

## Chapter 6

# Hypothesis Testing and Null Distributions



### 6.1. Introduction

It is assumed that the readers are familiar with the concept of testing statistical hypotheses on the parameters of a real scalar normal density or independent real scalar normal densities. Those who are not or require a refresher may consult the textbook: Mathai and Haubold (2017) on basic “Probability and Statistics” [De Gruyter, Germany, 2017, free download]. Initially, we will only employ the likelihood ratio criterion for testing hypotheses on the parameters of one or more real multivariate Gaussian (or normal) distributions. All of our tests will be based on a simple random sample of size  $n$  from a  $p$ -variate nonsingular Gaussian distribution, that is, the  $p \times 1$  vectors  $X_1, \dots, X_n$  constituting the sample are iid (independently and identically distributed) as  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ ,  $j = 1, \dots, n$ , when a single real Gaussian population is involved. The corresponding test criterion for the complex Gaussian case will also be mentioned in each section.

In this chapter, we will utilize the following notations. Lower-case letters such as  $x, y$  will be used to denote real scalar mathematical or random variables. No distinction will be made between mathematical and random variables. Capital letters such as  $X, Y$  will denote real vector/matrix-variate variables, whether mathematical or random. A tilde placed on a letter as for instance  $\tilde{x}, \tilde{y}, \tilde{X}$  and  $\tilde{Y}$  will indicate that the variables are in the complex domain. No tilde will be used for constant matrices unless the point is to be stressed that the matrix concerned is in the complex domain. The other notations will be identical to those utilized in the previous chapters.

First, we consider certain problems related to testing hypotheses on the parameters of a  $p$ -variate real Gaussian population. Only the likelihood ratio criterion, also referred to as  $\lambda$ -criterion, will be utilized. Let  $L$  denote the joint density of the sample values in a simple random sample of size  $n$ , namely,  $X_1, \dots, X_n$ , which are iid  $N_p(\mu, \Sigma)$ ,  $\Sigma > O$ . Then, as was previously established,

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2}(X_j - \mu)' \Sigma^{-1} (X_j - \mu)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{n}{2} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}, \quad (6.1.1)$$

where  $S = \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})'$  is the sample sum of products matrix and  $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$  is the sample average,  $n$  being the sample size. As well, we have already determined that the maximum likelihood estimators (MLE's) of  $\mu$  and  $\Sigma$  are  $\hat{\mu} = \bar{X}$  and  $\hat{\Sigma} = \frac{1}{n}S$ , the sample covariance matrix. Consider the parameter space

$$\Omega = \{(\mu, \Sigma) \mid \Sigma > O, \mu' = (\mu_1, \dots, \mu_p), -\infty < \mu_j < \infty, j = 1, \dots, p\}.$$

The maximum value of  $L$  within  $\Omega$  is obtained by substituting the MLE's of the parameters into  $L$ , and since  $(\bar{X} - \hat{\mu}) = (\bar{X} - \bar{X}) = O$  and  $\text{tr}(\hat{\Sigma}^{-1} S) = \text{tr}(nI_p) = np$ ,

$$\max_{\Omega} L = \frac{e^{-\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |\frac{1}{n}S|^{\frac{n}{2}}} = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}}. \quad (6.1.2)$$

Under any given hypothesis on  $\mu$  or  $\Sigma$ , the parameter space is reduced to a subspace  $\omega$  in  $\Omega$  or  $\omega \subset \Omega$ . For example, if  $H_o : \mu = \mu_o$  where  $\mu_o$  is a given vector, then the parameter space under this null hypothesis reduces to  $\omega = \{(\mu, \Sigma) \mid \mu = \mu_o, \Sigma > O\} \subset \Omega$ , “null hypothesis” being a technical term used to refer to the hypothesis being tested. The alternative hypothesis against which the null hypothesis is tested, is usually denoted by  $H_1$ . If  $\mu = \mu_o$  specifies  $H_o$ , then a natural alternative is  $H_1 : \mu \neq \mu_o$ . One of two things can happen when considering the maximum of the likelihood function under  $H_o$ . The overall maximum may occur in  $\omega$  or it may be attained outside of  $\omega$  but inside  $\Omega$ . If the null hypothesis  $H_o$  is actually true, then  $\omega$  and  $\Omega$  will coincide and the maxima in  $\omega$  and in  $\Omega$  will agree. If there are several local maxima, then the overall maximum or supremum is taken. The  $\lambda$ -criterion is defined as follows:

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L}, \quad 0 < \lambda \leq 1. \quad (6.1.3)$$

If the null hypothesis is true, then  $\lambda = 1$ . Accordingly, an observed value of  $\lambda$  that is close to 0 in a testing situation indicates that the null hypothesis  $H_o$  is incorrect and should then be rejected. Hence, the test criterion under the likelihood ratio test is to “reject  $H_o$  for  $0 < \lambda \leq \lambda_o$ ”, that is, for small values of  $\lambda$ , so that, under  $H_o$ , the coverage probability over this interval is equal to the significance level  $\alpha$  or the probability of rejecting  $H_o$  when  $H_o$  is true, that is,  $Pr\{0 < \lambda \leq \lambda_o \mid H_o\} = \alpha$  for a pre-assigned  $\alpha$ , which is also known as the size of the critical region or the size of the type-1 error. However, rejecting  $H_o$  when it is not actually true or when the alternative  $H_1$  is true is a correct decision

whose probability is known as the power of the test and written as  $1 - \beta$  where  $\beta$  is the probability of committing a type-2 error or the error of not rejecting  $H_o$  when  $H_o$  is not true. Thus we have

$$Pr\{0 < \lambda \leq \lambda_o | H_o\} = \alpha \text{ and } Pr\{0 < \lambda \leq \lambda_o | H_1\} = 1 - \beta. \quad (6.1.4)$$

When we preassign  $\alpha = 0.05$ , we are allowing a tolerance of 5% for the probability of committing the error of rejecting  $H_o$  when it is actually true and we say that we have a test at the 5% significance level. Usually, we set  $\alpha$  as 0.05 or 0.01. Alternatively, we can allow  $\alpha$  to vary and calculate what is known as the  $p$ -value when carrying out a test. Such is the principle underlying the likelihood ratio test, the resulting test criterion being referred to as the  $\lambda$ -criterion.

In the complex case, a tilde will be placed above  $\lambda$  and  $L$ , (6.1.3) and (6.1.4) remaining essentially the same:

$$\tilde{\lambda} = \frac{\sup_{\omega} \tilde{L}}{\sup_{\Omega} \tilde{L}}, \quad 0 < |\tilde{\lambda}| \leq 1, \quad (6.1a.1)$$

and

$$Pr\{0 < |\tilde{\lambda}| \leq \lambda_o | H_o\} = \alpha, \quad Pr\{0 < |\tilde{\lambda}| \leq \lambda_o | H_1\} = 1 - \beta \quad (6.1a.2)$$

where  $\alpha$  is the size or significance level of the test and  $1 - \beta$ , the power of the test.

## 6.2. Testing $H_o : \mu = \mu_o$ (Given) When $\Sigma$ is Known, the Real $N_p(\mu, \Sigma)$ Case

When  $\Sigma$  is known, the only parameter to estimate is  $\mu$ , its MLE being  $\bar{X}$ . Hence, the maximum in  $\Omega$  is the following:

$$\sup_{\Omega} L = \frac{e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}. \quad (6.2.1)$$

In this case,  $\mu$  is also specified under the null hypothesis  $H_o$ , so that there is no parameter to estimate. Accordingly,

$$\begin{aligned} \sup_{\omega} L &= \frac{e^{-\frac{1}{2} \sum_{j=1}^n (X_j - \mu_o)' \Sigma^{-1} (X_j - \mu_o)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} \\ &= \frac{e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S) - \frac{n}{2}(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}. \end{aligned} \quad (6.2.2)$$

Thus,

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = e^{-\frac{n}{2}(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o)}, \quad (6.2.3)$$

and small values of  $\lambda$  correspond to large values of  $\frac{n}{2}(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o)$ . When  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , it has already been established that  $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$ ,  $\Sigma > O$ . As well,  $n(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o)$  is the exponent in a  $p$ -variate real normal density under  $H_o$ , which has already been shown to have a real chisquare distribution with  $p$  degrees of freedom or

$$n(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o) \sim \chi_p^2.$$

Hence, the test criterion is

$$\text{Reject } H_o \text{ if } n(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o) \geq \chi_{p,\alpha}^2, \text{ with } Pr\{\chi_p^2 \geq \chi_{p,\alpha}^2\} = \alpha. \quad (6.2.4)$$

Under the alternative hypothesis, the distribution of the test statistic is a noncentral chisquare with  $p$  degrees of freedom and non-centrality parameter  $\lambda = \frac{n}{2}(\mu - \mu_o)' \Sigma^{-1}(\mu - \mu_o)$ .

**Example 6.2.1.** For example, suppose that we have a sample of size 5 from a population that has a trivariate normal distribution and let the significance level  $\alpha$  be 0.05. Let  $\mu_o$ , the hypothesized mean value vector specified by the null hypothesis, the known covariance matrix  $\Sigma$ , and the five observation vectors  $X_1, \dots, X_5$  be the following:

$$\mu_o = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix},$$

$$X_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad X_2 = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}, \quad X_3 = \begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix}, \quad X_4 = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}, \quad X_5 = \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix},$$

the inverse of  $\Sigma$  having been evaluated via elementary transformations. The sample average,  $\frac{1}{5}(X_1 + \dots + X_5)$  denoted by  $\bar{X}$ , is

$$\bar{X} = \frac{1}{5} \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix} \right\} = \frac{1}{5} \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix},$$

and

$$\bar{X} - \mu_o = \frac{1}{5} \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix}.$$

For testing  $H_o$ , the following test statistic has to be evaluated:

$$n(\bar{X} - \mu_o)' \Sigma^{-1} (\bar{X} - \mu_o) = \frac{5}{5^2} [4 \quad 4 \quad 8] \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix} = \frac{40}{5} = 8.$$

As per our criterion,  $H_o$  should be rejected if  $8 \geq \chi_{p,\alpha}^2$ . Since  $\chi_{p,\alpha}^2 = \chi_{3,0.05}^2 = 7.81$ , this critical value being available from a chisquare table,  $H_o : \mu = \mu_o$  should be rejected at the specified significance level. Moreover, in this case, the  $p$ -value is  $Pr\{\chi_3^2 \geq 8\} \approx 0.035$ , which can be evaluated by interpolation from the percentiles provided in a chi-square table or by making use of statistical packages such as R.

### 6.2.1. Paired variables and linear functions

Let  $Y_1, \dots, Y_k$  be  $p \times 1$  vectors having their own  $p$ -variate distributions which are not known. However, suppose that a certain linear function  $X = a_1 Y_1 + \dots + a_k Y_k$  is known to have a  $p$ -variate real Gaussian distribution with mean value vector  $E[X] = \mu$  and covariance matrix  $\text{Cov}(X) = \Sigma$ ,  $\Sigma > O$ , that is,  $X = a_1 Y_1 + \dots + a_k Y_k \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , where  $a_1, \dots, a_k$  are fixed known scalar constants. An example of this type is  $X = Y_1 - Y_2$  where  $Y_1$  consists of measurements on  $p$  attributes before subjecting those attributes to a certain process, such as administering a drug to a patient, and  $Y_2$  consists of the measurements on the same attributes after the process is completed. We would like to examine the difference  $Y_1 - Y_2$  to study the effect of the process on these characteristics. If it is reasonable to assume that this difference  $X = Y_1 - Y_2$  is  $N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , then we could test hypotheses on  $E[X] = \mu$ . When  $\Sigma$  is known, the general problem reduces to that discussed in Sect. 6.2. Assuming that we have iid variables on  $Y_1, \dots, Y_k$ , we would evaluate the corresponding values of  $X$ , which produces iid variables on  $X$ , that is, a simple random sample of size  $n$  from  $X = a_1 Y_1 + \dots + a_k Y_k$ . Thus, when  $\Sigma$  is known, letting  $u = n(\bar{X} - \mu_o)' \Sigma^{-1} (\bar{X} - \mu_o) \sim \chi_p^2$  where  $\bar{X}$  denote the sample average, the test would be carried out as follows at significance level  $\alpha$ :

$$\text{Reject } H_o : \mu = \mu_o \text{ (specified) when } u \geq \chi_{p,\alpha}^2, \text{ with } Pr\{\chi_p^2 \geq \chi_{p,\alpha}^2\} = \alpha, \quad (6.2.5)$$

the non-null distribution of the test statistic  $u$  being a non-central chisquare.

**Example 6.2.2.** Three variables  $x_1 =$  systolic pressure,  $x_2 =$  diastolic pressure and  $x_3 =$  weight are monitored after administering a drug for the reduction of all these  $p = 3$  variables. Suppose that a sample of  $n = 5$  randomly selected individuals are given the medication for one week. The following five pairs of observations on each of the three variables were obtained before and after the administration of the medication:

$$\begin{bmatrix} 150, 140 \\ 90, 90 \\ 70, 68 \end{bmatrix}, \begin{bmatrix} 180, 150 \\ 95, 90 \\ 75, 70 \end{bmatrix}, \begin{bmatrix} 160, 160 \\ 85, 80 \\ 70, 65 \end{bmatrix}, \begin{bmatrix} 140, 138 \\ 85, 90 \\ 70, 71 \end{bmatrix}, \begin{bmatrix} 130, 128 \\ 85, 85 \\ 75, 74 \end{bmatrix}.$$

Let  $X$  denote the difference, that is,  $X$  is equal to the reading before the medication was administered minus the reading after the medication could take effect. The observation vectors on  $X$  are then

$$X_1 = \begin{bmatrix} 150 - 140 \\ 90 - 90 \\ 70 - 68 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \\ 2 \end{bmatrix}, X_2 = \begin{bmatrix} 30 \\ 5 \\ 5 \end{bmatrix}, X_3 = \begin{bmatrix} 0 \\ 5 \\ 5 \end{bmatrix}, X_4 = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}.$$

In this case,  $X_1, \dots, X_5$  are observations on iid variables. We are going to assume that these iid variables are coming from a population whose distribution is  $N_3(\mu, \Sigma)$ ,  $\Sigma > O$ , where  $\Sigma$  is known. Let the sample average  $\bar{X} = \frac{1}{5}(X_1 + \dots + X_5)$ , the hypothesized mean value vector specified by the null hypothesis  $H_o : \mu = \mu_o$ , and the known covariance matrix  $\Sigma$  be as follows:

$$\bar{X} = \frac{1}{5} \begin{bmatrix} 44 \\ 5 \\ 12 \end{bmatrix}, \mu_o = \begin{bmatrix} 8 \\ 0 \\ 2 \end{bmatrix}, \Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Let us evaluate  $\bar{X} - \mu_o$  and  $n(\bar{X} - \mu_o)' \Sigma^{-1} (\bar{X} - \mu_o)$  which are needed for testing the hypothesis  $H_o : \mu = \mu_o$ :

$$\begin{aligned} \bar{X} - \mu_o &= \frac{1}{5} \begin{bmatrix} 44 \\ 5 \\ 12 \end{bmatrix} - \begin{bmatrix} 8 \\ 0 \\ 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 \\ 5 \\ 2 \end{bmatrix} \\ n(\bar{X} - \mu_o)' \Sigma^{-1} (\bar{X} - \mu_o) &= \frac{5}{5^2} [4 \ 5 \ 2] \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 2 \end{bmatrix} = 8.4. \end{aligned}$$

Let us test  $H_o$  at the significance level  $\alpha = 0.05$ . The critical value which can readily be found in a chisquare table is  $\chi_{p, \alpha}^2 = \chi_{3, 0.05}^2 = 7.81$ . As per our criterion, we reject  $H_o$  if  $8.4 \geq \chi_{p, \alpha}^2$ ; since  $8.4 > 7.81$ , we reject  $H_o$ . The  $p$ -value in this case is  $Pr\{\chi_p^2 \geq 8.4\} = Pr\{\chi_3^2 \geq 8.4\} \approx 0.04$ .

## 6.2.2. Independent Gaussian populations

Let  $Y_j \sim N_p(\mu_{(j)}, \Sigma_j)$ ,  $\Sigma_j > O$ ,  $j = 1, \dots, k$ , and let these  $k$  populations be independently distributed. Assume that a simple random sample of size  $n_j$  from  $Y_j$  is available for  $j = 1, \dots, k$ ; then these samples can be represented by the  $p$ -vectors  $Y_{jq}$ ,  $q = 1, \dots, n_j$ , which are iid as  $Y_{j1}$ , for  $j = 1, \dots, k$ . Consider a given linear

function  $X = a_1Y_1 + \cdots + a_kY_k$  where  $X$  is  $p \times 1$  and the  $Y_j$ 's are taken in a given order. Let  $U = a_1\bar{Y}_1 + \cdots + a_k\bar{Y}_k$  where  $\bar{Y}_j = \frac{1}{n_j} \sum_{q=1}^{n_j} Y_{jq}$  for  $j = 1, \dots, k$ . Then  $E[U] = a_1\mu_{(1)} + \cdots + a_k\mu_{(k)} = \mu$  (say), where  $a_1, \dots, a_k$  are given real scalar constants. The covariance matrix in  $U$  is  $\text{Cov}(U) = \frac{a_1^2}{n_1} \Sigma_1 + \cdots + \frac{a_k^2}{n_k} \Sigma_k = \frac{1}{n} \Sigma$  (say), where  $n$  is a symbol. Consider the problem of testing hypotheses on  $\mu$  when  $\Sigma$  is known or when  $a_j, \Sigma_j, j = 1, \dots, k$ , are known. Let  $H_o : \mu = \mu_o$  (specified), in the sense  $\mu_{(j)}$  is a known vector for  $j = 1, \dots, k$ , when  $\Sigma$  is known. Then, under  $H_o$ , all the parameters are known and the standardized  $U$  is observable, the test statistic being

$$\sum_{j=1}^k a_j n_j (\bar{Y}_j - \mu_{(j)})' \Sigma_j^{-1} (\bar{Y}_j - \mu_{(j)}) \sim \sum_{j=1}^k a_j \chi_p^{2(j)} \quad (6.2.6)$$

where  $\chi_p^{2(j)}$ ,  $j = 1, \dots, k$ , denote independent chisquares random variables, each having  $p$  degrees of freedom. However, since this is a linear function of independent chisquare variables, even the null distribution is complicated. Thus, only the case of two independent populations will be examined.

Consider the problem of testing the hypothesis  $\mu_1 - \mu_2 = \delta$  (a given vector) when there are two independent normal populations sharing a common covariance matrix  $\Sigma$  (known). Then  $U$  is  $U = \bar{Y}_1 - \bar{Y}_2$  with  $E[U] = \mu_1 - \mu_2 = \delta$  (given) under  $H_o$  and  $\text{Cov}(U) = (\frac{1}{n_1} + \frac{1}{n_2}) \Sigma = \frac{n_1+n_2}{n_1 n_2} \Sigma$ , the test statistic, denoted by  $v$ , being

$$v = \frac{n_1 n_2}{n_1 + n_2} (U - \delta)' \Sigma^{-1} (U - \delta) = \frac{n_1 n_2}{n_1 + n_2} (\bar{Y}_1 - \bar{Y}_2 - \delta)' \Sigma^{-1} (\bar{Y}_1 - \bar{Y}_2 - \delta) \sim \chi_p^2. \quad (6.2.7)$$

The resulting test criterion is

$$\text{Reject } H_o \text{ if the observed value of } v \geq \chi_{p, \alpha}^2 \text{ with } Pr\{\chi_p^2 \geq \chi_{p, \alpha}^2\} = \alpha. \quad (6.2.8)$$

**Example 6.2.3.** Let  $Y_1 \sim N_3(\mu_{(1)}, \Sigma)$  and  $Y_2 \sim N_3(\mu_{(2)}, \Sigma)$  represent independently distributed normal populations having a known common covariance matrix  $\Sigma$ . The null hypothesis is  $H_o : \mu_{(1)} - \mu_{(2)} = \delta$  where  $\delta$  is specified. Denote the observation vectors on  $Y_1$  and  $Y_2$  by  $Y_{1j}, j = 1, \dots, n_1$  and  $Y_{2j}, j = 1, \dots, n_2$ , respectively, and let the sample sizes be  $n_1 = 4$  and  $n_2 = 5$ . Let those observation vectors be

$$Y_{11} = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}, Y_{12} = \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix}, Y_{13} = \begin{bmatrix} 7 \\ 8 \\ 7 \end{bmatrix}, Y_{14} = \begin{bmatrix} 8 \\ 10 \\ 12 \end{bmatrix} \text{ and}$$

$$Y_{21} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, Y_{22} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}, Y_{23} = \begin{bmatrix} 7 \\ 10 \\ 8 \end{bmatrix}, Y_{24} = \begin{bmatrix} 6 \\ 5 \\ 6 \end{bmatrix}, Y_{25} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix},$$

and the common covariance matrix  $\Sigma$  be

$$\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Let the hypothesized vector under  $H_0 : \mu_{(1)} - \mu_{(2)} = \delta$  be  $\delta' = (1, 1, 2)$ . In order to test this null hypothesis, the following quantities must be evaluated:

$$\begin{aligned} \bar{Y}_1 &= \frac{1}{n_1}(Y_{11} + \cdots + Y_{1n_1}) = \frac{1}{4}(Y_{11} + Y_{12} + Y_{13} + Y_{14}), \\ \bar{Y}_2 &= \frac{1}{n_2}(Y_{21} + \cdots + Y_{2n_2}) = \frac{1}{5}(Y_{21} + \cdots + Y_{25}), \\ U &= \bar{Y}_1 - \bar{Y}_2, \quad v = \frac{n_1 n_2}{n_1 + n_2}(U - \delta)' \Sigma^{-1}(U - \delta). \end{aligned}$$

They are

$$\begin{aligned} \bar{Y}_1 &= \frac{1}{4} \left\{ \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} + \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} + \begin{bmatrix} 7 \\ 8 \\ 7 \end{bmatrix} + \begin{bmatrix} 8 \\ 10 \\ 12 \end{bmatrix} \right\} = \frac{1}{4} \begin{bmatrix} 22 \\ 24 \\ 27 \end{bmatrix}, \\ \bar{Y}_2 &= \frac{1}{5} \left\{ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 7 \\ 10 \\ 8 \end{bmatrix} + \begin{bmatrix} 6 \\ 5 \\ 6 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \right\} = \frac{1}{5} \begin{bmatrix} 20 \\ 20 \\ 21 \end{bmatrix}, \\ U &= \bar{Y}_1 - \bar{Y}_2 = \frac{1}{4} \begin{bmatrix} 22 \\ 24 \\ 27 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 20 \\ 20 \\ 21 \end{bmatrix} = \begin{bmatrix} 1.50 \\ 2.00 \\ 2.55 \end{bmatrix}. \end{aligned}$$

Then,

$$\begin{aligned} U - \delta &= \begin{bmatrix} 1.50 \\ 2.00 \\ 2.55 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.50 \\ 1.00 \\ 0.55 \end{bmatrix}. \\ v &= \frac{n_1 n_2}{n_1 + n_2} (U - \delta)' \Sigma^{-1} (U - \delta) \\ &= \frac{(4)(5)}{9} [0.50 \quad 1.00 \quad 0.55] \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0.50 \\ 1.00 \\ 0.55 \end{bmatrix} \\ &= 1.3275 \times \frac{20}{9} = 2.95. \end{aligned}$$



Let us test  $H_o$  at the significance level  $\alpha = 0.05$ . The critical value which is available from a chisquare table is  $\chi_{p, \alpha}^2 = \chi_{3, 0.05}^2 = 7.81$ . As per our criterion, we reject  $H_o$  if  $2.95 \geq \chi_{p, \alpha}^2$ ; however, since  $2.95 < 7.81$ , we cannot reject  $H_o$ . The  $p$ -value in this case is  $Pr\{\chi_p^2 \geq 2.95\} = Pr\{\chi_3^2 \geq 2.95\} \approx 0.096$ , which can be determined by interpolation.

### 6.2a. Testing $H_o : \mu = \mu_o$ (given) When $\Sigma$ is Known, Complex Gaussian Case

The derivation of the  $\lambda$ -criterion in the complex domain is parallel to that provided for the real case. In the parameter space,

$$\sup_{\Omega} \tilde{L} = \frac{e^{-\text{tr}(\Sigma^{-1} \tilde{S})}}{\pi^{np} |\det(\Sigma)|^n} \quad (6.2a.1)$$

and under  $H_o : \mu = \mu_o$ , a given vector,

$$\sup_{\omega} \tilde{L} = \frac{e^{-\text{tr}(\Sigma^{-1} \tilde{S}) - n(\bar{\bar{X}} - \mu_o)^* \Sigma^{-1} (\bar{\bar{X}} - \mu_o)}}{\pi^{np} |\det(\Sigma)|^n}. \quad (6.2a.2)$$

Accordingly,

$$\tilde{\lambda} = \frac{\sup_{\omega} \tilde{L}}{\sup_{\Omega} \tilde{L}} = e^{-n(\bar{\bar{X}} - \mu_o)^* \Sigma^{-1} (\bar{\bar{X}} - \mu_o)}. \quad (6.2a.3)$$

Here as well, small values of  $\tilde{\lambda}$  correspond to large values of  $\tilde{y} \equiv n(\bar{\bar{X}} - \mu_o)^* \Sigma^{-1} (\bar{\bar{X}} - \mu_o)$ , which has a real gamma distribution with the parameters ( $\alpha = p$ ,  $\beta = 1$ ) or a chisquare distribution with  $p$  degrees of freedom in the complex domain as described earlier so that  $2\tilde{y}$  has a real chisquare distribution having  $2p$  degrees of freedom. Thus, a real chisquare table can be utilized for testing the null hypothesis  $H_o$ , the criterion being

$$\text{Reject } H_o \text{ if } 2n(\bar{\bar{X}} - \mu_o)^* \Sigma^{-1} (\bar{\bar{X}} - \mu_o) \geq \chi_{2p, \alpha}^2, \text{ with } Pr\{\chi_{2p}^2 \geq \chi_{2p, \alpha}^2\} = \alpha. \quad (6.2a.4)$$

The test criteria as well as the decisions are parallel to those obtained for the real case in the situations of paired values and in the case of independent populations. Accordingly, such test criteria and associated decisions will not be further discussed.

**Example 6.2a.1.** Let  $p = 2$  and the  $2 \times 1$  complex vector  $\tilde{X} \sim \tilde{N}_2(\tilde{\mu}, \tilde{\Sigma})$ ,  $\tilde{\Sigma} = \tilde{\Sigma}^* > O$ , with  $\tilde{\Sigma}$  assumed to be known. Consider the null hypothesis  $H_o : \tilde{\mu} = \tilde{\mu}_o$  where  $\tilde{\mu}_o$  is specified. Let the known  $\tilde{\Sigma}$  and the specified  $\tilde{\mu}_o$  be the following where  $i = \sqrt{-1}$ :

$$\tilde{\mu}_o = \begin{bmatrix} 1+i \\ 1-2i \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix} \Rightarrow \tilde{\Sigma}^{-1} = \frac{1}{4} \begin{bmatrix} 3 & -(1+i) \\ -(1-i) & 2 \end{bmatrix}, \quad \tilde{\Sigma} = \tilde{\Sigma}^* > O.$$

Let the general  $\tilde{\mu}$  and general  $\tilde{X}$  be represented as follows for  $p = 2$ :

$$\tilde{\mu} = \begin{bmatrix} \mu_1 + i\nu_1 \\ \mu_2 + i\nu_2 \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} x_1 + iy_1 \\ x_2 + iy_2 \end{bmatrix}$$

so that, for the given  $\tilde{\Sigma}$ ,

$$\det(\tilde{\Sigma}) = (2)(3) - (1+i)(1-i) = 6 - (1^2 + 1^2) = 4 = \det(\tilde{X}^*) = |\det(\tilde{X})|.$$

The exponent of the general density for  $p = 2$ , excluding  $-1$ , is the form  $(\tilde{X} - \tilde{\mu})^* \tilde{\Sigma}^{-1} (\tilde{X} - \tilde{\mu})$ . Further,

$$[(\tilde{X} - \tilde{\mu})^* \tilde{\Sigma}^{-1} (\tilde{X} - \tilde{\mu})]^* = (\tilde{X} - \tilde{\mu})^* \tilde{\Sigma}^{-1} (\tilde{X} - \tilde{\mu})$$

since both  $\tilde{\Sigma}$  and  $\tilde{\Sigma}^{-1}$  are Hermitian. Thus, the exponent, which is  $1 \times 1$ , is real and negative definite. The explicit form, excluding  $-1$ , for  $p = 2$  and the given covariance matrix  $\tilde{\Sigma}$ , is the following:

$$Q = \frac{1}{4} \{ 3[(x_1 - \mu_1)^2 + (y_1 - \nu_1)^2] + 2[(x_2 - \mu_2)^2 + (y_2 - \nu_2)^2] \\ + 2[(x_1 - \mu_1)(x_2 - \mu_2) + (y_1 - \nu_1)(y_2 - \nu_2)] \},$$

and the general density for  $p = 2$  and this  $\tilde{\Sigma}$  is of the following form:

$$f(\tilde{X}) = \frac{1}{4\pi^2} e^{-Q}$$

where the  $Q$  is as previously given. Let the following be an observed sample of size  $n = 4$  from a  $\tilde{N}_2(\tilde{\mu}_o, \tilde{\Sigma})$  population whose associated covariance matrix  $\tilde{\Sigma}$  is as previously specified:

$$\tilde{X}_1 = \begin{bmatrix} 1 \\ 1+i \end{bmatrix}, \quad \tilde{X}_2 = \begin{bmatrix} 2-3i \\ i \end{bmatrix}, \quad \tilde{X}_3 = \begin{bmatrix} 2+i \\ 3 \end{bmatrix}, \quad \tilde{X}_4 = \begin{bmatrix} i \\ 2-i \end{bmatrix}.$$

Then,

$$\bar{\tilde{X}} = \frac{1}{4} \left\{ \begin{bmatrix} 1 \\ 1+i \end{bmatrix} + \begin{bmatrix} 2-3i \\ i \end{bmatrix} + \begin{bmatrix} 2+i \\ 3 \end{bmatrix} + \begin{bmatrix} i \\ 2-i \end{bmatrix} \right\} = \frac{1}{4} \begin{bmatrix} 5-i \\ 6+i \end{bmatrix}$$

$$\bar{\tilde{X}} - \tilde{\mu}_o = \frac{1}{4} \begin{bmatrix} 5-i \\ 6+i \end{bmatrix} - \begin{bmatrix} 1+i \\ 1-2i \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1-5i \\ 2+9i \end{bmatrix},$$

$$\begin{aligned}
& 2n(\bar{X} - \tilde{\mu}_o)' \tilde{\Sigma}^{-1} (\bar{X} - \tilde{\mu}_o) \\
&= (2)(4) \frac{1}{4^2} [1 + 5i \quad 2 - 9i] \frac{1}{4} \begin{bmatrix} 3 & -(1+i) \\ -(1-i) & 2 \end{bmatrix} \begin{bmatrix} 1 - 5i \\ 2 + 9i \end{bmatrix} \\
&= \frac{1}{8} \{3(1 + 5i)(1 - 5i) + 2(2 - 9i)(2 + 9i) - (1 + i)(1 + 5i)(2 + 9i) \\
&\quad - (1 - i)(2 - 9i)(1 - 5i)\} \\
&= \frac{1}{8} \{3 \times 26 + 2 \times 85 + 2 \times 62\} = 46.5.
\end{aligned}$$

Let us test the stated null hypothesis at the significance level  $\alpha = 0.05$ . Since  $\chi_{2p, \alpha}^2 = \chi_{4, 0.05}^2 = 9.49$  and  $46.5 > 9.49$ , we reject  $H_o$ . In this case, the  $p$ -value is  $Pr\{\chi_{2p}^2 \geq 46.5\} = Pr\{\chi_4^2 \geq 46.5\} \approx 0$ .

### 6.2.3. Test involving a subvector of a mean value vector when $\Sigma$ is known

Let the  $p \times 1$  vector  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ ,  $j = 1, \dots, n$ , and the  $X_j$ 's be independently distributed. Let the joint density of  $X_j$ ,  $j = 1, \dots, n$ , be denoted by  $L$ . Then, as was previously established,

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2}(X_j - \mu)' \Sigma^{-1} (X_j - \mu)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{n}{2} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} \quad (i)$$

where  $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$  and, letting  $\mathbf{X} = (X_1, \dots, X_n)$  of dimension  $p \times n$  and  $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$ ,  $S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$ . Let  $\bar{X}$ ,  $\Sigma^{-1}$  and  $\mu$  be partitioned as follows:

$$\bar{X} = \begin{bmatrix} \bar{X}^{(1)} \\ \bar{X}^{(2)} \end{bmatrix}, \quad \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}$$

where  $\bar{X}^{(1)}$  and  $\mu^{(1)}$  are  $r \times 1$ ,  $r < p$ , and  $\Sigma^{11}$  is  $r \times r$ . Consider the hypothesis  $\mu^{(1)} = \mu_o^{(1)}$  (specified) with  $\Sigma$  known. Thus, this hypothesis concerns only a subvector of the mean value vector, the population covariance matrix being assumed known. In the entire parameter space  $\Omega$ ,  $\mu$  is estimated by  $\bar{X}$  where  $\bar{X}$  is the maximum likelihood estimator (MLE) of  $\mu$ . The maximum of the likelihood function in the entire parameter space is then

$$\max_{\Omega} L = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}. \quad (ii)$$

Let us now determine the MLE of  $\mu^{(2)}$ , which is the only unknown quantity under the null hypothesis. To this end, we consider the following expansion:

$$\begin{aligned} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu) &= [(\bar{X}^{(1)} - \mu_o^{(1)})', (\bar{X}^{(2)} - \mu^{(2)})'] \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} \bar{X}^{(1)} - \mu_o^{(1)} \\ \bar{X}^{(2)} - \mu^{(2)} \end{bmatrix} \\ &= (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma^{11} (\bar{X}^{(1)} - \mu_o^{(1)}) + (\bar{X}^{(2)} - \mu^{(2)})' \Sigma^{22} (\bar{X}^{(2)} - \mu^{(2)}) \\ &\quad + 2(\bar{X}^{(2)} - \mu^{(2)})' \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}). \end{aligned} \quad (iii)$$

Noting that there are only two terms involving  $\mu^{(2)}$  in (iii), we have

$$\begin{aligned} \frac{\partial}{\partial \mu^{(2)}} \ln L = 0 &\Rightarrow 0 - 2\Sigma^{22}(\bar{X}^{(2)} - \mu^{(2)}) - 2\Sigma^{21}(\bar{X}^{(1)} - \mu_o^{(1)}) = 0 \\ &\Rightarrow \hat{\mu}^{(2)} = \bar{X}^{(2)} + (\Sigma^{22})^{-1} \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}). \end{aligned}$$

Then, substituting this MLE  $\hat{\mu}^{(2)}$  in the various terms in (iii), we have the following:

$$\begin{aligned} (\bar{X}^{(2)} - \hat{\mu}^{(2)})' \Sigma^{22} (\bar{X}^{(2)} - \hat{\mu}^{(2)}) &= (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}) \\ 2(\bar{X}^{(2)} - \hat{\mu}^{(2)})' \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}) &= -2(\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}) \Rightarrow \\ (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu) &= (\bar{X}^{(1)} - \mu_o^{(1)})' [\Sigma^{11} - \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21}] (\bar{X}^{(1)} - \mu_o^{(1)}) \\ &= (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)}), \end{aligned}$$

since, as established in Sect. 1.3,  $\Sigma_{11}^{-1} = \Sigma^{11} - \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21}$ . Thus, the maximum of  $L$  under the null hypothesis is given by

$$\max_{H_o} L = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{n}{2} (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)})}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}},$$

and the  $\lambda$ -criterion is then

$$\lambda = \frac{\max_{H_o} L}{\max_{\Omega} L} = e^{-\frac{n}{2} (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)})}. \quad (6.2.1)$$

Hence, we reject  $H_o$  for small values of  $\lambda$  or for large values of  $n(\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)}) \sim \chi_r^2$  since the expected value and covariance matrix of  $\bar{X}^{(1)}$  are respectively  $\mu_o^{(1)}$  and  $\Sigma_{11}/n$ . Accordingly, the criterion can be enunciated as follows:

$$\text{Reject } H_o : \mu^{(1)} = \mu_o^{(1)} \text{ (given) if } u \equiv n(\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)}) \geq \chi_{r, \alpha}^2 \quad (6.2.2)$$

with  $Pr\{\chi_r^2 \geq \chi_{r, \alpha}^2\} = \alpha$ . In the complex Gaussian case, the corresponding  $2\tilde{u}$  will be distributed as a real chisquare random variable having  $2r$  degrees of freedom; thus, the

criterion will consist of rejecting the corresponding null hypothesis whenever the observed value of  $2\tilde{u} \geq \chi_{2r, \alpha}^2$ .

**Example 6.2.4.** Let the  $4 \times 1$  vector  $X$  have a real normal distribution  $N_4(\mu, \Sigma)$ ,  $\Sigma > O$ . Consider the hypothesis that part of  $\mu$  is specified. For example, let the hypothesis  $H_o$  and  $\Sigma$  be the following:

$$H_o : \mu = \mu_o = \begin{bmatrix} 1 \\ -1 \\ \mu_3 \\ \mu_4 \end{bmatrix}, \Sigma = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix} = \Sigma' > O, X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \equiv \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}.$$

Since we are specifying the first two parameters in  $\mu$ , the hypothesis can be tested by computing the distribution of  $X^{(1)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ . Observe that  $X^{(1)} \sim N_2(\mu^{(1)}, \Sigma_{11})$ ,  $\Sigma_{11} > O$  where

$$\mu^{(1)} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \mu_o^{(1)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, H_o : \mu^{(1)} = \mu_o^{(1)}, \Sigma_{11} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \Rightarrow \Sigma_{11}^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Let the observed vectors from the original  $N_4(\mu, \Sigma)$  population be

$$X_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 4 \end{bmatrix}, X_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ 2 \end{bmatrix}, X_3 = \begin{bmatrix} 0 \\ 2 \\ 3 \\ 4 \end{bmatrix}, X_4 = \begin{bmatrix} 2 \\ 1 \\ -1 \\ 3 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 4 \end{bmatrix}.$$

Then the observations corresponding to the subvector  $X^{(1)}$ , denoted by  $X_j^{(1)}$ , are the following:

$$X_1^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, X_2^{(1)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, X_3^{(1)} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, X_4^{(1)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, X_5^{(1)} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

In this case, the sample size  $n = 5$  and the sample mean, denoted by  $\bar{X}^{(1)}$ , is

$$\begin{aligned} \bar{X}^{(1)} &= \frac{1}{5} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\} = \frac{1}{5} \begin{bmatrix} 4 \\ 3 \end{bmatrix} \Rightarrow \\ \bar{X}^{(1)} - \mu_o^{(1)} &= \frac{1}{5} \begin{bmatrix} 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} -1 \\ 8 \end{bmatrix}. \end{aligned}$$

Therefore

$$\begin{aligned} n(\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)}) &= \frac{5}{5^2} [-1 \quad 8] \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 8 \end{bmatrix} \\ &= \frac{1}{15} (146) = 9.73. \end{aligned}$$

If  $9.73 > \chi_{2,\alpha}^2$ , then we would reject  $H_o^{(1)} : \mu^{(1)} = \mu_o^{(1)}$ . Let us test this hypothesis at the significance level  $\alpha = 0.01$ . Since  $\chi_{2,0.01}^2 = 9.21$ , we reject the null hypothesis. In this instance, the  $p$ -value, which can be determined from a chisquare table, is  $Pr\{\chi_2^2 \geq 9.73\} \approx 0.007$ .

#### 6.2.4. Testing $\mu_1 = \dots = \mu_p$ , with $\Sigma$ known, real Gaussian case

Let  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ ,  $j = 1, \dots, n$ , and the  $X_j$  be independently distributed. Letting  $\mu' = (\mu_1, \dots, \mu_p)$ , consider the hypothesis

$$H_o : \mu_1 = \mu_2 = \dots = \mu_p = \nu,$$

where  $\nu$ , the common  $\mu_j$  is unknown. This implies that  $\mu_i - \mu_j = 0$  for all  $i$  and  $j$ . Consider the  $p \times 1$  vector  $J$  of unities,  $J' = (1, \dots, 1)$  and then take any non-null vector that is orthogonal to  $J$ . Let  $A$  be such a vector so that  $A'J = 0$ . Actually,  $p - 1$  linearly independent such vectors are available. For example, if  $p$  is even, then take  $1, -1, \dots, 1, -1$  as the elements of  $A$  and, when  $p$  is odd, one can start with  $1, -1, \dots, 1, -1$  and take the last three elements as  $1, -2, 1$ , or the last element as  $0$ , that is,

$$J = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ -1 \\ 1 \\ -1 \end{bmatrix} \quad \text{for } p \text{ even} \quad \text{and} \quad A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ 1 \\ -2 \\ 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 \\ -1 \\ \vdots \\ -1 \\ 0 \end{bmatrix} \quad \text{for } p \text{ odd}.$$

When the last element of the vector  $A$  is zero, we are simply ignoring the last element in  $X_j$ . Let the  $p \times 1$  vector  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ ,  $j = 1, \dots, n$ , and the  $X_j$ 's be independently distributed. Let the scalar  $y_j = A'X_j$  and the  $1 \times n$  vector  $Y = (y_1, \dots, y_n) = (A'X_1, \dots, A'X_n) = A'(X_1, \dots, X_n) = A'\mathbf{X}$ , where the  $p \times n$  matrix  $\mathbf{X} = (X_1, \dots, X_n)$ . Let  $\bar{y} = \frac{1}{n}(y_1 + \dots + y_n) = A'\frac{1}{n}(X_1 + \dots + X_n) = A'\bar{X}$ . Then,  $\sum_{j=1}^n (y_j - \bar{y})(y_j - \bar{y})' = A' \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})' A$  where  $\sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})' = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = S =$

the sample sum of products matrix in the  $X_j$ 's, where  $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$  the  $p \times n$  matrix whose columns are all equal to  $\bar{X}$ . Thus, one has  $\sum_{j=1}^n (y_j - \bar{y})^2 = A'SA$ . Consider the hypothesis  $\mu_1 = \dots = \mu_p = \nu$ . Then,  $A'\mu = \nu A'J = \nu 0 = 0$  under  $H_o$ . Since  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , we have  $y_j \sim N_1(A'\mu, A'\Sigma A)$ ,  $A'\Sigma A > 0$ . Under  $H_o$ ,  $y_j \sim N_1(0, A'\Sigma A)$ ,  $j = 1, \dots, n$ , the  $y_j$ 's being independently distributed. Consider the joint density of  $y_1, \dots, y_n$ , denoted by  $L$ :

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2A'\Sigma A}(y_j - A'\mu)^2}}{(2\pi)^{\frac{1}{2}}[A'\Sigma A]^{\frac{1}{2}}}. \quad (i)$$

Since  $\Sigma$  is known, the only unknown quantity in  $L$  is  $\mu$ . Differentiating  $\ln L$  with respect to  $\mu$  and equating the result to a null vector, we have

$$\sum_{j=1}^n (y_j - A'\hat{\mu}) = 0 \Rightarrow \sum_{j=1}^n y_j - nA'\hat{\mu} = 0 \Rightarrow \bar{y} - A'\hat{\mu} = 0 \Rightarrow A'(\bar{X} - \hat{\mu}) = 0.$$

However, since  $A$  is a fixed known vector and the equation holds for arbitrary  $\bar{X}$ ,  $\hat{\mu} = \bar{X}$ . Hence the maximum of  $L$ , in the entire parameter space  $\Omega = \mu$ , is the following:

$$\max_{\Omega} L = \frac{e^{-\frac{1}{2A'\Sigma A} \sum_{j=1}^n [A'(X_j - \bar{X})]^2}}{(2\pi)^{\frac{np}{2}} [A'\Sigma A]^{\frac{n}{2}}} = \frac{e^{-\frac{1}{2A'\Sigma A} A'SA}}{(2\pi)^{\frac{np}{2}} [A'\Sigma A]^{\frac{n}{2}}}. \quad (ii)$$

Now, noting that under  $H_o$ ,  $A'\mu = 0$ , we have

$$\max_{H_o} L = \frac{e^{-\frac{1}{2A'\Sigma A} \sum_{j=1}^n A'X_j X_j' A'}}{(2\pi)^{\frac{np}{2}} [A'\Sigma A]^{\frac{n}{2}}}. \quad (iii)$$

From (i) to (iii), the  $\lambda$ -criterion is as follows, observing that  $A'(\sum_{j=1}^n X_j X_j')A = \sum_{j=1}^n A'(X_j - \bar{X})(X_j - \bar{X})'A + nA'(\bar{X}\bar{X}'A) = A'SA + nA'\bar{X}\bar{X}'A$ :

$$\lambda = e^{-\frac{n}{2A'\Sigma A} A'\bar{X}\bar{X}'A}. \quad (6.2.3)$$

But since  $\sqrt{\frac{n}{A'\Sigma A}} A'\bar{X} \sim N_1(0, 1)$  under  $H_o$ , we may test this null hypothesis either by using the standard normal variable or a chisquare variable as  $\frac{n}{A'\Sigma A} A'\bar{X}\bar{X}'A \sim \chi_1^2$  under  $H_o$ . Accordingly, the criterion consists of rejecting  $H_o$

$$\text{when } \left| \sqrt{\frac{n}{A'\Sigma A}} A'\bar{X} \right| \geq z_{\frac{\alpha}{2}}, \text{ with } Pr\{z \geq z_{\beta}\} = \beta, z \sim N_1(0, 1)$$

or

$$\text{when } u \equiv \frac{n}{A'\Sigma A} (A'\bar{X}\bar{X}'A) \geq \chi_{1, \alpha}^2, \text{ with } Pr\{\chi_1^2 \geq \chi_{1, \alpha}^2\} = \alpha, u \sim \chi_1^2. \quad (6.2.4)$$

**Example 6.2.5.** Consider a 4-variate real Gaussian vector  $X \sim N_4(\mu, \Sigma)$ ,  $\Sigma > O$  with  $\Sigma$  as specified in Example 6.2.4 and the null hypothesis that the individual components of the mean value vector  $\mu$  are all equal, that is,

$$\Sigma = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}, \quad H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4 \equiv v \text{ (say)}, \quad \text{with } \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}.$$

Let  $L$  be a  $4 \times 1$  constant vector such that  $L' = (1, -1, 1, -1)$ . Then, under  $H_o$ ,  $L'\mu = 0$  and  $u = L'X$  is univariate normal; more specifically,  $u \sim N_1(0, L'\Sigma L)$  where

$$L'\Sigma L = [1 \quad -1 \quad 1 \quad -1] \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} = 7 \Rightarrow u \sim N_1(0, 7).$$

Let the observation vectors be the same as those used in Example 6.2.4 and let  $u_j = L'X_j$ ,  $j = 1, \dots, 5$ . Then, the five independent observations from  $u \sim N_1(0, 7)$  are the following:

$$u_1 = L'X_1 = [1 \quad -1 \quad 1 \quad -1] \begin{bmatrix} 1 \\ 0 \\ 2 \\ 4 \end{bmatrix} = -1, \quad u_2 = L' \begin{bmatrix} -1 \\ 1 \\ 1 \\ 2 \end{bmatrix} = -3, \quad u_3 = L' \begin{bmatrix} 0 \\ 2 \\ 3 \\ 4 \end{bmatrix} = -3,$$

$$u_4 = L' \begin{bmatrix} 2 \\ 1 \\ -1 \\ 3 \end{bmatrix} = -3, \quad u_5 = L' \begin{bmatrix} 2 \\ -1 \\ 0 \\ 4 \end{bmatrix} = -1,$$

the average  $\bar{u} = \frac{1}{5}(u_1 + \dots + u_5) = \frac{1}{5}(-1 - 3 - 3 - 3 - 1)$  being equal to  $-\frac{11}{5}$ . Then, the standardized sample mean  $z = \frac{\sqrt{n}}{\sigma_u}(\bar{u} - 0) \sim N_1(0, 1)$ . Let us test the null hypothesis at the significance level  $\alpha = 0.05$ . Referring to a  $N_1(0, 1)$  table, the required critical value, denoted by  $z_{\frac{\alpha}{2}} = z_{0.025}$  is 1.96. Therefore, we reject  $H_o$  in favor of the alternative hypothesis that at least two components of  $\mu$  are unequal at significance level  $\alpha$  if the observed value of

$$|z| = \left| \frac{\sqrt{n}}{\sigma_u}(\bar{u} - 0) \right| \geq 1.96.$$



Since the observed value of  $|z|$  is  $|\frac{\sqrt{5}}{\sqrt{7}}(-\frac{7}{5} - 0)| = \sqrt{1.4} = 1.18$  is less than 1.96, we do not reject  $H_o$  at the 5% significance level. Letting  $z \sim N_1(0, 1)$ , the  $p$ -value in this case is  $Pr\{|z| \geq 1.18\} = 0.238$ , this quantile being available from a standard normal table.

In the complex case, proceeding in a parallel manner to the real case, the lambda criterion will be the following:

$$\tilde{\lambda} = e^{-\frac{n}{A^* \tilde{\Sigma} A} A^* \tilde{X} \tilde{X}^* A} \quad (6.2a.5)$$

where an asterisk indicates the conjugate transpose. Letting  $\tilde{u} = \frac{2n}{A^* \tilde{\Sigma} A} (A^* \tilde{X} \tilde{X}^* A)$ , it can be shown that under  $H_o$ ,  $\tilde{u}$  is distributed as a real chisquare random variable having 2 degrees of freedom. Accordingly, the criterion will be as follows:

$$\text{Reject } H_o \text{ if the observed } \tilde{u} \geq \chi_{2, \alpha}^2 \text{ with } Pr\{\chi_2^2 \geq \chi_{2, \alpha}^2\} = \alpha. \quad (6.2a.6)$$

**Example 6.2a.2.** When  $p > 2$ , the computations become quite involved in the complex case. Thus, we will let  $p = 2$  and consider the bivariate complex  $\tilde{N}_2(\tilde{\mu}, \tilde{\Sigma})$  distribution that was specified in Example 6.2a.1, assuming that  $\tilde{\Sigma}$  is as given therein, the same set of observations being utilized as well. In this case, the null hypothesis is  $H_o : \tilde{\mu}_1 = \tilde{\mu}_2$ , the parameters and sample average being

$$\tilde{\mu} = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix}, \quad \tilde{X} = \frac{1}{4} \begin{bmatrix} 5-i \\ 6+i \end{bmatrix}.$$

Letting  $L' = (1, -1)$ ,  $L' \tilde{\mu} = 0$  under  $H_o$ , and

$$\tilde{u} = L' \tilde{X} = \frac{1}{4} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 5-i \\ 6+i \end{bmatrix} = -\frac{1}{4}(1+2i); \quad (L' \tilde{X})^* (L' \tilde{X}) = \frac{1}{16}(1-2i)(1+2i) = \frac{5}{16};$$

$$L' \tilde{\Sigma} L = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 3; \quad v = \frac{2n}{L' \tilde{\Sigma} L} [(L' \tilde{X})^* (L' \tilde{X})] = \frac{8}{3} \times \frac{5}{16} = \frac{5}{6}.$$

The criterion consists of rejecting  $H_o$  if the observed value of  $v \geq \chi_{2, \alpha}^2$ . Letting the significance level of the test be  $\alpha = 0.05$ , the critical value is  $\chi_{2, 0.05}^2 = 5.99$ , which is readily available from a chisquare table. The observed value of  $v$  being  $\frac{5}{6} < 5.99$ , we do not reject  $H_o$ . In this case, the  $p$ -value is  $Pr\{\chi_2^2 \geq \frac{5}{6}\} \approx 0.318$ .

### 6.2.5. Likelihood ratio criterion for testing $H_o : \mu_1 = \dots = \mu_p, \Sigma$ known

Consider again,  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ ,  $j = 1, \dots, n$ , with the  $X_j$ 's being independently distributed and  $\Sigma$ , assumed known. Letting the joint density of  $X_1, \dots, X_n$  be denoted by  $L$ , then, as determined earlier,

$$L = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{n}{2} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} \quad (i)$$

where  $n$  is the sample size and  $S$  is the sample sum of products matrix. In the entire parameter space

$$\Omega = \{(\mu, \Sigma) \mid \Sigma > O \text{ known, } \mu' = (\mu_1, \dots, \mu_p)\},$$

the MLE of  $\mu$  is  $\bar{X}$  = the sample average. Then

$$\max_{\Omega} L = \frac{e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}. \quad (ii)$$

Consider the following hypothesis on  $\mu' = (\mu_1, \dots, \mu_p)$ :

$$H_o : \mu_1 = \dots = \mu_p = \nu, \nu \text{ is unknown.}$$

Then, the MLE of  $\mu$  under  $H_o$  is  $\hat{\mu} = J\hat{\nu} = J\frac{1}{p}J'\bar{X}$ ,  $J' = (1, \dots, 1)$ . This  $\hat{\nu}$  is in fact the sum of all observations on all components of  $\bar{X}_j$ ,  $j = 1, \dots, n$ , divided by  $np$ , which is identical to the sum of all the coordinates of  $\bar{X}$  divided by  $p$  or  $\hat{\mu} = \frac{1}{p}JJ'\bar{X}$ . In order to evaluate the maximum of  $L$  under  $H_o$ , it suffices to substitute  $\hat{\mu}$  to  $\mu$  in (i). Accordingly, the  $\lambda$ -criterion is

$$\lambda = \frac{\max_{H_o} L}{\max_{\Omega} L} = e^{-\frac{n}{2}(\bar{X} - \hat{\mu})' \Sigma^{-1}(\bar{X} - \hat{\mu})}. \quad (6.2.5)$$

Thus, we reject  $H_o$  for small values of  $\lambda$  or for large values of  $w \equiv n(\bar{X} - \hat{\mu})' \Sigma^{-1}(\bar{X} - \hat{\mu})$ . Let us determine the distribution of  $w$ . First, note that

$$\bar{X} - \hat{\mu} = \bar{X} - \frac{1}{p}JJ'\bar{X} = (I_p - \frac{1}{p}JJ')\bar{X},$$

and let

$$\begin{aligned} w &= n(\bar{X} - \hat{\mu})' \Sigma^{-1}(\bar{X} - \hat{\mu}) = n\bar{X}'(I - \frac{1}{p}JJ')\Sigma^{-1}(I - \frac{1}{p}JJ')\bar{X} \quad (iii) \\ &= (\bar{X} - \mu)'(I - \frac{1}{p}JJ')\Sigma^{-1}(I - \frac{1}{p}JJ')(\bar{X} - \mu) \end{aligned}$$

since  $J'(I - \frac{1}{p}JJ') = O$ ,  $\mu = \nu J$  being the true mean value of the  $N_p(\mu, \Sigma)$  distribution. Observe that  $\sqrt{n}(\bar{X} - \mu) \sim N_p(O, \Sigma)$ ,  $\Sigma > O$ , and that  $\frac{1}{p}JJ'$  is idempotent. Since  $I - \frac{1}{p}JJ'$  is also idempotent and its rank is  $p - 1$ , there exists an orthonormal matrix  $P$ ,  $PP' = I$ ,  $P'P = I$ , such that

$$I - \frac{1}{p}JJ' = P' \begin{bmatrix} I_{p-1} & O \\ O' & 0 \end{bmatrix} P.$$

Letting  $U = P\sqrt{n}(\bar{X} - \hat{\mu})$ , with  $U' = (u_1, \dots, u_{p-1}, u_p)$ ,  $U \sim N_p(O, P\Sigma P')$ . Now, on noting that

$$U' \begin{bmatrix} I_{p-1} & O \\ O' & 0 \end{bmatrix} = (u_1, \dots, u_{p-1}, 0),$$

we have

$$n(\bar{X} - \hat{\mu})' \Sigma^{-1} (\bar{X} - \hat{\mu}) = [U_1', 0] P \Sigma^{-1} P' \begin{bmatrix} U_1 \\ 0 \end{bmatrix} = U_1' B^{-1} U_1, \quad U_1' = (u_1, \dots, u_{p-1}),$$

$B$  being the covariance matrix associated with  $U_1$ , so that  $U_1 \sim N_{p-1}(O, B)$ ,  $B > O$ . Thus,  $U_1' B^{-1} U_1 \sim \chi_{p-1}^2$ , a real scalar chisquare random variable having  $p - 1$  degrees of freedom. Hence, upon evaluating

$$w = \bar{X}' \left( I - \frac{1}{p} J J' \right) \Sigma^{-1} \left( I - \frac{1}{p} J J' \right) \bar{X},$$

one would reject  $H_o : \mu_1 = \dots = \mu_p = \nu$ ,  $\nu$  unknown, whenever the observed value of

$$w \geq \chi_{p-1, \alpha}^2, \quad \text{with } Pr\{\chi_{p-1}^2 \geq \chi_{p-1, \alpha}^2\} = \alpha. \quad (6.2.6)$$

Observe that the degrees of freedom of this chisquare variable, that is,  $p - 1$ , coincides with the number of parameters being restricted by  $H_o$ .

**Example 6.2.6.** Consider the trivariate real Gaussian population  $X \sim N_3(\mu, \Sigma)$ ,  $\Sigma > O$ , as already specified in Example 6.2.1 with the same  $\Sigma$  and the same observed sample vectors for testing  $H_o : \mu' = (\nu, \nu, \nu)$ , namely,

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \bar{X} = \frac{1}{5} \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

The following test statistic has to be evaluated for  $p = 3$ :

$$w = \bar{X}' \left( I - \frac{1}{p} J J' \right) \Sigma^{-1} \left( I - \frac{1}{p} J J' \right) \bar{X}, \quad J' = (1, 1, 1).$$

We have to evaluate the following quantities in order to determine the value of  $w$ :

$$\begin{aligned} \frac{1}{3}JJ'\Sigma^{-1} &= \frac{1}{3} \begin{bmatrix} \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 1 & 0 \end{bmatrix}, \quad \Sigma^{-1}\frac{1}{3}JJ' = \frac{1}{3} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \\ \frac{1}{3}JJ'\Sigma^{-1}\frac{1}{3}JJ' &= \frac{1}{9} \begin{bmatrix} \frac{3}{2} & \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & \frac{3}{2} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \\ \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 1 & 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \\ + \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} &= \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{2}{3} & -\frac{7}{6} \\ 0 & -\frac{7}{6} & \frac{7}{6} \end{bmatrix} = (I - \frac{1}{3}JJ')\Sigma^{-1}(I - \frac{1}{3}JJ'). \end{aligned}$$

Thus,

$$\begin{aligned} w &= \bar{X}'(I - \frac{1}{3}JJ')\Sigma^{-1}(I - \frac{1}{3}JJ')\bar{X}, \quad J' = (1, 1, 1) \\ &= \frac{1}{5^2} [9 \quad 4 \quad 3] \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{2}{3} & -\frac{7}{6} \\ 0 & -\frac{7}{6} & \frac{7}{6} \end{bmatrix} \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix} \\ &= \frac{1}{5^2} \left[ 9^2 \frac{1}{3} + 4^2 \frac{3}{2} + 3^2 \frac{7}{6} - \frac{2}{3} (9)(4) - \frac{14}{6} (4)(3) \right] \\ &= 0.38. \end{aligned}$$

We reject  $H_0$  whenever  $w \geq \chi_{p-1, \alpha}^2$ . Letting the significance level be  $\alpha = 0.05$ , the tabulated critical value is  $\chi_{p-1, \alpha}^2 = \chi_{2, 0.05}^2 = 5.99$ , and since  $0.38 < 5.99$ , we do not reject the null hypothesis. In this instance, the  $p$ -value is  $Pr\{\chi_2^2 \geq 0.38\} \approx 0.32$ .

### 6.3. Testing $H_0 : \mu = \mu_0$ (given) When $\Sigma$ is Unknown, Real Gaussian Case

In this case, both  $\mu$  and  $\Sigma$  are unknown in the entire parameter space  $\Omega$ ; however,  $\mu = \mu_0$  known while  $\Sigma$  is still unknown in the subspace  $\omega$ . The MLE under  $\Omega$  is the same as that obtained in Sect. 6.1.1., that is,

$$\sup_{\Omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}}. \quad (6.3.1)$$

When  $\mu = \mu_o$ ,  $\Sigma$  is estimated by  $\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^n (X_j - \mu_o)(X_j - \mu_o)'$ . As shown in Sect. 3.5,  $\hat{\Sigma}$  can be reexpressed as follows:

$$\begin{aligned}\hat{\Sigma} &= \frac{1}{n} \sum_{j=1}^n (X_j - \mu_o)(X_j - \mu_o) = \frac{1}{n} S + (\bar{X} - \mu_o)(\bar{X} - \mu_o)' \\ &= \frac{1}{n} [S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'].\end{aligned}$$

Then, under the null hypothesis, we have

$$\sup_{\omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'|^{\frac{n}{2}}}. \quad (6.3.2)$$

Thus,

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = \frac{|S|^{\frac{n}{2}}}{|S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'|^{\frac{n}{2}}}.$$

On applying results on the determinants of partitioned matrices which were obtained in Sect. 1.3, we have the following equivalent representations of the denominator:

$$\begin{aligned}\left| \begin{bmatrix} S & n(\bar{X} - \mu_o) \\ -(\bar{X} - \mu_o)' & 1 \end{bmatrix} \right| &= [1] |S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'| \\ &= |S| |1 + n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)|,\end{aligned}$$

that is,

$$|S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'| = |S| [1 + n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)],$$

which yields the following simplified representation of the likelihood ratio statistic:

$$\lambda = \frac{1}{[1 + n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)]^{\frac{n}{2}}}. \quad (6.3.3)$$

Small values of  $\lambda$  correspond to large values of  $u \equiv n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)$ , which is connected to Hotelling's  $T_n^2$  statistic. Hence the criterion is the following: "Reject  $H_o$  for large values of  $u$ ". The distribution of  $u$  can be derived by making use of the independence of the sample mean and sample sum of products matrix and the densities of these quantities. An outline of the derivation is provided in the next subsection.

### 6.3.1. The distribution of the test statistic

Let us examine the distribution of  $u = n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu)$ . We have already established in Theorems 3.5.3, that  $S$  and  $\bar{X}$  are independently distributed in the case of a real  $p$ -variate nonsingular Gaussian  $N_p(\mu, \Sigma)$  population. It was also determined in Corollary 3.5.1 that the distribution of the sample average  $\bar{X}$  is a  $p$ -variate real Gaussian vector with the parameters  $\mu$  and  $\frac{1}{n}\Sigma$ ,  $\Sigma > O$  and in the continuing discussion, it is shown that the distribution of  $S$  is a matrix-variate Wishart with  $m = n - 1$  degrees of freedom, where  $n$  is the sample size and parameter matrix  $\Sigma > O$ . Hence the joint density of  $S$  and  $\bar{X}$ , denoted by  $f(S, \bar{X})$ , is the product of the marginal densities. Letting  $\Sigma = I$ , this joint density is given by

$$f(S, \bar{X}) = \frac{n^{\frac{p}{2}}}{(2\pi)^{\frac{p}{2}} 2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} |S|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(S) - \frac{n}{2}\text{tr}((\bar{X} - \mu)(\bar{X} - \mu)')}, \quad m = n - 1. \quad (i)$$

Note that it is sufficient to consider the case  $\Sigma = I$ . Due to the presence of  $S^{-1}$  in  $u = (\bar{X} - \mu)'S^{-1}(\bar{X} - \mu)$ , the effect of any scaling matrix on  $X_j$  will disappear. If  $X_j$  goes to  $A^{\frac{1}{2}}X_j$  for any constant positive definite matrix  $A$  then  $S^{-1}$  will go to  $A^{-\frac{1}{2}}S^{-1}A^{-\frac{1}{2}}$  and thus  $u$  will be free of  $A$ .

Letting  $Y = S^{-\frac{1}{2}}(\bar{X} - \mu)$  for fixed  $S$ ,  $Y \sim N_p(O, S^{-1}/n)$ , so that the conditional density of  $Y$ , given  $S$ , is

$$g(Y|S) = \frac{n^{\frac{p}{2}} |S|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} e^{-\frac{n}{2}\text{tr}(SY Y')}.$$

Thus, the joint density of  $S$  and  $Y$ , denoted by  $f_1(S, Y)$ , is

$$f_1(S, Y) = \frac{n^{\frac{p}{2}}}{(2\pi)^{\frac{p}{2}} 2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} |S|^{\frac{m+1}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(S[I + nYY'])}, \quad m = n - 1. \quad (ii)$$

On integrating out  $S$  from (ii) by making use of a matrix-variate gamma integral, we obtain the following marginal density of  $Y$ , denoted by  $f_2(Y)$ :

$$f_2(Y)dY = \frac{n^{\frac{p}{2}}}{(\pi)^{\frac{p}{2}}} \frac{\Gamma_p(\frac{m+1}{2})}{\Gamma_p(\frac{m}{2})} |I + nYY'|^{-\frac{(m+1)}{2}} dY, \quad m = n - 1. \quad (iii)$$

However,  $|I + nYY'| = 1 + nY'Y$ , which can be established by considering two representations of the determinant

$$\begin{vmatrix} I & -\sqrt{n}Y \\ \sqrt{n}Y' & 1 \end{vmatrix},$$

similarly to what was done in Sect. 6.3 to obtain the likelihood ratio statistic given in (6.3.3). As well, it can easily be shown that

$$\frac{\Gamma_p(\frac{m+1}{2})}{\Gamma_p(\frac{m}{2})} = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m+1}{2} - \frac{p}{2})}$$

by expanding the matrix-variate gamma functions. Now, letting  $s = Y'Y$ , it follows from Theorem 4.2.3 that  $dY = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} s^{\frac{p}{2}-1} ds$ . Thus, the density of  $s$ , denoted by  $f_3(s)$ , is

$$f_3(s) ds = \frac{n^{\frac{p}{2}} \Gamma(\frac{m+1}{2})}{\Gamma(\frac{m+1}{2} - \frac{p}{2}) \Gamma(\frac{p}{2})} s^{\frac{p}{2}-1} (1 + ns)^{-(\frac{m+1}{2})} ds \quad (6.3.4)$$

$$= \frac{n^{\frac{p}{2}} \Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2} - \frac{p}{2}) \Gamma(\frac{p}{2})} s^{\frac{p}{2}-1} (1 + ns)^{-(\frac{n}{2})} ds, \quad m = n - 1, \quad (6.3.5)$$

for  $n = p + 1, p + 2, \dots, 0 \leq s < \infty$ , and zero elsewhere. It can then readily be seen from (6.3.5) that  $ns = nY'Y = n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu) = u$  is distributed as a real scalar type-2 beta random variable whose parameters are  $(\frac{p}{2}, \frac{n}{2} - \frac{p}{2})$ ,  $n = p + 1, \dots$ . Thus, the following result:

**Theorem 6.3.1.** *Consider a real  $p$ -variate normal population  $N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , and a simple random sample of size  $n$  from this normal population,  $X_j \sim N_p(\mu, \Sigma)$ ,  $j = 1, \dots, n$ , the  $X_j$ 's being independently distributed. Let the  $p \times n$  matrix  $\mathbf{X} = (X_1, \dots, X_n)$  be the sample matrix and the  $p$ -vector  $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$  denote the sample average. Let  $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$  be a  $p \times n$  matrix whose columns are all equal to  $\bar{X}$ , and  $S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$  be the sample sum of products matrix. Then,  $u = n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu)$  has a real scalar type-2 beta distribution with the parameters  $(\frac{p}{2}, \frac{n}{2} - \frac{p}{2})$ , so that  $u \sim \frac{p}{n-p} F_{p, n-p}$  where  $F_{p, n-p}$  denotes a real  $F$  random variable whose degrees of freedoms are  $p$  and  $n - p$ .*

Hence, in order to test the hypothesis  $H_o : \mu = \mu_o$ , the likelihood ratio statistic gives the test criterion: Reject  $H_o$  for large values of  $u = n(\bar{X} - \mu_o)'S^{-1}(\bar{X} - \mu_o)$ , which is equivalent to rejecting  $H_o$  for large values of an  $F$ -random variable having  $p$  and  $n - p$  degrees of freedom where  $F_{p, n-p} = \frac{n-p}{p} u = \frac{n-p}{p} n(\bar{X} - \mu_o)'S^{-1}(\bar{X} - \mu_o)$ , that is,

$$\begin{aligned} \text{reject } H_o \text{ if } \frac{n-p}{p} u = F_{p, n-p} \geq F_{p, n-p, \alpha}, \\ \text{with } \alpha = Pr\{F_{p, n-p} \geq F_{n, n-p, \alpha}\} \end{aligned} \quad (6.3.6)$$

at a given significance level  $\alpha$  where  $u = n(\bar{X} - \mu_o)' S^{-1}(\bar{X} - \mu_o) \sim \frac{p}{n-p} F_{p, n-p}$ ,  $n$  being the sample size.

**Example 6.3.1.** Consider a trivariate real Gaussian vector  $X \sim N_3(\mu, \Sigma)$ ,  $\Sigma > O$ , where  $\Sigma$  is unknown. We would like to test the following hypothesis on  $\mu$ :  $H_o : \mu = \mu_o$ , with  $\mu'_o = (1, 1, 1)$ . Consider the following simple random sample of size  $n = 5$  from this  $N_3(\mu, \Sigma)$  population:

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, X_3 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix},$$

so that

$$\begin{aligned} \bar{X} &= \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, X_1 - \bar{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix}, X_2 - \bar{X} = \frac{1}{5} \begin{bmatrix} 4 \\ -2 \\ -9 \end{bmatrix}, \\ X_3 - \bar{X} &= \frac{1}{5} \begin{bmatrix} -6 \\ 3 \\ 6 \end{bmatrix}, X_4 - \bar{X} = \frac{1}{5} \begin{bmatrix} -11 \\ 3 \\ 6 \end{bmatrix}, X_5 - \bar{X} = \frac{1}{5} \begin{bmatrix} 9 \\ -7 \\ -4 \end{bmatrix}. \end{aligned}$$

Let  $\mathbf{X} = [X_1, \dots, X_5]$  the  $3 \times 5$  sample matrix and  $\bar{\mathbf{X}} = [\bar{X}, \bar{X}, \dots, \bar{X}]$  be the  $3 \times 5$  matrix of sample means. Then,

$$\begin{aligned} \mathbf{X} - \bar{\mathbf{X}} &= [X_1 - \bar{X}, \dots, X_5 - \bar{X}] = \frac{1}{5} \begin{bmatrix} 4 & 4 & -6 & -11 & 9 \\ 3 & -2 & 3 & 3 & -7 \\ 1 & -9 & 6 & 6 & -4 \end{bmatrix}, \\ S &= (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = \frac{1}{5^2} \begin{bmatrix} 270 & -110 & -170 \\ -110 & 80 & 85 \\ -170 & 85 & 170 \end{bmatrix}. \end{aligned}$$

Let  $S = \frac{1}{5^2} A$ . In order to evaluate the test statistic, we need  $S^{-1} = 25A^{-1}$ . To obtain the correct inverse without any approximation, we will use the transpose of the cofactor matrix divided by the determinant. The determinant of  $A$ ,  $|A|$ , as obtained in terms of the elements of the first row and the corresponding cofactors is equal to 531250. The matrix of cofactors, denoted by  $\text{Cof}(A)$ , which is symmetric in this case, is the following:

$$\text{Cof}(A) = \begin{bmatrix} 6375 & 4250 & 4250 \\ 4250 & 17000 & -4250 \\ 4250 & -4250 & 9500 \end{bmatrix} \Rightarrow S^{-1} = \frac{25}{531250} \text{Cof}(A).$$



The null hypothesis is  $H_o : \mu = \mu_o = (1, 1, 1)'$ , so that

$$\bar{X} - \mu_o = \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = -\frac{1}{5} \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix},$$

the observed value of the test statistic being

$$\begin{aligned} w &= \frac{n-p}{p} n (\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o) = \frac{(5-3)}{3} 5 \frac{25}{5^2} [4 \ 3 \ 1] A^{-1} \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix} \\ &= \frac{2}{3} 5 \frac{25}{5^2} \frac{1}{531250} [(4)^2(6375) + (3)^2(17000) + (1)^2(9500) \\ &\quad + 2(4)(3)(4250) + 2(4)(1)(4250) - 2(3)(1)(4250)] \\ &= \frac{2}{3} 5 \frac{25}{5^2} \frac{1}{531250} [375000] = 2.35. \end{aligned}$$

The test statistic  $w$  under the null hypothesis is  $F$ -distributed, that is,  $w \sim F_{p, n-p}$ . Let us test  $H_o$  at the significance level  $\alpha = 0.05$ . Since the critical value as obtained from an  $F$ -table is  $F_{p, n-p, \alpha} = F_{3, 2, 0.05} = 19.2$  and  $2.35 < 19.2$ , we do not reject  $H_o$ .

**Note 6.3.1.** If  $S$  is replaced by  $\frac{1}{n-1}S$ , an unbiased estimator for  $\Sigma$ , then the test statistic  $\frac{1}{n-1}n(\bar{X} - \mu_o)'[\frac{1}{n-1}S]^{-1}(\bar{X} - \mu_o) = \frac{T_n^2}{n-1}$  where  $T_n^2$  denotes Hotelling's  $T^2$  statistic, which for  $p = 1$  corresponds to the square of a Student- $t$  statistic having  $n-1$  degrees of freedom.

Since  $u$  as defined in Theorem 6.3.1 is distributed as a type-2 beta random variable with the parameters  $(\frac{p}{2}, \frac{n-p}{2})$ , we have the following results:  $\frac{1}{u}$  is type-2 beta distributed with the parameters  $(\frac{n-p}{2}, \frac{p}{2})$ ,  $\frac{u}{1+u}$  is type-1 beta distributed with the parameters  $(\frac{p}{2}, \frac{n-p}{2})$ , and  $\frac{1}{1+u}$  is type-1 beta distributed with the parameters  $(\frac{n-p}{2}, \frac{p}{2})$ ,  $n$  being the sample size.

### 6.3.2. Paired values or linear functions when $\Sigma$ is unknown

Let  $Y_1, \dots, Y_k$  be  $p \times 1$  vectors having their own distributions which are unknown. However, suppose that it is known that a certain linear function  $X = a_1 Y_1 + \dots + a_k Y_k$  has a  $p$ -variate real Gaussian  $N_p(\mu, \Sigma)$  distribution with  $\Sigma > O$ . We would like to test hypotheses of the type  $E[X] = a_1 \mu_{(1)}^{(o)} + \dots + a_k \mu_{(k)}^{(o)}$  where the  $\mu_j^{(o)}$ 's,  $j = 1, \dots, k$ , are specified. Since we do not know the distributions of  $Y_1, \dots, Y_k$ , let us convert the iid variables on  $Y_j$ ,  $j = 1, \dots, k$ , to iid variables on  $X_j$ , say  $X_1, \dots, X_n$ ,  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , where  $\Sigma$  is unknown. First, the observations on  $Y_1, \dots, Y_k$  are transformed into observations on the  $X_j$ 's. The problem then involves a single normal population whose covariance

matrix is unknown. An example of this type is  $Y_1$  representing a  $p \times 1$  vector before a certain process, such as administering a drug to a patient; in this instance,  $Y_1$  could consist of measurements on  $p$  characteristics observed in a patient. Observations on  $Y_2$  will then be the measurements on the same  $p$  characteristics after the process such as after administering the drug to the patient. Then  $Y_{2q} - Y_{1q} = X_q$  will represent the variable corresponding to the difference in the measurements on the  $q$ -th characteristic. Let the hypothesis be  $H_o : \mu = \mu_o$  (given),  $\Sigma$  being unknown. Note that once the observations on  $X_j$  are taken, then the individual  $\mu_{(j)}$ 's are irrelevant as they no longer are of any use. Once the  $X_j$ 's are determined, one can compute the sum of products matrix  $S$  in  $X_j$ . In this case, the test statistic is  $u = n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)$ , which is distributed as a type-2 beta with parameters  $(\frac{p}{2}, \frac{n-p}{2})$ . Then,  $u \sim \frac{p}{n-p} F$  where  $F$  is an  $F$  random variable having  $p$  and  $n - p$  degrees of freedom, that is, an  $F_{p, n-p}$  random variable,  $n$  being the sample size. Thus, the test criterion is applied as follows: Determine the observed value of  $u$  and the corresponding observed value of  $F_{p, n-p}$  that is,  $\frac{n-p}{p} u$ , and then

$$\text{reject } H_o \text{ if } \frac{n-p}{p} u \geq F_{p, n-p, \alpha}, \text{ with } Pr\{F_{p, n-p} \geq F_{p, n-p, \alpha}\} = \alpha. \quad (6.3.7)$$

**Example 6.3.2.** Five motivated individuals were randomly selected and subjected to an exercise regimen for a month. The exercise program promoters claim that the subjects can expect a weight loss of 5 kg as well as a 2-in. reduction in lower stomach girth by the end of the month period. Let  $Y_1$  and  $Y_2$  denote the two component vectors representing weight and girth before starting the routine and at the end of the exercise program, respectively. The following are the observations on the five individuals:

$$(Y_1, Y_2) = \begin{bmatrix} (85, 85) \\ (40, 41) \end{bmatrix}, \begin{bmatrix} (80, 70) \\ (40, 45) \end{bmatrix}, \begin{bmatrix} (75, 73) \\ (36, 36) \end{bmatrix}, \begin{bmatrix} (70, 71) \\ (38, 38) \end{bmatrix}, \begin{bmatrix} (70, 68) \\ (35, 34) \end{bmatrix}.$$

Obviously,  $Y_1$  and  $Y_2$  are dependent variables having a joint distribution. We will assume that the difference  $X = Y_1 - Y_2$  has a real Gaussian distribution, that is,  $X \sim N_2(\mu, \Sigma)$ ,  $\Sigma > O$ . Under this assumption, the observations on  $X$  are

$$X_1 = \begin{bmatrix} 85 - 85 \\ 40 - 41 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, X_2 = \begin{bmatrix} 10 \\ -5 \end{bmatrix}, X_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, X_4 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Let  $\mathbf{X} = [X_1, X_2, \dots, X_5]$  and  $\mathbf{X} - \bar{\mathbf{X}} = [X_1 - \bar{X}, \dots, X_5 - \bar{X}]$ , both being  $2 \times 5$  matrix. The observed sample average  $\bar{X}$ , the claim of the exercise routine promoters  $\mu = \mu_o$  as

well as other relevant quantities are as follows:

$$\begin{aligned}\bar{X} &= \frac{1}{5}(X_1 + \cdots + X_5) = \frac{1}{5} \begin{bmatrix} 13 \\ -5 \end{bmatrix}, \mu_o = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \\ X_1 - \bar{X} &= \begin{bmatrix} 0 \\ -1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 13 \\ -5 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} -13 \\ 0 \end{bmatrix}, \\ X_2 - \bar{X} &= \frac{1}{5} \begin{bmatrix} 37 \\ -20 \end{bmatrix}, X_3 - \bar{X} = \frac{1}{5} \begin{bmatrix} -3 \\ 5 \end{bmatrix}, X_4 - \bar{X} = \frac{1}{5} \begin{bmatrix} -18 \\ 5 \end{bmatrix}, X_5 - \bar{X} = \frac{1}{5} \begin{bmatrix} -3 \\ 10 \end{bmatrix};\end{aligned}$$

$$\mathbf{X} - \bar{\mathbf{X}} = [X_1 - \bar{X}, \dots, X_5 - \bar{X}] = \frac{1}{5} \begin{bmatrix} -13 & 37 & -3 & -18 & -3 \\ 0 & -20 & 5 & 5 & 10 \end{bmatrix};$$

$$S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = \frac{1}{5^2} \begin{bmatrix} -13 & 37 & -3 & -18 & -3 \\ 0 & -20 & 5 & 5 & 10 \end{bmatrix} \begin{bmatrix} -13 & 0 \\ 37 & -20 \\ -3 & 5 \\ -18 & 5 \\ -3 & 10 \end{bmatrix}$$

$$= \frac{1}{5^2} \begin{bmatrix} 1880 & -875 \\ -875 & 550 \end{bmatrix} = \frac{1}{5^2} A, A = \begin{bmatrix} 1880 & -875 \\ -875 & 550 \end{bmatrix} \Rightarrow S^{-1} = 25A^{-1};$$

$$|A| = 1880(550) - (875)^2 = 214975; \text{Cof}(A) = \begin{bmatrix} 550 & 875 \\ 875 & 1880 \end{bmatrix}; A^{-1} = \frac{\text{Cof}(A)}{|A|};$$

$$A^{-1} = \frac{1}{268375} \begin{bmatrix} 550 & 875 \\ 875 & 1880 \end{bmatrix}, S^{-1} = 25A^{-1};$$

$$\bar{X} - \mu_o = \frac{1}{5} \begin{bmatrix} 13 \\ -5 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \end{bmatrix} = -\frac{1}{5} \begin{bmatrix} 12 \\ 15 \end{bmatrix}.$$

The test statistic being  $w = \frac{(n-p)}{p} n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)$ , its observed value is

$$\begin{aligned}v &= \frac{(5-2)}{2} 5 \frac{5^2}{5^2} \frac{1}{268375} [12 \quad 15] \begin{bmatrix} 550 & 875 \\ 875 & 1880 \end{bmatrix} \begin{bmatrix} 12 \\ 15 \end{bmatrix} \\ &= \frac{3}{2} \frac{5}{268375} [(12)^2(550) + (15)^2(1880) + 2(12)(15)(875)] = 22.84.\end{aligned}$$

Letting the significance level of the test be  $\alpha = 0.05$ , the critical value is  $F_{p, n-p, \alpha} = F_{2, 3, 0.05} = 9.55$ . Since  $22.84 > 9.55$ ,  $H_o$  is thus rejected.

### 6.3.3. Independent Gaussian populations

Consider  $k$  independent  $p$ -variate real Gaussian populations whose individual distribution is  $N_p(\mu_{(j)}, \Sigma_j)$ ,  $\Sigma_j > O$ ,  $j = 1, \dots, k$ . Given simple random samples of sizes

$n_1, \dots, n_k$  from these  $k$  populations, we may wish to test a hypothesis on a given linear functions of the mean values, that is,  $H_o : a_1\mu_{(1)} + \dots + a_k\mu_{(k)} = \mu_0$  where  $a_1, \dots, a_k$  are known constants and  $\mu_0$  is a given quantity under the null hypothesis. We have already discussed this problem for the case of known covariance matrices. When the  $\Sigma_j$ 's are all unknown or some of them are known while others are not, the MLE's of the unknown covariance matrices turn out to be the respective sample sums of products matrices divided by the corresponding sample sizes. This will result in a linear function of independent Wishart matrices whose distribution proves challenging to determine, even for the null case.

### Special case of two independent Gaussian populations

Consider the special case of two independent real Gaussian populations having identical covariance matrices. that is, let the populations be  $Y_{1q} \sim N_p(\mu_{(1)}, \Sigma)$ ,  $\Sigma > O$ , the  $Y_{1q}$ 's,  $q = 1, \dots, n_1$ , being iid, and  $Y_{2q} \sim N_p(\mu_{(2)}, \Sigma)$ ,  $\Sigma > O$ , the  $Y_{2q}$ 's,  $q = 1, \dots, n_2$ , being iid. Let the sample  $p \times n_1$  and  $p \times n_2$  matrices be denoted as  $\mathbf{Y}_1 = (Y_{11}, \dots, Y_{1n_1})$  and  $\mathbf{Y}_2 = (Y_{21}, \dots, Y_{2n_2})$  and let the sample averages be  $\bar{Y}_j = \frac{1}{n_j}(Y_{j1} + \dots + Y_{jn_j})$ ,  $j = 1, 2$ . Let  $\bar{\mathbf{Y}}_j = (\bar{Y}_j, \dots, \bar{Y}_j)$ , a  $p \times n_j$  matrix whose columns are equal to  $\bar{Y}_j$ ,  $j = 1, 2$ , and let

$$S_j = (\mathbf{Y}_j - \bar{\mathbf{Y}}_j)(\mathbf{Y}_j - \bar{\mathbf{Y}}_j)', \quad j = 1, 2,$$

be the corresponding sample sum of products matrices. Then,  $S_1$  and  $S_2$  are independently distributed as Wishart matrices having  $n_1 - 1$  and  $n_2 - 1$  degrees of freedom, respectively. As the sum of two independent  $p \times p$  real or complex matrices having matrix-variate gamma distributions with the same scale parameter matrix is again gamma distributed with the shape parameters summed up and the same scale parameter matrix, we observe that since the two populations are independently distributed,  $S_1 + S_2 \equiv S$  has a Wishart distribution having  $n_1 + n_2 - 2$  degrees of freedom. We now consider a hypothesis of the type  $\mu_{(1)} = \mu_{(2)}$ . In order to do away with the unknown common mean value, we may consider the real  $p$ -vector  $U = \bar{Y}_1 - \bar{Y}_2$ , so that  $E(U) = O$  and  $\text{Cov}(U) = \frac{1}{n_1}\Sigma + \frac{1}{n_2}\Sigma = (\frac{1}{n_1} + \frac{1}{n_2})\Sigma = \frac{n_1+n_2}{n_1n_2}\Sigma$ . The MLE of this pooled covariance matrix is  $\frac{1}{n_1+n_2}S$  where  $S$  is Wishart distributed with  $n_1 + n_2 - 2$  degrees of freedom. Then, following through the steps included in Sect. 6.3.1 with the parameter  $m$  now being  $n_1 + n_2 - 2$ , the power of  $S$  will become  $\frac{(n_1+n_2-2+1)}{2} - \frac{p+1}{2}$  when integrating out  $S$ . Letting the null hypothesis be  $H_o : E[Y_1] - E[Y_2] = \delta$  (specified), such as  $\delta = 0$ , the function resulting from integrating out  $S$  is

$$c \left[ \frac{(n_1 + n_2)}{n_1 n_2} u \right]^{\frac{p}{2}-1} \left[ 1 + \frac{(n_1 + n_2)}{n_1 n_2} u \right]^{-\frac{1}{2}(n_1+n_2-1)} \quad (6.3.8)$$

where  $c$  is the normalizing constant, so that  $w = \frac{(n_1+n_2)}{n_1n_2}(\bar{Y}_1 - \bar{Y}_2 - \delta)'S^{-1}(\bar{Y}_1 - \bar{Y}_2 - \delta)$  is distributed as a type-2 beta with the parameters  $(\frac{p}{2}, \frac{(n_1+n_2-1-p)}{2})$ . Writing  $w = \frac{p}{n_1+n_2-1-p}F_{p, n_1+n_2-1-p}$ , this  $F$  is seen to be an  $F$  statistic having  $p$  and  $n_1 + n_2 - 1 - p$  degrees of freedom. We will state these results as theorems.

**Theorem 6.3.2.** *Let the  $p \times p$  real positive definite matrices  $X_1$  and  $X_2$  be independently distributed as real matrix-variate gamma random variables with densities*

$$f_j(X_j) = \frac{|B|^{\alpha_j}}{\Gamma_p(\alpha_j)} |X_j|^{\alpha_j - \frac{p+1}{2}} e^{-\text{tr}(BX_j)}, \quad B > O, \quad X_j > O, \quad \Re(\alpha_j) > \frac{p-1}{2}, \quad (6.3.9)$$

$j=1,2$ , and zero elsewhere. Then, as can be seen from (5.2.6), the Laplace transform associated with  $X_j$  or that of  $f_j$ , denoted as  $L_{X_j}(*T)$ , is

$$L_{X_j}(*T) = |I + B^{-1} *T|^{-\alpha_j}, \quad I + B^{-1} *T > O, \quad j = 1, 2. \quad (i)$$

Accordingly,  $U_1 = X_1 + X_2$  has a real matrix-variate gamma density with the parameters  $(\alpha_1 + \alpha_2, B)$  whose associated Laplace transform is

$$L_{U_1}(*T) = |I + B^{-1} *T|^{-(\alpha_1+\alpha_2)}, \quad (ii)$$

and  $U_2 = a_1X_1 + a_2X_2$  has the Laplace transform

$$L_{U_2}(*T) = |I + a_1B^{-1} *T|^{-\alpha_1} |I + a_2B^{-1} *T|^{-\alpha_2}, \quad (iii)$$

whenever  $I + a_jB^{-1} *T > O, j = 1, 2$ , where  $a_1$  and  $a_2$  are real scalar constants.

It follows from (ii) that  $X_1 + X_2$  is also real matrix-variate gamma distributed. When  $a_1 \neq a_2$ , it is very difficult to invert (iii) in order to obtain the corresponding density. This can be achieved by expanding one of the determinants in (iii) in terms of zonal polynomials, say the second one, after having first taken  $|I + a_1B^{-1} *T|^{-(\alpha_1+\alpha_2)}$  out as a factor in this instance.

**Theorem 6.3.3.** *Let  $Y_j \sim N_p(\mu_{(j)}, \Sigma), \Sigma > O, j = 1, 2$ , be independent  $p$ -variate real Gaussian distributions sharing the same covariance matrix. Given a simple random sample of size  $n_1$  from  $Y_1$  and a simple random sample of size  $n_2$  from  $Y_2$ , let the sample averages be denoted by  $\bar{Y}_1$  and  $\bar{Y}_2$  and the sample sums of products matrices, by  $S_1$  and  $S_2$ , respectively. Consider the hypothesis  $H_o : \mu_{(1)} - \mu_{(2)} = \delta$  (given). Letting  $S = S_1 + S_2$  and*

$$w = \frac{(n_1 + n_2)}{n_1n_2}(\bar{Y}_1 - \bar{Y}_2 - \delta)'S^{-1}(\bar{Y}_1 - \bar{Y}_2 - \delta), \quad w \sim \frac{p}{n_1 + n_2 - 1 - p}F_{p, n_1+n_2-1-p} \quad (iv)$$

where  $F_{p, n_1+n_2-1-p}$  denotes an  $F$  distribution with  $p$  and  $n_1 + n_2 - 1 - p$  degrees of freedom, or equivalently,  $w$  is distributed as a type-2 beta variable with the parameters  $(\frac{p}{2}, \frac{(n_1+n_2-1-p)}{2})$ . We reject the null hypothesis  $H_0$  if  $\frac{n_1+n_2-1-p}{p} w \geq F_{p, n_1+n_2-1-p, \alpha}$  with

$$\Pr\{F_{p, n_1+n_2-1-p} \geq F_{p, n_1+n_2-1-p, \alpha}\} = \alpha \quad (vi)$$

at a given significance level  $\alpha$ .

**Theorem 6.3.4.** Let  $w$  be as defined in Theorem 6.3.3. Then  $w_1 = \frac{1}{w}$  is a real scalar type-2 beta variable with the parameters  $(\frac{n_1+n_2-1-p}{2}, \frac{p}{2})$ ;  $w_2 = \frac{w}{1+w}$  is a real scalar type-1 beta variable with the parameters  $(\frac{p}{2}, \frac{(n_1+n_2-1-p)}{2})$ ;  $w_3 = \frac{1}{1+w}$  is a real scalar type-1 beta variable with the parameters  $(\frac{n_1+n_2-1-p}{2}, \frac{p}{2})$ .

Those last results follow from the connections between real scalar type-1 and type-2 beta random variables. Results parallel to those appearing in (i) to (vi) and stated Theorems 6.3.1–6.3.4 can similarly be obtained for the complex case.

**Example 6.3.3.** Consider two independent populations whose respective distributions are  $N_2(\mu_{(1)}, \Sigma_1)$  and  $N_2(\mu_{(2)}, \Sigma_2)$ ,  $\Sigma_j > O$ ,  $j = 1, 2$ , and samples of sizes  $n_1 = 4$  and  $n_2 = 5$  from these two populations, respectively. Let the population covariance matrices be identical with  $\Sigma_1 = \Sigma_2 = \Sigma$ , the common covariance matrix being unknown, and let the observed sample vectors from the first population,  $X_j \sim N_2(\mu_{(1)}, \Sigma)$ , be

$$X_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, X_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, X_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

Denoting the sample mean from the first population by  $\bar{X}$  and the sample sum of products matrix by  $S_1$ , we have

$$\bar{X} = \frac{1}{4} \begin{bmatrix} 0 \\ 2 \end{bmatrix} \text{ and } S_1 = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})', \mathbf{X} = [X_1, X_2, X_3, X_4], \bar{\mathbf{X}} = [\bar{X}, \dots, \bar{X}],$$

the observations on these quantities being the following:

$$\begin{aligned} X_1 - \bar{X} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 4 \\ -2 \end{bmatrix}, \\ X_2 - \bar{X} &= \frac{1}{4} \begin{bmatrix} -4 \\ 2 \end{bmatrix}, \quad X_3 - \bar{X} = \frac{1}{4} \begin{bmatrix} 4 \\ 6 \end{bmatrix}, \quad X_4 - \bar{X} = \frac{1}{4} \begin{bmatrix} -4 \\ -6 \end{bmatrix}; \\ \mathbf{X} - \bar{\mathbf{X}} &= \frac{1}{4} \begin{bmatrix} 4 & -4 & 4 & -4 \\ -2 & 2 & 6 & -6 \end{bmatrix}, \quad S_1 = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' \\ S_1 &= \frac{1}{4^2} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix}. \end{aligned}$$

Let the sample vectors from the second population denoted as  $Y_1, \dots, Y_5$  be

$$Y_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad Y_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad Y_3 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad Y_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad Y_5 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Then, the sample average and the deviation vectors are the following:

$$\begin{aligned} \bar{Y} &= \frac{1}{5}[Y_1 + \dots + Y_5] = \frac{1}{5} \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \\ Y_1 - \bar{Y} &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 5 \\ 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} -5 \\ 3 \end{bmatrix}, \quad Y_2 - \bar{Y} = \frac{1}{5} \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \\ Y_3 - \bar{Y} &= \frac{1}{5} \begin{bmatrix} 0 \\ -7 \end{bmatrix}, \quad Y_4 - \bar{Y} = \frac{1}{5} \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \quad Y_5 - \bar{Y} = \frac{1}{5} \begin{bmatrix} 5 \\ 3 \end{bmatrix}, \\ \mathbf{Y} - \bar{\mathbf{Y}} &= \frac{1}{5} \begin{bmatrix} -5 & 0 & 0 & 0 & 5 \\ 3 & -2 & -7 & 3 & 3 \end{bmatrix}, \quad S_2 = (\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})'; \end{aligned}$$

$$\begin{aligned} S_2 &= \frac{1}{5^2} \begin{bmatrix} 50 & 0 \\ 0 & 80 \end{bmatrix}; \quad S = S_1 + S_2 = \frac{1}{16} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix} + \frac{1}{25} \begin{bmatrix} 50 & 0 \\ 0 & 80 \end{bmatrix} \\ &= \begin{bmatrix} 6.00 & 2.00 \\ 2.00 & 8.20 \end{bmatrix} \Rightarrow S^{-1} = \frac{\text{Cof}(S)}{|S|} \\ |S| &= 45.20; \quad S^{-1} = \frac{1}{45.20} \begin{bmatrix} 8.20 & -2.00 \\ -2.00 & 6.00 \end{bmatrix}. \end{aligned}$$

Letting the null hypothesis be

$$H_0 : \mu_{(1)} - \mu_{(2)} = \delta = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\bar{X} - \bar{Y} - \delta = \frac{1}{4} \begin{bmatrix} 0 \\ 2 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 5 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = - \begin{bmatrix} 2.0 \\ 0.9 \end{bmatrix}; \quad n_1 = 4, \quad n_2 = 5.$$

Thus, test statistic is  $u \sim F_{p, n_1+n_2-1-p}$  where

$$\begin{aligned} u &= \frac{(n_1 + n_2 - 1 - p)(n_1 + n_2)}{p n_1 n_2} (\bar{X} - \bar{Y} - \delta)' S^{-1} (\bar{X} - \bar{Y} - \delta) \\ &= \frac{(4 + 5 - 1 - 2)(4 + 5)}{2} \frac{1}{(4)(5) 45.2} \begin{bmatrix} -2.0 & -0.9 \end{bmatrix} \begin{bmatrix} 8.2 & -2.0 \\ -2.0 & 6.0 \end{bmatrix} \begin{bmatrix} -2.0 \\ -0.9 \end{bmatrix} \\ &\approx 0.91. \end{aligned}$$

Let us test  $H_0$  at the 5% significance level. Since the required critical value is  $F_{p, n_1+n_2-1-p, \alpha} = F_{2, 6, 0.05} = 5.14$  and  $0.91 < 5.14$ , the null hypothesis is not rejected.

#### 6.3.4. Testing $\mu_1 = \dots = \mu_p$ when $\Sigma$ is unknown in the real Gaussian case

Let the  $p \times 1$  vector  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ ,  $j = 1, \dots, n$ , the  $X_j$ 's being independently distributed. Let the  $p \times 1$  vector of unities be denoted by  $J$  or  $J' = (1, \dots, 1)$ , and let  $A$  be a vector that is orthogonal to  $J$  so that  $A'J = 0$ . For example, we can take

$$A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ 1 \\ -1 \end{bmatrix} \quad \text{when } p \text{ is even, } A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ 1 \\ -2 \\ 1 \end{bmatrix} \quad \text{or } A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ -1 \\ 0 \end{bmatrix} \quad \text{when } p \text{ is odd.}$$

If the last component of  $A$  is zero, we are then ignoring the last component of  $X_j$ . Let  $y_j = A'X_j$ ,  $j = 1, \dots, n$ , and the  $y_j$ 's be independently distributed. Then  $y_j \sim N_1(A'\mu, A'\Sigma A)$ ,  $A'\Sigma A > O$ , is a univariate normal variable with mean value  $A'\mu$  and variance  $A'\Sigma A$ . Consider the  $p \times n$  sample matrix comprising the  $X_j$ 's, that is,  $\mathbf{X} = (X_1, \dots, X_n)$ . Let the sample average of the  $X_j$ 's be  $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$  and  $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$ . Then, the sample sum of products matrix  $S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$ . Consider the  $1 \times n$  vector  $Y = (y_1, \dots, y_n) = (A'X_1, \dots, A'X_n) = A'\mathbf{X}$ ,  $\bar{y} = \frac{1}{n}(y_1 + \dots + y_n) = A'\bar{X}$ ,  $\sum_{j=1}^n (y_j - \bar{y})^2 = A'(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'A = A'SA$ . Let the null hypothesis be  $H_0 : \mu_1 = \dots = \mu_p = \nu$ , where  $\nu$  is unknown,  $\mu' = (\mu_1, \dots, \mu_p)$ . Thus,  $H_0$  is  $A'\mu = \nu A'J = 0$ . The joint density of  $y_1, \dots, y_n$ , denoted by  $L$ , is then



$$\begin{aligned}
L &= \prod_{j=1}^n \frac{e^{-\frac{1}{2A'\Sigma A}(y_j - A'\mu)^2}}{(2\pi)^{\frac{1}{2}}[A'\Sigma A]^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2(A'\Sigma A)}\sum_{j=1}^n (y_j - A'\mu)^2}}{(2\pi)^{\frac{n}{2}}[A'\Sigma A]^{\frac{n}{2}}} \\
&= \frac{e^{-\frac{1}{2A'\Sigma A}\{A'SA + nA'(\bar{X} - \mu)(\bar{X} - \mu)'A\}}}{(2\pi)^{\frac{n}{2}}[A'\Sigma A]^{\frac{n}{2}}} \quad (i)
\end{aligned}$$

where

$$\begin{aligned}
\sum_{j=1}^n (y_j - A'\mu)^2 &= \sum_{j=1}^n (y_j - \bar{y} + \bar{y} - A'\mu)^2 = \sum_{j=1}^n (y_j - \bar{y})^2 + n(\bar{y} - A'\mu)^2 \\
&= \sum_{j=1}^n A'(X_j - \bar{X})(X_j - \bar{X})'A + nA'(\bar{X} - \mu)(\bar{X} - \mu)'A \\
&= A'SA + nA'(\bar{X} - \mu)(\bar{X} - \mu)'A.
\end{aligned}$$

Let us determine the MLE's of  $\mu$  and  $\Sigma$ . We have

$$\ln L = -\frac{n}{2}(2\pi) - \frac{n}{2} \ln A'\Sigma A - \frac{1}{2A'\Sigma A}[A'SA + n(A'(\bar{X} - \mu)(\bar{X} - \mu)'A)].$$

On differentiating  $\ln L$  with respect to  $\mu_1$  and equating the result to zero, we have

$$\begin{aligned}
\frac{\partial}{\partial \mu_1} \ln L = 0 &\Rightarrow nA' \left[ \frac{\partial}{\partial \mu_1} \{ \bar{X}\bar{X}' - \bar{X}\mu' - \mu\bar{X}' + \mu\mu' \} \right] A = O \\
&\Rightarrow nA'[-\bar{X}[1, 0, \dots, 0] - \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \bar{X}' + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \mu' + \mu[1, 0, \dots, 0]]A = O \\
&\Rightarrow 2a_1A'(\bar{X} - \mu) = 0 \Rightarrow \hat{\mu} = \bar{X} \quad (ii)
\end{aligned}$$

since the equation holds for each  $\mu_j$ ,  $j = 1, \dots, p$ , and  $A' = (a_1, \dots, a_p)$ ,  $a_j \neq 0$ ,  $j = 1, \dots, p$ ,  $A$  being fixed. As well,  $(\bar{X} - \mu)(\bar{X} - \mu)' = \bar{X}\bar{X}' - \bar{X}\mu' - \mu\bar{X}' + \mu\mu'$ . Now, consider differentiating  $\ln L$  with respect to an element of  $\Sigma$ , say,  $\sigma_{11}$ , at  $\hat{\mu} = \bar{X}$ :

$$\begin{aligned}
\frac{\partial}{\partial \sigma_{11}} \ln L &= 0 \\
&\Rightarrow -\frac{2n a_1^2 \sigma_{11}}{2A'\Sigma A} + \frac{A'SA}{2(A'\Sigma A)^2} (2a_1^2 \sigma_{11}) = 0 \\
&\Rightarrow A'\hat{\Sigma}A = \frac{1}{n}A'SA
\end{aligned}$$

for each element in  $\Sigma$  and hence  $\hat{\Sigma} = \frac{1}{n}S$ . Thus,

$$\max_{\Omega} L = \frac{e^{-\frac{n}{2}} n^{\frac{n}{2}}}{(A'SA)^{\frac{n}{2}}}. \quad (iii)$$

Under  $H_o$ ,  $A'\mu = 0$  and consequently the maximum under  $H_o$  is the following:

$$\max_{H_o} L = \frac{e^{-\frac{n}{2}} n^{\frac{n}{2}}}{[A'(S + n\bar{X}\bar{X}')A]^{\frac{n}{2}}}. \quad (iv)$$

Accordingly, the  $\lambda$ -criterion is

$$\lambda = \frac{(A'SA)^{\frac{n}{2}}}{[A'(S + n\bar{X}\bar{X}')A]^{\frac{n}{2}}} = \frac{1}{[1 + n\frac{A'\bar{X}\bar{X}'A}{A'SA}]^{\frac{n}{2}}}. \quad (6.3.10)$$

We would reject  $H_o$  for small values of  $\lambda$  or for large values of  $u \equiv nA'\bar{X}\bar{X}'A/A'SA$  where  $\bar{X}$  and  $S$  are independently distributed. Observing that  $S \sim W_p(n-1, \Sigma)$ ,  $\Sigma > O$  and  $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$ ,  $\Sigma > O$ , we have

$$\frac{n}{A'\Sigma A} A'\bar{X}\bar{X}'A \sim \chi_1^2 \text{ and } \frac{A'SA}{A'\Sigma A} \sim \chi_{n-1}^2.$$

Hence,  $(n-1)u$  is a  $F$  statistic with 1 and  $n-1$  degrees of freedom, and the null hypothesis is to be rejected whenever

$$v \equiv n(n-1) \frac{A'\bar{X}\bar{X}'A}{A'SA} \geq F_{1, n-1, \alpha}, \text{ with } Pr\{F_{1, n-1} \geq F_{1, n-1, \alpha}\} = \alpha. \quad (6.3.11)$$

**Example 6.3.4.** Consider a real bivariate Gaussian  $N_2(\mu, \Sigma)$  population where  $\Sigma > O$  is unknown. We would like to test the hypothesis  $H_o : \mu_1 = \mu_2$ ,  $\mu' = (\mu_1, \mu_2)$ , so that  $\mu_1 - \mu_2 = 0$  under this null hypothesis. Let the sample be  $X_1, X_2, X_3, X_4$ , as specified in Example 6.3.3. Let  $A' = (1, -1)$  so that  $A'\mu = 0$  under  $H_o$ . With the same observation vectors as those comprising the first sample in Example 6.3.3,  $A'X_1 = (1)$ ,  $A'X_2 = (-2)$ ,  $A'X_3 = (-1)$ ,  $A'X_4 = (0)$ . Letting  $y = A'X_j$ , the observations on  $y_j$  are  $(1, -2, -1, 0)$  or  $A'\mathbf{X} = A'[X_1, X_2, X_3, X_4] = [1, -2, -1, 0]$ . The sample sum of products matrix as evaluated in the first part of Example 6.3.3 is

$$S_1 = \frac{1}{16} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix} \Rightarrow A'S_1A = \frac{1}{16} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{80}{16} = 5.$$

Our test statistic is

$$v = n(n-1) \frac{A'\bar{X}\bar{X}'A}{A'S_1A} \sim F_{1, n-1}, \quad n = 4.$$

Let the significance level be  $\alpha = 0.05$ . the observed values of  $A'\bar{X}\bar{X}'A$ ,  $A'S_1A$ ,  $v$ , and the tabulated critical value of  $F_{1, n-1, \alpha}$  are the following:

$$A'\bar{X}\bar{X}'A = \frac{1}{4}; \quad A'S_1A = \frac{80}{16} = 5;$$

$$v = 4(3)\left(\frac{1}{5 \times 4}\right) = 0.6; \quad F_{1, n-1, \alpha} = F_{1, 3, 0.05} = 10.13.$$

As  $0.6 < 10.13$ ,  $H_o$  is not rejected.

### 6.3.5. Likelihood ratio test for testing $H_o : \mu_1 = \dots = \mu_p$ when $\Sigma$ is unknown

In the entire parameter space  $\Omega$  of a  $N_p(\mu, \Sigma)$  population,  $\mu$  is estimated by the sample average  $\bar{X}$  and, as previously determined, the maximum of the likelihood function is

$$\max_{\Omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}} \quad (i)$$

where  $S$  is the sample sum of products matrix and  $n$  is the sample size. Under the hypothesis  $H_o : \mu_1 = \dots = \mu_p = v$ , where  $v$  is unknown, this  $v$  is estimated by  $\hat{v} = \frac{1}{np} \sum_{i,j} x_{ij} = \frac{1}{p} J' \bar{X}$ ,  $J' = (1, \dots, 1)$ , the  $p \times 1$  sample vectors  $X'_j = (x_{1j}, \dots, x_{pj})$ ,  $j = 1, \dots, n$ , being independently distributed. Thus, under the null hypothesis  $H_o$ , the population covariance matrix is estimated by  $\frac{1}{n}(S + n(\bar{X} - \hat{\mu})(\bar{X} - \hat{\mu})')$ , and, proceeding as was done to obtain Eq. (6.3.3), the  $\lambda$ -criterion reduces to

$$\lambda = \frac{|S|^{\frac{n}{2}}}{|S + n(\bar{X} - \hat{\mu})(\bar{X} - \hat{\mu})'|^{\frac{n}{2}}} \quad (6.3.12)$$

$$= \frac{1}{(1+u)^{\frac{n}{2}}}, \quad u = n(\bar{X} - \hat{\mu})' S^{-1} (\bar{X} - \hat{\mu}). \quad (6.3.13)$$

Given the structure of  $u$  in (6.3.13), we can take the Gaussian population covariance matrix  $\Sigma$  to be the identity matrix  $I$ , as was explained in Sect. 6.3.1. Observe that

$$(\bar{X} - \hat{\mu})' = (\bar{X} - \frac{1}{p} J J' \bar{X})' = \bar{X}' [I - \frac{1}{p} J J'] \quad (ii)$$

where  $I - \frac{1}{p} J J'$  is idempotent of rank  $p - 1$ ; hence there exists an orthonormal matrix  $P$ ,  $P P' = I$ ,  $P' P = I$ , such that

$$I - \frac{1}{p} J J' = P \begin{bmatrix} I_{p-1} & O \\ O' & 0 \end{bmatrix} P' \Rightarrow$$

$$\sqrt{n}(\bar{X} - \hat{\mu})' = \sqrt{n} \bar{X}' (I - \frac{1}{p} J J') = \sqrt{n} \bar{X}' P \begin{bmatrix} I_{p-1} & O \\ O' & 0 \end{bmatrix} = [V_1', 0], \quad V = \sqrt{n} \bar{X}' P,$$

where  $V_1$  is the subvector of the first  $p - 1$  components of  $V$ . Then the quadratic form  $u$ , which is our test statistic, reduces to the following:

$$u = n(\bar{X} - \hat{\mu})' S^{-1} (\bar{X} - \hat{\mu}) = [V_1', 0] S^{-1} \begin{bmatrix} V_1 \\ 0 \end{bmatrix} = V_1' S^{11} V_1,$$

$$S^{-1} = \begin{bmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{bmatrix}.$$

We note that the test statistic  $u$  has the same structure that of  $u$  in Theorem 6.3.1 with  $p$  replaced by  $p - 1$ . Accordingly,  $u = n(\bar{X} - \hat{\mu})' S^{-1} (\bar{X} - \hat{\mu})$  is distributed as a real scalar type-2 beta variable with the parameters  $\frac{p-1}{2}$  and  $\frac{n-(p-1)}{2}$ , so that  $\frac{n-p+1}{p-1} u \sim F_{p-1, n-p+1}$ . Thus, the test criterion consists of

$$\text{rejecting } H_o \text{ if the observed value of } \frac{n-p+1}{p-1} u \geq F_{p-1, n-p+1, \alpha},$$

$$\text{with } Pr\{F_{p-1, n-p+1} \geq F_{p-1, n-p+1, \alpha}\} = \alpha. \quad (6.3.14)$$

**Example 6.3.5.** Let the population be  $N_2(\mu, \Sigma)$ ,  $\Sigma > O$ ,  $\mu' = (\mu_1, \mu_2)$  and the null hypothesis be  $H_o : \mu_1 = \mu_2 = \nu$  where  $\nu$  and  $\Sigma$  are unknown. The sample values, as specified in Example 6.3.3, are

$$X_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, X_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, X_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \Rightarrow \bar{X} = \frac{1}{4} \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

The maximum likelihood estimate of  $\mu$  under  $H_o$ , is

$$\hat{\mu} = \frac{1}{p} J J' \bar{X}, \quad J = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

and

$$(\bar{X} - \hat{\mu})' = \bar{X}' [I - \frac{1}{p} J J'] = \bar{X}' [I - \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}] = \frac{1}{4} [-1 \quad 1].$$

As previously calculated, the sample sum of products matrix is

$$S_1 = \frac{1}{4^2} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix} \Rightarrow S_1^{-1} = \frac{16}{4096} \begin{bmatrix} 80 & -32 \\ -32 & 64 \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 80 & -32 \\ -32 & 64 \end{bmatrix}; \quad n = 4, \quad p = 2.$$

The test statistic  $v$  and its observed value are

$$\begin{aligned}
 v &= \frac{(n-p+1)}{(p-1)} n(\bar{X} - \hat{\mu})' S_1^{-1} (\bar{X} - \hat{\mu}) \sim F_{p-1, n-p+1} = F_{1, 3} \\
 &= \frac{(4-2+1)}{(2-1)} (4) \frac{1}{4^2} \frac{1}{256} \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 80 & -32 \\ -32 & 64 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\
 &= \frac{(3)(4)}{(4^2)(256)} [(-1)^2(80) + (1)^2(64) - 2(32)(-1)(1)] \\
 &= 0.61.
 \end{aligned}$$

At significance level  $\alpha = 0.05$ , the tabulated critical value  $F_{1, 3, 0.05}$  is 10.13, and since the observed value 0.61 is less than 10.13,  $H_o$  is not rejected.

#### 6.4. Testing Hypotheses on the Population Covariance Matrix

Let the  $p \times 1$  independent vectors  $X_j$ ,  $j = 1, \dots, n$ , have a  $p$ -variate real nonsingular  $N_p(\mu, \Sigma)$  distribution and the  $p \times n$  matrix  $\mathbf{X} = (X_1, \dots, X_n)$  be the sample matrix. Denoting the sample average by  $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$  and letting  $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$ , each column of  $\bar{\mathbf{X}}$  being equal to  $\bar{X}$ , the sample sum of products matrix is  $S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$ . We have already established that  $S$  is Wishart distributed with  $m = n - 1$  degrees of freedom, that is,  $S \sim W_p(m, \Sigma)$ ,  $\Sigma > O$ . Letting  $S_\mu = (\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})'$  where  $\mathbf{M} = (\mu, \dots, \mu)$ , each of its column being the  $p \times 1$  vector  $\mu$ ,  $S_\mu \sim W_p(n, \Sigma)$ ,  $\Sigma > O$ , where the number of degrees of freedom is  $n$  itself whereas the number of degrees of freedom associated with  $S$  is  $m = n - 1$ . Let us consider the hypothesis  $H_o : \Sigma = \Sigma_o$  where  $\Sigma_o$  is a given known matrix and  $\mu$  is unspecified. Then, the MLE's of  $\mu$  and  $\Sigma$  in the entire parameter space are  $\hat{\mu} = \bar{X}$  and  $\hat{\Sigma} = \frac{1}{n}S$ , and the joint density of the sample values  $X_1, \dots, X_n$ , denoted by  $L$ , is given by

$$L = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S) - \frac{n}{2}\text{tr}(\bar{X} - \mu)(\bar{X} - \mu)'}. \quad (6.4.1)$$

Thus, as previously determined, the maximum of  $L$  in the parameter space  $\Omega = \{(\mu, \Sigma) | \Sigma > O\}$  is

$$\max_{\Omega} L = \frac{n^{\frac{np}{2}} e^{-\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}}, \quad (i)$$

the maximum of  $L$  under the null hypothesis  $H_o : \Sigma = \Sigma_o$  being given by

$$\max_{H_o} L = \frac{e^{-\frac{1}{2}\text{tr}(\Sigma_o^{-1}S)}}{(2\pi)^{\frac{np}{2}} |\Sigma_o|^{\frac{n}{2}}}. \quad (ii)$$

Then, the  $\lambda$ -criterion is the following:

$$\lambda = \frac{e^{\frac{np}{2}}}{n^{\frac{np}{2}}} |\Sigma_o^{-1} S|^{\frac{n}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma_o^{-1} S)}. \quad (6.4.2)$$

Letting  $u = \lambda^{\frac{2}{n}}$ ,

$$u = \frac{e^p}{n^p} |\Sigma_o^{-1} S| e^{-\frac{1}{n}\text{tr}(\Sigma_o^{-1} S)}, \quad (6.4.3)$$

and we would reject  $H_o$  for small values of  $u$  since it is a monotonically increasing function of  $\lambda$ , which means that the null hypothesis ought to be rejected for large values of  $\text{tr}(\Sigma_o^{-1} S)$  as the exponential function dominates the polynomial function for large values of the argument. Let us determine the distribution of  $w = \text{tr}(\Sigma_o^{-1} S)$  whose Laplace transform with parameter  $s$  is

$$L_w(s) = E[e^{-sw}] = E[e^{-s \text{tr}(\Sigma_o^{-1} S)}]. \quad (iii)$$

This can be evaluated by integrating out over the density of  $S$  which has a Wishart distribution with  $m = n - 1$  degrees of freedom when  $\mu$  is estimated:

$$L_w(s) = \frac{1}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) |\Sigma|^{\frac{m}{2}}} \int_{S>O} |S|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1} S) - s \text{tr}(\Sigma_o^{-1} S)} dS. \quad (iv)$$

The exponential part is  $-\frac{1}{2}\text{tr}(\Sigma^{-1} S) - s \text{tr}(\Sigma_o^{-1} S) = -\frac{1}{2}\text{tr}[(\Sigma^{-\frac{1}{2}} S \Sigma^{-\frac{1}{2}})(I + 2s \Sigma^{\frac{1}{2}} \Sigma_o^{-1} \Sigma^{\frac{1}{2}})]$  and hence,

$$L_w(s) = |I + 2s \Sigma^{\frac{1}{2}} \Sigma_o^{-1} \Sigma^{\frac{1}{2}}|^{-\frac{m}{2}}. \quad (6.4.4)$$

**The null case,  $\Sigma = \Sigma_o$**

In this case,  $\Sigma^{\frac{1}{2}} \Sigma_o^{-1} \Sigma^{\frac{1}{2}} = I$ , so that

$$L_w(s) = |I + 2sI|^{-\frac{m}{2}} = (1 + 2s)^{-\frac{mp}{2}} \Rightarrow w \sim \chi_{mp}^2. \quad (6.4.5)$$

Thus, the test criterion is the following:

$$\text{Reject } H_o \text{ if } w \geq \chi_{mp, \alpha}^2, \text{ with } Pr\{\chi_{mp}^2 \geq \chi_{mp, \alpha}^2\} = \alpha. \quad (6.4.6)$$

When  $\mu$  is known, it is used instead of its MLE to determine  $S_\mu$ , and the resulting criterion consists of rejecting  $H_o$  whenever the observed  $w_\mu = \text{tr}(\Sigma_o^{-1} S_\mu) \geq \chi_{np, \alpha}^2$  where  $n$  is the sample size. These results are summarized in the following theorem.

**Theorem 6.4.1.** *Let the null hypothesis be  $H_o : \Sigma = \Sigma_o$  (given) and  $w = \text{tr}(\Sigma_o^{-1} S)$  where  $S$  is the sample sum of products matrix. Then, the null distribution of  $w = \text{tr}(\Sigma_o^{-1} S)$*

has a real scalar chisquare distribution with  $(n-1)p$  degrees of freedom when the estimate of  $\mu$ , namely  $\hat{\mu} = \bar{X}$ , is utilized to compute  $S$ ; when  $\mu$  is specified,  $w$  has a chisquare distribution having  $np$  degrees of freedom where  $n$  is the sample size.

### The non-null density of $w$

The non-null density of  $w$  is available from (6.4.4). Let  $\lambda_1, \dots, \lambda_p$  be the eigenvalues of  $\Sigma^{\frac{1}{2}} \Sigma_o^{-1} \Sigma^{\frac{1}{2}}$ . Then  $L_w(s)$  in (6.4.4) can be re-expressed as follows:

$$L_w(s) = \prod_{j=1}^p (1 + 2\lambda_j s)^{-\frac{m}{2}}. \quad (6.4.7)$$

This is the Laplace transform of a variable of the form  $w = \lambda_1 w_1 + \dots + \lambda_p w_p$  where  $w_1, \dots, w_p$  are independently distributed real scalar chisquare random variables, each having  $m = n - 1$  degrees of freedom, where  $\lambda_j > 0$ ,  $j = 1, \dots, p$ . The distribution of linear combinations of chisquare random variables corresponds to the distribution of quadratic forms; the reader may refer to Mathai and Provost (1992) for explicit representations of their density functions.

**Note 6.4.1.** If the population mean value  $\mu$  is known, then one can proceed by making use of  $\mu$  instead of the sample mean to determine  $S_\mu$ , in which case  $n$ , the sample size, ought to be used instead of  $m = n - 1$  in the above discussion.

#### 6.4.1. Arbitrary moments of $\lambda$

From (6.4.2), the  $h$ -th moment of the  $\lambda$ -criterion for testing  $H_o : \Sigma = \Sigma_o$  (given) in a real nonsingular  $N_p(\mu, \Sigma)$  population, is obtained as follows:

$$\begin{aligned} \lambda^h &= \frac{e^{\frac{np}{2}}}{n^{\frac{np}{2}}} |\Sigma_o^{-1}|^{\frac{nh}{2}} |S|^{\frac{nh}{2}} e^{-\frac{h}{2} \text{tr}(\Sigma_o^{-1} S)} \Rightarrow \\ E[\lambda^h] &= \frac{e^{\frac{np}{2}}}{n^{\frac{np}{2}} 2^{\frac{(n-1)p}{2}} |\Sigma_o|^{\frac{nh}{2}} |\Sigma|^{\frac{n-1}{2}} \Gamma_p(\frac{n-1}{2})} \int_{S>O} |S|^{\frac{nh}{2} + \frac{n-1}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{h}{2} \text{tr}(\Sigma_o^{-1} S)} dS \\ &= \frac{e^{\frac{np}{2}} 2^{p(\frac{nh}{2} + \frac{n-1}{2})} \Gamma_p(\frac{nh}{2} + \frac{n-1}{2})}{n^{\frac{np}{2}} 2^{\frac{(n-1)p}{2}} |\Sigma_o|^{\frac{nh}{2}} |\Sigma|^{\frac{n-1}{2}} \Gamma_p(\frac{n-1}{2})} |\Sigma^{-1} + h \Sigma_o^{-1}|^{-(\frac{nh}{2} + \frac{n-1}{2})} \\ &= e^{\frac{np}{2}} \left(\frac{2}{n}\right)^{\frac{np}{2}} \frac{|\Sigma|^{\frac{nh}{2}} \Gamma_p(\frac{nh}{2} + \frac{n-1}{2})}{|\Sigma_o|^{\frac{nh}{2}} \Gamma_p(\frac{n-1}{2})} |I + h \Sigma \Sigma_o^{-1}|^{-(\frac{nh}{2} + \frac{n-1}{2})} \end{aligned} \quad (6.4.8)$$

for  $\Re(\frac{nh}{2} + \frac{n-1}{2}) > \frac{p-1}{2}$ ,  $I + h\Sigma\Sigma_o^{-1} > O$ . Under  $H_o : \Sigma = \Sigma_o$ , we have  $|I + h\Sigma\Sigma_o^{-1}|^{-(\frac{nh}{2} + \frac{n-1}{2})} = (1+h)^{-p(\frac{nh}{2} + \frac{n-1}{2})}$ . Thus, the  $h$ -th null moment is given by

$$E[\lambda^h | H_o] = e^{\frac{np}{2}h} \left(\frac{2}{n}\right)^{\frac{np}{2}} \frac{\Gamma_p(\frac{nh}{2} + \frac{n-1}{2})}{\Gamma_p(\frac{n-1}{2})} (1+h)^{-p(\frac{nh}{2} + \frac{n-1}{2})} \quad (6.4.9)$$

for  $\Re(\frac{nh}{2} + \frac{n-1}{2}) > \frac{p-1}{2}$ .

#### 6.4.2. The asymptotic distribution of $-2 \ln \lambda$ when testing $H_o : \Sigma = \Sigma_o$

Let us determine the asymptotic distribution of  $-2 \ln \lambda$  where  $\lambda$  is the likelihood ratio statistic for testing  $H_o : \Sigma = \Sigma_o$  (specified) in a real nonsingular  $N_p(\mu, \Sigma)$  population, as  $n \rightarrow \infty$ ,  $n$  being the sample size. This distribution can be determined by expanding both real matrix-variate gamma functions appearing in (6.4.9) and applying Stirling's approximation formula as given in (6.5.14) by letting  $\frac{n}{2}(1+h) \rightarrow \infty$  in the numerator gamma functions and  $\frac{n}{2} \rightarrow \infty$  in the denominator gamma functions. Then, we have

$$\begin{aligned} \frac{\Gamma_p(\frac{nh}{2} + \frac{n-1}{2})}{\Gamma_p(\frac{n-1}{2})} &= \prod_{j=1}^p \frac{\Gamma(\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j-1}{2})}{\Gamma(\frac{n}{2} - \frac{1}{2} - \frac{j-1}{2})} \\ &\rightarrow \prod_{j=1}^p \frac{(2\pi)^{\frac{1}{2}} [\frac{n}{2}(1+h)]^{\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j}{2}} e^{-\frac{n}{2}(1+h)}}{(2\pi)^{\frac{1}{2}} [\frac{n}{2}]^{\frac{n}{2} - \frac{1}{2} - \frac{j}{2}} e^{-\frac{n}{2}}} \\ &= \left(\frac{n}{2}\right)^{\frac{np}{2}} e^{-\frac{np}{2}h} (1+h)^{\frac{np}{2}(1+h) - \frac{p}{2} - \frac{p(p+1)}{4}}. \end{aligned}$$

Hence, from (6.4.9)

$$E[\lambda^h | H_o] \rightarrow (1+h)^{-\frac{p(p+1)}{4}} \text{ as } n \rightarrow \infty, \quad (6.4.10)$$

where  $(1+h)^{-\frac{p(p+1)}{4}}$  is the  $h$ -th moment of the distribution of  $e^{-y/2}$  when  $y \sim \chi_{\frac{p(p+1)}{2}}^2$ . Thus, under  $H_o$ ,  $-2 \ln \lambda \rightarrow \chi_{\frac{p(p+1)}{2}}^2$  as  $n \rightarrow \infty$ . For general procedures leading to asymptotic normality, see Mathai (1982).

**Theorem 6.4.2.** *Letting  $\lambda$  be the likelihood ratio statistic for testing  $H_o : \Sigma = \Sigma_o$  (given) on the covariance matrix of a real nonsingular  $N_p(\mu, \Sigma)$  distribution, the null distribution of  $-2 \ln \lambda$  is asymptotically (as then sample size tends to  $\infty$ ) that of a real scalar chisquare random variable having  $\frac{p(p+1)}{2}$  degrees of freedom, where  $n$  denotes the sample size. This number of degrees of freedom is also equal to the number of parameters restricted by the null hypothesis.*



**Note 6.4.2.** Sugiura and Nagao (1968) have shown that the test based on the statistic  $\lambda$  as specified in (6.4.2) is biased whereas it becomes unbiased upon replacing  $n$ , the sample size, by the degrees of freedom  $n - 1$  in (6.4.2). Accordingly, percentage points are then computed for  $-2 \ln \lambda_1$ , where  $\lambda_1$  is the statistic  $\lambda$  given in (6.4.2) wherein  $n - 1$  is substituted to  $n$ . Korin (1968), Davis (1971), and Nagarsenker and Pillai (1973) computed 5% and 1% percentage points for this test statistic. Davis and Field (1971) evaluated the percentage points for  $p = 2(1)10$  and  $n = 6(1)30(5)50, 60, 120$  and Korin (1968), for  $p = 2(1)10$ .

**Example 6.4.1.** Let us take the same 3-variate real Gaussian population  $N_3(\mu, \Sigma)$ ,  $\Sigma > O$  and the same data as in Example 6.3.1, so that intermediate calculations could be utilized. The sample size is 5 and the sample values are the following:

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, X_3 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix},$$

the sample average and the sample sum of products matrix being

$$\bar{X} = \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, S = \frac{1}{5^2} \begin{bmatrix} 270 & -110 & -170 \\ -110 & 80 & 85 \\ -170 & 85 & 170 \end{bmatrix}.$$

Let us consider the hypothesis  $\Sigma = \Sigma_o$  where

$$\Sigma_o = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix} \Rightarrow |\Sigma_o| = 10, \text{Cof}(\Sigma_o) = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix};$$

$$\Sigma_o^{-1} = \frac{\text{Cof}(\Sigma_o)}{|\Sigma_o|} = \frac{1}{10} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix};$$

$$\begin{aligned} \text{tr}(\Sigma_o^{-1}S) &= \frac{1}{(10)(5^2)} \text{tr} \left\{ \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix} \begin{bmatrix} 270 & -110 & -170 \\ -110 & 80 & 85 \\ -170 & 85 & 170 \end{bmatrix} \right\} \\ &= \frac{1}{(10)(5^2)} [5(270) + (4(80) + 2(85)) + (2(85) + 6(170))] \\ &= 12.12; n = 5, p = 3. \end{aligned}$$

Let us test the null hypothesis at the significance level  $\alpha = 0.05$ . The distribution of the test statistic  $w$  and the tabulated critical value are as follows:

$$w = \text{tr}(\Sigma_o^{-1}S) \sim \chi_{(n-1)p}^2 \simeq \chi_{12}^2; \chi_{12, 0.05}^2 = 21.03.$$

As the observed value  $12.12 < 21.03$ ,  $H_0$  is not rejected. The asymptotic distribution of  $-2 \ln \lambda$ , as  $n \rightarrow \infty$ , is  $\chi_{p(p+1)/2}^2 \simeq \chi_6^2$  where  $\lambda$  is the likelihood ratio criterion statistic. Since  $\chi_{6, 0.05}^2 = 12.59$  and  $12.59 > 12.12$ , we still do not reject  $H_0$  as  $n \rightarrow \infty$ .

### 6.4.3. Tests on Wilks' concept of generalized variance

The concept of *generalized variance* was explained in Chap. 5. The sample generalized variance is simply the determinant of  $S$ , the sample sum of products matrix. When the population is  $p$ -variate Gaussian, it has already been shown in Chap. 5 that  $S$  is Wishart distributed with  $m = n - 1$  degrees of freedom,  $n$  being the sample size, and parameter matrix  $\Sigma > O$ , which is the population covariance matrix. When the population is multivariate normal, several types of tests of hypotheses involve the sample generalized variance. The first author has given the exact distributions of such tests, see Mathai (1972a,b) and Mathai and Rathie (1971).

### 6.5. The Sphericity Test or Testing if $H_0 : \Sigma = \sigma^2 I$ , Given a $N_p(\mu, \Sigma)$ Sample

When the covariance matrix  $\Sigma = \sigma^2 I$ , where  $\sigma^2 > 0$  is a real scalar quantity, the ellipsoid  $(X - \mu)' \Sigma^{-1} (X - \mu) = c > 0$ , which represents a specific contour of constant density for a nonsingular  $N_p(\mu, \Sigma)$  distribution, becomes the sphere defined by the equation  $\frac{1}{\sigma^2} (X - \mu)' (X - \mu) = c$  or  $\frac{1}{\sigma^2} ((x_1 - \mu_1)^2 + \dots + (x_p - \mu_p)^2) = c > 0$ , whose center is located at the point  $\mu$ ; hence the test's name, the sphericity test. Given a  $N_p(\mu, \Sigma)$  sample of size  $n$ , the maximum of the likelihood function in the entire parameter space is

$$\sup_{\Omega} L = \frac{n^{\frac{np}{2}} e^{-\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}},$$

as was previously established. However, under the null hypothesis  $H_0 : \Sigma = \sigma^2 I$ ,  $\text{tr}(\Sigma^{-1} S) = (\sigma^2)^{-1} (\text{tr}(S))$  and  $|\Sigma| = (\sigma^2)^p$ . Thus, if we let  $\theta = \sigma^2$  and substitute  $\hat{\mu} = \bar{X}$  in  $L$ , under  $H_0$  the loglikelihood function will be  $\ln L_{\omega} = -\frac{np}{2} \ln(2\pi) - \frac{np}{2} \ln \theta - \frac{1}{2\theta} \text{tr}(S)$ . Differentiating this function with respect to  $\theta$  and equating the result to zero produces the following estimator for  $\theta$ :

$$\hat{\theta} = \hat{\sigma}^2 = \frac{\text{tr}(S)}{np}. \quad (6.5.1)$$

Accordingly, the maximum of the likelihood function under  $H_0$  is the following:

$$\max_{\omega} L = \frac{n^{\frac{np}{2}} e^{-\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} \left[ \frac{\text{tr}(S)}{p} \right]^{\frac{n}{2}}}.$$

Thus, the  $\lambda$ -criterion for testing

$$H_o : \Sigma = \sigma^2 I, \sigma^2 > 0 \text{ (unknown)}$$

is

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = \frac{|S|^{\frac{n}{2}}}{[\frac{\text{tr}(S)}{p}]^{\frac{n}{2}}}. \quad (6.5.2)$$

In the complex Gaussian case when  $\tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$ ,  $\Sigma = \Sigma^* > O$  where an asterisk indicates the conjugate transpose,  $\tilde{X}_j = X_{j1} + iX_{j2}$  where  $X_{j1}$  and  $X_{j2}$  are real  $p \times 1$  vectors and  $i = \sqrt{-1}$ . The covariance matrix associated with  $\tilde{X}_j$  is then defined as

$$\begin{aligned} \text{Cov}(\tilde{X}_j) &= E[(\tilde{X}_j - \tilde{\mu})(\tilde{X}_j - \tilde{\mu})^*] \\ &= E[(X_{j1} - \mu_{(1)}) + i(X_{j2} - \mu_{(2)})][(X_{j1} - \mu_{(1)})' - i(X_{j2} - \mu_{(2)})'] \\ &= \Sigma_{11} + \Sigma_{22} + i(\Sigma_{21} - \Sigma_{12}) \equiv \Sigma, \text{ with } \mu = \mu_{(1)} + i\mu_{(2)}, \end{aligned}$$

where  $\Sigma$  is assumed to be Hermitian positive definite, with  $\Sigma_{11} = \text{Cov}(X_{j1})$ ,  $\Sigma_{22} = \text{Cov}(X_{j2})$ ,  $\Sigma_{12} = \text{Cov}(X_{j1}, X_{j2})$  and  $\Sigma_{21} = \text{Cov}(X_{j2}, X_{j1})$ . Thus, the hypothesis of sphericity in the complex Gaussian case is  $\Sigma = \sigma^2 I$  where  $\sigma$  is real and positive. Then, under the null hypothesis  $\tilde{H}_o : \Sigma = \sigma^2 I$ , the Hermitian form  $\tilde{Y}^* \Sigma \tilde{Y} = c > 0$  where  $c$  is real and positive, becomes  $\sigma^2 \tilde{Y}^* \tilde{Y} = c \Rightarrow |\tilde{y}_1|^2 + \dots + |\tilde{y}_p|^2 = \frac{c}{\sigma^2} > 0$ , which defines a sphere in the complex space, where  $|\tilde{y}_j|$  denotes the absolute value or modulus of  $\tilde{y}_j$ . If  $\tilde{y}_j = y_{j1} + iy_{j2}$  with  $i = \sqrt{-1}$ ,  $y_{j1}, y_{j2}$  being real, then  $|\tilde{y}_j|^2 = y_{j1}^2 + y_{j2}^2$ .

The joint density of the sample values in the real Gaussian case is the following:

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2}(X_j - \mu)' \Sigma^{-1} (X_j - \mu)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - n(\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}$$

where  $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ ,  $X_j, j = 1, \dots, n$  are iid  $N_p(\mu, \Sigma)$ ,  $\Sigma > O$ . We have already derived the maximum of  $L$  in the entire parameter space  $\Omega$ , which, in the real case, is

$$\sup_{\Omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}}, \quad (6.5.3)$$

where  $S$  is the sample sum of products matrix. Under  $H_o$ ,  $|\Sigma|^{\frac{n}{2}} = (\sigma^2)^{\frac{np}{2}}$  and  $\text{tr}(\Sigma^{-1} S) = \frac{1}{\sigma^2}(s_{11} + \dots + s_{pp}) = \frac{1}{\sigma^2} \text{tr}(S)$ . Thus, the maximum likelihood estimator of  $\sigma^2$  is  $\frac{1}{np} \text{tr}(S)$ . Accordingly, the  $\lambda$ -criterion is

$$\lambda = |S|^{\frac{n}{2}} / \left( \frac{\text{tr}(S)}{p} \right)^{\frac{np}{2}} \Rightarrow u_1 = \lambda^{\frac{2}{n}} = \frac{p^p |S|}{[\text{tr}(S)]^p}, \quad (6.5.4)$$

in the real case. Interestingly,  $(u_1)^{1/p}$  is the ratio of the geometric mean of the eigenvalues of  $S$  to their arithmetic mean. The structure remains the same in the complex domain, in which case  $\det(\tilde{S})$  is replaced by the absolute value  $|\det(\tilde{S})|$  so that

$$\tilde{u}_1 = \frac{p^p |\det(\tilde{S})|}{[\text{tr}(\tilde{S})]^p}. \quad (6.5a.1)$$

For arbitrary  $h$ , the  $h$ -th moment of  $u_1$  in the real case can be obtained by integrating out over the density of  $S$ , which, as explained in Sect. 5.5, 5.5a, is a Wishart density with  $n - 1 = m$  degrees of freedom and parameter matrix  $\Sigma > O$ . However, when the null hypothesis  $H_o$  holds,  $\Sigma = \sigma^2 I_p$ , so that the  $h$ -th moment in the real case is

$$E[u_1^h | H_o] = \frac{p^{ph}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})(\sigma^2)^{\frac{mp}{2}}} \int_{S>O} |S|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\frac{1}{2\sigma^2}\text{tr}(S)} (\text{tr}(S))^{-ph} dS. \quad (i)$$

In order to evaluate this integral, we replace  $[\text{tr}(S)]^{-ph}$  by an equivalent integral:

$$[\text{tr}(S)]^{-ph} = \frac{1}{\Gamma(ph)} \int_{x=0}^{\infty} x^{ph-1} e^{-x(\text{tr}(S))} dx, \quad \Re(h) > 0. \quad (ii)$$

Then, substituting (ii) in (i), the exponent becomes  $-\frac{1}{2\sigma^2}(1 + 2\sigma^2 x)(\text{tr}(S))$ . Now, letting  $S_1 = \frac{1}{2\sigma^2}(1 + 2\sigma^2 x)S \Rightarrow dS = (2\sigma^2)^{\frac{p(p+1)}{2}} (1 + 2\sigma^2 x)^{-\frac{p(p+1)}{2}} dS_1$ , and we have

$$\begin{aligned} E[u_1^h | H_o] &= \frac{(2\sigma^2)^{ph}}{\Gamma_p(\frac{m}{2})} \frac{p^{ph}}{\Gamma(ph)} \int_0^{\infty} x^{ph-1} (1 + 2\sigma^2 x)^{-\frac{(m}{2}+h)p} dx \\ &\quad \times \int_{S_1>O} |S_1|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\text{tr}(S_1)} dS_1 \\ &= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} \frac{p^{ph}}{\Gamma(ph)} \int_0^{\infty} y^{ph-1} (1 + y)^{-\frac{(m}{2}+h)p} dy, \quad y = 2\sigma^2 x \\ &= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} p^{ph} \frac{\Gamma(\frac{mp}{2})}{\Gamma(\frac{mp}{2} + ph)}, \quad \Re(h) > 0, \quad m = n - 1. \end{aligned} \quad (6.5.5)$$

The corresponding  $h$ -th moment in the complex case is the following:

$$E[\tilde{u}_1^h | H_o] = \frac{\tilde{\Gamma}_p(m + h)}{\tilde{\Gamma}_p(m)} p^{ph} \frac{\tilde{\Gamma}(mp)}{\tilde{\Gamma}(mp + ph)}, \quad \Re(h) > 0, \quad m = n - 1. \quad (6.5a.2)$$

By making use of the multiplication formula for gamma functions, one can expand the real gamma function  $\Gamma(mz)$  as follows:

$$\Gamma(mz) = (2\pi)^{\frac{1-m}{2}} m^{mz-\frac{1}{2}} \Gamma(z) \Gamma(z + \frac{1}{m}) \cdots \Gamma(z + \frac{m-1}{m}), \quad m = 1, 2, \dots, \quad (6.5.6)$$

and for  $m = 2$ , we have the duplication formula

$$\Gamma(2z) = \pi^{-\frac{1}{2}} 2^{2z-1} \Gamma(z) \Gamma\left(z + \frac{1}{2}\right). \quad (6.5.7)$$

Then on applying (6.5.6),

$$\frac{p^{ph} \Gamma\left(\frac{mp}{2}\right)}{\Gamma\left(\frac{mp}{2} + ph\right)} = \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m}{2} + h\right)} \prod_{j=1}^{p-1} \frac{\Gamma\left(\frac{m}{2} + \frac{j}{p}\right)}{\Gamma\left(\frac{m}{2} + h + \frac{j}{p}\right)}. \quad (iii)$$

Moreover, it follows from the definition of the real matrix-variate gamma functions that

$$\frac{\Gamma_p\left(\frac{m}{2} + h\right)}{\Gamma_p\left(\frac{m}{2}\right)} = \prod_{j=1}^p \frac{\Gamma\left(\frac{m}{2} - \frac{j-1}{2} + h\right)}{\Gamma\left(\frac{m}{2} - \frac{j-1}{2}\right)}. \quad (iv)$$

On canceling  $\Gamma\left(\frac{m}{2} + h\right)/\Gamma\left(\frac{m}{2}\right)$  when multiplying (iii) by (iv), we are left with

$$E[u_1^h | H_o] = \left\{ \prod_{j=1}^{p-1} \frac{\Gamma\left(\frac{m}{2} + \frac{j}{p}\right)}{\Gamma\left(\frac{m}{2} - \frac{j}{2}\right)} \right\} \left\{ \prod_{j=1}^{p-1} \frac{\Gamma\left(\frac{m}{2} - \frac{j}{2} + h\right)}{\Gamma\left(\frac{m}{2} + \frac{j}{p} + h\right)} \right\}, \quad m = n - 1. \quad (6.5.8)$$

The corresponding  $h$ -th moment in the complex case is the following:

$$E[\tilde{u}_1^h | H_o] = \left\{ \prod_{j=1}^{p-1} \frac{\tilde{\Gamma}\left(m + \frac{j}{p}\right)}{\tilde{\Gamma}(m - j)} \right\} \left\{ \prod_{j=1}^{p-1} \frac{\tilde{\Gamma}(m - j + h)}{\tilde{\Gamma}\left(m + \frac{j}{p} + h\right)} \right\}, \quad m = n - 1. \quad (6.5a.3)$$

For  $h = s - 1$ , one can treat  $E[u_1^{s-1} | H_o]$  as the Mellin transform of the density of  $u_1$  in the real case. Letting this density be denoted by  $f_{u_1}(u_1)$ , it can be expressed in terms of a G-function as follows:

$$f_{u_1}(u_1 | H_o) = c_1 G_{p-1, p-1}^{p-1, 0} \left[ u_1 \left| \begin{matrix} \frac{m}{2} + \frac{j}{p} - 1, j=1, \dots, p-1 \\ \frac{m}{2} - \frac{j}{2} - 1, j=1, \dots, p-1 \end{matrix} \right. \right], \quad 0 \leq u_1 \leq 1, \quad (6.5.9)$$

and  $f_{u_1}(u_1 | H_o) = 0$  elsewhere, where

$$c_1 = \left\{ \prod_{j=1}^{p-1} \frac{\Gamma\left(\frac{m}{2} + \frac{j}{p}\right)}{\Gamma\left(\frac{m}{2} - \frac{j}{2}\right)} \right\},$$

the corresponding density in the complex case being the following:

$$\tilde{f}_{\tilde{u}_1|H_o}(\tilde{u}_1) = \tilde{c}_1 \tilde{G}_{p-1,p-1}^{p-1,0} \left[ \tilde{u}_1 \left| \begin{matrix} m+\frac{j}{p}-1, & j=1,\dots,p-1 \\ m-j-1, & j=1,\dots,p-1 \end{matrix} \right. \right], \quad 0 \leq |\tilde{u}_2| \leq 1, \quad (6.5a.4)$$

and  $\tilde{f}_{\tilde{u}_1}(\tilde{u}_1) = 0$  elsewhere, where  $\tilde{G}$  is a real G-function whose parameters are different from those appearing in (6.5.9), and

$$\tilde{c}_1 = \left\{ \prod_{j=1}^{p-1} \frac{\tilde{\Gamma}(m + \frac{j}{p})}{\tilde{\Gamma}(m - j)} \right\}.$$

For computable series representation of a G-function with general parameters, the reader may refer to Mathai (1970a, 1993). Observe that  $u_1$  in the real case is structurally a product of  $p - 1$  mutually independently distributed real scalar type-1 beta random variables with the parameters  $(\alpha_j = \frac{m}{2} - \frac{j}{2}, \beta_j = \frac{j}{2} + \frac{j}{p})$ ,  $j = 1, \dots, p - 1$ . In the complex case,  $\tilde{u}_1$  is structurally a product of  $p - 1$  mutually independently distributed real scalar type-1 beta random variables with the parameters  $(\alpha_j = m - j, \beta_j = j + \frac{j}{p})$ ,  $j = 1, \dots, p - 1$ . This observation is stated as a result.

**Theorem 6.5.1.** *Consider the sphericity test statistic for testing the hypothesis  $H_o : \Sigma = \sigma^2 I$  where  $\sigma^2 > 0$  is an unknown real scalar. Let  $u_1$  and the corresponding complex quantity  $\tilde{u}_1$  be as defined in (6.5.4) and (6.5a.1) respectively. Then, in the real case,  $u_1$  is structurally a product of  $p - 1$  independently distributed real scalar type-1 beta random variables with the parameters  $(\alpha_j = \frac{m}{2} - \frac{j}{2}, \beta_j = \frac{j}{2} + \frac{j}{p})$ ,  $j = 1, \dots, p - 1$ , and, in the complex case,  $\tilde{u}_1$  is structurally a product of  $p - 1$  independently distributed real scalar type-1 beta random variables with the parameters  $(\alpha_j = m - j, \beta_j = j + \frac{j}{p})$ ,  $j = 1, \dots, p - 1$ , where  $m = n - 1$ ,  $n =$  the sample size.*

For certain special cases, one can represent (6.5.9) and (6.5a.4) in terms of known elementary functions. Some such cases are now being considered.

**Real case:  $p = 2$**

In the real case, for  $p = 2$

$$E[u_1^h | H_o] = \frac{\Gamma(\frac{m}{2} + \frac{1}{2}) \Gamma(\frac{m}{2} - \frac{1}{2} + h)}{\Gamma(\frac{m}{2} - \frac{1}{2}) \Gamma(\frac{m}{2} + \frac{1}{2} + h)} = \frac{\frac{m}{2} - \frac{1}{2}}{\frac{m}{2} - \frac{1}{2} + h}.$$

This means  $u_1$  is a real type-1 beta variable with the parameters  $(\alpha = \frac{m}{2} - \frac{1}{2}, \beta = 1)$ . The corresponding result in the complex case is that  $\tilde{u}_1$  is a real type-1 beta variable with the parameters  $(\alpha = m - 1, \beta = 1)$ .

**Real case:**  $p = 3$

In the real case

$$E[u_1^h | H_0] = \frac{\Gamma(\frac{m}{2} + \frac{1}{3})\Gamma(\frac{m}{2} + \frac{2}{3})}{\Gamma(\frac{m}{2} - \frac{1}{2})\Gamma(\frac{m}{2} - 1)} \frac{\Gamma(\frac{m}{2} - \frac{1}{2} + h)\Gamma(\frac{m}{2} - 1 + h)}{\Gamma(\frac{m}{2} + \frac{1}{3} + h)\Gamma(\frac{m}{2} + \frac{2}{3} + h)},$$

so that  $u_1$  is equivalent to the product of two independently distributed real type-1 beta random variables with the parameters  $(\alpha_j, \beta_j) = (\frac{m}{2} - \frac{j}{2}, \frac{j}{2} + \frac{j}{3})$ ,  $j = 1, 2$ . This density can be obtained by treating  $E[u_1^h | H_0]$  for  $h = s - 1$  as the Mellin transform of the density of  $u_1$ . The density is then available by taking the inverse Mellin transform. Thus, again denoting it by  $f_{u_1}(u_1)$ , we have

$$\begin{aligned} f_{u_1}(u_1 | H_0) &= c_3 \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \phi_3(s) ds, \quad c > \frac{1}{2} \\ &= c_3 \left[ \sum_{\nu=0}^{\infty} R_{\nu} + \sum_{\nu=0}^{\infty} R'_{\nu} \right], \\ c_3 &= \frac{\Gamma(\frac{m}{2} + \frac{1}{3})\Gamma(\frac{m}{2} + \frac{2}{3})}{\Gamma(\frac{m}{2} - \frac{1}{2})\Gamma(\frac{m}{2} - 1)}, \\ \phi_3(s) &= \frac{\Gamma(\frac{m}{2} - \frac{1}{2} - 1 + s)\Gamma(\frac{m}{2} - 1 - 1 + s)}{\Gamma(\frac{m}{2} + \frac{1}{3} - 1 + s)\Gamma(\frac{m}{2} + \frac{2}{3} - 1 + s)} u_1^{-s}, \end{aligned}$$

where  $R_{\nu}$  is the residue of the integrand  $\phi_3(s)$  at the poles of  $\Gamma(\frac{m}{2} - \frac{3}{2} + s)$  and  $R'_{\nu}$  is the residue of the integrand  $\phi_3(s)$  at the pole of  $\Gamma(\frac{m}{2} - 2 + s)$ . Letting  $s_1 = \frac{m}{2} - \frac{3}{2} + s$ ,

$$\begin{aligned} R_{\nu} &= \lim_{s \rightarrow -\nu + \frac{3}{2} - \frac{m}{2}} \phi_3(s) = \lim_{s_1 \rightarrow -\nu} [(s_1 + \nu) u_1^{\frac{m}{2} - \frac{3}{2}} \frac{\Gamma(s_1)\Gamma(-\frac{1}{2} + s_1)u_1^{-s_1}}{\Gamma(\frac{1}{3} + \frac{1}{2} + s_1)\Gamma(\frac{2}{3} + \frac{1}{2} + s_1)}] \\ &= u_1^{\frac{m}{2} - \frac{3}{2}} \frac{(-1)^{\nu}}{\nu!} \frac{\Gamma(-\frac{1}{2} - \nu)}{\Gamma(\frac{1}{3} + \frac{1}{2} - \nu)\Gamma(\frac{2}{3} + \frac{1}{2} - \nu)} u_1^{\nu}. \end{aligned}$$

We can replace negative  $\nu$  in the arguments of the gamma functions with positive  $\nu$  by making use of the following formula:

$$\Gamma(a - \nu) = \frac{(-1)^{\nu} \Gamma(a)}{(-a + 1)_{\nu}}, \quad a \neq 1, 2, \dots, \quad \nu = 0, 1, \dots, \quad (6.5.10)$$

where for example,  $(b)_{\nu}$  is the Pochhammer symbol

$$(b)_{\nu} = b(b+1) \cdots (b+\nu-1), \quad b \neq 0, \quad (b)_0 = 1, \quad (6.5.11)$$

so that

$$\begin{aligned} \Gamma\left(-\frac{1}{2} - \nu\right) &= \frac{(-1)^\nu \Gamma(-\frac{1}{2})}{(\frac{3}{2})_\nu}, \quad \Gamma\left(\frac{1}{3} + \frac{1}{2} - \nu\right) = \frac{(-1)^\nu \Gamma(\frac{1}{3} + \frac{1}{2})}{(-\frac{1}{3} + \frac{1}{2})_\nu}, \\ \Gamma\left(\frac{2}{3} + \frac{1}{2} - \nu\right) &= \frac{(-1)^\nu \Gamma(\frac{2}{3} + \frac{1}{2})}{(-\frac{2}{3} + \frac{1}{2})_\nu}. \end{aligned}$$

The sum of the residues then becomes

$$\sum_{\nu=0}^{\infty} R_\nu = \frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{1}{3} + \frac{1}{2})\Gamma(\frac{2}{3} + \frac{1}{2})} u_1^{\frac{m}{2}-\frac{3}{2}} {}_2F_1\left(-\frac{1}{3} + \frac{1}{2}; -\frac{2}{3} + \frac{1}{2}; \frac{3}{2}; u_1\right), \quad 0 \leq u_1 \leq 1.$$

It can be similarly shown that

$$\sum_{\nu=0}^{\infty} R'_\nu = \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{3} + 1)\Gamma(\frac{2}{3} + 1)} u_1^{\frac{m}{2}-2} {}_2F_1\left(-\frac{1}{3}, -\frac{2}{3}; \frac{1}{2}; u_1\right), \quad 0 \leq u_1 \leq 1.$$

Accordingly, the density of  $u_1$  for  $p = 3$  is the following:

$$\begin{aligned} f_1(u_1|H_o) &= c_3 \left\{ \frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{5}{6})\Gamma(\frac{7}{6})} u_1^{\frac{m}{2}-\frac{3}{2}} {}_2F_1\left(\frac{1}{6}, -\frac{1}{6}; \frac{3}{2}; u_1\right) \right. \\ &\quad \left. + \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{4}{3})\Gamma(\frac{5}{3})} u_1^{\frac{m}{2}-2} {}_2F_1\left(-\frac{1}{3}, -\frac{2}{3}; \frac{1}{2}; u_1\right) \right\}, \quad 0 \leq u_1 \leq 1 \quad (6.5.12) \end{aligned}$$

and  $f_{u_1}(u_1|H_o) = 0$  elsewhere.

**Real case:**  $p = 4$

In this case,

$$E[u_1^h|H_o] = c_4 \frac{\Gamma(\frac{m}{2} - \frac{3}{2} + s)\Gamma(\frac{m}{2} - 2 + s)\Gamma(\frac{m}{2} - \frac{5}{2} + s)}{\Gamma(\frac{m}{2} - \frac{3}{4} + s)\Gamma(\frac{m}{2} - \frac{2}{4} + s)\Gamma(\frac{m}{2} - \frac{1}{4} + s)},$$

where  $c_4$  is the normalizing constant. However, noting that

$$\frac{\Gamma(\frac{m}{2} - \frac{3}{2} + s)}{\Gamma(\frac{m}{2} - \frac{1}{2} + s)} = \frac{1}{\frac{m}{2} - \frac{3}{2} + s},$$



there is one pole at  $s = -\frac{m}{2} + \frac{3}{2}$ . The poles of  $\Gamma(\frac{m}{2} - \frac{5}{2} + s)$  occur at  $s = -\frac{m}{2} + \frac{5}{2} - \nu$ ,  $\nu = 0, 1, \dots$ , and hence at  $\nu = 1$ , the pole coincides with the earlier pole and there is a pole of order 2 at  $s = -\frac{m}{2} + \frac{3}{2}$ . Each one of other poles of the integrand is simple, that is, of order 1. The second order pole will bring in a logarithmic function. As all the cases for which  $p \geq 4$  will bring in poles of higher orders, they will not be herein discussed. The general expansion of a G-function of the type  $G_{m,m}^{m,0}(\cdot)$  is provided in Mathai (1970a, 1993). In the complex case, starting from  $p \geq 3$ , poles of higher orders are coming in, so that the densities can only be written in terms of logarithms, psi and zeta functions; hence, these will not be considered. Observe that  $\tilde{u}_1$  corresponds to product of independently distributed real type-1 beta random variables, even though the densities are available only in terms of logarithms, psi and zeta functions for  $p \geq 3$ . The null and non-null densities of the  $\lambda$ -criterion in the general case, were derived by the first author and some results obtained under the null distribution can also be found in Mathai and Saxena (1973). Several researchers have contributed to various aspects of the sphericity and multi-sample sphericity tests; for some of the first author's contributions, the reader may refer to Mathai and Rathie (1970) and Mathai (1977, 1984, 1986).

Gamma products such as those appearing in (6.5.8) and (6.5a.3) are frequently encountered when considering various types of tests on the parameters of a real or complex Gaussian or certain other types of distributions. Structural representations in the form of product of independently distributed real scalar type-1 beta random variables occur in numerous situations. Thus, a general asymptotic result on the  $h$ -th moment of such products of type-1 beta random variables will be derived. This is now stated as a result.

**Theorem 6.5.2.** *Let  $u$  be a real scalar random variable whose  $h$ -th moment is of the form*

$$E[u^h] = \frac{\Gamma_p(\alpha + \alpha h + \gamma)}{\Gamma_p(\alpha + \gamma)} \frac{\Gamma_p(\alpha + \gamma + \delta)}{\Gamma_p(\alpha + \alpha h + \gamma + \delta)} \quad (6.5.13)$$

where  $\Gamma_p(\cdot)$  is a real matrix-variate gamma function on  $p \times p$  real positive definite matrices,  $\alpha$  is real,  $\gamma$  is bounded,  $\delta$  is real,  $0 < \delta < \infty$  and  $h$  is arbitrary. Then, as  $\alpha \rightarrow \infty$ ,  $-2 \ln u \rightarrow \chi_{2p\delta}^2$ , a real chisquare random variable having  $2p\delta$  degrees of freedom, that is, a real gamma random variable with the parameters ( $\alpha = p\delta$ ,  $\beta = 2$ ).

**Proof:** On expanding the real matrix-variate gamma functions, we have the following:

$$\frac{\Gamma_p(\alpha + \gamma + \delta)}{\Gamma_p(\alpha + \gamma)} = \prod_{j=1}^p \frac{\Gamma(\alpha + \gamma + \delta - \frac{j-1}{2})}{\Gamma(\alpha + \gamma - \frac{j-1}{2})}. \quad (i)$$

$$\frac{\Gamma_p(\alpha(1+h) + \gamma)}{\Gamma_p(\alpha(1+h) + \gamma + \delta)} = \prod_{j=1}^p \frac{\Gamma(\alpha(1+h) + \gamma - \frac{j-1}{2})}{\Gamma(\alpha(1+h) + \gamma + \delta - \frac{j-1}{2})}. \quad (ii)$$

Consider the following form of Stirling's asymptotic approximation formula for gamma functions, namely,

$$\Gamma(z + \eta) \approx \sqrt{2\pi} z^{z+\eta-\frac{1}{2}} e^{-z} \text{ for } |z| \rightarrow \infty \text{ and } \eta \text{ bounded.} \quad (6.5.14)$$

On applying this asymptotic formula to the gamma functions appearing in (i) and (ii) for  $\alpha \rightarrow \infty$ , we have

$$\prod_{j=1}^p \frac{\Gamma(\alpha + \gamma + \delta - \frac{j-1}{2})}{\Gamma(\alpha + \gamma - \frac{j-1}{2})} \rightarrow \alpha^{p\delta}$$

and

$$\prod_{j=1}^p \frac{\Gamma(\alpha(1+h) + \gamma - \frac{j-1}{2})}{\Gamma(\alpha(1+h) + \gamma + \delta - \frac{j-1}{2})} \rightarrow [\alpha(1+h)]^{-p\delta}, \quad (iii)$$

so that

$$E[u^h] \rightarrow (1+h)^{-p\delta}. \quad (iv)$$

On noting that  $E[u^h] = E[e^{h \ln u}] \rightarrow (1+h)^{-p\delta}$ , it is seen that  $\ln u$  has the mgf  $(1+h)^{-p\delta}$  for  $1+h > 0$  or  $-2 \ln u$  has mgf  $(1-2h)^{-p\delta}$  for  $1-2h > 0$ , which happens to be the mgf of a real scalar chisquare variable with  $2p\delta$  degrees of freedom if  $2p\delta$  is a positive integer or a real gamma variable with the parameters  $(\alpha = p\delta, \beta = 2)$ . Hence the following result.

**Corollary 6.5.1.** Consider a slightly more general case than that considered in Theorem 6.5.2. Let the  $h$ -th moment of  $u$  be of the form

$$E[u^h] = \left\{ \prod_{j=1}^p \frac{\Gamma(\alpha(1+h) + \gamma_j)}{\Gamma(\alpha + \gamma_j)} \right\} \left\{ \prod_{j=1}^p \frac{\Gamma(\alpha + \gamma_j + \delta_j)}{\Gamma(\alpha(1+h) + \gamma_j + \delta_j)} \right\}. \quad (6.5.15)$$

Then as  $\alpha \rightarrow \infty$ ,  $E[u^h] \rightarrow (1+h)^{-(\delta_1 + \dots + \delta_p)}$ , which implies that  $-2 \ln u \rightarrow \chi_{2(\delta_1 + \dots + \delta_p)}^2$  whenever  $2(\delta_1 + \dots + \delta_p)$  is a positive integer or, equivalently,  $-2 \ln u$  tends to a real gamma variable with the parameters  $(\alpha = \delta_1 + \dots + \delta_p, \beta = 2)$ .

Let us examine the asymptotic distribution of the test statistic for the sphericity test in the light of Theorem 6.5.2. It is seen from (6.5.4) that  $\lambda^h = u^{h \frac{n}{2}}$ . Thus, by replacing  $h$  by  $\frac{n}{2}h$  in (6.5.8) with  $m = n - 1$ , we have

$$E[\lambda^h | H_0] = \left\{ \prod_{j=1}^{p-1} \frac{\Gamma(\frac{n-1}{2} + \frac{j}{p})}{\Gamma(\frac{n-1}{2} - \frac{j}{2})} \right\} \left\{ \prod_{j=1}^{p-1} \frac{\Gamma(\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j}{2})}{\Gamma(\frac{n}{2}(1+h) - \frac{1}{2} + \frac{j}{p})} \right\}. \quad (6.5.16)$$

Then, it follows from Corollary 6.5.1 that  $-2 \ln \lambda \rightarrow \chi_{2 \sum_{j=1}^{p-1} (\frac{j}{2} + \frac{j}{p})}^2$ , a chi-square random variable having  $2 \sum_{j=1}^{p-1} (\frac{j}{2} + \frac{j}{p}) = \frac{p(p-1)}{2} + (p-1) = \frac{(p-1)(p+2)}{2}$  degrees of freedom. Hence the following result:

**Theorem 6.5.3.** *Consider the  $\lambda$ -criterion for testing the hypothesis of sphericity. Then, under the null hypothesis,  $-2 \ln \lambda \rightarrow \chi_{\frac{(p-1)(p+2)}{2}}^2$ , as the sample size  $n \rightarrow \infty$ . In the complex case, as  $n \rightarrow \infty$ ,  $-2 \ln \lambda \rightarrow \chi_{(p-1)(p+1)}^2$ , a real scalar chisquare variable with  $2[\frac{p(p-1)}{2} + \frac{(p-1)}{2}] = p(p-1) + (p-1) = (p-1)(p+1)$  degrees of freedom.*

**Note 6.5.1.** We observe that the degrees of freedom of the real chisquare variable in the real scalar case is  $\frac{(p-1)(p+2)}{2}$ , which is also equal to the number of parameters restricted by the null hypothesis. Indeed, when  $\Sigma = \sigma^2 I$ , we have  $\sigma_{ij} = 0$ ,  $i \neq j$ , which produces  $\frac{p(p-1)}{2}$  restrictions and, since  $\sigma^2$  is unknown, requiring that the diagonal elements are such that  $\sigma_{11} = \dots = \sigma_{pp}$  produces  $p-1$  additional restrictions for a total of  $\frac{(p-1)(p+2)}{2}$  restrictions being imposed. Thus, the degrees of freedom of the asymptotic chisquare variable corresponds to the number of restrictions imposed by  $H_o$ , which, actually, is a general result.

## 6.6. Testing the Hypothesis that the Covariance Matrix is Diagonal

Consider the null hypothesis that  $\Sigma$ , the nonsingular covariance matrix of a  $p$ -variate real normal distribution, is diagonal, that is,

$$H_o : \Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp}).$$

Since the population is assumed to be normally distributed, this implies that the components of the  $p$ -variate Gaussian vector are mutually independently distributed as univariate normal random variables whose respective variances are  $\sigma_{jj}$ ,  $j = 1, \dots, p$ . Consider a simple random sample of size  $n$  from a nonsingular  $N_p(\mu, \Sigma)$  population or, equivalently, let  $X_1, \dots, X_n$  be independently distributed as  $N_p(\mu, \Sigma)$  vectors,  $\Sigma > 0$ . Under  $H_o$ ,  $\sigma_{jj}$  is estimated by its MLE which is  $\hat{\sigma}_{jj} = \frac{1}{n} s_{jj}$  where  $s_{jj}$  is the  $j$ -th diagonal element of  $S = (s_{ij})$ , the sample sum of products matrix. The maximum of the likelihood function under the null hypothesis is then

$$\max_{H_o} L = \prod_{j=1}^p \max_{H_o} L_j = \frac{1}{(2\pi)^{\frac{np}{2}} \prod_{j=1}^p [s_{jj}]^{\frac{n}{2}}} n^{\frac{np}{2}} e^{-\frac{1}{2}(np)},$$

the likelihood function being the joint density evaluated at an observed value of the sample. Observe that the overall maximum or the maximum in the entire parameter space remains the same as that given in (6.1.1). Thus, the  $\lambda$ -criterion is given by

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = \frac{|S|^{\frac{n}{2}}}{\prod_{j=1}^p s_{jj}^{\frac{n}{2}}} \Rightarrow u_2 = \lambda^{\frac{2}{n}} = \frac{|S|}{\prod_{j=1}^p s_{jj}} \quad (6.6.1)$$

where  $S \sim W_p(m, \Sigma)$ ,  $\Sigma > O$ , and  $m = n - 1$ ,  $n$  being the sample size. Under  $H_0$ ,  $\Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp})$ . Then for an arbitrary  $h$ , the  $h$ -th moment of  $u_2$  is available by taking the expected value of  $\lambda^{\frac{2}{n}}$  with respect to the density of  $S$ , that is,

$$E[u_2^h | H_0] = \int_{S>O} \frac{|S|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)} (\prod_{j=1}^p s_{jj})^{-h}}{2^{\frac{mp}{2}} |\Sigma|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})} dS \quad (6.6.2)$$

where, under  $H_0$ ,  $|\Sigma| = \sigma_{11} \cdots \sigma_{pp}$ . As was done in Sect. 6.1.1, we may replace  $s_{jj}^{-h}$  by the equivalent integral,

$$s_{jj}^{-h} = \frac{1}{\Gamma(h)} \int_0^{\infty} x_j^{h-1} e^{-x_j(s_{jj})} dx_j, \Re(h) > 0.$$

Thus,

$$\prod_{j=1}^p s_{jj}^{-h} = \frac{1}{[\Gamma(h)]^p} \int_0^{\infty} \cdots \int_0^{\infty} x_1^{h-1} \cdots x_p^{h-1} e^{-\text{tr}(YS)} dx_1 \wedge \cdots \wedge dx_p \quad (i)$$

where  $Y = \text{diag}(x_1, \dots, x_p)$ , so that  $\text{tr}(YS) = x_1 s_{11} + \cdots + x_p s_{pp}$ . Then, (6.6.2) can be reexpressed as follow:

$$\begin{aligned} E[u_2^h | H_0] &= \frac{\int_0^{\infty} \cdots \int_0^{\infty} x_1^{h-1} \cdots x_p^{h-1}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) (\prod_{j=1}^p \sigma_{jj})^{\frac{mp}{2}}} \int_{S>O} |S|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}((\Sigma^{-1}+2Y)S)} dS \\ &= \frac{\Gamma_p(\frac{m}{2}+h)}{\Gamma_p(\frac{m}{2})} \frac{\int_0^{\infty} \cdots \int_0^{\infty} x_1^{h-1} \cdots x_p^{h-1}}{2^{\frac{mp}{2}} (\prod_{j=1}^p \sigma_{jj})^{\frac{mp}{2}}} \\ &\quad \times \left| \frac{(\Sigma^{-1}+2Y)}{2} \right|^{-\frac{(m}{2}+h)} dx_1 \wedge \cdots \wedge dx_p, \end{aligned}$$

and observing that, under  $H_0$ ,

$$\begin{aligned} \left| \frac{(\Sigma^{-1}+2Y)}{2} \right| &= \left| \frac{\Sigma^{-1}}{2} \right| |I + 2\Sigma Y| \text{ with} \\ |I + 2\Sigma Y| &= (1 + 2\sigma_{11}y_1) \cdots (1 + 2\sigma_{pp}y_p), \end{aligned}$$

$$\begin{aligned}
E[u_2^h | H_0] &= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} \prod_{j=1}^p \frac{1}{\Gamma(h)} \int_0^\infty y_j^{h-1} (1 + y_j)^{-(\frac{m}{2} + h)} dy_j, \quad y_j = 2x_j \sigma_{jj} \\
&= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} \left[ \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2} + h)} \right]^p, \quad \Re(\frac{m}{2} + h) > \frac{p-1}{2}.
\end{aligned} \tag{6.6.3}$$

Thus,

$$\begin{aligned}
E[u_2^h | H_0] &= \left[ \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2} + h)} \right]^p \prod_{j=1}^p \frac{\Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\Gamma(\frac{m}{2} - \frac{j-1}{2})} \\
&= \frac{[\Gamma(\frac{m}{2})]^{p-1}}{\{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} - \frac{j}{2})\}} \frac{\{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} - \frac{j}{2} + h)\}}{[\Gamma(\frac{m}{2} + h)]^{p-1}}.
\end{aligned}$$

Denoting the density of  $u_2$  as  $f_{u_2}(u_2 | H_0)$ , we can express it as an inverse Mellin transform by taking  $h = s - 1$ . Then,

$$f_{u_2}(u_2 | H_0) = c_{2,p-1} G_{p-1,p-1}^{p-1,0} \left[ u_2 \left| \begin{matrix} \frac{m}{2}-1, \dots, \frac{m}{2}-1 \\ \frac{m}{2}-\frac{j}{2}-1, \quad j=1, \dots, p-1 \end{matrix} \right. \right], \quad 0 \leq u_2 \leq 1, \tag{6.6.4}$$

and zero elsewhere, where

$$c_{2,p-1} = \frac{[\Gamma(\frac{m}{2})]^{p-1}}{\{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} - \frac{j}{2})\}}.$$

Some special cases of this density are expounded below.

### Real and complex cases: $p = 2$

When  $p = 2$ ,  $u_2$  has a real type-1 beta density with the parameters  $(\alpha = \frac{m}{2} - \frac{1}{2}, \beta = \frac{1}{2})$  in the real case. In the complex case, it has a real type-1 beta density with the parameters  $(\alpha = m - 1, \beta = 1)$ .

### Real and complex cases: $p = 3$

In this case,  $f_{u_2}(u_2 | H_0)$  is given by

$$f_{u_2}(u_2 | H_0) = c_{2,2} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\Gamma(\frac{m}{2} - \frac{3}{2} + s) \Gamma(\frac{m}{2} - 2 + s)}{[\Gamma(\frac{m}{2} - 1 + s)]^2} u_2^{-s} ds.$$

The poles of the integrand are simple. Those coming from  $\Gamma(\frac{m}{2} - \frac{3}{2} + s)$  occur at  $s = -\frac{m}{2} + \frac{3}{2} - \nu$ ,  $\nu = 0, 1, \dots$ . The residue  $R_\nu$  is the following:

$$R_\nu = u_2^{\frac{m}{2} - \frac{3}{2} + \nu} \frac{(-1)^\nu \Gamma(-\frac{1}{2} - \nu)}{\nu! [\Gamma(\frac{1}{2} - \nu)]^2} = \frac{\Gamma(-\frac{1}{2}) (\frac{1}{2})_\nu (\frac{1}{2})_\nu (-1)^\nu u_2^{\frac{m}{2} - \frac{3}{2} + \nu}}{[\Gamma(\frac{1}{2})]^2 (\frac{3}{2})_\nu \nu!}.$$

Summing the residues, we have

$$\sum_{\nu=0}^{\infty} R_\nu = \frac{\Gamma(-\frac{1}{2})}{[\Gamma(\frac{1}{2})]^2} u_2^{\frac{m}{2} - \frac{3}{2}} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}; u_2\right), \quad 0 \leq u_2 \leq 1.$$

Now, consider the sum of the residues at the poles of  $\Gamma(\frac{m}{2} - 2 + s)$ . Observing that  $\Gamma(\frac{m}{2} - 2 + s)$  cancels out one of the gamma functions in the denominator, namely  $\Gamma(\frac{m}{2} - 1 + s) = (\frac{m}{2} - 2 - s)\Gamma(\frac{m}{2} - 2 + s)$ , the integrand becomes

$$\frac{\Gamma(\frac{m}{2} - \frac{3}{2} + s) u_2^{-s}}{(\frac{m}{2} - 2 + s)\Gamma(\frac{m}{2} - 1 + s)},$$

the residue at the pole  $s = -\frac{m}{2} + 2$  being  $\frac{\Gamma(\frac{1}{2}) u_2^{\frac{m}{2} - 2}}{\Gamma(1)}$ . Then, noting that  $\Gamma(-\frac{1}{2}) = -2\Gamma(\frac{1}{2}) = -2\sqrt{\pi}$ , the density is the following:

$$f_{u_2}(u_2|H_0) = c_{2,2} \left\{ \sqrt{\pi} u_2^{\frac{m}{2} - 2} - \frac{2}{\sqrt{\pi}} u_2^{\frac{m}{2} - \frac{3}{2}} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}; u_2\right) \right\}, \quad 0 \leq u_2 \leq 1, \quad (6.6.5)$$

and zero elsewhere.

In the complex case, the integrand is

$$\frac{\Gamma(m - 2 + s)\Gamma(m - 3 + s)}{[\Gamma(m - 1 + s)]^2} u_2^{-s} = \frac{1}{(m - 2 + s)^2(m - 3 + s)} u_2^{-s},$$

and hence there is a pole of order 1 at  $s = -m + 3$  and a pole of order 2 at  $s = -m + 2$ .

The residue at  $s = -m + 3$  is  $\frac{u_2^{m-3}}{(1)^2} = u_2^{m-3}$  and the residue at  $s = -m + 2$  is given by

$$\lim_{s \rightarrow -m+2} \frac{\partial}{\partial s} (m - 2 + s)^2 \left[ \frac{1}{(m - 2 + s)^2(m - 3 + s)} u_2^{-s} \right] = \lim_{s \rightarrow -m+2} \frac{\partial}{\partial s} \left[ \frac{u_2^{-s}}{(m - 3 + s)} \right],$$

which gives the residue as  $u_2^{m-2} \ln u_2 - u_2^{m-2}$ . Thus, the sum of the residues is  $u_2^{m-3} + u_2^{m-2} \ln u_2 - u_2^{m-2}$  and the constant part is

$$\frac{[\Gamma(m)]^2}{\Gamma(m-1)\Gamma(m-2)} = (m-1)^2(m-2), \quad m > 2,$$

so that the density is

$$f_{u_2}(u_2) = (m-1)^2(m-2)[u_2^{m-3} + u_2^{m-2} \ln u_2 - u_2^{m-2}], \quad 0 < u_2 \leq 1, \quad m \geq 3,$$

and zero elsewhere. Note that as  $u_2 \rightarrow 0$ , the limit of  $\frac{u_2^{m-1}}{m-1} \ln u_2$  is zero. By integrating out over  $0 < u_2 \leq 1$  while  $m \geq 3$ , it can be verified that  $f_{u_2}(\cdot)$  is indeed a density function.

**Real and complex cases:  $p \geq 4$**

As poles of higher orders are present when  $p \geq 4$ , both in the real and complex cases, the exact density function of the test statistic will not be herein explicitly given for those cases. Actually, the resulting densities would involve G-functions for which general expansions are for instance provided in Mathai (1993). The exact null and non-null densities of  $u = \lambda^{\frac{2}{n}}$  have been previously derived by the first author. Percentage points accurate to the 11th decimal place are available from Mathai and Katiyar (1979a, 1980) for the null case; as well, various aspects of the distribution of the test statistic are discussed in Mathai and Rathie (1971) and Mathai (1973, 1984, 1985)

Let us now consider the asymptotic distribution of the  $\lambda$ -criterion under the null hypothesis,

$$H_o : \Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp}).$$

Given the representation of the  $h$ -th moment of  $u_2$  provided in (6.6.3) and referring to Corollary 6.5.1, it is seen that the sum of the  $\delta_j$ 's is  $\sum_{j=1}^{p-1} \delta_j = \sum_{j=1}^{p-1} \frac{j}{2} = \frac{p(p-1)}{4}$ , so that the number of degrees of freedom of the asymptotic chisquare distribution is  $2[\frac{p(p-1)}{4}] = \frac{p(p-1)}{2}$  which, as it should be, is the number of restrictions imposed by  $H_o$ , noting that when  $\Sigma$  is diagonal,  $\sigma_{ij} = 0$ ,  $i \neq j$ , which produces  $\frac{p(p-1)}{2}$  restrictions. Hence, the following result:

**Theorem 6.6.1.** *Let  $\lambda$  be the likelihood ratio criterion for testing the hypothesis that the covariance matrix  $\Sigma$  of a nonsingular  $N_p(\mu, \Sigma)$  distribution is diagonal. Then, as  $n \rightarrow \infty$ ,  $-2 \ln \lambda \rightarrow \chi_{\frac{p(p-1)}{2}}^2$  in the real case. In the corresponding complex case, as  $n \rightarrow \infty$ ,  $-2 \ln \lambda \rightarrow \chi_{p(p-1)}^2$ , a real scalar chisquare variable having  $p(p-1)$  degrees of freedom.*

## 6.7. Equality of Diagonal Elements, Given that $\Sigma$ is Diagonal, Real Case

In the case of a  $p$ -variate real nonsingular  $N_p(\mu, \Sigma)$  population, whenever  $\Sigma$  is diagonal, the individual components are independently distributed as univariate normal random variables. Consider a simple random sample of size  $n$ , that is, a set of

$p \times 1$  vectors  $X_1, \dots, X_n$ , that are iid as  $X_j \sim N_p(\mu, \Sigma)$  where it is assumed that  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ . Letting  $X'_j = (x_{1j}, \dots, x_{pj})$ , the joint density of the  $x_{rj}$ 's,  $j = 1, \dots, n$ , in the above sample, which is denoted by  $L_r$ , is given by

$$L_r = \prod_{j=1}^n \frac{e^{-\frac{1}{2\sigma_r^2}(x_{rj}-\mu_r)^2}}{(2\pi)^{\frac{1}{2}}(\sigma_r^2)^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2\sigma_r^2} \sum_{j=1}^n (x_{rj}-\mu_r)^2}}{(2\pi)^{\frac{n}{2}}(\sigma_r^2)^{\frac{n}{2}}}.$$

Then, on substituting the maximum likelihood estimators of  $\mu_r$  and  $\sigma_r^2$  in  $L_r$ , its maximum is

$$\max L_r = \frac{n^{\frac{n}{2}} e^{-\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} (s_{rr})^{\frac{n}{2}}}, \quad s_{rr} = \sum_{j=1}^n (x_{rj} - \bar{x}_r)^2.$$

Under the null hypothesis  $H_0$ ,  $\sigma_1^2 = \dots = \sigma_p^2 \equiv \sigma^2$  and the MLE of  $\sigma^2$  is a pooled estimate which is equal to  $\frac{1}{np}(s_{11} + \dots + s_{pp})$ . Thus, the  $\lambda$ -criterion is the following in this case:

$$\lambda = \frac{\sup_{\omega}}{\sup_{\Omega}} = \frac{[s_{11}s_{22} \cdots s_{pp}]^{\frac{n}{2}}}{\left(\frac{s_{11} + \dots + s_{pp}}{p}\right)^{\frac{np}{2}}}. \quad (6.7.1)$$

If we let

$$u_3 = \lambda^{\frac{2}{n}} = \frac{p^p (\prod_{j=1}^p s_{jj})}{(\sum_{j=1}^p s_{jj})^p}, \quad (6.7.2)$$

then, for arbitrary  $h$ , the  $h$ -th moment of  $u_3$  is the following:

$$E[u_3^h | H_0] = E\left[\frac{p^{ph} (\prod_{j=1}^p s_{jj}^h)}{(s_{11} + \dots + s_{pp})^{ph}}\right] = E\left[p^{ph} \left(\prod_{j=1}^p s_{jj}^h\right) (s_{11} + \dots + s_{pp})^{-ph}\right]. \quad (6.7.3)$$

Observe that  $\frac{s_{jj}}{\sigma^2} \stackrel{iid}{\sim} \chi_{n-1}^2 = \chi_m^2$ ,  $m = n - 1$ , for  $j = 1, \dots, p$ , the density of  $s_{jj}$  being of the form

$$f_{s_{jj}}(s_{jj}) = \frac{s_{jj}^{\frac{m}{2}-1} e^{-\frac{s_{jj}}{2\sigma^2}}}{(2\sigma^2)^{\frac{m}{2}} \Gamma(\frac{m}{2})}, \quad 0 \leq s_{jj} < \infty, \quad m = n - 1 = 1, 2, \dots, \quad (i)$$

under  $H_0$ . Note that  $(s_{11} + \dots + s_{pp})^{-ph}$  can be replaced by an equivalent integral as

$$(s_{11} + \dots + s_{pp})^{-ph} = \frac{1}{\Gamma(ph)} \int_0^\infty x^{ph-1} e^{-x(s_{11} + \dots + s_{pp})} dx, \quad \Re(h) > 0. \quad (ii)$$



Due to independence of the  $s_{jj}$ 's, the joint density of  $s_{11}, \dots, s_{pp}$ , is the product of the densities appearing in (i), and on integrating out  $s_{11}, \dots, s_{pp}$ , we end up with the following:

$$\frac{1}{(2\sigma^2)^{\frac{mp}{2}}} \left\{ \prod_{j=1}^p \frac{\Gamma(\frac{m}{2} + h)}{\Gamma(\frac{m}{2})} \right\} \left\{ \prod_{j=1}^n \left[ \frac{(1 + 2\sigma^2 x)}{2\sigma^2} \right]^{-\frac{m}{2} + h} \right\} = \frac{[\Gamma(\frac{m}{2} + h)]^p}{[\Gamma(\frac{m}{2})]^p} \left[ \frac{(1 + 2\sigma^2 x)^{-p(\frac{m}{2} + h)}}{(2\sigma^2)^{-ph}} \right].$$

Now, the integral over  $x$  can be evaluated as follows:

$$\frac{(2\sigma^2)^{ph}}{\Gamma(ph)} \int_0^\infty x^{ph-1} (1 + 2\sigma^2 x)^{-p(\frac{m}{2} + h)} dx = \frac{\Gamma(\frac{mp}{2})}{\Gamma(\frac{mp}{2} + ph)}, \Re(h) > 0.$$

Thus,

$$E[u_3^h | H_0] = p^{ph} \frac{\Gamma^p(\frac{m}{2} + h)}{\Gamma^p(\frac{m}{2})} \frac{\Gamma(\frac{mp}{2})}{\Gamma(\frac{mp}{2} + ph)}, \Re(h) > 0. \quad (6.7.4)$$

The density of  $u_3$  can be written in terms of an H-function. Since  $p$  is a positive integer, we can expand one gamma ratio using Gauss' multiplication formula:

$$\frac{\Gamma(\frac{mp}{2})}{\Gamma(\frac{mp}{2} + ph)} = \frac{(2\pi)^{\frac{1-p}{2}} p^{\frac{pm}{2} - \frac{1}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{m}{2} + \frac{1}{p}) \cdots \Gamma(\frac{m}{2} + \frac{p-1}{p})}{(2\pi)^{\frac{1-p}{2}} p^{\frac{mp}{2} - \frac{1}{2} + ph} \Gamma(\frac{m}{2} + h) \cdots \Gamma(\frac{m}{2} + \frac{p-1}{p} + h)}$$

for  $p = 1, 2, \dots, m \geq p$ . Accordingly,

$$\begin{aligned} E[u_3^h | H_0] &= \frac{[\Gamma(\frac{m}{2} + h)]^p}{[\Gamma(\frac{m}{2})]^p} \prod_{j=0}^{p-1} \frac{\Gamma(\frac{m}{2} + \frac{j}{p})}{\Gamma(\frac{m}{2} + \frac{j}{p} + h)} \\ &= \frac{[\Gamma(\frac{m}{2} + h)]^{p-1}}{[\Gamma(\frac{m}{2})]^{p-1}} \prod_{j=1}^{p-1} \frac{\Gamma(\frac{m}{2} + \frac{j}{p})}{\Gamma(\frac{m}{2} + \frac{j}{p} + h)} = c_{3,p-1} \frac{[\Gamma(\frac{m}{2} + h)]^{p-1}}{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} + \frac{j}{p} + h)}, \end{aligned} \quad (6.7.5)$$

$$c_{3,p-1} = \frac{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} + \frac{j}{p})}{[\Gamma(\frac{m}{2})]^{p-1}}, \Re(\frac{m}{2} + h) > 0. \quad (6.7.6)$$

Hence, for  $h = s - 1$ , (6.7.5) is the Mellin transform of the density of  $u_3$ . Thus, denoting the density by  $f_{u_3}(u_3)$ , we have

$$f_{u_3}(u_3 | H_0) = c_{3,p-1} G_{p-1,p-1}^{p-1,0} \left[ u_3 \left| \begin{matrix} \frac{m}{2} - 1 + \frac{j}{p}, j=1, \dots, p-1 \\ \frac{m}{2} - 1, \dots, \frac{m}{2} - 1 \end{matrix} \right. \right], 0 \leq u_3 \leq 1, \quad (6.7.7)$$

and zero elsewhere.

In the complex case, the  $h$ -th moment is the following:

$$E[\tilde{u}_3^h | H_o] = \tilde{c}_{3,p-1} \frac{[\tilde{\Gamma}(m+h)]^{p-1}}{\prod_{j=1}^{p-1} \tilde{\Gamma}(m + \frac{j}{p} + h)}, \quad (6.7a.1)$$

$$\tilde{c}_{3,p-1} = \frac{\prod_{j=1}^{p-1} \tilde{\Gamma}(m + \frac{j}{p})}{[\tilde{\Gamma}(m)]^{p-1}}. \quad (6.7a.2)$$

and the corresponding density is given by

$$\tilde{f}_{\tilde{u}_3}(\tilde{u}_3 | H_o) = \tilde{c}_{3,p-1} G_{p-1,p-1}^{p-1,0} \left[ \tilde{u}_3 \left| \begin{matrix} m-1+\frac{j}{p}, j=1,\dots,p-1 \\ m-1,\dots,m-1 \end{matrix} \right. \right], \quad 0 \leq |\tilde{u}_3| \leq 1, \quad (6.7a.3)$$

and zero elsewhere,  $G$  denoting a real  $G$ -function.

### Real and complex cases: $p = 2$

It is seen from (6.7.5) that for  $p = 2$ ,  $u_3$  is a real type-1 beta with the parameters  $(\alpha = \frac{m}{2}, \beta = \frac{1}{p})$  in the real case. Whenever  $p \geq 3$ , poles of order 2 or more are occurring, and the resulting density functions which are expressible in terms generalized hypergeometric functions, will not be explicitly provided. For a general series expansion of the  $G$ -function, the reader may refer to Mathai (1970a, 1993).

In the complex case, when  $p = 2$ ,  $\tilde{u}_3$  has a real type-1 beta density with the parameters  $(\alpha = m, \beta = \frac{1}{p})$ . In this instance as well, poles of higher orders will be present when  $p \geq 3$ , and hence explicit forms of the corresponding densities will not be herein provided. The exact null and non-null distributions of the test statistic are derived for the general case in Mathai and Saxena (1973), and highly accurate percentage points are provided in Mathai (1979a,b).

An asymptotic result can also be obtained as  $n \rightarrow \infty$ . Consider the  $h$ -th moment of  $\lambda$ , which is available from (6.7.5) in the real case and from (6.7a.1) in the complex case. Then, referring to Corollary 6.5.2,  $\delta_j = \frac{j}{p}$  whether in the real or in the complex situations. Hence,  $2[\sum_{j=1}^{p-1} \delta_j] = 2 \sum_{j=1}^{p-1} \frac{j}{p} = (p-1)$  in both the real and the complex cases. As well, observe that in the complex case, the diagonal elements are real since  $\tilde{\Sigma}$  is Hermitian positive definite. Accordingly, the number of restrictions imposed by  $H_o$  in either the real or complex cases is  $p-1$ . Thus, the following result:

**Theorem 6.7.1.** *Consider the  $\lambda$ -criterion for testing the equality of the diagonal elements, given that the covariance matrix is already diagonal. Then, as  $n \rightarrow \infty$ , the null distribution of  $-2 \ln \lambda \rightarrow \chi_{p-1}^2$  in both the real and the complex cases.*

### 6.8. Hypothesis that the Covariance Matrix is Block Diagonal, Real Case

We will discuss a generalization of the problem examined in Sect. 6.6, considering again the case of real Gaussian vectors. Let  $X_1, \dots, X_n$  be iid as  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , and

$$X_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{pj} \end{bmatrix} = \begin{bmatrix} X_{(1j)} \\ \vdots \\ X_{(kj)} \end{bmatrix}, \quad X_{(1j)} = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{p_1j} \end{bmatrix}, \quad X_{(2j)} = \begin{bmatrix} x_{p_1+1,j} \\ \vdots \\ x_{p_1+p_2,j} \end{bmatrix}, \dots,$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & O & \cdots & O \\ O & \Sigma_{22} & \ddots & O \\ \vdots & \vdots & \cdots & \vdots \\ O & O & \cdots & \Sigma_{kk} \end{bmatrix}, \quad \Sigma_{jj} \text{ being } p_j \times p_j, \quad j = 1, \dots, k.$$

In this case, the  $p \times 1$  real Gaussian vector is subdivided into subvectors of orders  $p_1, \dots, p_k$ , so that  $p_1 + \dots + p_k = p$ , and, under the null hypothesis  $H_0$ ,  $\Sigma$  is assumed to be a block diagonal matrix, which means that the subvectors are mutually independently distributed  $p_j$ -variate real Gaussian vectors with corresponding mean value vector  $\mu_{(j)}$  and covariance matrix  $\Sigma_{jj}$ ,  $j = 1, \dots, k$ . Then, the joint density of the sample values under the null hypothesis can be written as  $L = \prod_{r=1}^k L_r$  where  $L_r$  is the joint density of the sample values corresponding to the subvector  $X_{(rj)}$ ,  $j = 1, \dots, n$ ,  $r = 1, \dots, k$ . Letting the  $p \times n$  general sample matrix be  $\mathbf{X} = (X_1, \dots, X_n)$ , we note that the sample representing the first  $p_1$  rows of  $\mathbf{X}$  corresponds to the sample from the first subvector  $X_{(1j)} \stackrel{iid}{\sim} N_p(\mu_{(1)}, \Sigma_{11})$ ,  $\Sigma_{11} > O$ ,  $j = 1, \dots, n$ . The MLE's of  $\mu_{(r)}$  and  $\Sigma_{rr}$  are the corresponding sample mean and sample covariance matrix. Thus, the maximum of  $L_r$  is available as

$$\max L_r = \frac{e^{-\frac{np_r}{2}} n^{\frac{np_r}{2}}}{(2\pi)^{\frac{np_r}{2}} |S_{rr}|^{\frac{n}{2}}} \stackrel{ind}{\Rightarrow} \prod_{r=1}^k \max L_r = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} \prod_{r=1}^k |S_{rr}|^{\frac{n}{2}}}.$$

Hence,

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = \frac{|S|^{\frac{n}{2}}}{\prod_{r=1}^k |S_{rr}|^{\frac{n}{2}}}, \quad (6.8.1)$$

and

$$u_4 \equiv \lambda^{\frac{2}{n}} = \frac{|S|}{\prod_{r=1}^k |S_{rr}|}. \quad (6.8.2)$$

Observe that the covariance matrix  $\Sigma = (\sigma_{ij})$  can be written in terms of the matrix of population correlations. If we let  $D = \text{diag}(\sigma_1, \dots, \sigma_p)$  where  $\sigma_t^2 = \sigma_{tt}$  denotes the

variance associated the component  $x_{ij}$  in  $X'_j = (x_{1j}, \dots, x_{pj})$  where  $\text{Cov}(X) = \Sigma$ , and  $R = (\rho_{rs})$  be the population correlation matrix, where  $\rho_{rs}$  is the population correlation between the components  $x_{rj}$  and  $x_{sj}$ , then,  $\Sigma = DRD$ . Consider a partitioning of  $\Sigma$  into  $k \times k$  blocks as well as the corresponding partitioning of  $D$  and  $R$ :

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & \Sigma_{kk} \end{bmatrix}, D = \begin{bmatrix} D_1 & O & \cdots & O \\ O & D_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & D_k \end{bmatrix}, R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1k} \\ R_{21} & R_{22} & \cdots & R_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ R_{k1} & R_{k2} & \cdots & R_{kk} \end{bmatrix},$$

where, for example,  $\Sigma_{jj}$  is  $p_j \times p_j$ ,  $p_1 + \dots + p_k = p$ , and the corresponding partitioning of  $D$  and  $R$ . Consider a corresponding partitioning of the sample sum of products matrix  $S = (S_{ij})$ ,  $D^{(s)}$  and  $R^{(s)}$  where  $R^{(s)}$  is the sample correlation matrix and  $D^{(s)} = \text{diag}(\sqrt{s_{11}}, \dots, \sqrt{s_{pp}})$ , where  $S_{jj}$ ,  $D_j^{(s)}$ ,  $R_{jj}^{(s)}$  are  $p_j \times p_j$ ,  $p_1 + \dots + p_k = p$ . Then,

$$\frac{|\Sigma|}{\prod_{j=1}^k |\Sigma_{jj}|} = \frac{|R|}{\prod_{j=1}^k |R_{jj}|} \quad (6.8.3)$$

and

$$u_4 \equiv \lambda^{\frac{2}{n}} = \frac{|S|}{\prod_{j=1}^k |S_{jj}|} = \frac{|R^{(s)}|}{\prod_{j=1}^k |R_{jj}^{(s)}|}. \quad (6.8.4)$$

An additional interesting property is now pointed out. Consider a linear function of the original  $p \times 1$  vector  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , in the form  $CX_j$  where  $C$  is the diagonal matrix,  $\text{diag}(c_1, \dots, c_p)$ . In this case, the product  $CX_j$  is such that the  $r$ -th component of  $X_j$  is weighted or multiplied by  $c_r$ . Let  $C$  be a block diagonal matrix that is partitioned similarly to  $D$  so that its  $j$ -th diagonal block matrix be the  $p_j \times p_j$  diagonal submatrix  $C_j$ . Then,

$$u_c = \frac{|CSC'|}{\prod_{j=1}^k |C_j S_{jj} C'_j|} = \frac{|S|}{\prod_{j=1}^k |S_{jj}|} = u_4. \quad (6.8.5)$$

In other words,  $u_4$  is invariant under linear transformations on  $X_j \stackrel{iid}{\sim} N_p(\mu, \Sigma)$ ,  $\Sigma > O$ ,  $j = 1, \dots, n$ . That is, if  $Y_j = CX_j + d$  where  $d$  is a constant column vector, then the  $p \times n$  sample matrix on  $Y_j$ , namely,  $\bar{Y} = (Y_1, \dots, Y_n) = (CX_1 + d, \dots, CX_n + d)$ ,

$$\mathbf{Y} - \bar{\mathbf{Y}} = C(\mathbf{X} - \bar{\mathbf{X}}) \Rightarrow S_y = (\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})' = C(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'C' = CSC'.$$

Letting  $S_y$  be partitioned as  $S$  into  $k \times k$  blocks and  $S_y = (S_{ijy})$ , we have

$$u_y \equiv \frac{|S_y|}{\prod_{j=1}^k |S_{jyy}|} = \frac{|CSC'|}{\prod_{j=1}^k |C_j S_{jj} C'_j|} = \frac{|S|}{\prod_{j=1}^k |S_{jj}|} = u_4. \quad (6.8.6)$$

Arbitrary moments of  $u_4$  can be derived by proceeding as in Sect. 6.6. The  $h$ -th null moment, that is, the  $h$ -th moment under the null hypothesis  $H_o$ , is then

$$E[u_4^h | H_o] = \frac{1}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) |\Sigma_o|^{\frac{m}{2}}} \int_{S>O} |S|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma_o^{-1}S)} \left\{ \prod_{r=1}^k |S_{rr}|^{-h} \right\} dS \quad (i)$$

where  $m = n - 1$ ,  $n$  being the sample size, and

$$\Sigma_o = \begin{bmatrix} \Sigma_{11} & O & \cdots & O \\ O & \Sigma_{22} & \cdots & O \\ \vdots & \vdots & \ddots & O \\ O & O & \cdots & \Sigma_{kk} \end{bmatrix}, \quad \text{tr}(\Sigma_o^{-1}S) = \text{tr}(\Sigma_{11}^{-1}S_{11}) + \cdots + \text{tr}(\Sigma_{kk}^{-1}S_{kk}), \quad (ii)$$

where  $S_{rr}$  is the  $r$ -th diagonal block of  $S$ , corresponding to  $\Sigma_{rr}$  of  $\Sigma$  whose order  $p_r \times p_r$ ,  $r = 1, \dots, k$ ,  $p_1 + \cdots + p_k = p$ . On noting that

$$|S_{rr}|^{-h} = \frac{1}{\Gamma_{p_r}(h)} \int_{Y_r>O} |Y_r|^{h-\frac{p_r+1}{2}} e^{-\text{tr}(Y_r S_{rr})} dY_r, \quad r = 1, \dots, k, \quad (iii)$$

where  $Y_r > O$  is a  $p_r \times p_r$  real positive definite matrix, and replacing each  $|S_{rr}|^{-h}$  by its integral representation as given in (iii), the exponent of  $e$  in (i) becomes

$$-\frac{1}{2}[\text{tr}(\Sigma_o^{-1}S) + 2\text{tr}(YS)], \quad Y = \begin{bmatrix} Y_1 & O & \cdots & O \\ O & Y_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & Y_k \end{bmatrix}, \quad \text{tr}(YS) = \text{tr}(Y_1 S_{11}) + \cdots + \text{tr}(Y_k S_{kk}).$$

The right-hand side of equation (i) then becomes

$$E[u_4^h | H_o] = \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2}) |\Sigma_o|^{\frac{m}{2}}} 2^{ph} \left\{ \prod_{r=1}^k \frac{1}{\Gamma_{p_r}(h)} \int_{Y_r>O} |Y_r|^{h-\frac{p_r+1}{2}} \right\} \\ \times |\Sigma_o^{-1} + 2Y|^{-\frac{m}{2}+h} dY_1 \wedge \dots \wedge dY_k, \quad \Re(h) > -\frac{m}{2} + \frac{p-1}{2}. \quad (iv)$$

It should be pointed out that the non-null moments of  $u_4$  can be obtained by substituting a general  $\Sigma$  to  $\Sigma_o$  in (iv). Note that if we replace  $2Y$  by  $Y$ , the factor containing 2, namely  $2^{ph}$ , will disappear. Further, under  $H_o$ ,

$$|\Sigma_o^{-1} + 2Y|^{-\frac{m}{2}+h} = \left\{ \prod_{r=1}^k |\Sigma_{rr}|^{\frac{m}{2}+h} \right\} \left\{ \prod_{r=1}^k |I + 2\Sigma_{rr} Y_r|^{-\frac{m}{2}+h} \right\}. \quad (v)$$

Then, each  $Y_r$ -integral can be evaluated as follows:

$$\begin{aligned} & \frac{1}{\Gamma_{p_r}(h)} \int_{Y_r > 0} |Y_r|^{h - \frac{p_r+1}{2}} |I + 2\Sigma_{rr} Y_r|^{-(\frac{m}{2}+h)} dY_r \\ &= 2^{-p_r h} |\Sigma_{rr}|^{-h} \frac{1}{\Gamma_{p_r}(h)} \int_{Z_r > 0} |Z_r|^{h - \frac{p_r+1}{2}} |I + Z_r|^{-(\frac{m}{2}+h)} dZ_r, \quad Z_r = 2\Sigma_{rr}^{\frac{1}{2}} Y_r \Sigma_{rr}^{\frac{1}{2}} \\ &= 2^{-p_r h} |\Sigma_{rr}|^{-h} \frac{\Gamma_{p_r}(h)}{\Gamma_{p_r}(h)} \frac{\Gamma_{p_r}(\frac{m}{2})}{\Gamma_{p_r}(\frac{m}{2} + h)}. \end{aligned} \quad (vi)$$

On combining equations (i) to (vi), we have

$$E[u_4^h | H_0] = \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} \prod_{r=1}^k \frac{\Gamma_{p_r}(\frac{m}{2})}{\Gamma_{p_r}(\frac{m}{2} + h)}, \quad \Re(\frac{m}{2} + h) > \frac{p-1}{2}, \quad (6.8.7)$$

$$= c_{4,p} \frac{\Gamma_p(\frac{m}{2} + h)}{\prod_{r=1}^k \Gamma_{p_r}(\frac{m}{2} + h)} = c_{4,p} c^* \frac{\prod_{j=1}^p \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\prod_{r=1}^k [\prod_{i=1}^{p_r} \Gamma(\frac{m}{2} - \frac{i-1}{2} + h)]}, \quad (6.8.8)$$

$$c_{4,p} = \frac{\prod_{r=1}^k \Gamma_{p_r}(\frac{m}{2})}{\Gamma_p(\frac{m}{2})} = \frac{[\prod_{r=1}^k \pi^{\frac{p_r(p_r-1)}{4}}] \prod_{r=1}^k [\prod_{i=1}^{p_r} \Gamma(\frac{m}{2} - \frac{i-1}{2})]}{\pi^{\frac{p(p-1)}{4}} \prod_{j=1}^p \Gamma(\frac{m}{2} - \frac{j-1}{2})}, \quad (6.8.9)$$

$$c^* = \frac{\pi^{\frac{p(p-1)}{4}}}{\prod_{r=1}^k \pi^{\frac{p_r(p_r-1)}{4}}}$$

so that when  $h = 0$ ,  $E[u_4^h | H_0] = 1$ . Observe that one set of gamma products can be canceled in (6.8.8) and (6.8.9). When that set is the product of the first  $p_1$  gamma functions, the  $h$ -th moment of  $u_4$  is given by

$$E[u_4^h | H_0] = c_{4,p-p_1} \frac{\prod_{j=p_1+1}^p \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\prod_{r=2}^k [\prod_{i=1}^{p_r} \Gamma(\frac{m}{2} - \frac{i-1}{2} + h)]}, \quad (6.8.10)$$

where  $c_{4,p-p_1}$  is such that  $E[u_4^h | H_0] = 1$  when  $h = 0$ . Since the structure of the expression given in (6.8.10) is that of the  $h$ -th moment of a product of  $p - p_1$  independently distributed real scalar type-1 beta random variables, it can be inferred that the distribution of  $u_4 | H_0$  is also that of a product of  $p - p_1$  independently distributed real scalar type-1 beta random variables whose parameters can be determined from the arguments of the gamma functions appearing in (6.8.10).

Some of the gamma functions appearing in (6.8.10) will cancel out for certain values of  $p_1, \dots, p_k$ , thereby simplifying the representation of the moments and enabling one to

express the density of  $u_4$  in terms of elementary functions in such instances. The exact null density in the general case was derived by the first author. For interesting representations of the exact density, the reader is referred to Mathai and Rathie (1971) and Mathai and Saxena (1973), some exact percentage points of the null distribution being included in Mathai and Katiyar (1979a). As it turns out, explicit forms are available in terms of elementary functions for the following special cases, see also Anderson (2003):  $p_1 = p_2 = p_3 = 1$ ;  $p_1 = p_2 = p_3 = 2$ ;  $p_1 = p_2 = 1, p_3 = p - 2$ ;  $p_1 = 1, p_2 = p_3 = 2$ ;  $p_1 = 1, p_2 = 2, p_3 = 3$ ;  $p_1 = 2, p_2 = 2, p_3 = 4$ ;  $p_1 = p_2 = 2, p_3 = 3$ ;  $p_1 = 2, p_2 = 3, p$  is even.

### 6.8.1. Special case: $k = 2$

Let us consider a certain  $2 \times 2$  partitioning of  $S$ , which corresponds to the special case  $k = 2$ . When  $p_1 = 1$  and  $p_2 = p - 1$  so that  $p_1 + p_2 = p$ , the test statistic is

$$\begin{aligned} u_4 &= \frac{|S|}{|S_{11}| |S_{22}|} = \frac{|S_{11} - S_{12} S_{22}^{-1} S_{21}|}{|S_{11}|} \\ &= \frac{s_{11} - S_{12} S_{22}^{-1} S_{21}}{s_{11}} = 1 - r_{1.(2\dots p)}^2 \end{aligned} \quad (6.8.11)$$

where  $r_{1.(2\dots p)}$  is the multiple correlation between  $x_1$  and  $(x_2, \dots, x_p)$ . As stated in Theorem 5.6.3,  $1 - r_{1.(2\dots p)}^2$  is distributed as a real scalar type-1 beta variable with the parameters  $(\frac{n-1}{2} - \frac{p-1}{2}, \frac{p-1}{2})$ . The simplifications in (6.8.11) are achieved by making use of the properties of determinants of partitioned matrices, which are discussed in Sect. 1.3. Since  $s_{11}$  is  $1 \times 1$  in this case, the numerator determinant is a real scalar quantity. Thus, this yields a type-2 beta distribution for  $w = \frac{u_4}{1-u_4}$  and thereby  $\frac{n-p}{p-1} w$  has an  $F$ -distribution, so that the test can be based on an  $F$  statistic having  $(n-1) - (p-1) = n-p$  and  $p-1$  degrees of freedom.

### 6.8.2. General case: $k = 2$

If in a  $2 \times 2$  partitioning of  $S$ ,  $S_{11}$  is of order  $p_1 \times p_1$  and  $S_{22}$  is of order  $p_2 \times p_2$  with  $p_2 = p - p_1$ . Then  $u_4$  can be expressed as

$$\begin{aligned} u_4 &= \frac{|S|}{|S_{11}| |S_{22}|} = \frac{|S_{11} - S_{12} S_{22}^{-1} S_{21}|}{|S_{11}|} \\ &= |I - S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}| = |I - U|, \quad U = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}} \end{aligned} \quad (6.8.12)$$

where  $U$  is called the *multiple correlation matrix*. It will be shown that  $U$  has a real matrix-variate type-1 beta distribution when  $S_{11}$  is of general order rather than being a scalar.

**Theorem 6.8.1.** Consider  $u_4$  for  $k = 2$ . Let  $S_{11}$  be  $p_1 \times p_1$  and  $S_{22}$  be  $p_2 \times p_2$  so that  $p_1 + p_2 = p$ . Without any loss of generality, let us assume that  $p_1 \leq p_2$ . Then, under  $H_0 : \Sigma_{12} = O$ , the multiple correlation matrix  $U$  has a real matrix-variate type-1 beta distribution with the parameters  $(\frac{p_2}{2}, \frac{m}{2} - \frac{p_2}{2})$ , with  $m = n - 1$ ,  $n$  being sample size, and thereby  $(I - U) \sim$  type-1 beta  $(\frac{m}{2} - \frac{p_2}{2}, \frac{p_2}{2})$ , the determinant of  $I - U$  being  $u_4$  under the null hypothesis when  $k = 2$ .

**Proof:** Since  $\Sigma$  under  $H_0$  can readily be eliminated from a structure such as  $u_4$ , we will take a Wishart matrix  $S$  having  $m = n - 1$  degrees of freedom,  $n$  denoting the sample size, and parameter matrix  $I$ , the identity matrix. At first, assume that  $\Sigma$  is a block diagonal matrix and make the transformation  $S_1 = \Sigma^{-\frac{1}{2}} S \Sigma^{-\frac{1}{2}}$ . As a result,  $u_4$  will be free of  $\Sigma_{11}$  and  $\Sigma_{22}$ , and so, we may take  $S \sim W_p(m, I)$ . Now, consider the submatrices  $S_{11}$ ,  $S_{22}$ ,  $S_{12}$  so that  $dS = dS_{11} \wedge dS_{22} \wedge dS_{12}$ . Let  $f(S)$  denote the  $W_p(m, I)$  density. Then,

$$f(S)dS = \frac{|S|^{\frac{m-p+1}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} e^{-\frac{1}{2}\text{tr}(S)} dS, \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad S_{21} = S'_{12}.$$

However, appealing to a result stated in Sect. 1.3, we have

$$\begin{aligned} |S| &= |S_{22}| |S_{11} - S_{12} S_{22}^{-1} S_{21}| \\ &= |S_{22}| |S_{11}| |I - S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}|. \end{aligned}$$

The joint density of  $S_{11}$ ,  $S_{22}$ ,  $S_{12}$  denoted by  $f_1(S_{11}, S_{22}, S_{12})$  is then

$$\begin{aligned} f_1(S_{11}, S_{22}, S_{12}) &= |S_{11}|^{\frac{m-p+1}{2}} |S_{22}|^{\frac{m-p+1}{2}} |I - U|^{\frac{m-p+1}{2}} \\ &\quad \times \frac{e^{-\frac{1}{2}\text{tr}(S_{11}) - \frac{1}{2}\text{tr}(S_{22})}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})}, \quad U = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}. \end{aligned}$$

Letting  $Y = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-\frac{1}{2}}$ , it follows from a result on Jacobian of matrix transformation, previously established in Chap. 1, that  $dY = |S_{11}|^{-\frac{p_2}{2}} |S_{22}|^{-\frac{p_1}{2}} dS_{12}$ . Thus, the joint density of  $S_{11}$ ,  $S_{22}$ ,  $Y$ , denoted by  $f_2(S_{11}, S_{22}, Y)$ , is given by

$$\begin{aligned} f_2(S_{11}, S_{22}, Y) &= |S_{11}|^{\frac{m+p_2-p+1}{2}} |S_{22}|^{\frac{m+p_1-p+1}{2}} |I - YY'|^{\frac{m-p+1}{2}} \\ &\quad \times \frac{e^{-\frac{1}{2}\text{tr}(S_{11}) - \frac{1}{2}\text{tr}(S_{22})}}{2^{\frac{pm}{2}} \Gamma_p(\frac{m}{2})}, \end{aligned}$$



Note that  $S_{11}$ ,  $S_{22}$ ,  $Y$  are independently distributed as  $f_2(\cdot)$  can be factorized into functions of  $S_{11}$ ,  $S_{22}$ ,  $Y$ . Now, letting  $U = YY'$ , it follows from Theorem 4.2.3 that

$$dY = \frac{\pi^{\frac{p_1 p_2}{2}}}{\Gamma_{p_1}(\frac{p_2}{2})} |U|^{\frac{p_2}{2} - \frac{p_1 + 1}{2}} dU,$$

and the density of  $U$ , denoted by  $f_3(U)$ , can then be expressed as follows:

$$f_3(U) = c |U|^{\frac{p_2}{2} - \frac{p_1 + 1}{2}} |I - U|^{\frac{m}{2} - \frac{p_2}{2} - \frac{p_1 + 1}{2}}, \quad 0 < U < I, \quad (6.8.13)$$

which is a real matrix-variate type-1 beta density with the parameters  $(\frac{p_2}{2}, \frac{m}{2} - \frac{p_2}{2})$ , where  $c$  is the normalizing constant. As a result,  $I - U$  has a real matrix-variate type-1 beta distribution with the parameters  $(\frac{m}{2} - \frac{p_2}{2}, \frac{p_2}{2})$ . Finally, observe that  $u_4$  is the determinant of  $I - U$ .

**Corollary 6.8.1.** Consider  $u_4$  as given in (6.8.12) and the determinant  $|I - U|$  where  $U$  and  $I - U$  are defined in Theorem 6.8.1. Then for  $k = 2$  and an arbitrary  $h$ ,  $E[u_4^h | H_o] = |I - U|^h$ .

**Proof:** On letting  $k = 2$  in (6.8.8), we obtain the  $h$ -th moment of  $u_4 | H_o$  as

$$E[u_4^h | H_o] = c_{4,p} \frac{\prod_{j=1}^p \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\{\prod_{j=1}^{p_1} \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)\} \{\prod_{j=1}^{p_2} \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)\}}. \quad (i)$$

After canceling  $p_2$  of the gamma functions, the remaining gamma product in the numerator of (i) is

$$\Gamma\left(\alpha - \frac{p_2}{2}\right) \Gamma\left(\alpha - \frac{p_2 + 1}{2}\right) \cdots \Gamma\left(\alpha - \frac{p-1}{2}\right) = \Gamma_{p_1}\left(\alpha - \frac{p_2}{2}\right), \quad \alpha = \frac{m}{2} + h,$$

excluding  $\pi^{\frac{p_1(p_1-1)}{4}}$ . The remainder of the gamma product present in the denominator is comprised of the gamma functions coming from  $\Gamma_{p_1}(\frac{m}{2} + h)$ , excluding  $\pi^{\frac{p_1(p_1-1)}{4}}$ . The normalizing constant will automatically take care of the factors containing  $\pi$ . Now, the resulting part containing  $h$  is  $\Gamma_{p_1}(\frac{m}{2} - \frac{p_2}{2} + h) / \Gamma_{p_1}(\frac{m}{2} + h)$ , which is the gamma ratio in the  $h$ -th moment of a  $p_1 \times p_1$  real matrix-variate type-1 beta distribution with the parameters  $(\frac{m}{2} - \frac{p_2}{2}, \frac{p_2}{2})$ .

Since this happens to be  $E[|I - U|^h]$  for  $I - U$  distributed as is specified in Theorem 6.8.1, the Corollary is established.

An asymptotic result can be established from Corollary 6.5.1 and the  $\lambda$ -criterion for testing block-diagonality or equivalently the independence of subvectors in a  $p$ -variate

Gaussian population. The resulting chisquare variable will have  $2 \sum_j \delta_j$  degrees of freedom where  $\delta_j$  is as defined in Corollary 6.5.1 for the second parameter of the real scalar type-1 beta distribution. Referring to (6.8.10), we have

$$\begin{aligned} \sum_j \delta_j &= \sum_{j=p_1+1}^p \frac{j-1}{2} - \sum_{j=2}^k \sum_{i=1}^{p_j} \frac{i-1}{2} = \sum_{j=p_1+1}^p \frac{j-1}{2} - \sum_{j=2}^k \frac{p_j(p_j-1)}{4} \\ &= \sum_{j=1}^p \frac{j-1}{2} - \sum_{j=1}^k \frac{p_j(p_j-1)}{4} = \frac{p(p-1)}{4} - \sum_{j=1}^k \frac{p_j(p_j-1)}{4} = \sum_{j=1}^k \frac{p_j(p-p_j)}{4}. \end{aligned}$$

Accordingly, the degrees of freedom of the resulting chisquare is  $2[\sum_{j=1}^k \frac{p_j(p-p_j)}{4}] = \sum_{j=1}^k \frac{p_j(p-p_j)}{2}$  in the real case. It can also be observed that the number of restrictions imposed by the null hypothesis  $H_o$  is obtained by first letting all the off-diagonal elements of  $\Sigma = \Sigma'$  equal to zero and subtracting the off-diagonal elements of the  $k$  diagonal blocks which produces  $\frac{p(p-1)}{2} - \sum_{j=1}^k \frac{p_j(p_j-1)}{2} = \sum_{j=1}^k \frac{p_j(p-p_j)}{2}$ . In the complex case, the number of degrees of freedom will be twice that obtained for the real case, the chisquare variable remaining a real scalar chisquare random variable. This is now stated as a theorem.

**Theorem 6.8.2.** Consider the  $\lambda$ -criterion given in (6.8.1) in the real case and let the corresponding  $\lambda$  in the complex case be  $\tilde{\lambda}$ . Then  $-2 \ln \lambda \rightarrow \chi_{\delta}^2$  as  $n \rightarrow \infty$  where  $n$  is the sample size and  $\delta = \sum_{j=1}^k \frac{p_j(p-p_j)}{2}$ , which is also the number of restrictions imposed by  $H_o$ . Analogously, in the complex case,  $-2 \ln \tilde{\lambda} \rightarrow \chi_{\tilde{\delta}}^2$  as  $n \rightarrow \infty$ , where the chisquare variable remains a real scalar chisquare random variable,  $\tilde{\delta} = \sum_{j=1}^k p_j(p-p_j)$  and  $n$  denotes the sample size.

## 6.9. Hypothesis that the Mean Value and Covariance Matrix are Given

Consider a real  $p$ -variate Gaussian population  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , and a simple random sample,  $X_1, \dots, X_n$ , from this population, the  $X_i$ 's being iid as  $X_j$ . Let the sample mean and the sample sum of products matrix be denoted by  $\bar{X}$  and  $S$ , respectively. Consider the hypothesis  $H_o : \mu = \mu_o, \Sigma = \Sigma_o$  where  $\mu_o$  and  $\Sigma_o$  are specified. Let us examine the likelihood ratio test for testing  $H_o$  and obtain the resulting  $\lambda$ -criterion. Let the parameter space be  $\Omega = \{(\mu, \Sigma) | \Sigma > O, -\infty < \mu_j < \infty, j = 1, \dots, p, \mu' = (\mu_1, \dots, \mu_p)\}$ . Let the joint density of  $X_1, \dots, X_n$  be denoted by  $L$ . Then, as previously obtained, the maximum value of  $L$  is

$$\max_{\Omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}} \quad (6.9.1)$$

and the maximum under  $H_o$  is

$$\max_{H_o} L = \frac{e^{-\frac{1}{2}\text{tr}(\sum_{j=1}^n (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o))}}{(2\pi)^{\frac{np}{2}} |\Sigma_o|^{\frac{n}{2}}}. \quad (6.9.2)$$

Thus,

$$\lambda = \frac{\max_{H_o} L}{\max_{\Omega} L} = \frac{e^{\frac{np}{2}} |S|^{\frac{n}{2}}}{n^{\frac{np}{2}} |\Sigma_o|^{\frac{n}{2}}} e^{-\frac{1}{2}\text{tr}(\sum_{j=1}^n (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o))}. \quad (6.9.3)$$

We reject  $H_o$  for small values of  $\lambda$ . Since the exponential part dominates the polynomial part for large values, we reject for large values of the exponent, excluding  $(-1)$ , which means for large values of  $\sum_{j=1}^n (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o) \sim \chi_{np}^2$  since  $(X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o) \stackrel{iid}{\sim} \chi_p^2$  for each  $j$ . Hence the criterion consists of

$$\text{rejecting } H_o \text{ if the observed values of } \sum_{j=1}^n (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o) \geq \chi_{np, \alpha}^2$$

with

$$Pr\{\chi_{np}^2 \geq \chi_{np, \alpha}^2\} = \alpha. \quad (6.9.4)$$

Let us determine the  $h$ -th moment of  $\lambda$  for an arbitrary  $h$ . Note that

$$\lambda^h = \frac{e^{\frac{np}{2}}}{n^{\frac{np}{2}} |\Sigma_o|^{\frac{n}{2}}} |S|^{\frac{nh}{2}} e^{-\frac{h}{2}\text{tr}(\Sigma_o^{-1} S) - \frac{hn}{2}(\bar{X} - \mu_o)' \Sigma_o^{-1} (\bar{X} - \mu_o)}. \quad (6.9.5)$$

Since  $\lambda$  contains  $S$  and  $\bar{X}$  and these quantities are independently distributed, we can integrate out the part containing  $S$  over a Wishart density having  $m = n - 1$  degrees of freedom and the part containing  $\bar{X}$  over the density of  $\bar{X}$ . Thus, for  $m = n - 1$ ,

$$\begin{aligned} E \left[ (|S|^{\frac{nh}{2}} / |\Sigma_o|^{\frac{nh}{2}}) e^{-\frac{h}{2}\text{tr}(\Sigma_o^{-1} S)} | H_o \right] &= \frac{\int_{S>O} |S|^{\frac{m}{2} + \frac{nh}{2} - \frac{p+1}{2}} e^{-\frac{(1+h)}{2}\text{tr}(\Sigma_o^{-1} S)} dS}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) |\Sigma_o|^{\frac{n}{2}(1+h) - \frac{1}{2}}} \\ &= 2^{\frac{np}{2}} \frac{\Gamma_p(\frac{n}{2}(1+h) - \frac{1}{2})}{\Gamma_p(\frac{n}{2} - \frac{1}{2})} (1+h)^{-[\frac{n}{2}(1+h) - \frac{1}{2}]p}. \quad (i) \end{aligned}$$

Under  $H_o$ , the integral over  $\bar{X}$  gives

$$\int_{\bar{X}} \frac{\sqrt{n}}{(2\pi)^{\frac{p}{2}} |\Sigma_o|^{\frac{1}{2}}} e^{-(1+h)\frac{n}{2}(\bar{X} - \mu_o)' \Sigma_o^{-1} (\bar{X} - \mu_o)} d\bar{X} = (1+h)^{-\frac{p}{2}}. \quad (ii)$$

From (i) and (ii), we have

$$E[\lambda^h | H_o] = \frac{e^{\frac{np}{2}} 2^{\frac{np}{2}}}{n^{\frac{np}{2}}} \frac{\Gamma_p(\frac{n}{2}(1+h) - \frac{1}{2})}{\Gamma_p(\frac{n-1}{2})} (1+h)^{-[\frac{n}{2}(1+h)]p}. \quad (6.9.6)$$

The inversion of this expression is quite involved due to branch points. Let us examine the asymptotic case as  $n \rightarrow \infty$ . On expanding the gamma functions by making use of the version of Stirling's asymptotic approximation formula for gamma functions given in (6.5.14), namely  $\Gamma(z + \eta) \approx \sqrt{2\pi} z^{z+\eta-\frac{1}{2}} e^{-z}$  for  $|z| \rightarrow \infty$  and  $\eta$  bounded, we have

$$\begin{aligned} \frac{\Gamma_p(\frac{n}{2}(1+h) - \frac{1}{2})}{\Gamma_p(\frac{n-1}{2})} &= \prod_{j=1}^p \frac{\Gamma(\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j-1}{2})}{\Gamma(\frac{n-1}{2} - \frac{j-1}{2})} \\ &= \prod_{j=1}^p \frac{\sqrt{2\pi} [\frac{n}{2}(1+h)]^{\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j-1}{2}} e^{-\frac{n}{2}(1+h)}}{\sqrt{2\pi} [\frac{n}{2}]^{\frac{n}{2} - \frac{1}{2} - \frac{j-1}{2}} e^{-\frac{n}{2}}} \\ &= \left[\frac{n}{2}\right]^{\frac{np}{2}} e^{-\frac{np}{2}} (1+h)^{\frac{n}{2}(1+h)p - \frac{p}{2} - \frac{p(p+1)}{4}}. \end{aligned} \quad (iii)$$

Thus, as  $n \rightarrow \infty$ , it follows from (6.9.6) and (iii) that

$$E[\lambda^h | H_o] = (1+h)^{-\frac{1}{2}(p + \frac{p(p+1)}{2})}, \quad (6.9.7)$$

which implies that, asymptotically,  $-2 \ln \lambda$  has a real scalar chisquare distribution with  $p + \frac{p(p+1)}{2}$  degrees of freedom in the real Gaussian case. Hence the following result:

**Theorem 6.9.1.** *Given a  $N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , population, consider the hypothesis  $H_o : \mu = \mu_o, \Sigma = \Sigma_o$  where  $\mu_o$  and  $\Sigma_o$  are specified. Let  $\lambda$  denote the  $\lambda$ -criterion for testing this hypothesis. Then, in the real case,  $-2 \ln \lambda \rightarrow \chi_{\delta}^2$  as  $n \rightarrow \infty$  where  $\delta = p + \frac{p(p+1)}{2}$  and, in the corresponding complex case,  $-2 \ln \lambda \rightarrow \chi_{\delta_1}^2$  as  $n \rightarrow \infty$  where  $\delta_1 = 2p + p(p+1)$ , the chisquare variable remaining a real scalar chisquare random variable.*

**Note 6.9.1.** In the real case, observe that the hypothesis  $H_o : \mu = \mu_o, \Sigma = \Sigma_o$  imposes  $p$  restrictions on the  $\mu$  parameters and  $\frac{p(p+1)}{2}$  restrictions on the  $\Sigma$  parameters, for a total of  $p + \frac{p(p+1)}{2}$  restrictions, which corresponds to the degrees of freedom for the asymptotic chisquare distribution in the real case. In the complex case, there are twice as many restrictions.

**Example 6.9.1.** Consider the real trivariate Gaussian distribution  $N_3(\mu, \Sigma)$ ,  $\Sigma > O$  and the hypothesis  $H_o : \mu = \mu_o, \Sigma = \Sigma_o$  where  $\mu_o, \Sigma_o$  and an observed sample of size 5 are as follows:

$$\mu_o = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \Sigma_o = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & -2 \\ 0 & -2 & 3 \end{bmatrix} \Rightarrow |\Sigma_o| = 24, \text{Cof}(\Sigma_o) = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 9 & 6 \\ 0 & 6 & 12 \end{bmatrix},$$

$$\Sigma_o^{-1} = \frac{\text{Cof}(\Sigma_o)}{|\Sigma_o|} = \frac{1}{24} \begin{bmatrix} 8 & 0 & 0 \\ 0 & 9 & 6 \\ 0 & 6 & 12 \end{bmatrix};$$

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, X_3 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}.$$

Now,

$$\begin{aligned} (X_1 - \mu_o)' \Sigma_o^{-1} (X_1 - \mu_o) &= \frac{36}{24}, & (X_2 - \mu_o)' \Sigma_o^{-1} (X_2 - \mu_o) &= \frac{33}{24}, \\ (X_3 - \mu_o)' \Sigma_o^{-1} (X_3 - \mu_o) &= \frac{104}{24}, & (X_4 - \mu_o)' \Sigma_o^{-1} (X_4 - \mu_o) &= \frac{144}{24}, \\ (X_5 - \mu_o)' \Sigma_o^{-1} (X_5 - \mu_o) &= \frac{20}{24}, \end{aligned}$$

and

$$\sum_{j=1}^5 (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o) = \frac{1}{24} [36 + 33 + 104 + 144 + 20] = \frac{337}{24} = 14.04.$$

Note that, in this example,  $n = 5$ ,  $p = 3$  and  $np = 15$ . Letting the significance level of the test be  $\alpha = 0.05$ ,  $H_o$  is not rejected since  $14.04 < \chi_{15, 0.05}^2 = 25$ .

## 6.10. Testing Hypotheses on Linear Regression Models or Linear Hypotheses

Let the  $p \times 1$  real vector  $X_j$  have an expected value  $\mu$  and a covariance matrix  $\Sigma > O$  for  $j = 1, \dots, n$ , and the  $X_j$ 's be independently distributed. Let  $X_j, \mu, \Sigma$  be partitioned as follows where  $x_{1j}, \mu_1$  and  $\sigma_{11}$  are  $1 \times 1$ ,  $\mu_{(2)}, \Sigma_{21}$  are  $(p-1) \times 1$ ,  $\Sigma_{12} = \Sigma'_{21}$  and  $\Sigma_{22}$  is  $(p-1) \times (p-1)$ :

$$X_j = \begin{bmatrix} x_{1j} \\ X_{(2)j} \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_{(2)} \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \quad (i)$$

If the conditional expectation of  $x_{1j}$ , given  $X_{(2)j}$  is linear in  $X_{(2)j}$ , then omitting the subscript  $j$  since the  $X_j$ 's are iid, it was established in Eq. (3.3.5) that

$$E[x_1|X_{(2)}] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_{(2)} - \mu_{(2)}). \quad (6.10.1)$$

When the regression is linear, the best linear predictor of  $x_1$  in terms of  $X_{(2)}$  will be of the form

$$E[x_1|X_{(2)}] - E(x_1) = \beta'(X_{(2)} - E(X_{(2)})), \quad \beta' = (\beta_2, \dots, \beta_p). \quad (6.10.2)$$

Then, by appealing to properties of the conditional expectation and conditional variance, it was shown in Chap. 3 that  $\beta' = \Sigma_{12}\Sigma_{22}^{-1}$ . Hypothesizing that  $X_{(2)}$  is not random, or equivalently that the predictor function is a function of the preassigned values of  $X_{(2)}$ , amounts to testing whether  $\Sigma_{12}\Sigma_{22}^{-1} = O$ . Noting that  $\Sigma_{22} > O$  since  $\Sigma > O$ , the null hypothesis thus reduces to  $H_o : \Sigma_{12} = O$ . If the original population  $X$  is  $p$ -variate real Gaussian, this hypothesis is then equivalent to testing the independence of  $x_1$  and  $X_{(2)}$ . Actually, this has already been discussed in Sect. 6.8.2 for the case of  $k = 2$ , and is also tantamount to testing whether the population multiple correlation  $\rho_{1.(2..k)} = 0$ . Assuming that the population is Gaussian and letting of  $u = \lambda^{\frac{2}{n}}$  where  $\lambda$  is the lambda criterion,  $u \sim$  type-1 beta( $\frac{n-p}{2}$ ,  $\frac{p-1}{2}$ ) under the null hypothesis; this was established in Theorem 6.8.1 for  $p_1 = 1$  and  $p_2 = p - 1$ . Then,  $v = \frac{u}{1-u} \sim$  type-2 beta ( $\frac{n-p}{2}$ ,  $\frac{p-1}{2}$ ), that is,  $v \sim \frac{n-p}{p-1} F_{n-p, p-1}$  or  $\frac{(p-1)}{(n-p)} \frac{u}{1-u} \sim F_{n-p, p-1}$ . Hence, in order to test  $H_o : \beta = O$ ,

$$\text{reject } H_o \text{ if } F_{n-p, p-1} \geq F_{n-p, p-1, \alpha}, \text{ with } Pr\{F_{n-p, p-1} \geq F_{n-p, p-1, \alpha}\} = \alpha \quad (6.10.3)$$

The test statistic  $u$  is of the form

$$u = \frac{|S|}{s_{11}|S_{22}|}, \quad S \sim W_p(n-1, \Sigma), \quad \Sigma > O, \quad \frac{(p-1)}{(n-p)} \frac{u}{1-u} \sim F_{n-p, p-1},$$

where the submatrices of  $S$ ,  $s_{11}$  is  $1 \times 1$  and  $S_{22}$  is  $(p-1) \times (p-1)$ . Observe that the number of parameters being restricted by the hypothesis  $\Sigma_{12} = O$  is  $p_1 p_2 = 1(p-1) = p-1$ . Hence as  $n \rightarrow \infty$ , the null distribution of  $-2 \ln \lambda$  is a real scalar chisquare having  $p-1$  degrees of freedom. Thus, the following result:

**Theorem 6.10.1.** *Let the  $p \times 1$  vector  $X_j$  be partitioned into the subvectors  $x_{1j}$  of order 1 and  $X_{(2)j}$  of order  $p-1$ . Let the regression of  $x_{1j}$  on  $X_{(2)j}$  be linear in  $X_{(2)j}$ , that is,  $E[x_{1j}|X_{(2)j}] - E(x_{1j}) = \beta'(X_{(2)j} - E(X_{(2)j}))$ . Consider the hypothesis  $H_o : \beta = O$ . Let  $X_j \sim N_p(\mu, \Sigma)$ ,  $\Sigma > O$ , for  $j = 1, \dots, n$ , the  $X_j$ 's being independently distributed, and let  $\lambda$  be the  $\lambda$ -criterion for testing this hypothesis. Then, as  $n \rightarrow \infty$ ,  $-2 \ln \lambda \rightarrow \chi_{p-1}^2$ .*

**Example 6.10.1.** Let the population be  $N_3(\mu, \Sigma)$ ,  $\Sigma > O$ , and the observed sample of size  $n = 5$  be

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, X_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, X_4 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, X_5 = \begin{bmatrix} -1 \\ -1 \\ -2 \end{bmatrix}.$$

The resulting sample average  $\bar{X}$  and deviation vectors are then

$$\begin{aligned} \bar{X} &= \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, X_1 - \bar{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 \\ 3 \\ 3 \end{bmatrix}, \\ X_2 - \bar{X} &= \frac{1}{5} \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix}, X_3 - \bar{X} = \frac{1}{5} \begin{bmatrix} -6 \\ 3 \\ -2 \end{bmatrix}, \\ X_4 - \bar{X} &= \frac{1}{5} \begin{bmatrix} 4 \\ 3 \\ 8 \end{bmatrix}, X_5 - \bar{X} = \frac{1}{5} \begin{bmatrix} -6 \\ -7 \\ -12 \end{bmatrix}. \end{aligned}$$

Letting

$$\mathbf{X} = [X_1, \dots, X_5], \bar{\mathbf{X}} = [\bar{X}, \dots, \bar{X}] \text{ and } S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})',$$

$$\mathbf{X} - \bar{\mathbf{X}} = \frac{1}{5} \begin{bmatrix} 4 & 4 & -6 & 4 & -6 \\ 3 & -2 & 3 & 3 & -7 \\ 3 & 3 & -2 & 8 & -12 \end{bmatrix},$$

$$S = \frac{1}{5^2} \begin{bmatrix} 120 & 40 & 140 \\ 40 & 80 & 105 \\ 140 & 105 & 230 \end{bmatrix} = \begin{bmatrix} s_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix},$$

$$s_{11} = \frac{120}{25}, S_{12} = \frac{1}{25}[40, 140],$$

$$S_{22} = \frac{1}{25} \begin{bmatrix} 80 & 105 \\ 105 & 230 \end{bmatrix}, S_{22}^{-1} = \frac{25}{7375} \begin{bmatrix} 230 & -105 \\ -105 & 80 \end{bmatrix},$$

$$S_{12}S_{22}^{-1}S_{21} = \left[ \frac{1}{(25)^2} \right] \left[ \frac{25}{7375} \right] [40 \quad 140] \begin{bmatrix} 230 & -105 \\ -105 & 80 \end{bmatrix} \begin{bmatrix} 40 \\ 140 \end{bmatrix} = \frac{760000}{(25)(7375)},$$

so that the test statistic is

$$\begin{aligned} u &= \frac{s_{11} - S_{12}S_{22}^{-1}S_{21}}{s_{11}} = 1 - \frac{S_{12}S_{22}^{-1}S_{21}}{s_{11}} = 1 - r_{1.(2,3)}^2 \\ &= 1 - \frac{760000}{(120)(7375)} = 1 - 0.859 = 0.141 \Rightarrow \\ v &= \frac{(p-1)}{(n-p)} \frac{u}{1-u} = \left(\frac{2}{2}\right) \left(\frac{0.141}{0.859}\right) = 0.164, \quad v \sim F_{n-p, p-1}. \end{aligned}$$

Let us test  $H_o$  at the significance level  $\alpha = 0.05$ . In that case, the critical value which is available from  $F$  tables is  $F_{n-p, p-1, \alpha} = F_{2, 2, 0.05} = 19$ . Since the observed value of  $v$  is  $0.164 < 19$ ,  $H_o$  is not rejected.

**Note 6.10.1.** Observe that

$$u = 1 - r_{1.(2,3)}^2, \quad r_{1.(2,3)}^2 = \frac{S_{12}S_{22}^{-1}S_{21}}{s_{11}}$$

where  $r_{1.(2,3)}$  is the sample multiple correlation between the first component of  $X_j \sim N_3(\mu, \Sigma)$ ,  $\Sigma > O$ , and the other two components of  $X_j$ . If the population covariance matrix  $\Sigma$  is similarly partitioned, that is,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \text{where } \sigma_{11} \text{ is } 1 \times 1, \quad \Sigma_{22} \text{ is } (p-1) \times (p-1),$$

then, the population multiple correlation coefficient is  $\rho_{1.(2,3)}$  where

$$\rho_{1.(2,3)}^2 = \frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}{\sigma_{11}}.$$

Thus, if  $\Sigma_{12} = \Sigma_{21}' = O$ ,  $\rho_{1.(2,3)} = 0$  and conversely since  $\sigma_{11} > 0$  and  $\Sigma_{22} > O \Rightarrow \Sigma_{22}^{-1} > O$ . The regression coefficient  $\beta$  being equal to the transpose of  $\Sigma_{12}\Sigma_{22}^{-1}$ ,  $\Sigma_{12} = O$  also implies that the regression coefficient  $\beta = O$  and conversely. Accordingly, the hypothesis that the regression coefficient vector  $\beta = O$  is equivalent to hypothesizing that the population multiple correlation  $\rho_{1.(2,\dots,p)} = 0$ , which also implies the hypothesis that the two subvectors are independently distributed in the multivariate normal case, or that the covariance matrix  $\Sigma_{12} = O$ , the only difference being that the test on regression coefficients is in the conditional space whereas testing the independence of the subvectors or whether the population multiple correlation equals zero is carried out in the entire space. The numerical example included in this section also illustrates the main result presented in Sect. 6.8.1 in connection with testing whether a population multiple correlation coefficient is equal to zero.



### 6.10.1. A simple linear model

Consider a linear model of the following form where a real scalar variable  $y$  is estimated by a linear function of pre-assigned real scalar variables  $z_1, \dots, z_q$ :

$$y_j = \beta_o + \beta_1 z_{1j} + \dots + \beta_q z_{qj} + e_j, \quad j = 1, \dots, n \quad (i)$$

where  $y_1, \dots, y_n$  are  $n$  observations on  $y$ ,  $z_{i1}, z_{i2}, \dots, z_{in}$ ,  $i = 1, \dots, q$ , are preassigned values on  $z_1, \dots, z_q$ , and  $\beta_o, \beta_1, \dots, \beta_q$  are unknown parameters. The random components  $e_j$ ,  $j = 1, \dots, n$ , are the corresponding sum total contributions coming from all unknown factors. There are two possibilities with respect to this model:  $\beta_o = 0$  or  $\beta_o \neq 0$ . If  $\beta_o = 0$ ,  $\beta_o$  is omitted in model (i) and we let  $y_j = x_j$ . If  $\beta_o \neq 0$ , the model is modified by taking  $x_j = y_j - \bar{y}$  where  $\bar{y} = \frac{1}{n}(y_1 + \dots + y_n)$ , then becoming

$$x_j = y_j - \bar{y} = \beta_1(z_{1j} - \bar{z}_1) + \dots + \beta_q(z_{qj} - \bar{z}_q) + \epsilon_j \quad (ii)$$

for some error term  $\epsilon_j$ , where  $\bar{z}_i = \frac{1}{n}(z_{i1} + \dots + z_{in})$ ,  $i = 1, \dots, q$ . Letting  $Z'_j = (z_{1j}, \dots, z_{qj})$  if  $\beta_o = 0$  and  $Z'_j = (z_{1j} - \bar{z}_1, \dots, z_{qj} - \bar{z}_q)$ , otherwise, equation (ii) can be written in vector/matrix notation as follows:

$$\begin{aligned} \epsilon_j &= (x_j - \beta_1 z_{1j} - \dots - \beta_q z_{qj}) = (x_j - \beta' Z_j), \quad \beta' = (\beta_1, \dots, \beta_q), \\ \sum_{j=1}^n \epsilon_j^2 &= \sum_{j=1}^n (x_j - \beta_1 z_{1j} - \dots - \beta_q z_{qj})^2 = \sum_{j=1}^n (x_j - \beta' Z_j)^2. \end{aligned}$$

Letting

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} z_{11} & \dots & z_{1n} \\ z_{21} & \dots & z_{2n} \\ \vdots & \ddots & \vdots \\ z_{q1} & \dots & z_{qn} \end{bmatrix}, \quad (iii)$$

$$\epsilon' = (X' - \beta' Z) \Rightarrow \sum_{j=1}^n \epsilon_j^2 = \epsilon' \epsilon = (X' - \beta' Z)(X - Z' \beta). \quad (iv)$$

The least squares minimum is thus available by differentiating  $\epsilon' \epsilon$  with respect to  $\beta$ , equating the resulting expression to a null vector and solving, which will produce a single critical point that corresponds to the minimum as the maximum occurs at  $+\infty$ :

$$\begin{aligned} \frac{\partial}{\partial \beta} \left( \sum_{j=1}^n \epsilon_j^2 \right) = 0 &\Rightarrow \sum_{j=1}^n x_j Z_j' - \hat{\beta}' \sum_{j=1}^n Z_j Z_j' \\ &\Rightarrow \hat{\beta} = \left( \sum_{j=1}^n Z_j Z_j' \right)^{-1} \left( \sum_{j=1}^n x_j Z_j' \right); \end{aligned} \quad (v)$$

$$\frac{\partial}{\partial \beta} (\epsilon' \epsilon) = 0 \Rightarrow \hat{\beta} = (ZZ')^{-1} ZX, \quad (vi)$$

that is,

$$\hat{\beta} = \left( \sum_{j=1}^n Z_j Z_j' \right)^{-1} \left( \sum_{j=1}^n x_j Z_j' \right). \quad (6.10.4)$$

Since the  $z_{ij}$ 's are preassigned quantities, it can be assumed without any loss of generality that  $(ZZ')$  is nonsingular, and thereby that  $(ZZ')^{-1}$  exists, so that the least squares minimum, usually denoted by  $s^2$ , is available by substituting  $\hat{\beta}$  for  $\beta$  in  $\epsilon' \epsilon$ . Then, at  $\beta = \hat{\beta}$ ,

$$\begin{aligned} \epsilon' |_{\beta=\hat{\beta}} &= X' - X' Z' (ZZ')^{-1} Z = X' [I - Z' (ZZ')^{-1} Z] \text{ so that} \\ s^2 = \epsilon' \epsilon |_{\beta=\hat{\beta}} &= X' [I - Z' (ZZ')^{-1} Z] X \end{aligned} \quad (6.10.5)$$

where  $I - Z' (ZZ')^{-1} Z$  is idempotent and of rank  $(n - 1) - q$ . Observe that if  $\beta_o \neq 0$  in (i) and we had proceeded without eliminating  $\beta_o$ , then  $\beta$  would have been of order  $(k + 1) \times 1$  and  $I - Z' (ZZ')^{-1} Z$ , of rank  $n - (q + 1) = n - 1 - q$ , whereas if  $\beta_o \neq 0$  and we had eliminated  $\beta_o$  from the model, then the rank of  $I - Z' (ZZ')^{-1} Z$  would have been  $(n - 1) - q$ , that is, unchanged, since  $\sum_{j=1}^n (x_j - \bar{x})^2 = X' [I - \frac{1}{n} J J'] X$ ,  $J' = (1, \dots, 1)$  and the rank of  $I - \frac{1}{n} J J' = n - 1$ .

Some distributional assumptions on  $\epsilon_j$  are required in order to test hypotheses on  $\beta$ . Let  $\epsilon_j \sim N_1(0, \sigma^2)$ ,  $\sigma^2 > 0$ ,  $j = 1, \dots, n$ , be independently distributed. Then  $x_j \sim N_1(\beta' Z_j, \sigma^2)$ ,  $j = 1, \dots, n$  are independently distributed but not identically distributed as the mean value depends on  $j$ . Under the normality assumption for the  $\epsilon_j$ 's, it can readily be seen that the least squares estimators of  $\beta$  and  $\sigma^2$  coincide with the maximum likelihood estimators. It can also be observed that  $\sigma^2$  is estimated by  $\frac{s^2}{n}$  where  $n$  is the sample size. In this simple linear regression context, the parameter space  $\Omega = \{(\beta, \sigma^2) | \sigma^2 > 0\}$ . Thus, under the normality assumption, the maximum of the likelihood function  $L$  is given by

$$\max_{\Omega} L = \frac{e^{-\frac{n}{2} \frac{s^2}{\sigma^2}}}{(2\pi)^{\frac{n}{2}} [s^2]^{\frac{n}{2}}}. \quad (vi)$$

Under the hypothesis  $H_o : \beta = O$  or  $\beta_1 = 0 = \dots = \beta_q$ , the least squares minimum, usually denoted as  $s_o^2$ , is  $X'X$  and, assuming normality, the maximum of the likelihood function under  $H_o$  is the following:

$$\max_{H_o} L = \frac{e^{-\frac{n}{2}} n^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} [s_o^2]^{\frac{n}{2}}}. \quad (\text{vii})$$

Thus, the  $\lambda$ -criterion is

$$\begin{aligned} \lambda &= \frac{\max_{H_o} L}{\max_{\Omega} L} = \left[ \frac{s^2}{s_o^2} \right]^{\frac{n}{2}} \\ \Rightarrow u &= \lambda^{\frac{2}{n}} = \frac{X'[I - Z'(ZZ')^{-1}Z]X}{X'X} \\ &= \frac{X'[I - Z'(ZZ')^{-1}Z]X}{X'[I - Z'(ZZ')^{-1}Z]X + X'Z'(ZZ')^{-1}ZX} \\ &= \frac{1}{1 + u_1} \end{aligned} \quad (6.10.6)$$

where

$$u_1 = \frac{X'Z'(ZZ')^{-1}ZX}{X'[I - Z'(ZZ')^{-1}Z]X} = \frac{s_o^2 - s^2}{s^2} \quad (\text{viii})$$

with the matrices  $Z'(ZZ')^{-1}Z$  and  $I - Z'(ZZ')^{-1}Z$  being idempotent, mutually orthogonal, and of ranks  $q$  and  $(n - 1) - q$ , respectively. We can interpret  $s_o^2 - s^2$  as the sum of squares due to the hypothesis and  $s^2$  as the residual part. Under the normality assumption,  $s_o^2 - s^2$  and  $s^2$  are independently distributed in light of independence of quadratic forms that was discussed in Sect. 3.4.1; moreover, their representations as quadratic forms in idempotent matrices of ranks  $q$  and  $(n - 1) - q$  implies that  $\frac{s_o^2 - s^2}{s^2} \sim \chi_q^2$  and  $\frac{s^2}{s^2} \sim \chi_{(n-1)-q}^2$ . Accordingly, under the null hypothesis,

$$u_2 = \frac{(s_o^2 - s^2)/q}{s^2/[(n - 1) - q]} \sim F_{q, n-1-q}, \quad (6.10.7)$$

that is, an  $F$ -statistic having  $q$  and  $n - 1 - q$  degrees of freedom. Thus, we reject  $H_o$  for small values of  $\lambda$  or equivalently for large values of  $u_2$  or large values of  $F_{q, n-1-q}$ . Hence, the following criterion:

$$\text{Reject } H_o \text{ if the observed value of } u_2 \geq F_{q, n-1-q, \alpha}, \quad Pr\{F_{q, n-1-q} \geq F_{q, n-1-q, \alpha}\} = \alpha. \quad (6.10.8)$$

A detailed discussion of the real scalar variable case is provided in Mathai and Haubold (2017).

### 6.10.2. Hypotheses on individual parameters

Denoting the expected value of  $(\cdot)$  by  $E[(\cdot)]$ , it follows from (6.10.4) that

$$\begin{aligned} E[\hat{\beta}] &= (ZZ')^{-1}ZE(X) = (ZZ')^{-1}ZZ'\beta = \beta, \quad E[X] = Z'\beta, \quad \text{and} \\ \text{Cov}(\hat{\beta}) &= (ZZ')^{-1}Z[\text{Cov}(X)]Z'(ZZ')^{-1} = (ZZ')^{-1}Z(\sigma^2I)Z'(ZZ')^{-1} \\ &= \sigma^2(ZZ')^{-1}. \end{aligned}$$

Under the normality assumption on  $x_j$ , we have  $\hat{\beta} \sim N_q(\beta, \sigma^2(ZZ')^{-1})$ . Letting the  $(r, r)$ -th diagonal element of  $(ZZ')^{-1}$  be  $b_{rr}$ , then  $\hat{\beta}_r$ , the estimator of the  $r$ -th component of the parameter vector  $\beta$ , is distributed as  $\hat{\beta}_r \sim N_1(\beta_r, \sigma^2b_{rr})$ , so that

$$\frac{\hat{\beta}_r - \beta_r}{\hat{\sigma} \sqrt{b_{rr}}} \sim t_{n-1-q} \quad (6.10.9)$$

where  $t_{n-1-q}$  denotes a Student- $t$  distribution having  $n - 1 - q$  degrees of freedom and  $\hat{\sigma}^2 = \frac{s^2}{n-1-q}$  is an unbiased estimator for  $\sigma^2$ . On writing  $s^2$  in terms of  $\epsilon$ , it is easily seen that  $E[s^2] = (n - 1 - q)\sigma^2$  where  $s^2$  is the least squares minimum in the entire parameter space  $\Omega$ . Thus, one can test hypotheses on  $\beta_r$  and construct confidence intervals for that parameter by means of the Student- $t$  statistic specified in (6.10.9) or its square  $t_{n-1-q}^2$  which has an  $F$  distribution having 1 and  $n - 1 - q$  degrees of freedom, that is,  $t_{n-1-q}^2 \sim F_{1, n-1-q}$ .

**Example 6.10.2.** Let us consider a linear model of the following form:

$$y_j = \beta_o + \beta_1 z_{1j} + \cdots + \beta_q z_{qj} + e_j, \quad j = 1, \dots, n,$$

where the  $z_{ij}$ 's are preassigned numbers of the variable  $z_i$ . Let us take  $n = 5$  and  $q = 2$ , so that the sample is of size 5 and, excluding  $\beta_o$ , the model has two parameters. Let the observations on  $y$  and the preassigned values on the  $z_i$ 's be the following:

$$\begin{aligned} 1 &= \beta_o + \beta_1(0) + \beta_2(1) + e_1 \\ 2 &= \beta_o + \beta_1(1) + \beta_2(-1) + e_2 \\ 4 &= \beta_o + \beta_1(-1) + \beta_2(2) + e_3 \\ 6 &= \beta_o + \beta_1(2) + \beta_2(-2) + e_4 \\ 7 &= \beta_o + \beta_1(-2) + \beta_2(5) + e_5. \end{aligned}$$

The averages on  $y$ ,  $z_1$  and  $z_2$  are then

$$\bar{y} = \frac{1}{5}[1 + 2 + 4 + 6 + 7] = 4, \quad \bar{z}_1 = \frac{1}{5}[0 + 1 + (-1) + 2 + (-2)] = 0 \quad \text{and}$$

$$\bar{z}_2 = \frac{1}{5}[1 + (-1) + 2 + (-2) + 5] = 1,$$

and, in terms of deviations, the model becomes

$$x_j = y_j - \bar{y} = \beta_1(z_{1j} - \bar{z}_1) + \beta_2(z_{2j} - \bar{z}_2) + \epsilon_j, \quad \epsilon_j = e_j - \bar{e}.$$

That is,

$$\begin{bmatrix} -3 \\ -2 \\ 0 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -2 \\ -1 & 1 \\ 2 & -3 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} \Rightarrow X = Z'\beta + \epsilon.$$

When minimizing  $\epsilon'\epsilon = (X - Z'\beta)'(X - Z'\beta)$ , we determined that  $\hat{\beta}$ , the least squares estimate of  $\beta$ , the least squares minimum  $s^2$  and  $s_o^2 - s^2 =$  corresponding to the sum of squares due to  $\beta$ , could be express as

$$\hat{\beta} = (ZZ')^{-1}ZX, \quad s_o^2 - s^2 = X'Z'(ZZ')^{-1}ZX,$$

$$s^2 = X'[I - Z'(ZZ')^{-1}Z]X \quad \text{and that}$$

$$Z'(ZZ')^{-1}Z = [Z'(ZZ')^{-1}Z]^2, \quad I - Z'(ZZ')^{-1}Z = [I - Z'(ZZ')^{-1}Z]^2,$$

$$[I - Z'(ZZ')^{-1}Z][Z'(ZZ')^{-1}Z] = O.$$

Let us evaluate those quantities:

$$Z = \begin{bmatrix} 0 & 1 & -1 & 2 & -2 \\ 0 & -2 & 1 & -3 & 4 \end{bmatrix}, \quad ZZ' = \begin{bmatrix} 10 & -17 \\ -17 & 30 \end{bmatrix},$$

$$|ZZ'| = 11, \quad \text{Cof}(ZZ') = \begin{bmatrix} 30 & 17 \\ 17 & 10 \end{bmatrix}, \quad (ZZ')^{-1} = \frac{1}{11} \begin{bmatrix} 30 & 17 \\ 17 & 10 \end{bmatrix},$$

$$(ZZ')^{-1}Z = \frac{1}{11} \begin{bmatrix} 30 & 17 \\ 17 & 10 \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 & 2 & -2 \\ 0 & -2 & 1 & -3 & 4 \end{bmatrix}$$

$$= \frac{1}{11} \begin{bmatrix} 0 & -4 & -13 & 9 & 8 \\ 0 & -3 & -7 & 4 & 6 \end{bmatrix};$$

$$\begin{aligned}\hat{\beta} &= (ZZ')^{-1}ZX = \frac{1}{11} \begin{bmatrix} 0 & -4 & -13 & 9 & 8 \\ 0 & -3 & -7 & 4 & 6 \end{bmatrix} \begin{bmatrix} -3 \\ -2 \\ 0 \\ 2 \\ 3 \end{bmatrix} \\ &= \frac{1}{11} \begin{bmatrix} 50 \\ 32 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix};\end{aligned}$$

$$\begin{aligned}Z'(ZZ')^{-1}Z &= \frac{1}{11} \begin{bmatrix} 0 & 0 \\ 1 & -2 \\ -1 & 1 \\ 2 & -3 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 0 & -4 & -13 & 9 & 8 \\ 0 & -3 & -7 & 4 & 6 \end{bmatrix} \\ &= \frac{1}{11} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 & -4 \\ 0 & 1 & 6 & -5 & -2 \\ 0 & 1 & -5 & 6 & -2 \\ 0 & -4 & -2 & -2 & 8 \end{bmatrix}.\end{aligned}$$

Then,

$$\begin{aligned}X'Z'(ZZ')^{-1}ZX &= \frac{1}{11} \begin{bmatrix} -3 & -2 & 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 & -4 \\ 0 & 1 & 6 & -5 & -2 \\ 0 & 1 & -5 & 6 & -2 \\ 0 & -4 & -2 & -2 & 8 \end{bmatrix} \begin{bmatrix} -3 \\ -2 \\ 0 \\ 2 \\ 3 \end{bmatrix} \\ &= \frac{120}{11}, \quad X'X = 26,\end{aligned}$$

$$X'[I - Z'(ZZ')^{-1}Z]X = 26 - \frac{120}{11} = \frac{166}{11}, \quad \text{with } n = 5 \text{ and } q = 2.$$

The test statistics  $u_1$  and  $u_2$  and their observed values are the following:

$$\begin{aligned} u_1 &= \frac{X'Z'(ZZ')^{-1}ZX}{X'[I - Z'(ZZ')^{-1}Z]X} = \frac{s_o^2 - s^2}{s^2} \\ &= \frac{120}{166} = 0.72; \\ u_2 &= \frac{(s_o^2 - s^2)/q}{s^2/(n - 1 - q)} \sim F_{q, n-1-q} \\ &= \frac{120/2}{166/2} = 0.72. \end{aligned}$$

Letting the significance level be  $\alpha = 0.05$ , the required tabulated critical value is  $F_{q, n-1-q, \alpha} = F_{2, 2, 0.05} = 19$ . Since  $0.72 < 19$ , the hypothesis  $H_o : \beta = O$  is not rejected. Thus, we will not proceed to test individual hypotheses on the regression coefficients  $\beta_1$  and  $\beta_2$ . For tests on general linear models, refer for instance to Mathai (1971).

### 6.11. Problem Involving Two or More Independent Gaussian Populations

Consider  $k$  independent  $p$ -variate real normal populations  $X_j \sim N_p(\mu_{(j)}, \Sigma)$ ,  $\Sigma > O$ ,  $j = 1, \dots, k$ , having the same nonsingular covariance matrix  $\Sigma$  but possibly different mean values. We consider the problem of testing hypotheses on linear functions of the mean values. Let  $b = a_1\mu_{(1)} + \dots + a_k\mu_{(k)}$  where  $a_1, \dots, a_k$ , are real scalar constants, and the null hypothesis be  $H_o : b = b_o$  (given), which means that the  $a_j$ 's and  $\mu_{(j)}$ 's,  $j = 1, \dots, k$ , are all specified. It is also assumed that  $\Sigma$  is known. Suppose that simple random samples of sizes  $n_1, \dots, n_k$  from these  $k$  independent normal populations can be secured, and let the sample values be  $X_{jq}$ ,  $q = 1, \dots, n_j$ , where  $X_{j1}, \dots, X_{jn_j}$  are iid as  $N_p(\mu_{(j)}, \Sigma)$ ,  $\Sigma > O$ . Let the sample averages be denoted by  $\bar{X}_j = \frac{1}{n_j} \sum_{q=1}^{n_j} X_{jq}$ ,  $j = 1, \dots, k$ . Consider the test statistic  $U_k = a_1\bar{X}_1 + \dots + a_k\bar{X}_k$ . Since the populations are independent and  $U_k$  is a linear function of independent vector normal variables,  $U_k$  is normally distributed with the mean value  $b = a_1\mu_{(1)} + \dots + a_k\mu_{(k)}$  and covariance matrix  $\frac{1}{n}\Sigma$ , where  $\frac{1}{n} = (\frac{a_1^2}{n_1} + \dots + \frac{a_k^2}{n_k})$  and so,  $\sqrt{n}\Sigma^{-\frac{1}{2}}(U_k - b) \sim N_p(O, I)$ . Then, under the hypothesis  $H_o : b = b_o$  (given), which is being tested against the alternative  $H_1 : b \neq b_o$ , the test criterion is obtained by proceeding as was done in the single population case. Thus, the test statistic is  $z = n(U_k - b_o)' \Sigma^{-1}(U_k - b_o) \sim \chi_p^2$  and the criterion will be to reject the null hypothesis for large values of the  $z$ . Accordingly, the criterion is

$$\begin{aligned} \text{Reject } H_o : b = b_o \text{ if the observed value of } n(U_k - b_o)' \Sigma^{-1}(U_k - b_o) &\geq \chi_{p, \alpha}^2, \\ \text{with } Pr\{\chi_p^2 \geq \chi_{p, \alpha}^2\} &= \alpha. \end{aligned} \quad (6.11.1)$$

In particular, suppose that we wish to test the hypothesis  $H_o : \delta = \mu_{(1)} - \mu_{(2)} = \delta_o$ , such as  $\delta_o = 0$  as is often the case, against the natural alternative. In this case, when  $\delta_o = 0$ , the null hypothesis is that the mean value vectors are equal, that is,  $\mu_{(1)} = \mu_{(2)}$ , and the test statistic is  $z = n(\bar{X}_1 - \bar{X}_2)' \Sigma^{-1} (\bar{X}_1 - \bar{X}_2) \sim \chi_p^2$  with  $\frac{1}{n} = \frac{1}{n_1} + \frac{1}{n_2}$ , the test criterion being

$$\begin{aligned} \text{Reject } H_o : \mu_{(1)} - \mu_{(2)} = 0 \text{ if the observed value of } z &\geq \chi_{p, \alpha}^2, \\ \text{with } Pr\{\chi_p^2 \geq \chi_{p, \alpha}^2\} &= \alpha. \end{aligned} \quad (6.11.2)$$

For a numerical example, the reader is referred to Example 6.2.3. One can also determine the power of the test or the probability of rejecting  $H_o : \delta = \mu_{(1)} - \mu_{(2)} = \delta_o$ , under an alternative hypothesis, in which case the distribution of  $z$  is a noncentral chisquare variable with  $p$  degrees of freedom and non-centrality parameter  $\lambda = \frac{1}{2} \frac{n_1 n_2}{n_1 + n_2} (\delta - \delta_o)' \Sigma^{-1} (\delta - \delta_o)$ ,  $\delta = \mu_{(1)} - \mu_{(2)}$ , where  $n_1$  and  $n_2$  are the sample sizes. Under the null hypothesis, the non-centrality parameter  $\lambda$  is equal to zero. The power is given by

$$\text{Power} = Pr\{\text{reject } H_o | H_1\} = Pr\{\chi_p^2(\lambda) \geq \chi_{p, \alpha}^2(\lambda)\}. \quad (6.11.3)$$

When the population covariance matrices are identical and the common covariance matrix is unknown, one can also construct a statistic for testing hypotheses on linear functions of the mean value vectors by making use of steps parallel to those employed in the single population case, with the resulting criterion being based on Hotelling's  $T^2$  statistic for testing  $H_o : \mu_{(1)} = \mu_{(2)}$ .

### 6.11.1. Equal but unknown covariance matrices

Let us consider the same procedure as in Sect. 6.11 to test a hypothesis on a linear function  $b = a_1 \mu_{(1)} + \cdots + a_k \mu_{(k)}$  where  $a_1, \dots, a_k$  are known real scalar constants and  $\mu_{(j)}$ ,  $j = 1, \dots, k$ , are the population mean values. We wish to test the hypothesis  $H_o : b = b_o$  (given) in the sense all the mean values  $\mu_{(j)}$ ,  $j = 1, \dots, k$  and  $a_1, \dots, a_k$ , are specified. Let  $U_k = a_1 \bar{X}_1 + \cdots + a_k \bar{X}_k$  as previously defined. Then,  $E[U_k] = b$  and  $\text{Cov}(U_k) = (\frac{a_1^2}{n_1} + \cdots + \frac{a_k^2}{n_k}) \Sigma$ , where  $(\frac{a_1^2}{n_1} + \cdots + \frac{a_k^2}{n_k}) \equiv \frac{1}{n}$  for some symbol  $n$ . The common covariance matrix  $\Sigma$  has the MLE  $\frac{1}{n_1 + \cdots + n_k} (S_1 + \cdots + S_k)$  where  $S_j$  is the sample sum of products matrix for the  $j$ -th Gaussian population. It has been established that  $S = S_1 + \cdots + S_k$  has a Wishart distribution with  $(n_1 - 1) + \cdots + (n_k - 1) = N - k$ ,  $N = n_1 + \cdots + n_k$ , degrees of freedom, that is,

$$S \sim W_p(N - k, \Sigma), \quad \Sigma > O, \quad N = n_1 + \cdots + n_k. \quad (6.11.4)$$



Then, when  $\Sigma$  is unknown, it follows from a derivation parallel to that provided in Sect. 6.3 for the single population case that

$$w \equiv n(U_k - b)'S^{-1}(U_k - b) \sim \text{type-2 beta} \left( \frac{p}{2}, \frac{N - k - p}{2} \right), \quad (6.11.5)$$

or,  $w$  has a real scalar type-2 beta distribution with the parameters  $(\frac{p}{2}, \frac{N-k-p}{2})$ . Letting  $w = \frac{p}{N-k-p}F$ , this  $F$  is an  $F$ -statistic with  $p$  and  $N - k - p$  degrees of freedom.

**Theorem 6.11.1.** *Let  $U_k, n, N, b, S$  be as defined above. Then  $w = n(U_k - b)'S^{-1}(U_k - b)$  has a real scalar type-2 beta distribution with the parameters  $(\frac{p}{2}, \frac{N-k-p}{2})$ . Letting  $w = \frac{p}{N-k-p}F$ , this  $F$  is an  $F$ -statistic with  $p$  and  $N - k - p$  degrees of freedom.*

Hence for testing the hypothesis  $H_o : b = b_o$  (given), the criterion is the following:

$$\text{Reject } H_o \text{ if the observed value of } F = \frac{N - k - p}{p}w \geq F_{p, N-k-p, \alpha}, \quad (6.11.6)$$

$$\text{with } Pr\{F_{p, N-k-p} \geq F_{p, N-k-p, \alpha}\} = \alpha. \quad (6.11.7)$$

Note that by exploiting the connection between type-1 and type-2 real scalar beta random variables, one can obtain a number of properties on this  $F$ -statistic.

This situation has already been covered in Theorem 6.3.4 for the case  $k = 2$ .

## 6.12. Equality of Covariance Matrices in Independent Gaussian Populations

Let  $X_j \sim N_p(\mu_{(j)}, \Sigma_j)$ ,  $\Sigma_j > O$ ,  $j = 1, \dots, k$ , be independently distributed real  $p$ -variate Gaussian populations. Consider simple random samples of sizes  $n_1, \dots, n_k$  from these  $k$  populations, whose sample values, denoted by  $X_{jq}$ ,  $q = 1, \dots, n_j$ , are iid as  $X_{j1}$ ,  $j = 1, \dots, k$ . The sample sums of products matrices denoted by  $S_1, \dots, S_k$ , respectively, are independently distributed as Wishart matrix random variables with  $n_j - 1$ ,  $j = 1, \dots, k$ , degrees of freedoms. The joint density of all the sample values is then given by

$$L = \prod_{j=1}^k L_j, \quad L_j = \frac{e^{-\frac{1}{2}\text{tr}(\Sigma_j^{-1}S_j) - \frac{n_j}{2}(\bar{X}_j - \mu_{(j)})' \Sigma_j^{-1}(\bar{X}_j - \mu_{(j)})}}{(2\pi)^{\frac{n_j p}{2}} |\Sigma_j|^{\frac{n_j}{2}}}, \quad (6.12.1)$$

the MLE's of  $\mu_{(j)}$  and  $\Sigma_j$  being  $\hat{\mu}_{(j)} = \bar{X}_j$  and  $\hat{\Sigma}_j = \frac{1}{n_j}S_j$ . The maximum of  $L$  in the entire parameter space  $\Omega$  is

$$\max_{\Omega} L = \prod_{j=1}^k \max_{\Omega} L_j = \frac{\left[ \prod_{j=1}^k n_j^{\frac{n_j p}{2}} \right] e^{-\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} \prod_{j=1}^k |S_j|^{\frac{n_j}{2}}}, \quad N = n_1 + \dots + n_k. \quad (i)$$

Let us test the hypothesis of equality of covariance matrices:

$$H_o : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k = \Sigma$$

where  $\Sigma$  is unknown. Under this null hypothesis, the MLE of  $\mu_{(j)}$  is  $\bar{X}_j$  and the MLE of the common  $\Sigma$  is  $\frac{1}{N}(S_1 + \dots + S_k) = \frac{1}{N}S$ ,  $N = n_1 + \dots + n_k$ ,  $S = S_1 + \dots + S_k$ . Thus, the maximum of  $L$  under  $H_o$  is

$$\max_{H_o} L = \frac{N^{\frac{Np}{2}} e^{-\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} \prod_{j=1}^k |S|^{\frac{n_j}{2}}} = \frac{N^{\frac{Np}{2}} e^{-\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} |S|^{\frac{N}{2}}}, \quad (ii)$$

and the  $\lambda$ -criterion is the following:

$$\lambda = \frac{N^{\frac{Np}{2}} \left\{ \prod_{j=1}^k |S_j|^{\frac{n_j}{2}} \right\}}{|S|^{\frac{N}{2}} \left\{ \prod_{j=1}^k n_j^{\frac{n_j p}{2}} \right\}}. \quad (6.12.2)$$

Let us consider the  $h$ -th moment of  $\lambda$  for an arbitrary  $h$ . Letting  $c = \frac{N^{\frac{Np}{2}}}{\left\{ \prod_{j=1}^k n_j^{\frac{n_j p}{2}} \right\}}$ ,

$$\lambda^h = c^h \frac{\left\{ \prod_{j=1}^k |S_j|^{\frac{n_j h}{2}} \right\}}{|S|^{\frac{Nh}{2}}} = c^h \left\{ \prod_{j=1}^k |S_j|^{\frac{n_j h}{2}} \right\} |S|^{-\frac{Nh}{2}}. \quad (6.12.3)$$

The factor causing a difficulty, namely  $|S|^{-\frac{Nh}{2}}$ , will be replaced by an equivalent integral. Letting  $Y > O$  be a real  $p \times p$  positive definite matrix, we have the identity

$$|S|^{-\frac{Nh}{2}} = \frac{1}{\Gamma_p(\frac{Nh}{2})} \int_{Y>O} |Y|^{\frac{Nh}{2} - \frac{p+1}{2}} e^{-\text{tr}(YS)} dY, \quad \Re\left(\frac{Nh}{2}\right) > \frac{p-1}{2}, \quad S > O, \quad (6.12.4)$$

where

$$\text{tr}(YS) = \text{tr}(YS_1) + \dots + \text{tr}(YS_k). \quad (iii)$$

Thus, once (6.12.4) is substituted in (6.12.3),  $\lambda^h$  splits into products involving  $S_j$ ,  $j = 1, \dots, k$ ; this enables one to integrate out over the densities of  $S_j$ , which are Wishart densities with  $m_j = n_j - 1$  degrees of freedom. Noting that the exponent involving  $S_j$  is  $-\frac{1}{2}\text{tr}(\Sigma_j^{-1}S_j) - \text{tr}(YS_j) = -\frac{1}{2}\text{tr}[S_j(\Sigma_j^{-1} + 2Y)]$ , the integral over the Wishart density of  $S_j$  gives the following:

$$\begin{aligned}
 & \prod_{j=1}^k \frac{1}{2^{\frac{m_j p}{2}} \Gamma_p(\frac{m_j}{2}) |\Sigma_j|^{\frac{m_j}{2}}} \int_{S_j > O} |S_j|^{\frac{m_j}{2} + \frac{n_j h}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \text{tr}[S_j(\Sigma^{-1} + 2Y)]} dS_j \\
 &= \prod_{j=1}^k \frac{2^{\frac{n_j h p}{2}} \Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2}) |\Sigma_j^{-1} + 2Y|^{-\frac{(m_j + n_j h)}{2}}}{\Gamma_p(\frac{m_j}{2}) |\Sigma_j|^{\frac{m_j}{2}}} \\
 &= 2^{\frac{N p h}{2}} \prod_{j=1}^k \frac{|\Sigma_j|^{\frac{n_j h}{2}} |I + 2\Sigma_j Y|^{-\frac{(m_j + n_j h)}{2}} \Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{m_j}{2})}. \tag{iv}
 \end{aligned}$$

Thus, on substituting (iv) in  $E[\lambda^h]$ , we have

$$\begin{aligned}
 E[\lambda^h] &= \frac{c^h 2^{\frac{N p h}{2}}}{\Gamma_p(\frac{N h}{2})} \int_{Y > O} |Y|^{\frac{N h}{2} - \frac{p+1}{2}} \\
 &\quad \times \left\{ \prod_{j=1}^k \frac{|\Sigma_j|^{\frac{n_j h}{2}} |I + 2\Sigma_j Y|^{-\frac{(m_j + n_j h)}{2}} \Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{m_j}{2})} \right\} dY, \tag{6.12.5}
 \end{aligned}$$

which is the non-null  $h$ -th moment of  $\lambda$ . The  $h$ -th null moment is available when  $\Sigma_1 = \dots = \Sigma_k = \Sigma$ . In the null case,

$$\begin{aligned}
 & \prod_{j=1}^k |I + 2Y \Sigma_j|^{-\frac{(m_j + n_j h)}{2}} |\Sigma_j|^{\frac{n_j h}{2}} \frac{\Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{m_j}{2})} \\
 &= |\Sigma|^{\frac{N h}{2}} |I + 2Y \Sigma|^{-\frac{(N-k) + N h}{2}} \left\{ \prod_{j=1}^k \frac{\Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{m_j}{2})} \right\}. \tag{v}
 \end{aligned}$$

Then, substituting (v) in (6.12.5) and integrating out over  $Y$  produces

$$E[\lambda^h | H_o] = c^h \frac{\Gamma_p(\frac{N-k}{2})}{\Gamma_p(\frac{N-k}{2} + \frac{N h}{2})} \left\{ \prod_{j=1}^k \frac{\Gamma_p(\frac{n_j-1}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{n_j-1}{2})} \right\}. \tag{6.12.6}$$

Observe that when  $h = 0$ ,  $E[\lambda^h | H_o] = 1$ . For  $h = s - 1$  where  $s$  is a complex parameter, we have the Mellin transform of the density of  $\lambda$ , denoted by  $f(\lambda)$ , which can be expressed as follows in terms of an H-function:

$$f(\lambda) = \frac{1}{c} \frac{\Gamma_p(\frac{N-k}{2})}{\left\{ \prod_{j=1}^k \Gamma_p(\frac{n_j-1}{2}) \right\}} H_{1,k}^{k,0} \left[ \lambda \left| \begin{matrix} (-\frac{k}{2}, \frac{N}{2}) \\ (-\frac{1}{2}, \frac{n_j}{2}), j=1, \dots, k \end{matrix} \right. \right], \quad 0 < \lambda < 1, \tag{6.12.7}$$

and zero elsewhere, where the H-function is defined in Sect. 5.4.3, more details being available from Mathai and Saxena (1978) and Mathai et al. (2010). Since the coefficients of  $\frac{h}{2}$  in the gammas, that is,  $n_1, \dots, n_k$  and  $N$ , are all positive integers, one can expand all gammas by using the multiplication formula for gamma functions, and then,  $f(\lambda)$  can be expressed in terms of a G-function as well. It may be noted from (6.12.5) that for obtaining the non-null moments, and thereby the non-null density, one has to integrate out  $Y$  in (6.12.5). This has not yet been worked out for a general  $k$ . For  $k = 2$ , one can obtain a series form in terms of zonal polynomials for the integral in (6.12.5). The rather intricate derivations are omitted.

### 6.12.1. Asymptotic behavior

We now investigate the asymptotic behavior of  $-2 \ln \lambda$  as  $n_j \rightarrow \infty$ ,  $j = 1, \dots, k$ ,  $N = n_1 + \dots + n_k$ . On expanding the real matrix-variate gamma functions in the the gamma ratio involving  $h$  in (6.12.6), we have

$$\frac{\prod_{j=1}^k \Gamma_p\left(\frac{n_j}{2}(1+h) - \frac{1}{2}\right)}{\Gamma_p\left(\frac{N}{2}(1+h) - \frac{k}{2}\right)} \rightarrow \frac{\prod_{j=1}^k \left\{ \prod_{i=1}^p \Gamma\left(\frac{n_j}{2}(1+h) - \frac{1}{2} - \frac{i-1}{2}\right) \right\}}{\prod_{i=1}^p \Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} - \frac{i-1}{2}\right)}, \quad (i)$$

excluding the factor containing  $\pi$ . Letting  $\frac{n_j}{2}(1+h) \rightarrow \infty$ ,  $j = 1, \dots, k$ , and  $\frac{N}{2}(1+h) \rightarrow \infty$  as  $n_j \rightarrow \infty$ ,  $j = 1, \dots, k$ , with  $N \rightarrow \infty$ , we now express all the gamma functions in (i) in terms of Sterling's asymptotic formula. For the numerator, we have

$$\begin{aligned} & \prod_{j=1}^k \left\{ \prod_{i=1}^p \Gamma\left(\frac{n_j}{2}(1+h) - \frac{1}{2} - \frac{i-1}{2}\right) \right\} \\ & \rightarrow \prod_{j=1}^k \left\{ \prod_{i=1}^p \sqrt{(2\pi)} \left[\frac{n_j}{2}(1+h)\right]^{\frac{n_j}{2}(1+h) - \frac{1}{2} - \frac{i}{2}} e^{-\frac{n_j}{2}(1+h)} \right\} \\ & = \prod_{j=1}^k (\sqrt{2\pi})^p \left[\frac{n_j}{2}(1+h)\right]^{\frac{n_j}{2}(1+h)p - \frac{p}{2} - \frac{p(p+1)}{4}} e^{-\frac{n_j}{2}(1+h)p} \\ & = (\sqrt{2\pi})^{kp} \left[ \prod_{j=1}^k \left(\frac{n_j}{2}\right)^{\frac{n_j}{2}(1+h)p - \frac{p}{2} - \frac{p(p+1)}{4}} \right] e^{-\frac{N}{2}(1+h)p} \\ & \quad \times (1+h)^{\frac{N}{2}(1+h) - \frac{kp}{2} - k\frac{p(p+1)}{4}}, \end{aligned} \quad (ii)$$

and the denominator in (i) has the following asymptotic representation:

$$\begin{aligned} \prod_{i=1}^p \Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} - \frac{i-1}{2}\right) &\rightarrow \prod_{i=1}^p (\sqrt{2\pi}) \left[\frac{N}{2}(1+h)\right]^{\frac{N}{2}(1+h) - \frac{k}{2} - \frac{i}{2}} e^{-\frac{N}{2}(1+h)} \\ &= (\sqrt{2\pi})^p \left[\frac{N}{2}\right]^{\frac{N}{2}(1+h)p - \frac{kp}{2} - \frac{p(p+1)}{4}} e^{-\frac{N}{2}(1+h)p} \\ &\quad \times (1+h)^{\frac{N}{2}(1+h)p - \frac{kp}{2} - \frac{p(p+1)}{4}}. \end{aligned} \tag{iii}$$

Now, expanding the gammas in the constant part  $\Gamma_p(\frac{N-k}{2}) / \prod_{j=1}^k \Gamma_p(\frac{n_j-1}{2})$  and then taking care of  $c^h$ , we see that the factors containing  $\pi$ , the  $n_j$ 's and  $N$  disappear leaving

$$(1+h)^{-(k-1)\frac{p(p+1)}{4}}.$$

Hence  $-2 \ln \lambda \rightarrow \chi_{(k-1)\frac{p(p+1)}{2}}^2$  and hence we have the following result:

**Theorem 6.12.1.** *Consider the  $\lambda$ -criterion in (6.12.2) or the null density in (6.12.7). When  $n_j \rightarrow \infty, j = 1, \dots, k$ , the asymptotic null density of  $-2 \ln \lambda$  is a real scalar chisquare with  $(k-1)\frac{p(p+1)}{2}$  degrees of freedom.*

Observe that the number of parameters restricted by the null hypothesis  $H_o : \Sigma_1 = \dots = \Sigma_k = \Sigma$  where  $\Sigma$  is unknown, is  $(k-1)$  times the number of distinct parameters in  $\Sigma$ , which is  $\frac{p(p+1)}{2}$ , which coincides with the number of degrees of freedom of the asymptotic chisquare distribution under  $H_o$ .

**6.13. Testing the Hypothesis that  $k$  Independent  $p$ -variate Real Gaussian Populations are Identical and Multivariate Analysis of Variance**

Consider  $k$  independent  $p$ -variate real Gaussian populations  $X_{ij} \sim N_p(\mu^{(i)}, \Sigma_i), \Sigma_i > O, i = 1, \dots, k$ , and  $j = 1, \dots, n_i$ , where the  $p \times 1$  vector  $X_{ij}$  is the  $j$ -th sample value belonging to the  $i$ -th population, these samples (iid variables) being of sizes  $n_1, \dots, n_k$  from these  $k$  populations. The joint density of all the sample values, denoted by  $L$ , can be expressed as follows:

$$\begin{aligned} L &= \prod_{i=1}^k L_i, \quad L_i = \prod_{j=1}^{n_i} \frac{e^{-\frac{1}{2} \sum_{j=1}^{n_i} (X_{ij} - \mu^{(i)})' \Sigma_i^{-1} (X_{ij} - \mu^{(i)})}}{(2\pi)^{\frac{n_i p}{2}} |\Sigma_i|^{\frac{n_i}{2}}} \\ &= \frac{e^{-\frac{1}{2} \text{tr}(\Sigma_i^{-1} S_i) - \frac{n_i}{2} (\bar{X}_i - \mu^{(i)})' \Sigma_i^{-1} (\bar{X}_i - \mu^{(i)})}}{(2\pi)^{\frac{n_i}{2}} |\Sigma_i|^{\frac{n_i}{2}}} \end{aligned}$$

where  $\bar{X}_i = \frac{1}{n_i}(X_{i1} + \cdots + X_{in_i})$ ,  $i = 1, \dots, k$ , and  $E[X_{ij}] = \mu^{(i)}$ ,  $j = 1, \dots, n_i$ . Then, letting  $N = n_1 + \cdots + n_k$ ,

$$\begin{aligned} \max_{\Omega} L &= \prod_{i=1}^k \max_{\Omega} L_i = \prod_{i=1}^k \frac{n_i^{\frac{n_i p}{2}} e^{-\frac{n_i p}{2}}}{(2\pi)^{\frac{n_i p}{2}} |S_i|^{\frac{n_i}{2}}} \\ &= \frac{\{\prod_{i=1}^k n_i^{\frac{n_i p}{2}}\} e^{-\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} \{\prod_{i=1}^k |S_i|^{\frac{n_i}{2}}\}}. \end{aligned}$$

Consider the hypothesis  $H_o : \mu^{(1)} = \cdots = \mu^{(k)} = \mu$ ,  $\Sigma_1 = \cdots = \Sigma_k = \Sigma$ , where  $\mu$  and  $\Sigma$  are unknown. This corresponds to the hypothesis of equality of these  $k$  populations. Under  $H_o$ , the maximum likelihood estimator (MLE) of  $\mu$ , denoted by  $\hat{\mu}$ , is given by  $\hat{\mu} = \frac{1}{N}[n_1 \bar{X}_1 + \cdots + n_k \bar{X}_k]$  where  $N$  and  $\bar{X}_i$  are as defined above. As for the common  $\Sigma$ , its MLE is

$$\begin{aligned} \hat{\Sigma} &= \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \hat{\mu})(X_{ij} - \hat{\mu})' \\ &= \frac{1}{N} [S_1 + \cdots + S_k + \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'] \end{aligned}$$

where  $S_i$  is the sample sum of products matrix for the  $i$ -th sample, observing that

$$\begin{aligned} \sum_{j=1}^{n_i} (X_{ij} - \hat{\mu})(X_{ij} - \hat{\mu})' &= \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i + \bar{X}_i - \hat{\mu})(X_{ij} - \bar{X}_i + \bar{X}_i - \hat{\mu})' \\ &= \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' + \sum_{j=1}^{n_i} (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})' \\ &= S_i + n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'. \end{aligned}$$

Hence the maximum of the likelihood function under  $H_o$  is the following:

$$\max_{H_o} L = \frac{e^{-\frac{Np}{2}} N^{\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} |S + \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'|^{\frac{N}{2}}}$$

where  $S = S_1 + \cdots + S_k$ . Therefore the  $\lambda$ -criterion is given by

$$\lambda = \frac{\max_{H_o}}{\max_{\Omega}} = \frac{\{\prod_{i=1}^k |S_i|^{\frac{n_i}{2}}\} N^{\frac{Np}{2}}}{\{\prod_{i=1}^k n_i^{\frac{n_i p}{2}}\} |S + \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'|^{\frac{N}{2}}}. \quad (6.13.1)$$

### 6.13.1. Conditional and marginal hypotheses

For convenience, we may split  $\lambda$  into the product  $\lambda_1\lambda_2$  where  $\lambda_1$  is the  $\lambda$ -criterion for the conditional hypothesis  $H_{o1} : \mu^{(1)} = \dots = \mu^{(k)} = \mu$  given that  $\Sigma_1 = \dots = \Sigma_k = \Sigma$  and  $\lambda_2$  is the  $\lambda$ -criterion for the marginal hypothesis  $H_{o2} : \Sigma_1 = \dots = \Sigma_k = \Sigma$  where  $\mu$  and  $\Sigma$  are unknown. The conditional hypothesis  $H_{o1}$  is actually the null hypothesis usually being made when establishing the multivariate analysis of variance (MANOVA) procedure. We will only consider  $H_{o1}$  since the marginal hypothesis  $H_{o2}$  has already been discussed in Sect. 6.12. When the  $\Sigma_i$ 's are assumed to be equal, the common  $\Sigma$  is estimated by the MLE  $\frac{1}{N}(S_1 + \dots + S_k)$  where  $S_i$  is the sample sum of products matrix in the  $i$ -th population. The common  $\mu$  is estimated by  $\frac{1}{N}(n_1\bar{X}_1 + \dots + n_k\bar{X}_k)$ . Accordingly, the  $\lambda$ -criterion for this conditional hypothesis is the following:

$$\lambda_1 = \frac{|S|^{\frac{N}{2}}}{|S + \sum_{i=1}^k n_i(\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'|^{\frac{N}{2}}} \quad (6.13.2)$$

where  $S = S_1 + \dots + S_k$ ,  $\hat{\mu} = \frac{1}{N}(n_1\bar{X}_1 + \dots + n_k\bar{X}_k)$ ,  $N = n_1 + \dots + n_k$ . Note that the  $S_i$ 's are independently Wishart distributed with  $n_i - 1$  degrees of freedom, that is,  $S_i \stackrel{ind}{\sim} W_p(n_i - 1, \Sigma)$ ,  $i = 1, \dots, k$ , and hence  $S \sim W_p(N - k, \Sigma)$ . Let

$$Q = \sum_{i=1}^k n_i(\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'.$$

Since  $Q$  only contains sample averages and the sample averages and the sample sum of products matrices are independently distributed,  $Q$  and  $S$  are independently distributed. Moreover, since we can write  $\bar{X}_i - \hat{\mu}$  as  $(\bar{X}_i - \mu) - (\hat{\mu} - \mu)$ , where  $\mu$  is the common true mean value vector, without any loss of generality we can deem the  $\bar{X}_i$ 's to be independently  $N_p(O, \frac{1}{n_i}\Sigma)$  distributed,  $i = 1, \dots, k$ , and letting  $Y_i = \sqrt{n_i}\bar{X}_i$ , one has  $Y_i \stackrel{iid}{\sim} N_p(O, \Sigma)$  under the hypothesis  $H_{o1}$ . Now, observe that

$$\begin{aligned} \bar{X}_i - \hat{\mu} &= \bar{X}_i - \frac{1}{N}(n_1\bar{X}_1 + \dots + n_k\bar{X}_k) \\ &= -\frac{n_1}{N}\bar{X}_1 - \dots - \frac{n_{i-1}}{N}\bar{X}_{i-1} + \left(1 - \frac{n_i}{N}\right)\bar{X}_i - \frac{n_{i+1}}{N}\bar{X}_{i+1} - \dots - \frac{n_k}{N}\bar{X}_k, \quad (i) \end{aligned}$$

$$\begin{aligned}
Q &= \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})' \\
&= \sum_{i=1}^k n_i \bar{X}_i \bar{X}_i' - N \hat{\mu} \hat{\mu}' \\
&= \sum_{i=1}^k Y_i Y_i' - \frac{1}{N} (\sqrt{n_1} Y_1 + \cdots + \sqrt{n_k} Y_k) (\sqrt{n_1} Y_1 + \cdots + \sqrt{n_k} Y_k)',
\end{aligned} \tag{ii}$$

where  $\sqrt{n_1} Y_1 + \cdots + \sqrt{n_k} Y_k = (Y_1, \dots, Y_k) D J$  with  $J$  being the  $k \times 1$  vector of unities,  $J' = (1, \dots, 1)$ , and  $D = \text{diag}(\sqrt{n_1}, \dots, \sqrt{n_k})$ . Thus, we can express  $Q$  as follows:

$$Q = (Y_1, \dots, Y_k) \left[ I - \frac{1}{N} D J J' D \right] (Y_1, \dots, Y_k)'. \tag{iii}$$

Let  $B = \frac{1}{N} D J J' D$  and  $A = I - B$ . Then, observing that  $J' D^2 J = N$ , both  $B$  and  $A$  are idempotent matrices, where  $B$  is of rank 1 since the trace of  $B$  or equivalently the trace of  $\frac{1}{N} J' D^2 J$  is equal to one, so that the trace of  $A$  which is also its rank, is  $k - 1$ . Then, there exists an orthonormal matrix  $P$ ,  $P P' = I_k$ ,  $P' P = I_k$ , such that  $A = P' \begin{bmatrix} I_{k-1} & O \\ O' & 0 \end{bmatrix} P$  where  $O$  is a  $(k - 1) \times 1$  null vector,  $O'$  being its transpose. Letting  $(U_1, \dots, U_k) = (Y_1, \dots, Y_k) P'$ , the  $U_i$ 's are still independently  $N_p(O, \Sigma)$  distributed under  $H_{01}$ , so that

$$\begin{aligned}
Q &= (U_1, \dots, U_k) \begin{bmatrix} I_{k-1} & O \\ O' & 0 \end{bmatrix} (U_1, \dots, U_k)' = (U_1, \dots, U_{k-1}, O) (U_1, \dots, U_{k-1}, O)' \\
&= \sum_{i=1}^{k-1} U_i U_i' \sim W_p(k - 1, \Sigma).
\end{aligned} \tag{iv}$$

Thus,  $S + \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})' \sim W_p(N - 1, \Sigma)$ , which clearly is not independently distributed of  $S$ , referring to the ratio in (6.13.2).

### 6.13.2. Arbitrary moments of $\lambda_1$

Given (6.13.2), we have

$$\lambda_1 = \frac{|S|^{\frac{N}{2}}}{|S + Q|^{\frac{N}{2}}} \Rightarrow \lambda_1^h = |S|^{\frac{Nh}{2}} |S + Q|^{-\frac{Nh}{2}} \tag{v}$$



where  $|S + Q|^{-\frac{Nh}{2}}$  will be replaced by the equivalent integral

$$|S + Q|^{-\frac{Nh}{2}} = \frac{1}{\Gamma_p(\frac{Nh}{2})} \int_{T > O} |T|^{\frac{Nh}{2} - \frac{p+1}{2}} e^{-\text{tr}((S+Q)T)} dT \quad (vi)$$

with the  $p \times p$  matrix  $T > O$ . Hence, the  $h$ -th moment of  $\lambda_1$ , for arbitrary  $h$ , is the following expected value:

$$E[\lambda_1^h | H_{o1}] = E\{|S|^{\frac{Nh}{2}} |S + Q|^{-\frac{Nh}{2}}\}. \quad (vii)$$

We now evaluate (vii) by integrating out over the Wishart density of  $S$  and over the joint multinormal density for  $U_1, \dots, U_{k-1}$ :

$$\begin{aligned} E[\lambda_1^h | H_{o1}] &= \frac{1}{\Gamma_p(\frac{Nh}{2})} \int_{T > O} |T|^{\frac{Nh}{2} - \frac{p+1}{2}} \\ &\quad \times \frac{\int_{S > O} |S|^{\frac{Nh}{2} + \frac{N-k}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S) - \text{tr}(ST)} dS}{2^{\frac{(N-k)p}{2}} \Gamma_p(\frac{N-k}{2}) |\Sigma|^{\frac{N-k}{2}}} \\ &\quad \times \int_{U_1, \dots, U_{k-1}} \frac{e^{-\frac{1}{2} \sum_{i=1}^{k-1} U_i' \Sigma^{-1} U_i - \sum_{i=1}^{k-1} \text{tr}(T U_i U_i')}}{(2\pi)^{\frac{(k-1)p}{2}} |\Sigma|^{\frac{k-1}{2}}} dU_1 \wedge \dots \wedge dU_{k-1} \wedge dT. \end{aligned}$$

The integral over  $S$  is evaluated as follows:

$$\begin{aligned} &\frac{\int_{S > O} |S|^{\frac{Nh}{2} + \frac{N-k}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}(S+2\sigma T))} dS}{2^{\frac{(N-k)p}{2}} \Gamma_p(\frac{N-k}{2}) |\Sigma|^{\frac{N-k}{2}}} \\ &= 2^{\frac{Nhp}{2}} |\Sigma|^{\frac{Nh}{2}} \frac{\Gamma_p(\frac{N-k}{2} + \frac{Nh}{2})}{\Gamma_p(\frac{N-k}{2})} |I + 2\Sigma T|^{-\frac{(Nh + N-k)}{2}} \end{aligned} \quad (viii)$$

for  $I + 2\Sigma T > O$ ,  $\Re(\frac{N-k}{2} + \frac{Nh}{2}) > \frac{p-1}{2}$ . The integral over  $U_1, \dots, U_{k-1}$  is the following, denoted by  $\delta$ :

$$\delta = \int_{U_1, \dots, U_{k-1}} \frac{e^{-\frac{1}{2} \sum_{i=1}^{k-1} U_i' \Sigma^{-1} U_i - \text{tr}(\sum_{i=1}^{k-1} T U_i U_i')}}{(2\pi)^{\frac{(k-1)p}{2}} |\Sigma|^{\frac{k-1}{2}}} dU_1 \wedge \dots \wedge dU_{k-1}$$

where

$$\text{tr}\left(\sum_{i=1}^{k-1} T U_i U_i'\right) = \text{tr}\left(\sum_{i=1}^{k-1} U_i' T U_i\right) = \sum_{i=1}^{k-1} U_i' T U_i$$

since  $U_i' T U_i$  is scalar; thus, the exponent becomes  $-\frac{1}{2} \sum_{i=1}^{k-1} U_i' [\Sigma^{-1} + 2T] U_i$  and the integral simplifies to

$$\delta = |I + 2\Sigma T|^{-\frac{k-1}{2}}, \quad I + 2\Sigma T > O. \quad (ix)$$

Now the integral over  $T$  is the following:

$$\begin{aligned} & \frac{1}{\Gamma_p(\frac{Nh}{2})} \int_{T>O} |T|^{\frac{Nh}{2} - \frac{p+1}{2}} |I + 2\Sigma T|^{-\left(\frac{Nh}{2} + \frac{N-k}{2} + \frac{k-1}{2}\right)} \\ &= \frac{\Gamma_p\left(\frac{N-k}{2} + \frac{k-1}{2}\right)}{\Gamma_p\left(\frac{Nh}{2} + \frac{N-k}{2} + \frac{k-1}{2}\right)} 2^{-\frac{Nhp}{2}} |\Sigma|^{-\frac{Nh}{2}}, \quad \Re\left(\frac{N-k}{2} + \frac{Nh}{2}\right) > \frac{p-1}{2}. \end{aligned} \quad (x)$$

Therefore,

$$E[\lambda_1^h | H_{o1}] = \frac{\Gamma_p\left(\frac{N-k}{2} + \frac{Nh}{2}\right)}{\Gamma_p\left(\frac{N-k}{2}\right)} \frac{\Gamma_p\left(\frac{N-k}{2} + \frac{k-1}{2}\right)}{\Gamma_p\left(\frac{Nh}{2} + \frac{N-k}{2} + \frac{k-1}{2}\right)} \quad (6.13.3)$$

for  $\Re\left(\frac{N-k}{2} + \frac{Nh}{2}\right) > \frac{p-1}{2}$ .

### 6.13.3. The asymptotic distribution of $-2 \ln \lambda_1$

An asymptotic distribution of  $-2 \ln \lambda_1$  as  $N \rightarrow \infty$  can be derived from (6.13.3). First, on expanding the real matrix-variate gamma functions in (6.13.3), we obtain the following representation of the  $h$ -th null moment of  $\lambda_1$ :

$$\begin{aligned} E[\lambda_1^h | H_{o1}] &= \left\{ \prod_{j=1}^p \frac{\Gamma\left(\frac{N-k}{2} + \frac{k-1}{2} - \frac{j-1}{2}\right)}{\Gamma\left(\frac{N-k}{2} - \frac{j-1}{2}\right)} \right\} \\ &\times \left\{ \prod_{j=1}^p \frac{\Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} - \frac{j-1}{2}\right)}{\Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} + \frac{k-1}{2} - \frac{j-1}{2}\right)} \right\}. \end{aligned} \quad (xi)$$

Let us now express all the gamma functions in terms of Sterling's asymptotic formula by taking  $\frac{N}{2} \rightarrow \infty$  in the constant part and  $\frac{N}{2}(1+h) \rightarrow \infty$  in the part containing  $h$ . Then,

$$\begin{aligned} \frac{\Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} - \frac{j-1}{2}\right)}{\Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} + \frac{k-1}{2} - \frac{j-1}{2}\right)} &\rightarrow \frac{(2\pi)^{\frac{1}{2}} \left[\frac{N}{2}(1+h)\right]^{\frac{N}{2}(1+h) - \frac{k}{2} - \frac{j}{2}} e^{\frac{N}{2}(1+h)}}{(2\pi)^{\frac{1}{2}} \left[\frac{N}{2}(1+h)\right]^{\frac{N}{2}(1+h) - \frac{k}{2} + \frac{k-1}{2} - \frac{j}{2}} e^{\frac{N}{2}(1+h)}} \\ &= \frac{1}{\left[\frac{N}{2}(1+h)\right]^{\frac{k-1}{2}}}, \end{aligned} \quad (xii)$$

$$\frac{\Gamma\left(\frac{N}{2} - \frac{k}{2} + \frac{k-1}{2} - \frac{j-1}{2}\right)}{\Gamma\left(\frac{N}{2} - \frac{k}{2} - \frac{j-1}{2}\right)} \rightarrow \left(\frac{N}{2}\right)^{\frac{(k-1)}{2}}. \quad (xiii)$$

Hence,

$$E[\lambda_1^h | H_{o1}] \rightarrow (1+h)^{-\frac{(k-1)p}{2}} \text{ as } N \rightarrow \infty. \quad (6.13.4)$$

Thus, the following result:

**Theorem 6.13.1.** For the test statistic  $\lambda_1$  given in (6.13.2),  $-2 \ln \lambda_1 \rightarrow \chi_{(k-1)p}^2$ , that is,  $-2 \ln \lambda_1$  tends to a real scalar chisquare variable having  $(k-1)p$  degrees of freedom as  $N \rightarrow \infty$  with  $N = n_1 + \dots + n_k$ ,  $n_j$  being the sample size of the  $j$ -th  $p$ -variate real Gaussian population.

Under the marginal hypothesis  $H_{o2} : \Sigma_1 = \dots = \Sigma_k = \Sigma$  where  $\Sigma$  is unknown, the  $\lambda$ -criterion is denoted by  $\lambda_2$ , and its  $h$ -th moment, which is available from (6.12.6) of Sect. 6.12, is given by

$$E[\lambda_2^h | H_{o2}] = c^h \frac{\Gamma_p(\frac{N-k}{2})}{\Gamma_p(\frac{N-k}{2} + \frac{Nh}{2})} \left\{ \prod_{j=1}^p \frac{\Gamma_p(\frac{n_j-1}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{n_j-1}{2})} \right\} \quad (6.13.5)$$

for  $\Re(\frac{n_j-1}{2} + \frac{n_j h}{2}) > \frac{p-1}{2}$ ,  $j = 1, \dots, k$ . Hence the  $h$ -th null moment of the  $\lambda$  criterion for testing the hypothesis  $H_o$  of equality of the  $k$  independent  $p$ -variate real Gaussian populations is the following:

$$\begin{aligned} E[\lambda^h | H_o] &= E[\lambda_1^h | H_{o1}] E[\lambda_2^h | H_{o2}] \\ &= \frac{\Gamma_p(\frac{N-k}{2} + \frac{k-1}{2})}{\Gamma_p(\frac{Nh}{2} + \frac{N-k}{2} + \frac{k-1}{2})} c^h \left\{ \prod_{j=1}^p \frac{\Gamma_p(\frac{n_j-1}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{n_j-1}{2})} \right\} \end{aligned} \quad (6.13.6)$$

for  $\Re(\frac{n_j-1}{2} + \frac{n_j h}{2}) > \frac{p-1}{2}$ ,  $j = 1, \dots, k$ ,  $N = n_1 + \dots + n_k$ , where  $c$  is the constant associated with the  $h$ -th moment of  $\lambda_2$ . Combining Theorems 6.13.1 and Theorem 6.12.1, the asymptotic distribution of  $-2 \ln \lambda$  of (6.13.6) is a real scalar chisquare with  $(k-1)p + (k-1)\frac{p(p+1)}{2}$  degrees of freedom. Thus, the following result:

**Theorem 6.13.2.** For the  $\lambda$ -criterion for testing the hypothesis of equality of  $k$  independent  $p$ -variate real Gaussian populations,  $-2 \ln \lambda \rightarrow \chi_v^2$ ,  $v = (k-1)p + (k-1)\frac{p(p+1)}{2}$  as  $n_j \rightarrow \infty$ ,  $j = 1, \dots, k$ .

**Note 6.13.1.** Observe that for the conditional hypothesis  $H_{o1}$  in (6.13.2), the degrees of freedom of the asymptotic chisquare distribution of  $-2 \ln \lambda_1$  is  $(k-1)p$ , which is also the number of parameters restricted by the hypothesis  $H_{o1}$ . For the hypothesis  $H_{o2}$ , the corresponding degrees of freedom of the asymptotic chisquare distribution of  $-2 \ln \lambda_2$  is

$(k-1)\frac{p(p+1)}{2}$ , which as well is the number of parameters restricted by the hypothesis  $H_{o2}$ . The asymptotic chisquare distribution of  $-2 \ln \lambda$  for the hypothesis  $H_o$  of equality of  $k$  independent  $p$ -variate Gaussian populations the degrees of freedom is the sum of these two quantities, that is,  $(k-1)p + (k-1)\frac{p(p+1)}{2} = p(k-1)\frac{(p+3)}{2}$ , which also coincides with the number of parameters restricted under the hypothesis  $H_o$ .

### Exercises

**6.1.** Derive the  $\lambda$ -criteria for the following tests in a real univariate Gaussian population  $N_1(\mu, \sigma^2)$ , assuming that a simple random sample of size  $n$ , namely  $x_1, \dots, x_n$ , which are iid as  $N_1(\mu, \sigma^2)$ , is available: (1):  $\mu = \mu_o$  (given),  $\sigma^2$  is known; (2):  $\mu = \mu_o$ ,  $\sigma^2$  unknown; (3):  $\sigma^2 = \sigma_o^2$  (given), also you may refer to Mathai and Haubold (2017).

In all the following problems, it is assumed that a simple random sample of size  $n$  is available. The alternative hypotheses are the natural alternatives.

**6.2.** Repeat Exercise 6.1 for the corresponding complex Gaussian.

**6.3.** Construct the  $\lambda$ -criteria in the complex case for the tests discussed in Sects. 6.2–6.4.

**6.4.** In the real  $p$ -variate Gaussian case, consider the hypotheses (1):  $\Sigma$  is diagonal or the individual components are independently distributed; (2): The diagonal elements are equal, given that  $\Sigma$  is diagonal (which is a conditional test). Construct the  $\lambda$ -criterion in each case.

**6.5.** Repeat Exercise 6.4 for the complex Gaussian case.

**6.6.** Let the population be real  $p$ -variate Gaussian  $N_p(\mu, \Sigma)$ ,  $\Sigma = (\sigma_{ij}) > O$ ,  $\mu' = (\mu_1, \dots, \mu_p)$ . Consider the following tests and compute the  $\lambda$ -criteria: (1):  $\sigma_{11} = \dots = \sigma_{pp} = \sigma^2$ ,  $\sigma_{ij} = \nu$  for all  $i$  and  $j$ ,  $i \neq j$ . That is, all the variances are equal and all the covariances are equal; (2): In addition to (1),  $\mu_1 = \mu_2 = \dots = \mu$  or all the mean values are equal. Construct the  $\lambda$ -criterion in each case. The first one is known as  $L_{vc}$  criterion and the second one is known  $L_{mvc}$  criterion. Repeat the same exercise for the complex case. Some distributional aspects are examined in Mathai (1970b) and numerical tables are available in Mathai and Katiyar (1979b).

**6.7.** Let the population be real  $p$ -variate Gaussian  $N_p(\mu, \Sigma)$ ,  $\Sigma > O$ . Consider the hypothesis (1):

$$\Sigma = \begin{bmatrix} a & b & b & \dots & b \\ b & a & b & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \dots & a \end{bmatrix}, \quad a \neq 0, \quad b \neq 0, \quad a \neq b.$$

(2):

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Sigma_{11} = \begin{bmatrix} a_1 & b_1 & \cdots & b_1 \\ b_1 & a_1 & \cdots & b_1 \\ \vdots & \vdots & \ddots & \vdots \\ b_1 & b_1 & \cdots & a_1 \end{bmatrix}, \quad \Sigma_{22} = \begin{bmatrix} a_2 & b_2 & \cdots & b_2 \\ b_2 & a_2 & \cdots & b_2 \\ \vdots & \vdots & \ddots & \vdots \\ b_2 & b_2 & \cdots & a_2 \end{bmatrix}$$

where  $a_1 \neq 0$ ,  $a_2 \neq 0$ ,  $b_1 \neq 0$ ,  $b_2 \neq 0$ ,  $a_1 \neq b_1$ ,  $a_2 \neq b_2$ ,  $\Sigma_{12} = \Sigma'_{21}$  and all the elements in  $\Sigma_{12}$  and  $\Sigma_{21}$  are each equal to  $c \neq 0$ . Construct the  $\lambda$ -criterion in each case. (These are hypotheses on patterned matrices).

**6.8.** Repeat Exercise 6.7 for the complex case.

**6.9.** Consider  $k$  independent real  $p$ -variate Gaussian populations with different parameters, distributed as  $N_p(M_j, \Sigma_j)$ ,  $\Sigma_j > O$ ,  $M'_j = (\mu_{1j}, \dots, \mu_{pj})$ ,  $j = 1, \dots, k$ . Construct the  $\lambda$ -criterion for testing the hypothesis  $\Sigma_1 = \dots = \Sigma_k$  or the covariance matrices are equal. Assume that simple random samples of sizes  $n_1, \dots, n_k$  are available from these  $k$  populations.

**6.10.** Repeat Exercise 6.9 for the complex case.

**6.11.** For the second part of Exercise 6.7, which is also known as Wilks'  $L_{mvc}$  criterion, show that if  $u = \lambda^{\frac{2}{n}}$  where  $\lambda$  is the likelihood ratio criterion and  $n$  is the sample size, then

$$u = \frac{|S|}{[s + (p-1)s_1][s - s_1 + \frac{n}{p-1} \sum_{j=1}^p (\bar{x}_j - \bar{x})^2]} \quad (i)$$

where  $S = (s_{ij})$  is the sample sum of products matrix,  $s = \frac{1}{p} \sum_{i=1}^p s_{ii}$ ,  $s_1 = \frac{1}{p(p-1)} \sum_{i \neq j=1}^p s_{ij}$ ,  $\bar{x} = \frac{1}{p} \sum_{i=1}^p \bar{x}_i$ ,  $\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ik}$ . For the statistic  $u$  in (i), show that the  $h$ -th null moment or the  $h$ -th moment when the null hypothesis is true, is given by the following:

$$E[u^h | H_0] = \prod_{j=0}^{p-2} \frac{\Gamma(\frac{n-1}{2} + h - \frac{j}{2})}{\Gamma(\frac{n-1}{2} - \frac{j}{2})} \frac{\Gamma(\frac{n+1}{2} + \frac{j}{p-1})}{\Gamma(\frac{n+1}{2} + h + \frac{j}{p-1})}. \quad (ii)$$

Write down the conditions for the existence of the moment in (ii). [For the null and non-null distributions of Wilks'  $L_{mvc}$  criterion, see Mathai (1978).]

**6.12.** Let the  $(p + q) \times 1$  vector  $X$  have a  $(p + q)$ -variate nonsingular real Gaussian distribution,  $X \sim N_{p+q}(\mu, \Sigma)$ ,  $\Sigma > O$ . Let

$$\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2' & \Sigma_3 \end{bmatrix}$$

where  $\Sigma_1$  is  $p \times p$  with all its diagonal elements equal to  $\sigma_{aa}$  and all other elements equal to  $\sigma_{aa'}$ ,  $\Sigma_2$  has all elements equal to  $\sigma_{ab}$ ,  $\Sigma_3$  has all diagonal elements equal to  $\sigma_{bb}$  and all other elements equal to  $\sigma_{bb'}$  where  $\sigma_{aa}$ ,  $\sigma_{aa'}$ ,  $\sigma_{bb}$ ,  $\sigma_{bb'}$  are unknown. Then,  $\Sigma$  is known as bipolar. Let  $\lambda$  be the likelihood ratio criterion for testing the hypothesis that  $\Sigma$  is bipolar. Then show that the  $h$ -th null moment is the following:

$$E[\lambda^h | H_0] = [(p-1)^{p-1} (q-1)^{q-1}] \frac{\Gamma[\frac{(q-1)(n-1)}{2}] \Gamma[\frac{(p-1)(n-1)}{2}]}{\Gamma[(p-1)(h + \frac{n-1}{2})] \Gamma[(q-1)(h + \frac{n-1}{2})]} \\ \times \prod_{j=0}^{p+q-3} \frac{\Gamma[h + \frac{n-3}{2} - \frac{j}{2}]}{\Gamma[\frac{n-3}{2} - \frac{j}{2}]}$$

where  $n$  is the sample size. Write down the conditions for the existence of this  $h$ -th null moment.

**6.13.** Let  $X$  be  $m \times n$  real matrix having the matrix-variate Gaussian density

$$f(X) = \frac{1}{|2\pi \Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2} \text{tr}[\Sigma^{-1}(X-M)(X-M)']}, \quad \Sigma > O.$$

Letting  $S = XX'$ ,  $S$  is a non-central Wishart matrix. Derive the density of  $S$  and show that this density, denoted by  $f_s(S)$ , is the following:

$$f_s(S) = \frac{1}{\Gamma_m(\frac{n}{2}) |2\Sigma|^{\frac{n}{2}}} |S|^{\frac{n}{2} - \frac{m+1}{2}} e^{-\text{tr}(\Omega) - \frac{1}{2} \text{tr}(\Sigma^{-1}S)} \\ \times {}_0F_1\left(\ ; \frac{n}{2}; \frac{1}{2} \Sigma^{-1}S\right)$$

where  $\Omega = \frac{1}{2} MM' \Sigma^{-1}$  is the non-centrality parameter and  ${}_0F_1$  is a Bessel function of matrix argument.

**6.14.** Show that the  $h$ -th moment of the determinant of  $S$ , the non-central Wishart matrix specified in Exercise 6.13, is given by

$$E[|S|^h] = \frac{\Gamma_m(h + \frac{n}{2})}{\Gamma(\frac{n}{2}) |2\Sigma|^h} e^{-\text{tr}(\Omega)} {}_1F_1\left(h + \frac{n}{2}; \frac{n}{2}; \Omega\right)$$

where  ${}_1F_1$  is a hypergeometric function of matrix argument and  $\Omega$  is the non-centrality parameter defined in Exercise 6.13.

**6.15.** Letting  $v = \lambda^{\frac{2}{n}}$  in Eq. (6.3.10), show that, under the null hypothesis  $H_0$ ,  $v$  is distributed as a real scalar type-1 beta with the parameters  $(\frac{n-1}{2}, \frac{1}{2})$  and that  $\frac{nA'\bar{X}\bar{X}'A}{A'SA}$  is real scalar type-2 beta distributed with the parameters  $(\frac{1}{2}, \frac{n-1}{2})$ .

**6.16.** Show that for an arbitrary  $h$ , the  $h$ -th null moment of the test statistic  $\lambda$  specified in Eq. (6.3.10) is

$$E[\lambda^h | H_0] = \frac{\Gamma(\frac{n-1}{2} + \frac{nh}{2})}{\Gamma(\frac{n-1}{2})} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2} + \frac{nh}{2})}, \Re(h) > -\frac{n-1}{n}.$$

### References

T.W. Anderson (2003): *An Introduction to Multivariate Statistical Analysis*, Third Edition, Wiley, New York.

A.W. Davis (1971): Percentile approximations for a class of likelihood ratio criteria, *Biometrika*, **58**, 349–356.

A.W. Davis and J.B. Field (1971): Tables of some multivariate test criteria, *Technical Report No. 32*, Division of Mathematical Statistics, CSIRO, Canberra, Australia.

B.P. Korin (1968): On the distribution of a statistic used for testing a covariance matrix, *Biometrika*, **55**, 171–178.

A.M. Mathai (1970a): An expansion of Meijer's G-function in the logarithmic case with applications, *Mathematische Nachrichten*, **48**, 129–139.

A.M. Mathai (1970b): The exact distribution of Bartlett's criterion, *Publ. L'ISUP, Paris*, **19**, 1–15.

A.M. Mathai (1971): On the distribution of the likelihood ratio criterion for testing linear hypotheses on regression coefficients, *Annals of the Institute of Statistical Mathematics*, **23**, 181–197.

A.M. Mathai (1972a): The exact distribution of three tests associated with Wilks' concept of generalized variance, *Sankhya Series A*, **34**, 161–170.

A.M. Mathai (1972b): The exact non-central distribution of the generalized variance, *Annals of the Institute of Statistical Mathematics*, **24**, 53–65.

A.M. Mathai (1973): A review of the various methods of obtaining the exact distributions of multivariate test criteria, *Sankhya Series A*, **35**, 39–60.

A.M. Mathai (1977): A short note on the non-null distribution for the problem of sphericity, *Sankhya Series B*, **39**, 102.

A.M. Mathai (1978): On the non-null distribution of Wilks'  $L_{mvc}$ , *Sankhya Series B*, **40**, 43–48.

A.M. Mathai (1979a): The exact distribution and the exact percentage points for testing equality of variances in independent normal population, *Journal of Statistical Computations and Simulation*, **9**, 169–182.

A.M. Mathai (1979): On the non-null distributions of test statistics connected with exponential populations, *Communications in Statistics (Theory & Methods)*, **8(1)**, 47–55.

A.M. Mathai (1982): On a conjecture in geometric probability regarding asymptotic normality of a random simplex, *Annals of Probability*, **10**, 247–251.

A.M. Mathai (1984): On multi-sample sphericity (in Russian), *Investigations in Statistics (Russian)*, Tom **136**, 153–161.

A.M. Mathai (1985): Exact and asymptotic procedures in the distribution of multivariate test statistic, In *Statistical Theory and Data Analysis*, K. Matusita Editor, North-Holland, pp. 381–418.

A.M. Mathai (1986): Hypothesis of multisample sphericity, *Journal of Mathematical Sciences*, **33(1)**, 792–796.

A.M. Mathai (1993): *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*, Oxford University Press, Oxford.

A.M. Mathai and H.J. Haubold (2017): *Probability and Statistics: A Course for Physicists and Engineers*, De Gruyter, Germany.

A.M. Mathai and R.S. Katiyar (1979a): Exact percentage points for testing independence, *Biometrika*, **66**, 353–356.

A.M. Mathai and R. S. Katiyar (1979b): The distribution and the exact percentage points for Wilks'  $L_{mvc}$  criterion, *Annals of the Institute of Statistical Mathematics*, **31**, 215–224.

A.M. Mathai and R.S. Katiyar (1980): Exact percentage points for three tests associated with exponential populations, *Sankhya Series B*, **42**, 333–341.

A.M. Mathai and P.N. Rathie (1970): The exact distribution of Votaw's criterion, *Annals of the Institute of Statistical Mathematics*, **22**, 89–116.



A.M. Mathai and P.N. Rathie (1971): Exact distribution of Wilks' criterion, *The Annals of Mathematical Statistics*, **42**, 1010–1019.

A.M. Mathai and R.K. Saxena (1973): *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences*, Springer-Verlag, Lecture Notes No. 348, Heidelberg and New York.

A.M. Mathai and R.K. Saxena (1978): *The H-function with Applications in Statistics and Other Disciplines*, Wiley Halsted, New York and Wiley Eastern, New Delhi.

A.M. Mathai, R.K. Saxena and H.J. Haubold (2010): *The H-function: Theory and Applications*, Springer, New York.

B.N. Nagarsenker and K.C.S. Pillai (1973): The distribution of the sphericity test criterion, *Journal of Multivariate Analysis*, **3**, 226–2235.

N. Sugiura and H. Nagao (1968): Unbiasedness of some test criteria for the equality of one or two covariance matrices, *Annals of Mathematical Statistics*, **39**, 1686–1692.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

