned}$$

for all $x \in R^r$. Thus

$$x' B_1 x = \langle v_2 \Box x, v_2 \Box B_1 x - v_3 \Box (F_1 x) \rangle = x' B_2 x,$$

which implies that $B_1 = B_2$. Further,

$$-y' F_1 x = \langle v_3 \Box y, v_2 \Box (B_1 x) - v_3 \Box F_1 x \rangle = -y' F_2 x$$

for all $x, y \in R^r$. Thus $F_1 = F_2$. □

In summary, we have produced a covariance

$$T_0 = P_e \otimes A + Q_e \otimes B + J \otimes F$$

that is $(G_0 \otimes I_r)$-invariant but is not $(\mathcal{P}_3 \otimes I_r)$-invariant when $r > 1$. Of course, when $r = 1$, the two symmetry models $\gamma_{\mathcal{P}_3}$ and γ_{G_0} are the same. At this point, it is instructive to write out the matrix of T_0 in a special ordered basis for $\mathcal{L}_{r,3}$. Let $\varepsilon_1, \ldots, \varepsilon_n$ be the standard basis for R^r so

$$\{u_1 \square \varepsilon_1, \ldots, u_1 \square \varepsilon_r, v_2 \square \varepsilon_1, \ldots, v_2 \square \varepsilon_r, v_3 \square \varepsilon_1, \ldots, v_3 \square \varepsilon_r\}$$

is an orthonormal basis for $(\mathcal{L}_{r,3}, \langle \cdot, \cdot \rangle)$. A straightforward calculation shows that the matrix of T_0 in this basis is

$$[T_0] = \begin{pmatrix} A & 0 & 0 \\ 0 & B & F \\ 0 & -F & B \end{pmatrix}.$$

Since $[T_0]$ is symmetric and positive definite, the $2r \times 2r$ matrix

$$\Sigma = \begin{pmatrix} B & F \\ -F & B \end{pmatrix}$$

has these properties also. In other words, for each positive definite B, there is a nonzero skew-symmetric F (in fact, there exist infinitely many such skew-symmetric F's) such that Σ is positive definite. This special type of structured covariance has not arisen heretofore. However, it arises again in a very natural way in the next section where we discuss the complex normal distribution. It is not proved here, but the symmetry model of $G_0 \otimes I_r$ when $p = 3$ consists of all covariances of the form

$$T_0 = P_e \otimes A + Q_e \otimes B + J \otimes F$$

where A and B are positive definite and F is skew-symmetric.

9.5. COMPLEX COVARIANCE STRUCTURES

This section contains an introduction to complex covariance structures. One situation where this type of covariance structure arises was described at the end of the last section. To provide further motivation for the study of such models, we begin this section with a brief discussion of the complex normal distribution. The complex normal distribution arises in a variety of contexts and it seems appropriate to include the definition and the elementary properties of this distribution.

The notation introduced in Section 1.6 is used here. In particular, $\mathbb{C}$ is the field of complex numbers, $\mathbb{C}^n$ is the n-dimensional complex vector space of n-tuples (columns) of complex numbers, and $\mathcal{C}_n$ is the set of all $n \times n$ complex matrices. For $x, y \in \mathbb{C}^n$, the *inner product* between x and y is

$$(x, y) \equiv \sum_{j=1}^{n} \bar{x}_j y_j = x^* y.$$

where x^* denotes the conjugate transpose of x. Each $x \in \mathbb{C}^n$ has the unique representation $x = u + iv$ with $u, v \in R^n$. Of course, u is the *real part* of x, v is the *imaginary part* of x, and $i = \sqrt{-1}$ is the imaginary unit. This representation of x defines a real vector space isomorphism between $\mathbb{C}^n$ and R^{2n}. More precisely, for $x \in \mathbb{C}^n$, let

$$[x] = \begin{pmatrix} u \\ v \end{pmatrix} \in R^{2n}$$

where $x = u + iv$. Then $[ax + by] = a[x] + b[y]$ for $x, y \in \mathbb{C}^n$, $a, b \in R$, and obviously, $[\cdot]$ is a one-to-one onto function. In particular, this shows that $\mathbb{C}^n$ is a $2n$-dimensional real vector space. If $C \in \mathcal{C}_n$, then $C = A + iB$ where A and B are $n \times n$ real matrices. Thus for $x = u + iv \in \mathbb{C}^n$,

$$Cx = (A + iB)(u + iv) = Au - Bv + i(Av + Bu)$$

so

$$[Cx] = \begin{pmatrix} Au - Bv \\ Av + Bu \end{pmatrix} = \begin{pmatrix} A & -B \\ B & A \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

This suggests that we let $\{C\}$ be the $(2n) \times (2n)$ partitioned matrix given by

$$\{C\} = \begin{pmatrix} A & -B \\ B & A \end{pmatrix} : (2n) \times (2n).$$

With this definition, $[Cx] = \{C\}[x]$. The whole point is that the matrix $C \in \mathcal{C}_n$ applied to $x \in \mathbb{C}^n$ can be represented by applying the real matrix $\{C\}$ to the real vector $[x] \in R^{2n}$.

A complex matrix $C \in \mathcal{C}_n$ is called Hermitian if $C = C^*$. Writing $C = A + iB$ with A and B both real, C is Hermitian iff

$$A + iB = A' - iB',$$

which is equivalent to the two conditions

$$A = A', \qquad B = -B'.$$

Thus C is Hermitian iff $\{C\}$ is a symmetric real matrix. A Hermitian matrix C is positive definite if $x^*Cx > 0$ for all $x \in \mathbb{C}^n$, $x \neq 0$. However, for Hermitian C,

$$x^*Cx = [x]'\{C\}[x]$$

so C is positive definite iff $\{C\}$ is a positive definite real matrix. Of course, a Hermitian matrix C is positive semidefinite if $x^*Cx \geqslant 0$ for $x \in \mathbb{C}^n$ and C is positive semidefinite iff $\{C\}$ is positive semidefinite.

Now consider a random variable X with values in $\mathbb{C}$. Then $X = U + iV$ where U and V are real random variables. It is clear that the mean value of X must be defined by

$$\mathcal{E}X = \mathcal{E}U + i\mathcal{E}V$$

assuming $\mathcal{E}U$ and $\mathcal{E}V$ both exist. The variance of X, assuming it exists, is defined by

$$\operatorname{var}(X) = \mathcal{E}\left[(X - \mathcal{E}(X))\left(\overline{X - \mathcal{E}(X)}\right)\right]$$

where the bar denotes complex conjugate. Since X is a complex random variable, the complex conjugate is necessary if we want the variance of X to be a nonnegative real number. In terms of U and V,

$$\operatorname{var}(X) = \operatorname{var}(U) + \operatorname{var}(V).$$

It also follows that

$$\operatorname{var}(aX + b) = a\bar{a}\operatorname{var}(X)$$

for $a, b \in \mathbb{C}$. For two random variables X_1 and X_2 in $\mathbb{C}$, define the covariance between X_1 and X_2 (in that order) to be

$$\operatorname{cov}\{X_1, X_2\} \equiv \mathcal{E}\left[(X_1 - \mathcal{E}(X_1))\left(\overline{X_2 - \mathcal{E}(X_2)}\right)\right],$$

assuming the expectations in question exist. With this definition it is clear that $\operatorname{cov}\{X_1, X_1\} = \operatorname{var}(X_1)$, $\operatorname{cov}\{X_2, X_1\} = \overline{\operatorname{cov}\{X_1, X_2\}}$, and

$$\operatorname{cov}\{X_1, X_2 + X_3\} = \operatorname{cov}\{X_1, X_2\} + \operatorname{cov}\{X_1, X_3\}.$$

Further,

$$\operatorname{cov}\{a_1X_1 + b_1, a_2X_2 + b_2\} = a_1\bar{a}_2 \operatorname{cov}\{X_1, X_2\}$$

for $a_1, a_2, b_1, b_2 \in \mathbb{C}$.

We now turn to the problem of defining a normal distribution on $\mathbb{C}^n$. Basically, the procedure is the same as defining a normal distribution on R^n. Step one is to define a normal distribution with mean zero and variance one on $\mathbb{C}$, then define an arbitrary normal distribution on $\mathbb{C}$ by an affine transformation of the distribution defined in step one, and finally we say that $Z \in \mathbb{C}^n$ has a complex normal distribution if $(a, Z) = a^*Z$ has a normal distribution in $\mathbb{C}$ for each $a \in \mathbb{C}^n$. However it is not entirely obvious how to carry out step one. Consider $X \in \mathbb{C}$ and let $\mathbb{C}N(0,1)$ denote the distribution, yet to be defined, called the complex normal distribution with mean zero and variance one. Writing $X = U + iV$, we have

$$[X] = \begin{pmatrix} U \\ V \end{pmatrix} \in R^2$$

so the distribution of X on $\mathbb{C}$ determines the joint distribution of U and V on R^2 and, conversely, as $[\cdot]$ is one-to-one and onto. If $\mathcal{L}(X) = \mathbb{C}N(0,1)$, then the following two conditions should hold:

(i) $\mathcal{L}(aX) = \mathbb{C}N(0,1)$ for $a \in \mathbb{C}$ with $a\bar{a} = 1$.
(ii) $[X]$ has a bivariate normal distribution on R^2.

When $a\bar{a} = 1$ and X has mean zero and variance one, then aX has mean zero and variance one so condition (i) simply says that a scalar multiple of a complex normal is again complex normal. Condition (ii) is the requirement that a normal distribution on $\mathbb{C}$ be transformed into a normal distribution on R^2 under the real linear mapping $[\cdot]$. It can now be shown that conditions (i) and (ii) uniquely define the distribution of $[X]$ and hence provide us with the definition of a $\mathbb{C}N(0,1)$ distribution. Since $\mathcal{E}X = 0$, we have $\mathcal{E}[X] = 0$. Condition (i) implies that

$$\mathcal{L}([X]) = \mathcal{L}([aX]), \qquad a\bar{a} = 1.$$

However, writing $a = \alpha + i\beta$,

$$[aX] = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}[X] \equiv \Gamma[X]$$

where Γ is a 2×2 orthogonal matrix with determinant equal to one since $a\bar{a} = \alpha^2 + \beta^2 = 1$. Therefore,

$$\mathcal{L}([X]) = \mathcal{L}(\Gamma[X])$$

for all such orthogonal matrices. Using this together with the fact that $1 = \text{var}(X) = \text{var}(U) + \text{var}(V)$ implies that

$$\text{Cov}([X]) = \tfrac{1}{2}I_2.$$

Hence

$$\mathcal{L}([X]) = N_2(0, \tfrac{1}{2}I_2)$$

so the real and imaginary parts of X are independent normals with mean zero and variance one half.

Definition 9.1. A random variable $X = U + iV \in \mathbb{C}$ has a complex normal distribution with mean zero and variance one, written $\mathcal{L}(X) = \mathbb{C}N(0, 1)$, if

$$\mathcal{L}\left(\begin{pmatrix} U \\ V \end{pmatrix}\right) = N_2(0, \tfrac{1}{2}I_2).$$

With this definition, it is clear that when $\mathcal{L}(X) = \mathbb{C}N(0,1)$, the density of X on $\mathbb{C}$ with respect to two-dimensional Lebesgue measure on $\mathbb{C}$ is

$$p(x) = \frac{1}{\pi}\exp[-x\bar{x}], \qquad x \in \mathbb{C}.$$

Given $\mu \in \mathbb{C}$ and σ^2, $\sigma > 0$, a random variable $X_1 \in \mathbb{C}$ has a complex normal distribution with mean μ and variance σ^2 if $\mathcal{L}(X_1) = \mathcal{L}(\sigma X + \mu)$ where $\mathcal{L}(X) = \mathbb{C}N(0, 1)$. In such a case, we write $\mathcal{L}(X_1) = \mathbb{C}N(\mu, \sigma^2)$. It is clear that $X_1 = U_1 + iV_1$ has a $\mathbb{C}N(\mu, \sigma^2)$ distribution iff U_1 and V_1 are independent and normal with variance $\frac{1}{2}\sigma^2$ and means $\mathcal{E}U_1 = \mu_1$, $\mathcal{E}V_1 = \mu_2$, where $\mu = \mu_1 + i\mu_2$. As in the real case, a basic result is the following.

Proposition 9.15. Suppose $X_1, \ldots, X_m$ are independent random variables in $\mathbb{C}$ with $\mathcal{L}(X_j) = \mathbb{C}N(\mu_j, \sigma_j^2)$, $j = 1, \ldots, m$. Then

$$\mathcal{L}\left(\sum_{j=1}^{m}(a_jX_j + b_j)\right) = \mathbb{C}N\left(\sum_{j=1}^{m}(a_j\mu_j + b_j), \sum_{j=1}^{m} a_j\bar{a}_j\sigma_j^2\right)$$

for $a_j, b_j \in \mathbb{C}$, $j = 1, \ldots, m$.

Proof. This is proved by considering the real and imaginary parts of each X_j. The details are left to the reader. □

Suppose Y is a random vector in $\mathbb{C}^n$ with coordinates $Y_1, \dots, Y_n$ and that $\text{var}(Y_j) < +\infty$ for $j = 1, \dots, n$. Define a complex matrix H with elements h_{jk} given by

$$h_{jk} \equiv \text{cov}\{Y_j, Y_k\}.$$

Since $h_{jk} = \overline{h_{kj}}$, H is a Hermitian matrix. For $a, b \in \mathbb{C}^n$, a bit of algebra shows that

$$\text{cov}\{a^*Y, b^*Y\} = a^*Hb = (a, Hb).$$

As in the real case, H is the *covariance matrix* of Y and is denoted by $\text{Cov}(Y) \equiv H$. Since $a^*Ha = \text{var}(a^*Y) \geqslant 0$, H is positive semidefinite. If $H = \text{Cov}(Y)$ and $A \in \mathcal{C}_n$, it is readily verified that $\text{Cov}(AY) = AHA^*$.

We now turn to the definition of a complex normal distribution on the n-dimensional complex vector space $\mathbb{C}^n$.

Definition 9.2. A random vector $X \in \mathbb{C}^n$ has a complex normal distribution if, for each $a \in \mathbb{C}^n$, $(a, X) = a^*X$ has a complex normal distribution on $\mathbb{C}$.

If $X \in \mathbb{C}_n$ has a complex normal distribution and if $A \in \mathcal{C}_n$, it is clear that AX also has a complex normal distribution since $(a, AX) = (A^*a, X)$. In order to describe all the complex normal distributions on $\mathbb{C}^n$, we proceed as in the real case. Let $X_1, \dots, X_n$ be independent with $\mathcal{L}(X_j) = \mathbb{C}N(0, 1)$ on $\mathbb{C}$ and let $X \in \mathbb{C}^n$ have coordinates $X_1, \dots, X_n$. Since

$$a^*X = \sum_{j=1}^{n} \bar{a}_j X_j,$$

Proposition 9.15 shows that $\mathcal{L}(a^*X) = \mathbb{C}N(0, \Sigma \bar{a}_j a_j)$. Thus X has a complex normal distribution. Further, $\mathcal{E}X = 0$ and

$$\text{cov}\{X_j, X_k\} = \delta_{jk}$$

so $\text{Cov}(X) = I$. For $A \in \mathcal{C}_n$ and $\mu \in \mathbb{C}^n$, it follows that $Y = AX + \mu$ has a complex normal distribution and

$$\mathcal{E}Y = \mu, \qquad \text{Cov}(Y) = AA^* \equiv H.$$

Since every nonnegative definite Hermitian matrix can be written as AA^* for some $A \in \mathcal{C}_n$, we have produced a complex normal distribution on $\mathbb{C}^n$ with an arbitrary mean vector $\mu \in \mathbb{C}^n$ and an arbitrary nonnegative definite Hermitian covariance matrix. However, it still must be shown that, if X and $\tilde{X}$ in $\mathbb{C}^n$ are complex normal with $\mathcal{E}X = \mathcal{E}\tilde{X}$ and $\text{Cov}(X) = \text{Cov}(\tilde{X})$, then $\mathcal{L}(X) = \mathcal{L}(\tilde{X})$. The proof of this assertion is left to the reader. Given this fact, it makes sense to speak of the complex normal distribution on $\mathbb{C}^n$ with mean vector μ and covariance matrix H as this specifies a unique probability distribution. If X has such a distribution, the notation

$$\mathcal{L}(X) = \mathbb{C}N(\mu, H)$$

is used. Writing $X = U + iV$, it is useful to describe the joint distribution of U and V when $\mathcal{L}(X) = \mathbb{C}N(\mu, H)$ on $\mathbb{C}^n$. First, consider $\tilde{X} = \tilde{U} + i\tilde{V}$ where $\mathcal{L}(\tilde{X}) = \mathbb{C}N(\mu, I)$. Then the coordinates of $\tilde{X}$ are independent and it follows that

$$\mathcal{L}\begin{pmatrix} \tilde{U} \\ \tilde{V} \end{pmatrix} = N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \tfrac{1}{2}I_{2n}\right)$$

where $\mu = \mu_1 + i\mu_2$. For a general nonnegative definite Hermitian matrix H, write $H = AA^*$ with $A \in \mathcal{C}_n$. Then

$$\mathcal{L}(X) = \mathcal{L}(A\tilde{X} + \mu).$$

Since

$$[X] = \begin{pmatrix} U \\ V \end{pmatrix}$$

and

$$[AX + \mu] = \{A\}[\tilde{X}] + [\mu] = \begin{pmatrix} B & -C \\ C & B \end{pmatrix}\begin{pmatrix} \tilde{U} \\ \tilde{V} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

where $A = B + iC$, it follows that

$$\mathcal{L}([X]) = \mathcal{L}(\{A\}[\tilde{X}] + [\mu]).$$

But $H = \Sigma + iF$ where Σ is positive semidefinite, F is skew-symmetric, and the real matrix

$$\{H\} = \begin{pmatrix} \Sigma & -F \\ F & \Sigma \end{pmatrix}$$

is positive semidefinite. Since $H = AA^*$, $\{H\} = \{A\}\{A\}'$, and therefore,

$$\mathcal{L}([X]) = \mathcal{L}(\{A\}[\tilde{X}] + [\mu]) = N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \tfrac{1}{2}\{H\}\right)$$

$$= N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \tfrac{1}{2}\begin{pmatrix}\Sigma & -F\\ F & \Sigma\end{pmatrix}\right).$$

In summary, we have the following result.

Proposition 9.16. Suppose $\mathcal{L}(X) = \mathbb{C}N(\mu, H)$ and write $X = U + iV$, $\mu = \mu_1 + i\mu_2$, and $H = \Sigma + iF$. Then

$$\mathcal{L}\begin{pmatrix}U\\ V\end{pmatrix} = N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \tfrac{1}{2}\begin{pmatrix}\Sigma & -F\\ F & \Sigma\end{pmatrix}\right).$$

Conversely, with U and V jointly distributed as above, set $X = U + iV$, $\mu = \mu_1 + i\mu_2$, and $H = \Sigma + iF$. Then $\mathcal{L}(X) = \mathbb{C}N(\mu, H)$.

The above proposition establishes a one-to-one correspondence between n-dimensional complex normal distributions, say $\mathbb{C}N(\mu, H)$, and $2n$-dimensional real normal distributions with a special covariance structure given by

$$\tfrac{1}{2}\{H\} = \tfrac{1}{2}\begin{pmatrix}\Sigma & -F\\ F & \Sigma\end{pmatrix}$$

where $H = \Sigma + iF$. Given a sample of independent complex normal random vectors, the above correspondence provides us with the option of either analyzing the sample in the complex domain or representing everything in the real domain and performing the analysis there. Of course, the advantage of the real domain analysis is that we have developed a large body of theory that can be applied to this problem. However, this advantage is a bit illusory. As it turns out, many results for the complex normal distribution are clumsy to prove and difficult to understand when expressed in the real domain. From the point of view of understanding, the proper approach is simply to develop a theory of the complex normal distribution that parallels the development already given for the real normal distribution. Because of space limitations, this theory is not given in detail. Rather, we provide a brief list of results for the complex normal with the hope that the reader can see the parallel development. The proofs of many of these results are minor modifications of the corresponding real results.

Consider $X \in \mathbb{C}^p$ such that $\mathcal{L}(X) = \mathbb{C}N(\mu, H)$ where H is nonsingular. Then the density of X with respect to Lebesgue measure on $\mathbb{C}^p$ is

$$f(x) = \pi^{-p}(\det H)^{-1}\exp\left[-(x-\mu)^*H^{-1}(x-\mu)\right].$$

When $H = I$, then

$$\mathcal{L}(X^*X) = \tfrac{1}{2}\chi^2_{2p}(\mu^*\mu).$$

With this result and the spectral theorem for Hermitian matrices (see Halmos, 1958, Section 79), the distribution of quadratic forms, say X^*AX for a Hermitian, can be described in terms of linear combinations of independent noncentral chi-square random variables.

As in the real case, independence of jointly complex normal random vectors is equivalent to the absense of correlation. More precisely, if $\mathcal{L}(X) = \mathbb{C}N(\mu, H)$ and if $A: q \times p$ and $B: r \times p$ are complex matrices, then AX and BX are independent iff $AHB^* = 0$. In particular, if X is partitioned as

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \qquad X_j \in \mathbb{C}^{p_j}, j = 1,2$$

and H is partitioned similarly as

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$$

where H_{jk} is $p_j \times p_k$, then X_1 and X_2 are independent iff $H_{12} = 0$. When H_{22} is nonsingular, this implies that $X_1 - H_{12}H_{22}^{-1}X_2$ and X_2 are independent. This result yields the conditional distribution of X_1 given X_2, namely,

$$\mathcal{L}(X_1|X_2) = \mathbb{C}N\big(\mu_1 + H_{12}H_{22}^{-1}(X_2 - \mu_2), H_{11\cdot 2}\big)$$

where $H_{11\cdot 2} = H_{11} - H_{12}H_{22}^{-1}H_{21}$ and $\mu_j = \mathcal{E}X_j, j = 1,2.$

The complex Wishart distribution arises in a natural way, just as the real Wishart distribution did.

Definition 9.3. A $p \times p$ random Hermitian matrix S has a complex Wishart distribution with parameters H, p, and n if

$$\mathcal{L}(S) = \mathcal{L}\left(\sum_{j=1}^{n} X_j X_j^*\right)$$

where $X_1, \ldots, X_n \in \mathbb{C}^p$ are independent with

$$\mathcal{L}(X_j) = \mathbb{C}N(0, H).$$

In such a case, we write

$$\mathcal{L}(S) = \mathbb{C}W(H, p, n).$$

In this definition, p is the dimension, n is the degrees of freedom and H is a $p \times p$ nonnegative definite Hermitian matrix. It is clear that S is always nonnegative definite and, as in the real case, S is positive definite with probability one iff H is positive definite and $n \geqslant p$. When $p = 1$ and $H = 1$, it is clear that

$$\mathbb{C}W(1, 1, n) = \tfrac{1}{2}\chi^2_{2n}.$$

Further, complex analogues of Proposition 8.8, 8.9, and 8.13 show that if $\mathcal{L}(S) = W(H, p, n)$ with $n \geqslant p$ and H positive definite, and if $\mathcal{L}(X) = N(0, H)$ with X and S independent, then

$$\mathcal{L}(X^*S^{-1}X) = F_{2p, 2(n-p+1)}.$$

We now turn to a brief discussion of one special case of the complex MANOVA problem. Suppose $X_1, \ldots, X_n \in \mathbb{C}^p$ are independent with

$$\mathcal{L}(X_j) = \mathbb{C}N(\mu, H)$$

and assume that $H > 0$—that is, H is positive definite. The joint density of $X_1, \ldots, X_n$ with respect to $2np$-dimensional Lebesgue measure is

$$\begin{aligned} p(X|\mu, H) &= \prod_{j=1}^{n} \pi^{-p}|H|^{-1}\exp\left[-(X_j - \mu)^*H^{-1}(X_j - \mu)\right] \\ &= \pi^{-np}|H|^{-n}\exp\left[-\sum_{j=1}^{n}(X_j - \mu)^*H^{-1}(X_j - \mu)\right] \\ &= \pi^{-np}|H|^{-n}\exp\left[-n(\bar{X} - \mu)^*H^{-1}(\bar{X} - \mu)\right. \\ &\qquad \left. - \operatorname{tr}\left(\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})^*\right)H^{-1}\right] \end{aligned}$$

where $\bar{X} = n^{-1}\Sigma X_j$ and tr denote the trace. Here, X is the np-dimensional vector in $\mathbb{C}^{np}$ consisting of $X_1, X_2, \ldots, X_n$. Setting

$$S = \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})^*,$$

we have

$$p(X|\mu, H) = \pi^{-np}|H|^{-n}\exp[-n(\bar{X}-\mu)^*H^{-1}(\bar{X}-\mu) - \operatorname{tr} SH^{-1}].$$

It follows that $(\bar{X}, S)$ is a sufficient statistic for this parametric family and $\hat{\mu} \equiv \bar{X}$ is the maximum likelihood estimator of μ. Thus

$$p(X|\hat{\mu}, H) = \pi^{-np}|H|^{-n}\exp[-\operatorname{tr} SH^{-1}].$$

A minor modification of the argument given in Example 7.10 shows that when $S > 0$, $p(X|\hat{\mu}, H)$ is maximized uniquely, over all positive definite H, at $\hat{H} = n^{-1}S$. When $n \geqslant p + 1$, then S is positive definite with probability one so in this case, the maximum likelihood estimator of H is $\hat{H} = n^{-1}S$. If $\mu = 0$, then

$$\begin{aligned} p(X|0, H) &= \pi^{-np}|H|^{-n}\exp\left[-\sum_{j=1}^{n} X_j^*H^{-1}X_j\right] \\ &= \pi^{-np}|H|^{-n}\exp[-\operatorname{tr} \tilde{S}H^{-1}] \end{aligned}$$

where

$$\tilde{S} = \sum_{j=1}^{n} X_j X_j^* = S + n\bar{X}\bar{X}^*.$$

Obviously, $p(X|0, H)$ is maximized at $\tilde{H} = n^{-1}\tilde{S}$. Thus the likelihood ratio test for testing $\mu = 0$ versus $\mu \neq 0$ rejects for small values of

$$\Lambda = \frac{p(X|0, \tilde{H})}{p(X|\hat{\mu}, \hat{H})} = \frac{|\tilde{S}|^{-n}}{|S|^{-n}} = \frac{|S|^n}{|S + n\bar{X}\bar{X}^*|^n}.$$

As in the real case, $\bar{X}$ and S are independent,

$$\mathcal{L}(S) = \mathbb{C}W(H, p, n-1)$$

and

$$\mathcal{L}(\sqrt{n}\,\bar{X}) = \mathbb{C}N(\sqrt{n}\,\mu, H).$$

Setting $Y = \sqrt{n}\,\bar{X}$,

$$\Lambda^{1/n} = \frac{|S|}{|S + YY^*|} = \frac{1}{1 + Y^* S^{-1} Y}$$

so the likelihood ratio test rejects for large values of $Y^* S^{-1} Y \equiv T^2$. Arguments paralleling those in the real case can be used to show that

$$\mathcal{L}(T^2) = F(2p, 2(n-p), \delta)$$

where $\delta = n\mu^* H^{-1}\mu$ is the noncentrality parameter in the F distribution. Further, the monotone likelihood ratio property of the F- distribution can be used to show that the likelihood ratio test is uniformly most powerful among tests that are invariant under the group of complex linear transformations that preserve the above testing problem.

In the preceeding discussion, we have outlined one possible analysis of the one-sample problem for the complex normal distribution. A theory for the complex MANOVA problem similar to that given in Section 9.1 for the real MANOVA problem would require complex analogues of many results given in the first eight chapters of this book. Of course, it is possible to represent everything in terms of real random vectors. This representation consists of an $n \times 2p$ random matrix $Y \in \mathcal{L}_{2p,n}$ where

$$\mathcal{L}(Y) = N(ZB, I_n \otimes \Psi).$$

As usual, Z is $n \times r$ of rank r and $B: r \times 2p$ is a real matrix of unknown parameters. The distinguishing feature of the model is that Ψ is assumed to have the form

$$\Psi = \begin{pmatrix} \Sigma & -F \\ F & \Sigma \end{pmatrix}$$

where $\Sigma: p \times p$ is positive definite and $F: p \times p$ is skew-symmetric. For reasons that should be obvious by now, Ψ's of the above form are said to have complex covariance structure. This model can now be analyzed using the results developed for the real normal linear model. However, as stated earlier, certain results are clumsy to prove and more difficult to understand when expressed in the real domain rather than the complex domain. Although not at all obvious, these models are not equivalent to a product of real MANOVA models of the type discussed in Section 9.1.

9.6. ADDITIONAL EXAMPLES OF LINEAR MODELS

The examples of this section have been chosen to illustrate how conditioning can sometimes be helpful in finding maximum likelihood estimators and also to further illustrate the use of invariance in analyzing linear models. The linear models considered now are not products of MANOVA models and the regression subspaces are not invariant under the covariance transformations of the model. Thus finding the maximum likelihood estimator of mean vector is not just a matter of computing the orthogonal projection onto the regression subspace. For the models below, we derive maximum likelihood estimators and likelihood ratio tests and then discuss the problem of finding a good invariant test.

The first model we consider consists of a variation on the one-sample problem. Suppose $X_1, \ldots, X_n$ are independent with $\mathcal{L}(X_i) = N(\mu, \Sigma)$ where $X_i \in R^p$, $i = 1, \ldots, n$. As usual, form the $n \times p$ matrix X whose rows are X_i', $i = 1, \ldots, n$. Then

$$\mathcal{L}(X) = N(e\mu', I_n \otimes \Sigma)$$

where $e \in R^n$ is the vector of ones. When μ and Σ are unknown, the linear model for X is a MANOVA model and the results in Section 9.1 apply directly. To transform this model to canonical form, let Γ be an $n \times n$ orthogonal matrix with first row $e'/\sqrt{n}$. Setting $Y = \Gamma X$ and $\beta = \sqrt{n}\,\mu'$,

$$\mathcal{L}(Y) = N(\varepsilon_1 \beta, I_n \otimes \Sigma)$$

where ε_1 is the first unit vector in R^n and $\beta \in \mathcal{L}_{p,1}$. Partition Y as

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

where $Y_1 \in \mathcal{L}_{p,1}$, $Y_2 \in \mathcal{L}_{p,m}$, and $m = n - 1$. Then

$$\mathcal{L}(Y_1) = N(\beta, \Sigma)$$

and

$$\mathcal{L}(Y_2) = N(0, I_m \otimes \Sigma).$$

For testing $H_0 : \beta = 0$, the results of Section 9.1 show that the test that rejects for large values of $Y_1(Y_2'Y_2)^{-1}Y_1'$ (assuming $m \geqslant p$) is equivalent to the likelihood ratio test and this test is most powerful within the class of invariant tests.

We now turn to a testing problem to which the MANOVA results do not apply. With the above discussion in mind, consider $U \in \mathcal{L}_{p,1}$ and $Z \in \mathcal{L}_{p,m}$ where U and Z are independent with

$$\mathcal{L}(U) = N(\beta, \Sigma)$$

and

$$\mathcal{L}(Z) = N(0, I_m \otimes \Sigma).$$

Here, $\beta \in \mathcal{L}_{p,1}$ and $\Sigma > 0$ is a completely unknown $p \times p$ covariance matrix. Partition β into β_1 and β_1 where

$$\beta_i \in \mathcal{L}_{p_i,1}, \qquad i = 1,2, \quad p_1 + p_2 = p.$$

Consider the problem of testing the null hypothesis $H_0: \beta_1 = 0$ versus $H_1: \beta_1 \neq 0$ where β_2 and Σ are unknown. Under H_0, the regression subspace of the random matrix

$$\begin{pmatrix} U \\ Z \end{pmatrix} \in \mathcal{L}_{p,m+1}$$

is

$$M_0 = \left\{ \mu | \mu = \begin{pmatrix} 0 & \beta_2 \\ 0 & 0 \end{pmatrix} \in \mathcal{L}_{p,m+1}, \beta_2 \in \mathcal{L}_{p_2,1} \right\},$$

and the set of covariances is

$$\gamma = \{ I_{m+1} \otimes \Sigma | \Sigma \in \mathcal{S}_p^+ \}.$$

It is easy to verify that M_0 is not invariant under all the elements of γ so the maximum likelihood estimator of β_2 under H_0 cannot be found by least-squares (ignoring Σ). To calculate the likelihood ratio test for H_0 versus H_1, it is convenient to partition U and Z as

$$U = (U_1, U_2), \qquad U_i \in \mathcal{L}_{p_i,1}, \qquad i = 1,2$$

$$Z = (Z_1, Z_2), \qquad Z_i \in \mathcal{L}_{p_i,m}, \qquad i = 1,2$$

and then condition on U_1 and Z_1. Since U and Z are independent, the joint distribution of U and Z is specified by the two conditional distributions, $\mathcal{L}(U_2|U_1)$ and $\mathcal{L}(Z_2|Z_1)$, together with the two marginal distributions, $\mathcal{L}(U_1)$ and $\mathcal{L}(Z_1)$. Our results for the normal distribution show that these distributions are

$$\mathcal{L}(U_2|U_1) = N\big(\beta_2 + (U_1 - \beta_1)\Sigma_{11}^{-1}\Sigma_{12}, \Sigma_{22\cdot1}\big)$$

$$\mathcal{L}(U_1) = N(\beta_1, \Sigma_{11})$$

$$\mathcal{L}(Z_2|Z_1) = N\big(Z_1\Sigma_{11}^{-1}\Sigma_{12}, I_m \otimes \Sigma_{22\cdot1}\big)$$

$$\mathcal{L}(Z_1) = N(0, I_m \otimes \Sigma_{11})$$

where Σ is partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

with Σ_{ij} being $p_i \times p_j$, $i, j = 1, 2$. As usual, $\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. By Proposition 5.8, the reparameterization defined by $\Psi_{11} = \Sigma_{11}$, $\Psi_{12} = \Sigma_{11}^{-1}\Sigma_{12}$, and $\Psi_{22} = \Sigma_{22\cdot 1}$ is one-to-one and onto. To calculate the likelihood ratio test for H_0 versus H_1, we need to find the maximum likelihood estimators under H_0 and H_1.

Proposition 9.17. The likelihood ratio test of $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ rejects H_0 if the statistic

$$\Lambda = U_1 S_{11}^{-1} U_1'$$

is too large. Here, $S = Z'Z$ and

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

where S_{ij} is $p_i \times p_j$.

Proof. Let $f_1(U_1|\beta_1, \Psi_{11})$ be the density of $\mathcal{L}(U_1)$, let $f_2(U_2|U_1, \beta_1, \beta_2, \Psi_{12}, \Psi_{22})$ be the conditional density of $\mathcal{L}(U_2|U_1)$, let $f_3(Z_1|\Psi_{11})$ be the density of $\mathcal{L}(Z_1)$, and let $f_4(Z_2|Z_1, \Psi_{12}, \Psi_{22})$ be the density of $\mathcal{L}(Z_2|Z_1)$. Under H_0, $\beta_1 = 0$ and the unique value of β_2 that maximizes $f_2(U_2|U_1, 0, \beta_2, \Psi_{12}, \Psi_{22})$ is

$$\hat{\beta}_2 = U_2 - U_1\Psi_{12},$$

for Ψ_{12} fixed. It is clear that

$$f_2\left(U_2|U_1, 0, \hat{\beta}_2, \Psi_{12}, \Psi_{22}\right) \,\dot{\alpha}\, |\Psi_{22}|^{-1/2}$$

where the symbol $\dot{\alpha}$ means "is proportional to." We now maximize with respect to Ψ_{12}. With $\beta_2 = \hat{\beta}_2$, Ψ_{12} occurs only in the density of Z_2 given Z_1. Since $\mathcal{L}(Z_2|Z_1) = N(Z_1\Psi_{12}, I_m \otimes \Psi_{22})$, it follows from our treatment of the MANOVA problem that

$$\hat{\Psi}_{12} = (Z_1'Z_1)^{-1}Z_1'Z_2 = S_{11}^{-1}S_{12}$$

and

$$f_4(Z_2|Z_1, \hat{\Psi}_{12}, \Psi_{22}) \,\dot{\alpha}\, |\Psi_{22}|^{-m/2} \exp[-\tfrac{1}{2} \operatorname{tr} S_{22\cdot 1} \Psi_{22}^{-1}].$$

Since $\beta_1 = 0$, it is now clear that

$$\hat{\Psi}_{11} = \frac{1}{m+1}[Z_1'Z_1 + U_1'U_1] = \frac{1}{m+1}[S_{11} + U_1'U_1]$$

and

$$\hat{\Psi}_{22} = \frac{1}{m+1} S_{22\cdot 1}.$$

Substituting these values into the product of the four densities shows that the maximum under H_0 is proportional to

$$\Lambda_0 = |S_{22\cdot 1}|^{-(m+1)/2} |S_{11} + U_1'U_1|^{-(m+1)/2}$$

Under the alternative H_1, we again maximize the likelihood function by first noting that

$$\tilde{\beta}_2 \equiv U_2 - (U_1 - \beta_1)\Psi_{12}$$

maximizes the density of U_2 given U_1. Also,

$$f_2(U_2|U_1, \beta_1, \tilde{\beta}_2, \Psi_{12}, \Psi_{22}) \,\dot{\alpha}\, |\Psi_{22}|^{-1/2}.$$

With this choice of $\tilde{\beta}_2$, β_1 occurs only in the density of U_1 so $\tilde{\beta}_1 = U_1$ maximizes the density of U_1 and

$$f_1(U_1|\tilde{\beta}_1, \Psi_{11}) \,\dot{\alpha}\, |\Psi_{11}|^{-1/2}.$$

It now follows easily that the maximum likelihood estimators of Ψ_{12}, Ψ_{11}, and Ψ_{22} are

$$\tilde{\Psi}_{12} = S_{11}^{-1} S_{12}$$

$$\tilde{\Psi}_{11} = \frac{1}{m+1} Z_1 Z_1' = \frac{1}{m+1} S_{11}$$

$$\tilde{\Psi}_{22} = \frac{1}{m+1} S_{22\cdot 1}.$$

Substituting these into the product of the four densities shows that the maximum under H_1 is proportional to

$$\Lambda_1 = |S_{22\cdot 1}|^{-(m+1)/2}|S_{11}|^{-(m+1)/2}.$$

Hence the likelihood ratio test will reject H_0 for small values of

$$\Lambda_2 = \frac{\Lambda_0}{\Lambda_1} = \frac{|S_{11}|^{(m+1)/2}}{|S_{11} + U_1'U_1|^{(m+1)/2}} = \frac{1}{\left(1 + U_1S_{11}^{-1}U_1'\right)^{(m+1)/2}}.$$

Thus the likelihood ratio test rejects for large values of

$$\Lambda = U_1S_{11}^{-1}U_1'$$

and the proof is complete. □

We now want to show that the test derived above is a uniformly most powerful invariant test under a suitable group of affine transformations. Recall that U and Z are independent and

$$\mathcal{L}(U) = N(\beta, \Sigma), \qquad \mathcal{L}(Z) = N(0, I_m \otimes \Sigma).$$

The problem is to test $H_0: \beta_1 = 0$ where $\beta = (\beta_1, \beta_2)$ with $\beta_i \in \mathcal{L}_{p_i,1}$, $i = 1,2$. Consider the group G with elements $g = (\Gamma, A, (0, a))$ where

$$\Gamma \in \mathcal{O}_m, \qquad (0, a) \in \mathcal{L}_{p,1}, \qquad a \in \mathcal{L}_{p_2,1}$$

and

$$A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{ij} is $p_i \times p_j$ and A_{ii} is nonsingular for $i = 1,2$. The action of $g = (\Gamma, A, (0, a))$ is

$$g\begin{pmatrix} U \\ Z \end{pmatrix} \equiv \begin{pmatrix} UA' + (0, a) \\ \Gamma ZA' \end{pmatrix}.$$

The group operation, defined so G acts on the left of the sample space, is

$$(\Gamma_1, A_1, (0, a_1))(\Gamma_2, A_2, (0, a_2)) = (\Gamma_1\Gamma_2, A_1A_2, (0, a_2)A_1' + (0, a_1)).$$

It is routine to verify that the testing problem is invariant under G. Further, it is clear that the induced action of G on the parameter space is

$$(\Gamma, A, (0, a))(\beta, \Sigma) = (\beta A' + (0, a), A\Sigma A').$$

To characterize the invariant tests for the testing problem, a maximal invariant under the action of G on the sample space is needed.

Proposition 9.18. In the notation of Proposition 9.17, a maximal invariant is

$$\Lambda = U_1 S_{11}^{-1} U_1'.$$

Proof. As usual, the proof consists of showing that $\Lambda = U_1 S_{11}^{-1} U_1'$ is an orbit index. Since $m \geqslant p$, we deal with those Z's that have rank p, a set of probability one. The first claim is that for a given $U \in \mathcal{L}_{p,1}$ and $Z \in \mathcal{L}_{p,m}$ of rank p, there exists a $g \in G$ such that

$$g\begin{pmatrix} U \\ Z \end{pmatrix} = \begin{pmatrix} \Lambda^{1/2}\varepsilon_1' \\ Z_0 \end{pmatrix}$$

where $\varepsilon_1' = (1, 0, \ldots, 0) \in \mathcal{L}_{p,1}$ and

$$Z_0 = \begin{pmatrix} I_p \\ 0 \end{pmatrix} \in \mathcal{L}_{p,m}.$$

Write $Z = \Psi V$ where $\Psi \in \mathcal{F}_{p,m}$ and V is a $p \times p$ upper triangular matrix so $S = Z'Z = V'V$. Then consider

$$A = \begin{pmatrix} \xi_1 & 0 \\ 0 & \xi_2 \end{pmatrix} (V')^{-1}$$

where $\xi_i \in \mathcal{O}_{p_i}$, $i = 1, 2$, and note that A is of the form

$$A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$$

since $(V')^{-1}$ is lower triangular. The values of ξ_i, $i = 1, 2$, are specified in a moment. With this choice of A,

$$ZA' = \Psi V V^{-1} \begin{pmatrix} \xi_1 & 0 \\ 0 & \xi_2 \end{pmatrix}' = \Psi \begin{pmatrix} \xi_1 & 0 \\ 0 & \xi_2 \end{pmatrix}'$$

which is in $\mathcal{F}_{p,m}$ for any choice of $\xi_i \in \mathcal{O}_{p_i}$, $i = 1, 2$. Hence there is a $\Gamma \in \mathcal{O}_m$ such that

$$\Gamma Z A' = Z_0.$$

Since V is upper triangular, write

$$V^{-1} = \begin{pmatrix} V^{11} & V^{12} \\ 0 & V^{22} \end{pmatrix}$$

with V^{ij} being $p_i \times p_j$. Then

$$\begin{aligned} UA' &= UV^{-1}\begin{pmatrix} \xi_1' & 0 \\ 0 & \xi_2' \end{pmatrix} = (U_1, U_2)\begin{pmatrix} V^{11} & V^{12} \\ 0 & V^{22} \end{pmatrix}\begin{pmatrix} \xi_1' & 0 \\ 0 & \xi_2' \end{pmatrix} \\ &= \left(U_1 V^{11}\xi_1', U_1 V^{12}\xi'_2 + U_2 V^{22}\xi_2'\right). \end{aligned}$$

As $S = V'V$ and $V \in G_U^+$, it follows that $S_{11}^{-1} = V^{11}(V^{11})'$ so the vector $U_1 V^{11}$ has squared length $\Lambda = U_1 V^{11}(V^{11})'U_1' = U_1 S_{11}^{-1} U_1'$. Thus there exists $\xi_1' \in \mathcal{O}_{p_1}$ such that

$$U_1 V^{11}\xi_1' = \Lambda^{1/2}\tilde{\varepsilon}_1'$$

where $\tilde{\varepsilon}_1' = (1, 0, \ldots, 0) \in \mathcal{L}_{p_1,1}$. Now choose $a \in \mathcal{L}_{p_2,1}$ to be

$$a = U_1 V^{12}\xi_2' - U_2 V^{22}\xi_2'$$

so

$$UA' + (0, a) = \Lambda^{1/2}\varepsilon_1'.$$

The above choices for A, ξ_1, Γ, and a yield $g = (\Gamma, A, (0, a))$, which satisfies

$$g\begin{pmatrix} U \\ Z \end{pmatrix} = \begin{pmatrix} \Lambda^{1/2}\varepsilon_1' \\ Z_0 \end{pmatrix}$$

and this establishes the claim. To show that

$$\Lambda = U_1 S_{11}^{-1} U_1'$$

is maximal invariant, first notice that Λ is invariant. Further, if

$$\begin{pmatrix} U_1 \\ Z_1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} U_2 \\ Z_2 \end{pmatrix}$$

both yield the same value of Λ, then there exists $g_i \in G$ such that

$$g_i \begin{pmatrix} U_i \\ Z_i \end{pmatrix} = \begin{pmatrix} \Lambda^{1/2}\varepsilon_1' \\ Z_0 \end{pmatrix}, \qquad i = 1, 2.$$

Therefore,

$$g_2^{-1} g_1 \begin{pmatrix} U_1 \\ Z_1 \end{pmatrix} = \begin{pmatrix} U_2 \\ Z_2 \end{pmatrix}$$

and Λ is maximal invariant.

To show that a uniformly most powerful G-invariant test exists, the distribution of $\Lambda = U_1 S_{11}^{-1} U_1'$ is needed. However,

$$\mathcal{L}(U_1) = N(\beta_1, \Sigma_{11})$$

$$\mathcal{L}(S_{11}) = W(\Sigma_{11}, p_1, m)$$

and U_1 and S_{11} are independent. From Proposition 8.14, we see that

$$\mathcal{L}(\Lambda) = F(p_1, m - p_1 + 1, \delta)$$

where $\delta = \beta_1 \Sigma_{11}^{-1} \beta_1'$ and the null hypothesis is $H_0 : \delta = 0$. Since the noncentral F distribution has a monotone likelihood ratio, the test that rejects for large values of Λ is uniformly most powerful within the class of tests based on Λ. Since all G-invariant tests are functions of Λ, we conclude that the likelihood ratio test is uniformly most powerful invariant. □

The final problem to be considered in this chapter is a variation of the problem just solved. Again, the testing problem of interest is $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$, but it is assumed that the value of β_2 is known to be zero under both H_0 and H_1. Thus our model for U and Z is that U and Z are independent with

$$\mathcal{L}(U) = N((\beta_1, 0), \Sigma)$$

$$\mathcal{L}(Z) = N(0, I_m \otimes \Sigma)$$

where $U \in \mathcal{L}_{p,1}$, $\beta_1 \in \mathcal{L}_{p_1,1}$, $Z \in \mathcal{L}_{p,m}$, and $m \geq p$. In what follows, the likelihood ratio test of H_0 versus H_1 is derived and an invariance argument shows that there is no uniformly most powerful invariant test under a natural group that leaves the problem invariant. As usual, we partition U

into U_1 and U_2, $U_i \in \mathcal{L}_{p_i, 1}$, and Z is partitioned into $Z_1 \in \mathcal{L}_{p_1, m}$ and $Z_2 \in \mathcal{L}_{p_2, m}$ so

$$U = (U_1, U_2), \qquad Z = (Z_1, Z_2).$$

Also

$$S = Z'Z = \begin{pmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{pmatrix} \equiv \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

and $S_{11 \cdot 2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$.

Proposition 9.19. The likelihood ratio test of H_0 versus H_1 rejects for large values of the statistic

$$\Lambda = \frac{(U_1 - U_2S_{22}^{-1}S_{21})S_{11\cdot 2}^{-1}(U_1 - U_2S_{22}^{-1}S_{21})'}{1 + U_2S_{22}^{-1}U_2'}.$$

Proof. Under H_0,

$$\mathcal{L}(U) = N(0, \Sigma)$$

$$\mathcal{L}(Z) = N(0, I_m \otimes \Sigma)$$

so the maximum likelihood estimator of Σ is

$$\hat{\Sigma} = \frac{1}{m+1}(Z'Z + U'U) = \frac{1}{m+1}(S + U'U).$$

The value of the maximized likelihood function is proportional to

$$\Lambda_0 \equiv |\hat{\Sigma}|^{-(m+1)/2}.$$

Under H_1, the situation is a bit more complicated and it is helpful to consider conditional distributions. Under H_1,

$$\mathcal{L}(U_1|U_2) = N(\beta_1 + U_2\Sigma_{22}^{-1}\Sigma_{21}, \Sigma_{11\cdot 2})$$

$$\mathcal{L}(U_2) = N(0, \Sigma_{22})$$

$$\mathcal{L}(Z_1|Z_2) = N(Z_2\Sigma_{22}^{-1}\Sigma_{21}, I_m \otimes \Sigma_{11\cdot 2})$$

and

$$\mathcal{L}(Z_2) = N(0, I_m \otimes \Sigma_{22}).$$

The reparameterization defined by $\Psi_{11} = \Sigma_{11\cdot 2}$, $\Psi_{21} = \Sigma_{22}^{-1}\Sigma_{21}$, and $\Psi_{22} = \Sigma_{22}$ is one-to-one and onto. Let $f_1(U_1|U_2, \beta_1, \Psi_{21}, \Psi_{11})$, $f_2(U_2|\Psi_{22})$, $f_3(Z_1|Z_2, \Psi_{21}, \Psi_{11})$, and $f_4(Z_2|\Psi_{22})$ be the density functions with respect to Lebesgue measure $dU_1\, dU_2\, dZ_1\, dZ_2$ of the four distributions above. It is clear that

$$\tilde{\beta}_1 = U_1 - U_2\Psi_{21}$$

maximizes $f_1(U_1|U_2, \beta_1, \Psi_{21}, \Psi_{11})$ and $f_1(U_1|U_2, \tilde{\beta}_1, \Psi_{21}, \Psi_{11}) \,\dot{\alpha}\, |\Psi_{11}|^{-1/2}$. With $\tilde{\beta}_2$ substituted into f_1, the parameter Ψ_{21} only occurs in the density $f_3(Z_1|Z_2, \Psi_{21}, \Psi_{11})$. Since

$$\mathcal{L}(Z_1|Z_2) = N(Z_2\Psi_{21}, I_m \otimes \Psi_{11}),$$

our results for the MANOVA model show that

$$\tilde{\Psi}_{21} = (Z_2'Z_2)^{-1}Z_2'Z_1 = S_{22}^{-1}S_{21}$$

maximizes $f_3(Z_1|Z_2, \Psi_{21}, \Psi_{11})$ for each value of Ψ_{11}. When $\tilde{\Psi}_{21}$ is substituted into f_3, an inspection of the resulting four density functions shows that the maximum likelihood estimators of Ψ_{11} and Ψ_{22} are

$$\tilde{\Psi}_{11} = \frac{1}{m+1}S_{11\cdot 2}$$

and

$$\tilde{\Psi}_{22} = \frac{1}{m+1}(Z_2'Z_2 + U_2'U_2) = \frac{1}{m+1}(S_{22} + U_2'U_2).$$

Under H_1, this yields a maximized likelihood function proportional to

$$\Lambda_1 = |\tilde{\Psi}_{11}|^{-(m+1)/2}|\tilde{\Psi}_{22}|^{-(m+1)/2}.$$

Therefore the likelihood ratio test rejects H_0 for small values of

$$\Lambda_3 \equiv \frac{\Lambda_0}{\Lambda_1} = \left[\frac{|S_{22} + U_2'U_2||S_{11\cdot 2}|}{|S + U'U|}\right]^{(m+1)/2}.$$

However,

$$|S_{22} + U_2'U_2| = |S_{22}|\left(1 + U_2 S_{22}^{-1} U_2'\right)$$

and

$$|S| = |S_{22}||S_{11\cdot 2}|.$$

Thus

$$[\Lambda_3]^{2/(m+1)} = \frac{|S|\left(1 + U_2 S_{22}^{-1} U'\right)}{|S + U'U|} = \frac{1 + U_2 S_{22}^{-1} U_2'}{1 + US^{-1}U}.$$

Now, the identity

$$US^{-1}U' = \left(U_1 - U_2 S_{22}^{-1} S_{21}\right) S_{11\cdot 2}^{-1} \left(U_1 - U_2 S_{22}^{-1} S_{21}\right)' + U_2 S_{22}^{-1} U_2'$$

follows from the problems in Chapter 5. Hence rejecting for small values of

$$[\Lambda_3]^{2/(m+1)} = \frac{1}{1 + \Lambda},$$

where Λ is given in the statement of this proposition, is equivalent to rejecting for large values of Λ. □

The above testing problem is now analyzed via invariance. The group G consists of elements $g = (\Gamma, A)$ where $\Gamma \in \mathcal{O}_m$ and

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \qquad A_{ii} \in Gl_{p_i}, \quad i = 1,2.$$

The group action is

$$(\Gamma, A)\begin{pmatrix} U \\ Z \end{pmatrix} = \begin{pmatrix} UA' \\ \Gamma Z A' \end{pmatrix}$$

and group composition is

$$(\Gamma_1, A_1)(\Gamma_2, A_2) = (\Gamma_1\Gamma_2, A_1 A_2).$$

The action of the group on the parameter space is

$$(\Gamma, A)(\beta_1, \Sigma) = (\beta_1 A_{11}', A\Sigma A').$$

It is clear that the testing problem is invariant under the group G.

Proposition 9.20. Under the action of G on the sample space, a maximal invariant is the pair (W_1, W_2) where

$$W_1 = \frac{(U_1 - U_2S_{22}^{-1}S_{21})S_{11\cdot 2}^{-1}(U_1 - U_2S_{22}^{-1}S_{21})'}{1 + U_2S_{22}^{-1}U_2'}$$

and

$$W_2 = U_2S_{22}^{-1}U_2'.$$

A maximal invariant in the parameter space is

$$\delta = \beta_1\Sigma_{11\cdot 2}^{-1}\beta_1'.$$

Proof. As usual, the method of proof is a reduction argument that provides a convenient index for thc orbits in the sample space. Since $m \geqslant p$, a set of measure zero can be deleted from the sample space so that Z has rank p on the complement of this set. Let

$$Z_0 = \begin{pmatrix} I_p \\ 0 \end{pmatrix} \in \mathcal{L}_{p,m}$$

and set $u_1 = \varepsilon_1' \in \mathcal{L}_{p,1}$ and $u_2 = (0,\ldots,0,1,0,\ldots,0) \in \mathcal{L}_{p,1}$ where the one occurs in the $(p_1 + 1)$ coordinate of u_2. Now, given U and Z, we claim that there exists a $g = (\Gamma, A) \in G$ such that

$$g\begin{pmatrix} U \\ Z \end{pmatrix} = \begin{pmatrix} X_1u_1 + X_2u_2 \\ Z_0 \end{pmatrix}, \qquad X_i \geqslant 0,$$

where

$$X_1^2 = (U_1 - U_2S_{22}^{-1}S_{21})S_{11\cdot 2}(U_1 - U_2S_{22}^{-1}S_{21})'$$

and

$$X_2^2 = U_2S_{22}^{-1}U_2'.$$

To establish this claim, write $Z = \Psi T$ where $\Psi \in \mathcal{F}_{p,m}$ and $T \in G_T^+$ is a $p \times p$ lower triangular matrix. A modification of the proof of Proposition 5.2 establishes this representation for Z. Consider

$$A = \xi(T^{-1})' = \begin{pmatrix} \xi_1 & 0 \\ 0 & \xi_2 \end{pmatrix}(T^{-1})'$$

where $\xi_i \in \mathcal{O}_{p_i}$, $i = 1, 2$, so $\xi \in \mathcal{O}_p$ and

$$ZA' = \Psi TA' = \Psi \xi' \in \mathcal{F}_{p,m}.$$

Thus for any such ξ and $\Gamma \in \mathcal{O}_m$, $(\Gamma, A) \in G$. Also, Γ can be chosen so that

$$\Gamma ZA' = Z_0 \in \mathcal{F}_{p,m}.$$

Now,

$$UA' = (U_1, U_2)T^{-1}\xi' = (U_1, U_2)\begin{pmatrix} T^{11} & 0 \\ T^{21} & T^{22} \end{pmatrix}\begin{pmatrix} \xi_1' & 0 \\ 0 & \xi_2' \end{pmatrix}$$

$$= \left((U_1T^{11} + U_2T^{21})\xi_1', U_2T^{22}\xi_2'\right)$$

where T^{ij} is $p_i \times p_j$ and

$$T^{-1} \equiv \begin{pmatrix} T^{11} & 0 \\ T^{21} & T^{22} \end{pmatrix}.$$

Since

$$S = Z'Z = T'T,$$

a bit of algebra shows that

$$\begin{aligned}(U_1T^{11} + U_2T^{21})&(U_1T^{11} + U_2T^{21})' \\ &= (U_1 - U_2S_{22}^{-1}S_{21})S_{11\cdot 2}^{-1}(U_1 - U_2S_{22}^{-1}S_{21})' = X_1^2\end{aligned}$$

and

$$(U_2T^{22})(U_2T^{22})' = U_2S_{22}^{-1}U_2' = X_2^2.$$

Let $\tilde{\varepsilon}_1 = (1, 0, \ldots, 0) \in \mathcal{L}_{p_1,1}$ and $\tilde{\varepsilon}_2 = (1, 0, \ldots, 0) \in \mathcal{L}_{p_2,1}$. Since the vectors $X_1\tilde{\varepsilon}_1$ and $U_1T^{11} + U_2T^{21}$ have the same length, there exists $\xi_1' \in \mathcal{O}_{p_1}$ such that

$$(U_1T^{11} + U_2T^{22})\xi_1' = X_1\tilde{\varepsilon}_1.$$

For similar reasons, there exists a $\xi_2' \in \mathcal{O}_{p_2}$ such that

$$U_2T_{22}\xi_2' = X_2\tilde{\varepsilon}_2.$$

With these choices for ξ_1 and ξ_2,

$$\left((U_1T^{11} + U_2T^{21})\xi_1', U_2T^{22}\xi_2'\right) = (X_1u_1 + X_2u_2).$$

Thus there is a $g = (\Gamma, A) \in G$ such that

$$g\begin{pmatrix} U \\ Z \end{pmatrix} = \begin{pmatrix} X_1 u_1 + X_2 u_2 \\ Z_0 \end{pmatrix}.$$

This establishes the claim. It is now routine to show that $(X_1, X_2) = (X_1(U, Z), X_2(U, Z))$ is an invariant function. To show that (X_1, X_2) is maximal invariant, suppose (U, Z) and $(\tilde{U}, \tilde{Z})$ yield the same (X_1, X_2) values. Then there exist g and $\tilde{g}$ in G such that

$$g\begin{pmatrix} U \\ Z \end{pmatrix} = \begin{pmatrix} X_1 u_1 + X_2 u_2 \\ Z_0 \end{pmatrix} = \tilde{g}\begin{pmatrix} \tilde{U} \\ \tilde{Z} \end{pmatrix}$$

so

$$g^{-1}\tilde{g}\begin{pmatrix} \tilde{U} \\ \tilde{Z} \end{pmatrix} = \begin{pmatrix} U \\ Z \end{pmatrix}.$$

This shows that (X_1, X_2) is maximal invariant. Since the pair (W_1, W_2) is a one-to-one function of (X_1, X_2), it follows that (W_1, W_2) is maximal invariant. The proof that δ is a maximal invariant in the parameter space is similar and is left to the reader. □

In order to suggest an invariant test for $H_0 : \beta_1 = 0$ based on (W_1, W_2), the distribution of (W_1, W_2) is needed. Since

$$\mathcal{L}((U_1, U_2)) = N((\beta_1, 0), \Sigma)$$

and

$$\mathcal{L}(S) = W(\Sigma, p, m)$$

with S and U independent,

$$\mathcal{L}(W_2) = \mathcal{L}\left(U_2 S_{22}^{-1} U_2'\right) = F_{p_2, m-p_2+1}.$$

Therefore, W_2 is an ancillary statistic as its distribution does not depend on any parameters under H_0 or H_1. We now compute the conditional distribution of W_1 given W_2. Proposition 8.7 shows that

$$\mathcal{L}(S_{11\cdot 2}) = W(\Sigma_{11\cdot 2}, p_1, m - p_2)$$

$$\mathcal{L}\left(S_{22}^{-1} S_{21} | S_{22}\right) = N\left(\Sigma_{22}^{-1}\Sigma_{21}, S_{22}^{-1} \otimes \Sigma_{11\cdot 2}\right)$$

and

$$\mathcal{L}(S_{22}) = W(\Sigma_{22}, p_2, m)$$

where $S_{11\cdot 2}$ is independent of (S_{21}, S_{22}). Thus

$$\mathcal{L}\big(U_2 S_{22}^{-1} S_{21} | S_{22}, U_2\big) = N\big(U_2 \Sigma_{22}^{-1} \Sigma_{21}, \big(U_2 S_{22}^{-1} U_2'\big)\Sigma_{11\cdot 2}\big)$$

and conditional on (S_{22}, U_2),

$$\mathcal{L}(U_1 | S_{22}, U_2) = N\big(\beta_1 + U_2 \Sigma_{22}^{-1} \Sigma_{21}, \Sigma_{11\cdot 2}\big).$$

Further, U_1 and $U_2 S_{22}^{-1} S_{21}$ are conditionally independent—given (S_{22}, U_2). Therefore,

$$\mathcal{L}\big(U_1 - U_2 S_{22}^{-1} S_{21} | S_{22}, U_2\big) = N\big(\beta_1, \big(1 + U_2 S_{22}^{-1} U_2'\big)\Sigma_{11\cdot 2}\big)$$

so

$$\mathcal{L}\left(\frac{U_1 - U_2 S_{22}^{-1} S_{21}}{\sqrt{1 + W_2}} \Big| S_{22}, U_2\right) = N\left(\frac{\beta_1}{\sqrt{1 + W_2}}, \Sigma_{11\cdot 2}\right).$$

Since $S_{11\cdot 2}$ is independent of all other variables under consideration, and since

$$W_1 = \frac{\big(U_1 - U_2 S_{22}^{-1} S_{21}\big) S_{11\cdot 2} \big(U_1 - U_2 S_{22}^{-1} S_{21}\big)'}{1 + W_2},$$

it follows from Proposition 8.14 that

$$\mathcal{L}(W_1 | S_{22}, U_2) = F\left(p_1, m - p + 1; \frac{\delta}{1 + W_2}\right)$$

where $\delta = \beta_1 \Sigma_{11\cdot 2}^{-1} \beta_1'$. However, the conditional distribution of W_1 given (S_{22}, U_2) depends on (S_{22}, U_2) only through the function $W_2 = U_2 S_{22}^{-1} U_2'$. Thus

$$\mathcal{L}(W_1 | W_2) = F\left(p_1, m - p + 1; \frac{\delta}{1 + W_2}\right),$$

and

$$\mathcal{L}(W_2) = F_{p_2, m - p_2 + 1}.$$

Further, the null hypothesis is $H_0: \delta = 0$ versus the alternative $H_1: \delta > 0$. Under H_0, it is clear that W_1 and W_2 are independent. The likelihood ratio test rejects H_0 for large values of W_1 and ignores W_2. Of course, the level of this test is computed from a standard F-table, but the power of the test involves the marginal distribution of W_1 when $\delta > 0$. This marginal distribution, obtained by averaging the conditional distribution $\mathcal{L}(W_1|W_2)$ with respect to the distribution of W_2, is rather complicated.

To show that a uniformly most powerful test of H_0 versus H_1 does not exist, consider a particular alternative $\delta = \delta_0 > 0$. Let $f_1(w_1|w_2, \delta)$ denote the conditional density function of W_1 given W_2 and let $f_2(w_2)$ denote the density of W_2. For testing $H_0: \delta = 0$ versus $H_1: \delta = \delta_0$, the Neyman–Pearson Lemma asserts that the most powerful test of level α is to reject if

$$\frac{f_1(w_1|w_2, \delta_0)}{f_1(w_1|w_2, 0)} > c(\alpha)$$

where $c(\alpha)$ is chosen to make the test have level α. However, the rejection region for this test depends on the particular alternative δ_0 so a uniformly most powerful test cannot exist. Since W_2 is ancillary, we can argue that the test of H_0 should be carried out conditional on W_2, that is, the level and the power of tests should be compared only for the conditional distribution of W_1 given W_2. In this case, for w_2 fixed, the ratio

$$\frac{f_1(w_1|w_2, \delta_0)}{f_1(w_1|w_2, 0)}$$

is an increasing function of w_1 so rejecting for large values of the ratio (w_2 fixed) is equivalent to rejecting for $W_1 > k$. If k is chosen to make the test have level α, this argument leads to the level α likelihood ratio test.

PROBLEMS

1. Consider independent random vectors X_{ij} with $\mathcal{L}(X_{ij}) = N(\mu_i, \Sigma)$ for $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$. For scalars $a_1, \ldots, a_k$ consider testing $H_0: \Sigma a_i\mu_i = 0$ versus $H_1: \Sigma a_i\mu_i \neq 0$. With $\tau^2 = \Sigma a_i^2 n_i^{-1}$, let $b_i = \tau^{-1}a_i$, set $\overline{X}_i = n_i^{-1}\Sigma_j X_{ij}$ and let $S_i = \Sigma_j(X_{ij} - \overline{X}_i)(X_{ij} - \overline{X}_i)'$. Write this problem in the canonical form of Section 9.1 and prove that the test that rejects for large values of $\Lambda = (\Sigma_i b_i\overline{X}_i)'S^{-1}(\Sigma_i b_i\overline{X}_i)$ is UMP invariant. Here $S = \Sigma_i S_i$. What is the distribution of Λ under H_0?

2. Given $Y \in \mathcal{L}_{p,n}$ and $X \in \mathcal{L}_{k,n}$ of rank k, the least-squares estimate $\hat{B} = (X'X)^{-1}X'Y$ of B can be characterized as the B that mini-

mizes $\operatorname{tr}(Y - XB)'(Y - XB)$ over all $k \times p$ matrices.

(i) Show that for any $k \times p$ matrix B,

$$(Y - XB)'(Y - XB) = (Y - X\hat{B})'(Y - X\hat{B}) + \big(X(B - \hat{B})\big)'\big(X(B - \hat{B})\big).$$

(ii) A real-valued function ϕ defined for $p \times p$ nonnegative definite matrices is *nondecreasing* if $\phi(S_1) \leqslant \phi(S_1 + S_2)$ for any S_1 and S_2 ($S_i \geqslant 0$, $i = 1, 2$). Using (i), show that, if ϕ is nondecreasing, then $\phi((Y - XB)'(Y - XB))$ is minimized by $B = \hat{B}$.

(iii) For A that is $p \times p$ and nonnegative definite, show that $\phi(S) = \operatorname{tr} AS$ is nondecreasing. Also, show that $\phi(S) = \det(A + S)$ is nondecreasing.

(iv) Suppose $\phi(S) = \phi(\Gamma S \Gamma')$ for $S \geqslant 0$ and $\Gamma \in \mathcal{O}_p$ so $\phi(S)$ can be written as $\phi(S) = \psi(\lambda(S))$ where $\lambda(S)$ is the vector of ordered characteristic roots of S. Show that, if ψ is nondecreasing in each argument, then ϕ is nondecreasing.

3. (The MANOVA model under non-normality.) Let E be a random $n \times p$ matrix that satisfies $\mathcal{L}(\Gamma E \psi') = \mathcal{L}(E)$ for all $\Gamma \in \mathcal{O}_n$ and $\psi \in \mathcal{O}_p$. Assume that $\operatorname{Cov}(E) = I_n \otimes I_p$ and consider a linear model for $Y \in \mathcal{L}_{p,n}$ generated by $Y = ZB + EA'$ where Z is a fixed $n \times k$ matrix of rank k, B is a $k \times p$ matrix of unknown parameters, and A is an element of Gl_p.

(i) Show that the distribution of Y depends on (B, A) only through (B, AA').

(ii) Let $M = \{\mu | \mu = ZB, B \in \mathcal{L}_{p,k}\}$ and $\gamma = \{I_n \otimes \Sigma | \Sigma > 0, \Sigma \text{ is } p \times p\}$. Show that $\{M, \gamma\}$ serves as a parameter space for the linear model (the distribution of E is assumed fixed).

(iii) Consider the problem of testing $H_0: RB = 0$ where R is $r \times k$ of rank r. Show that the reduction to canonical form given in Section 9.1 can be used here to give a model of the form

$$\begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \tilde{Y}_3 \end{pmatrix} = \begin{pmatrix} \tilde{B}_1 \\ \tilde{B}_2 \\ 0 \end{pmatrix} + \tilde{E}A' \tag{9.6}$$

where $\tilde{Y}_1$ is $r \times p$, $\tilde{Y}_2$ is $(k - r) \times p$, $\tilde{Y}_3$ is $(n - k) \times p$, $\tilde{B}_1$ is $r \times p$, $\tilde{B}_2$ is $(k - r) \times p$, $\tilde{E}$ is $n \times p$, and A is as in the original model. Further, E and $\tilde{E}$ have the same distribution and the null hypothesis is $H_0: \tilde{B}_1 = 0$.

(iv) Now, assume the form of the model in (9.6) and drop the tildas. Using the invariance argument given in Section 9.1, the testing problem is invariant and any invariant test is a function of the t largest eigenvalues of $Y_1(Y_3'Y_3)^{-1}Y_1'$ where $t = \min\{r, p\}$. Assume $n - k \geqslant p$ and partition E as Y is partitioned. Under H_0, show that

$$W \equiv Y_1(Y_3'Y_3)^{-1}Y_1' = E_1(E_3'E_3)^{-1}E_1'.$$

(v) Using Proposition 7.3 show that W has the same distribution no matter what the distribution of E as long as $\mathcal{L}(\Gamma E) = \mathcal{L}(E)$ for all $\Gamma \in \mathcal{O}_n$ and E_3 has rank p with probability one. This distribution of W is the distribution obtained by assuming the elements of E are i.i.d. $N(0, 1)$. In particular, any invariant test of H_0 has the same distribution under H_0 as when E is $N(0, I_n \otimes I_p)$.

4. When Y_1 is $N(B_1, I_r \otimes \Sigma)$ and Y_3 is $N(0, I_m \otimes \Sigma)$ with $m \geqslant p + 2$, verify the claim that

$$\mathcal{E} Y_1'Y_1(Y_3'Y_3)^{-1} = \frac{r}{m - p - 1} I_p + \frac{1}{m - p - 1} B_1'B_1\Sigma^{-1}.$$

5. Consider a data matrix $Y: n \times 2$ and assume $\mathcal{L}(Y) = N(ZB, I_n \otimes \Sigma)$ where Z is $n \times 2$ of rank two so B is 2×2. In some situations, it is reasonable to assume that $\sigma_{11} = \sigma_{22}$—that is, the diagonal elements of Σ are the same. Under this assumption, use the results of Section 9.2 to derive the likelihood ratio test for $H_0: b_{11} = b_{12}, b_{21} = b_{22}$ versus $H_1: b_{11} \neq b_{12}$ or $b_{21} \neq b_{22}$. Is this test UMP invariant?

6. Consider a "two-way layout" situation with observations Y_{ij}, $i = 1, \ldots, m$ and $j = 1, \ldots, r$. Assume $Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$ where μ, α_i, and β_j are constants that satisfy $\Sigma\alpha_i = \Sigma\beta_j = 0$. The e_{ij} are random errors with mean zero (but not necessarily uncorrelated). Let Y be the $m \times n$ matrix of Y_{ij}'s, u_1 be the vector of ones in R^m, u_2 be the vector of ones in R^n, $\alpha \in R^m$ be the vector with coordinates α_i, and $\beta \in R^n$ be the vector with coordinates β_j. Let E be the matrix of e_{ij}'s.

(i) Show the model is $Y = \mu u_1 u_2' + \alpha u_2' + u_1\beta' + E$ in the vector space $\mathcal{L}_{n,m}$. Here, $\alpha \in R^m$ with $\alpha' u_1 = 0$ and $\beta \in R^n$ with $\beta' u_2$

= 0. Let

$$M_1 = \{x|x \in \mathcal{L}_{n,m}, x = \mu u_1 u_2', \mu \in R^1\}$$

$$M_2 = \{x|x = \alpha u_2', \alpha \in R^m, \alpha' u_1 = 0\}$$

$$M_3 = \{x|x = u_1\beta', \beta \in R^n, \beta' u_2 = 0\}.$$

Also, let $\langle \cdot, \cdot \rangle$ be the usual inner product on $\mathcal{L}_{n,m}$.

(ii) Show $M_1 \perp M_2 \perp M_3 \perp M_1$ in $(\mathcal{L}_{n,m}, \langle \cdot, \cdot \rangle)$.

Now, assume $\text{Cov}(E) = I_m \otimes A$ where $A = \gamma P + \delta Q$ with $P = n^{-1}u_2u_2'$, $Q = I - P$, and $\gamma > 0$ and $\delta > 0$ are unknown parameters.

(iii) Show the regression subspace $M = M_1 \oplus M_2 \oplus M_3$ is invariant under each $I_m \otimes A$. Find the Gauss–Markov estimates for μ, α, and β.

(iv) Now, assume E is $N(0, I_m \otimes A)$. Use an invariance argument to show that for testing $H_0: \alpha = 0$ versus $H_1: \alpha \neq 0$, the test that rejects for large values of $W = \|P_{M_2}Y\|^2/\|Q_M Y\|^2$ is a UMP invariant test. Here, $Q_M = I - P_M$. What is the distribution of W?

7. The regression subspace for the MANOVA model was described as $M = \{\mu|\mu = ZB, B \in \mathcal{L}_{p,k}\} \subseteq \mathcal{L}_{p,n}$ where Z is $n \times k$ of rank k. The subspace of M associated with the null hypothesis $H_0: RB = 0$ (R is $r \times r$ of rank r) is $\omega = \{\mu|\mu = ZB, B \in \mathcal{L}_{p,k}, RB = 0\}$. We know that $P_M = P_Z \otimes I_p$ where $P_Z = Z(Z'Z)^{-1}Z'$ (P_M is the orthogonal projection onto M in $(\mathcal{L}_{p,n}, \langle \cdot, \cdot \rangle)$). This problem gives one form for P_ω. Let $W = Z(Z'Z)^{-1}R'$.

(i) Show that W has rank r.

Let $P_W = W(W'W)^{-1}W'$ so $P_W \otimes I_p$ is an orthogonal projection.

(ii) Show that $\mathcal{R}(P_W \otimes I_p) \subseteq M - \omega$ where $M - \omega = M \cap \omega^\perp$. Also, show $\dim(\mathcal{R}(P_W \otimes I_p)) = rp$.

(iii) Show that $\dim \omega = (k - r)p$.

(iv) Now, show that $P_W \otimes I_p$ is the orthogonal projection onto $M - \omega$ so $P_Z \otimes I_p - P_W \otimes I_p$ is the orthogonal projection onto ω.

8. Assume $X_1, \ldots, X_n$ are i.i.d. from a five-dimensional $N(0, \Sigma)$ where Σ is a cyclic covariance matrix (Σ is written out explicitly at the beginning of Section 9.4). Find the maximum likelihood estimators of σ^2, ρ_1, ρ_2.

9. Suppose $X_1, \ldots, X_n$ are i.i.d. $N(0, \Psi)$ of dimension $2p$ and assume Ψ has the complex form

$$\Psi = \begin{pmatrix} \Sigma & F \\ -F & \Sigma \end{pmatrix}.$$

Let $S = \Sigma_1^n X_i X_i'$ and partition S as Ψ is partitioned. show that $\hat{\Sigma} = (2n)^{-1}(S_{11} + S_{22})$ and $\hat{F} = (2n)^{-1}(S_{12} - S_{21})$ are the maximum likelihood estimates of Σ and F.

10. Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \Sigma)$ p-dimensional random vectors where μ and Σ are unknown, $\Sigma > 0$. Suppose R is $r \times p$ of rank r and consider testing $H_0: R\mu = 0$ versus $H_1: R\mu \neq 0$. Let $\bar{X} = (1/n)\Sigma_1^n X_i$ and $S = \Sigma_1^n (X_i - \bar{X})(X_i - \bar{X})'$. Show that the test that rejects for large values of $T = (R\bar{X})'(RSR')^{-1}(R\bar{X})$ is equivalent to the likelihood ratio test. Also, show this test is UMP invariant under a suitable group of transformations. Apply this to the problem of testing $\mu_1 = \mu_2 = \cdots = \mu_p$ where $\mu_1, \ldots, \mu_p$ are the coordinates of μ.

11. Consider a linear model of the form $Y = ZB + E$ with $Z: n \times k$ of rank k, $B: k \times p$ unknown, and E a matrix of errors. Assume the first column of Z is the vector e of ones (the regression equation has the constant term in it). Assume $\text{Cov}(E) = A(\rho) \otimes \Sigma$ where $A(\rho)$ has ones on the diagonal and ρ off the diagonal $(-1/(n-1) < \rho < 1)$.
    (i) Show that the GM and least-squares estimates of B are the same.
    (ii) When $\mathcal{L}(E) = N(0, A(\rho) \otimes \Sigma)$ with Σ and ρ unknown, argue via invariance to construct tests for hypotheses of the form $\dot{R}\dot{B} = 0$ where $\dot{R}$ is $r \times k - 1$ of rank r and $\dot{B}: (k-1) \times p$ consists of the last $k - 1$ rows of B.

NOTES AND REFERENCES

1. The material in Section 9.1 is fairly standard and can be found in many texts on multivariate analysis although the treatment and emphasis may be different than here. The likelihood ratio test in the MANOVA setting was originally proposed by Wilks (1932). Various competitors to the likelihood ratio test were proposed in Lawley (1938), Hotelling (1947), Roy (1953), and Pillai (1955).

2. Arnold (1973) applied the theory of products of problems (which he had developed in his Ph.D. dissertation at Stanford) to situations involving patterned covariance matrices. This notion appears in both this chapter and Chapter 10.

3. Given the covariance structure assumed in Section 9.2, the regression subspaces considered there are not the most general for which the Gauss–Markov and least-squares estimators are the same. See Eaton (1970) for a discussion.

4. Andersson (1975) provides a complete description of all symmetry models.

5. Cyclic covariance models were first studied systematically in Olkin and Press (1969).

6. For early papers on the complex normal distribution, see Goodman (1963) and Giri (1965a). Also, see Andersson (1975).

7. Some of the material in Section 9.6 comes from Giri (1964, 1965b).

8. In Proposition 9.5, when $r = 1$, the statistic λ_1 is commonly known as Hotelling's T^2 (see Hotelling (1931)).