TOPICS IN MULTIVARIATE STATISTICS
and
STATISTICAL METHODS IN MULTIVARIATE ANALYSIS

Carlos A. Coelho

Contents

Chapter 1

Introduction

In real life, and as such also in most research problems, we often deal with a set of variables which try to assess a given phenomenon which we are trying to understand or explain. Therefore, in general, we will have to resort to the use of statistical methods that study the relations among groups of variables, rather than among individual variables and that are able to model groups of variables and nit just individual variables.

We have thus the need to learn and to use multivariate distributions, techniques and statistical models used in Multivariate Analysis.

First we will address in a very brief, mostly intuitive and heuristic manner some of the existing multivariate models and their interrelations.

Let us suppose that our data of n observations taken on p variables is organized in the $n{\times}p$ X matrix

$$X \; = \; \begin{array}{cc} \begin{array}{cccccc} p_1 & p_2 & & p_k & & p_m \end{array} & \\ \left[\begin{array}{c|c|c|c|c|c} X_1 & X_2 & \ldots & X_k & \ldots & X_m \end{array}\right] & \begin{array}{c} 1 \\ \vdots \\ n \end{array} \end{array}$$

with

$$p \; = \; \sum_{k=1}^{m} p_k \; .$$

Then, some of the multivariate (and univariate) methods and models used to study the relations among the m groups of variables are shown in Fig. 1.1., as well as their interrelations.

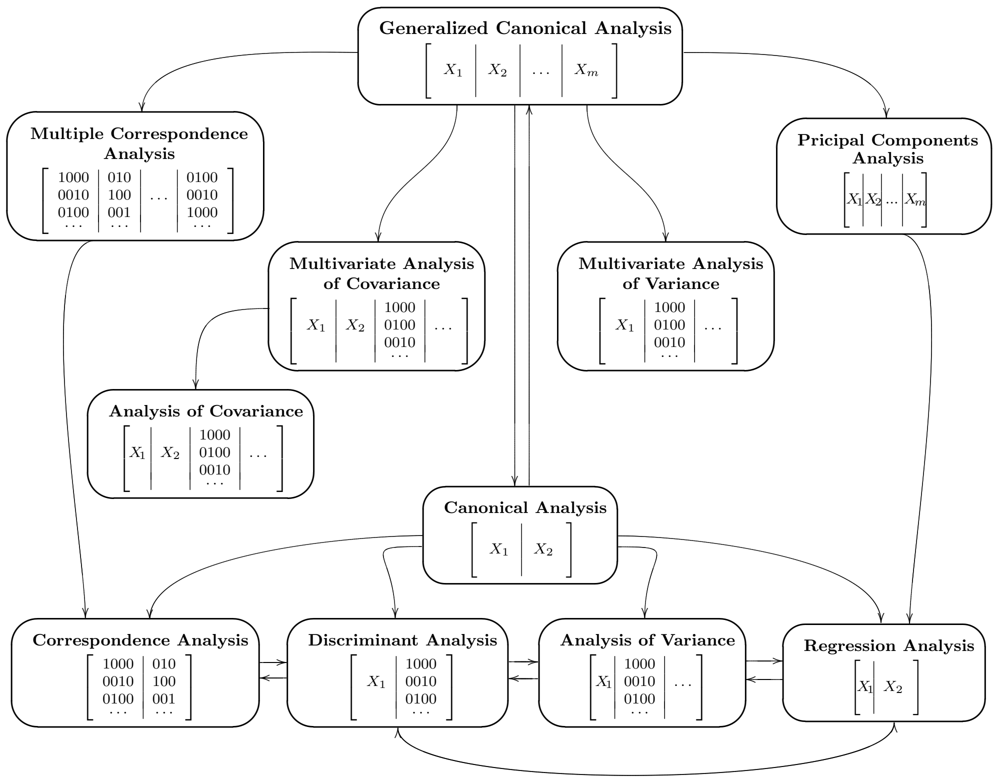

Figure 1.1 – Some multivariate (and univariate) statistical linear models and their interrelations.

Chapter 2

The Multivariate Normal Distribution

2.1 The p-multivariate Normal distribution

Definition 2.1: Let $\underline{X}$ ($p\times 1$) be the vector

$$\underline{X} \;=\; [\; X_1 \; X_2 \; X_3 \; ... \; X_p \;]' \tag{2.1}$$

where $X_1,\; X_2,\; X_3,\; ...,\; X_p$ are p random variables.

We say that the vector $\underline{X}$ has a p-multivariate Normal distribution, if

$$f_{\underline{X}}(\underline{x}) \;=\; \frac{1}{(2\pi)^{p/2}\;|\Sigma|^{1/2}}\; e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})} \tag{2.2}$$

where Σ is a positive-definite (symmetric) matrix.

The distribution whose p.d.f. (probability density function) is shown in (2.2) is thus the joint distribution of the random variables X_1, X_2, X_3, ..., X_p.

The mean or expected value of $\underline{X}$ is

$$E(\underline{X}) \;=\; \underline{\mu}\,,$$

where

$$\underline{\mu} \;=\; [\; \mu_1 \; \mu_2 \; \mu_3 \; ... \; \mu_p \;]'$$

with

$$E(X_j) \;=\; \mu_j \qquad (j = 1, ..., p)\ .$$

And the variance of $\underline{X}$ is

$$Var(\underline{X}) \;=\; \Sigma$$

where

$$\Sigma \;=\; \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \dots & \sigma_{2p} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \dots & \sigma_{3p} \\ \vdots & \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \sigma_{p3} & \dots & \sigma_{pp} \end{bmatrix}$$

with

$$Var(X_j) \;=\; \sigma_{jj} \qquad (j = 1, ..., p)$$

and

$$Cov(X_j,\; X_k) \;=\; \sigma_{jk} \;=\sigma_{kj} \;=\; Cov(X_k, X_j) \qquad (j, k = 1, ..., p,\;\; j \neq k)\,.$$

If $\underline{X}$ has a distribution with p.d.f. given by (2.2) we will denote this fact by

$$\underline{X} \;\sim\; N_p\left(\underline{\mu},\; \Sigma\right). \quad \square \tag{2.3}$$

2.2 Some properties of the multivariate Normal distribution

Let us suppose that

$$\underline{X} \;\sim\; N_p\left(\underline{\mu},\; \Sigma\right)\;\; .$$

Then

I) for any split of the vector $\underline{X}$, defined in (2.1), of the form

$$\underline{X} \;=\; \left[\begin{array}{c} \underline{X}_1 \\ \hline \underline{X}_2 \end{array}\right] \begin{array}{l} p_1 \\ p_2 \end{array}$$

where

$$\underline{X}_1 \;=\; [\; X_1\; X_2\; ...\; X_{p_1}\;]' \qquad (p_1 < p)$$

and

$$\underline{X}_2 \;=\; [\; X_{p_1+1} \; X_{p_1+2} \; ... \; X_p \;]'$$

we have

$$\underline{X}_1 \;\sim\; N_{p_1}\left(\underline{\mu}_1,\; \Sigma_{11}\right)$$

and

$$\underline{X}_2 \;\sim\; N_{p_2}\left(\underline{\mu}_2,\; \Sigma_{22}\right)$$

where

$$p_2 \;=\; p - p_1 \;,$$

$$\underline{\mu}_1 \;=\; [\; \mu_1 \; \mu_2 \; ... \; \mu_{p_1} \;]' \;, \qquad \underline{\mu}_2 \;=\; [\; \mu_{p_1+1} \; \mu_{p_1+2} \; ... \; \mu_p \;]'$$

$$\Sigma_{11} \;=\; \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p_1} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p_1} \\ \vdots & \vdots & & \vdots \\ \sigma_{p_1 1} & \sigma_{p_1 2} & \cdots & \sigma_{p_1 p_1} \end{bmatrix}$$

$$\Sigma_{22} \;=\; \begin{bmatrix} \sigma_{p_1+1,p_1+1} & \sigma_{p_1+1,p_1+2} & \cdots & \sigma_{p_1+1,p} \\ \sigma_{p_1+2,p_1+1} & \sigma_{p_1+2,p_1+2} & \cdots & \sigma_{p_1+2,p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p,p_1+1} & \sigma_{p,p_1+2} & \cdots & \sigma_{pp} \end{bmatrix}$$

with

$$\Sigma \;=\; \begin{array}{c c} \begin{array}{cc} p_1 & p_2 \end{array} & \\ \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} & \begin{array}{c} p_1 \\ p_2 \end{array} \end{array} \qquad (\Sigma_{21} = \Sigma_{12}')$$

e

$$\underline{\mu} \;=\; \begin{array}{cc} \left[\begin{array}{c} \underline{\mu}_1' \\ \hline \underline{\mu}_2' \end{array}\right] & \begin{array}{c} p_1 \\ p_2 \end{array} \end{array}$$

II) the vectors (or sets) of random variables $\underline{X}_1$ e $\underline{X}_2$ are said to be independent if and only if

$$\Sigma_{12} \;=\; \underset{p_1 \times p_2}{0} \;;$$

III)

$$X_j \;\sim\; N\left(\mu_j,\; \sigma_{jj}\right) \qquad (j = 1, ..., p) \;;$$

IV)

$$
\begin{aligned}
(\underline{X}_1|\underline{X}_2) &\sim N_{p_1}\left(\underline{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underline{X}_2 - \underline{\mu}_2),\ \Sigma_{11.2}\right)\\
(\underline{X}_2|\underline{X}_1) &\sim N_{p_2}\left(\underline{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\underline{X}_1 - \underline{\mu}_1),\ \Sigma_{22.1}\right)
\end{aligned}
$$

where

$$
\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \qquad \text{e} \qquad \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\ ;
$$

V) let

$$
\underline{Y} = A\underline{X} + \underline{b} \tag{2.4}
$$

where A $(r \times p)$ and $\underline{b}$ $(r \times 1)$ $(r \leq p)$ are respectively a fixed (non-random) matrix of *rank* r and a fixed (non-random) vector, i.e., formed by real non-random values; then

$$
\underline{Y} \sim N_r\left(A\underline{\mu} + \underline{b},\ A\Sigma A'\right)\ ; \tag{2.5}
$$

VI) if in (2.1) we have $p = 2$ then

$$
\underline{X} \sim N_2\left(\underline{\mu},\ \Sigma\right) \tag{2.6}
$$

with

$$
\underline{\mu} = [\mu_1\ \mu_2]'
$$

and

$$
\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}\ ;
$$

and in this case, we call bivariate Normal distribution to the distribution to the distribution in (2.6) and its p.d.f. is also commonly written in the form

$$
f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\left(\sigma_{11}\sigma_{22}(1-\rho^2)\right)^{1/2}} \times e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_{11}} + \frac{(x_2-\mu_2)^2}{\sigma_{22}} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}\right]} \tag{2.7}
$$

where

$$
\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}
$$

is the correlation between X_1 and X_2. We may note that (2.7) may be obtained directly from (2.2), for $p = 2$;

VII) the moment generating function of $\underline{X}$ is

$$M_{\underline{X}}(t) \;=\; E\left[e^{\underline{t}'\underline{X}}\right] \;=\; e^{\underline{t}'\underline{\mu}+\frac{1}{2}\underline{t}'\Sigma\underline{t}}\;;$$

and the characteristic function of $\underline{X}$ is

$$\Phi_{\underline{X}}(t) \;=\; E\left[e^{i\underline{t}'\underline{X}}\right] \;=\; e^{i\underline{t}'\underline{\mu}-\frac{1}{2}\underline{t}'\Sigma\underline{t}}\;;$$

VIII) from V) above, i.e., from (2.4) and (2.5), we may conclude that

$$(\underline{t}'\underline{X}) \;\sim\; N(\underline{t}'\underline{\mu},\; \underline{t}'\Sigma\underline{t})$$

for any vector $\underline{t} \in I\!R^p$;

IX)

$$(\underline{X}-\underline{\mu})'\Sigma^{-1}(\underline{X}-\underline{\mu}) \;\sim\; \chi^2_p$$

and

$$\underline{X}'\Sigma^{-1}\underline{X} \;\sim\; \chi^2_p(\delta)$$

with

$$\delta \;=\; \underline{\mu}'\Sigma^{-1}\underline{\mu}\;.$$

(For a proof see Theorem 1.4.1 in Muirhead (2005).)

The following is not exactly a property of the multivariate Normal distribution, but rather a necessary and sufficient condition in order that (2.3) holds, i.e., in order that $\underline{X}$ has a p-multivariate Normal distribution.

X) If for any vector $\underline{b} \in I\!R^p$, $\underline{b}'\underline{X}$ has a (univariate) Normal distribution, then $\underline{X}$ has a p-multivariate Normal distribution.

❑❑❑❑❑

Exercise: Prove property IV) above.

❑❑❑❑❑

Bibliography

Johnson, R. A. e Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6.ª ed.. Pearson Education Inc., Prentice Hall, New Jersey. (Cap.4, Sec. 4.2-4.4)

Kshirsagar, A. M. (1972). *Multivariate Analysis.* Marcel Dekker. New York.
(Cap. 2 pp. 23-27, 34-37 e 39-45.)

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*, 2nd ed., J. Wiley & Sons, New York. (Cap. 1 pp. 1-20 e 26-32.)

Chapter 3

Random samples from a multivariate Normal distribution

3.1 Presentation of a random sample from a multivariate Normal distribution

Let us consider a random sample of size n of a p-multivariate Normal distribution with expected value $\underline{\mu}$ and variance-covariance matrix Σ.

Such random sample will have $n{\times}p$ elements (n elements of each one of the p variables) and may be organized in the matrix X with n rows and p columns, each row corresponding to one of the n observations in the random sample and each column corresponding to each one of the p variables.

We have thus the matrix

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ X_{31} & X_{32} & \dots & X_{3p} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

where X_{ij} represents the i-th observation on the variable X_j ($i = 1, ..., n;\ j = 1, ..., p$).

We should note that the n rows of X are independent, given that each one of them corresponds to one of the elements in a random sample. Each one of those rows has expected value $\underline{\mu}$ and variance Σ.

The common data matrix, associated to a given random sample, of a p-multivariate Normal distribution, is a realization of a random matrix X as the one depicted above, and it will be denoted by $\mathcal{X}$.

3.2 The distribution of the matrix X

Let us consider the following notation, where E_{np} represents a matrix of dimensions $n{\times}p$ filled with 1's, and

$$\begin{aligned} vec(X) \; = \; & [X_{11}\ X_{21}\ ...\ X_{n1}\ X_{12}\ X_{22}\ ...\ X_{n2}\ X_{13}\ X_{23}\ ...\ X_{n3}\ \ ... \\ & \qquad\qquad\qquad\qquad X_{1p}\ X_{2p}\ ...\ X_{np}]' \end{aligned}$$

represents the vectorization of the matrix X, that is, the column vector resulting from piling up on top of each other the successive columns of X.

Then, from what was stated in the previous section, we have

$$E(X) \; = \; E_{n1}\underline{\mu}'$$

and

$$Var(X) \; = \; Var\left(vec(X)\right) \; = \; \Sigma \otimes I_n \ ,$$

where $\otimes$ denotes the Kronecker product.

The Kronecker product is defined in such a way that if

$$A \; = \; [a_{ij}] \quad , \quad i=1,...,p;\ j=1,...,q$$

and

$$B \; = \; [b_{kl}] \quad , \quad k=1,...,r;\ l=1,...,s$$

then

$$\underset{(pr\times qs)}{A \otimes B} \quad = \quad \begin{bmatrix} a_{11}B & a_{12}B & ... & a_{1q}B \\ a_{21}B & a_{22}B & ... & a_{2q}B \\ \vdots & \vdots & & \vdots \\ a_{p1}B & a_{p2}B & ... & a_{pq}B \end{bmatrix} ,$$

so that

$$(A \otimes B)' \; = \; A' \otimes B' \ ,$$

and, if $p = q$ and $r = s$, then

$$|A \otimes B| \;=\; |A|^r |B|^p \;,$$

and if A^{-1} and B^{-1} exist, then

$$(A \otimes B)^{-1} \;=\; A^{-1} \otimes B^{-1} \;.$$

Therefore,

$$X \;\sim\; N_{np}\left(E_{n1}\underline{\mu}',\; \Sigma \otimes I_n\right) \;, \tag{3.1}$$

being the p.d.f. of X given by

$$f_X(\mathcal{X}) \;=\; \frac{1}{(2\pi)^{np/2}\;|\Sigma|^{n/2}}\; e^{-\frac{1}{2}\,tr\left[\Sigma^{-1}(\mathcal{X}-E_{n1}\underline{\mu}')'(\mathcal{X}-E_{n1}\underline{\mu}')\right]} \;. \tag{3.2}$$

Result 3.1: A result with interest, related with the distribution presented in (3.1) is that if (3.1) happens, then if we take

$$\underset{n\times p}{Y} \;=\; \underset{n\times n}{A}\;\underset{n\times p}{X}\;\underset{p\times p}{B}$$

where A and B are real non-random matrices, we have

$$Y \;\sim\; N_{np}\left(AE_{n1}\underline{\mu}'B,\; B'\Sigma B \otimes AA'\right) \;,$$

that is,

$$f_Y(\mathcal{Y}) \;=\; \frac{1}{(2\pi)^{np/2}\;|A'A|^{p/2}\;|B\Sigma B'|^{n/2}} \; e^{tr\left[-\frac{1}{2}(B'\Sigma B)^{-1}(\mathcal{Y}-AE_{n1}\underline{\mu}'B)'(AA')^{-1}(\mathcal{Y}-AE_{n1}\underline{\mu}'B)\right]} \;,$$

a result that comes out of the fact that

$$vec(AXB) \;=\; (B' \otimes A)vec(X) \;.$$

If we use the result that says that if C is $q{\times}r$, D is $s{\times}t$, G is $r{\times}u$ and H is $t{\times}v$, then

$$(C \otimes D)(G \otimes H) \;=\; CG \otimes DH \;;$$

we have in fact

$$\begin{aligned} Var(Y) &= Var(vec(AXB)) \;=\; Var\left((B'\otimes A)vec(X)\right) \\ &= (B'\otimes A)\; Var(vec(X))\; (B'\otimes A)' \\ &= (B'\otimes A)\; (\Sigma\otimes I_n)\; (B\otimes A') \\ &= (B'\Sigma\otimes A)\; (B\otimes A') \\ &= B'\Sigma B\otimes AA'\;. \end{aligned}$$

❏❏❏❏❏

Exercises:

1. Let us suppose that the matrix X $(n{\times}p)$ has the distribution in (3.1). Then, working directly from the p.d.f. of X, obtain the p.d.f. of $Y = AXB$, where A and B are non-random matrices. What is the meaning and what represent the matrices A and B? Give some examples of matrices A and B.

2. Let X $(n{\times}p)$ be a random matrix, with $Var(X) = \Sigma\otimes A$. Obtain $Var(PXQ)$, where P and Q are real non-random matrices.

❏❏❏❏❏

3.3 How to obtain the p.d.f. in (3.2)

As we saw, the distribution with p.d.f. in (3.2) corresponds to the distribution of a random sample of size n from a p-multivariate Normal distribution.

Let us denote by $\underline{X}_i$ $(i = 1, ..., n)$ the i-th observation of that random sample. Note that

$$\underline{X}_i \;=\; [X_{i1}\; X_{i2}\; X_{i3}\; ...\; X_{ip}]'$$

is the column vector that corresponds to the i-th row of the matrix X. And note then also that, as already mentioned in section 3.2, relating to the rows of the matrix X, that $\underline{X}_i$ and $\underline{X}_{i'}$ $(i \neq i')$ are independent, given that they correspond to two different observations in a random sample. Hence, the joint distribution of the n random vectors $\underline{X}_i$ $(i = 1, ..., n)$ is given by

$$f_{\underline{X}_1,\underline{X}_2,...,\underline{X}_n}(\underline{x}_1,\underline{x}_2,...,\underline{x}_n) \;=\; \prod_{i=1}^{n} f_{\underline{X}_i}(\underline{x}_i)$$

with $\underline{X}_i \sim N_p(\underline{\mu},\ \Sigma)$, that is, with $f_{\underline{X}_i}(\underline{x}_i)$ given by (2.2), that is, with

$$f_{\underline{X}_i}(\underline{x}_i) \;=\; \frac{1}{(2\pi)^{p/2}\ |\Sigma|^{1/2}}\ e^{-\frac{1}{2}(\underline{x}_i-\underline{\mu})'\Sigma^{-1}(\underline{x}_i-\underline{\mu})}\ ,$$

so that

$$\begin{aligned} f_{\underline{X}_1,\underline{X}_2,\ldots,\underline{X}_n}(\underline{x}_1,\underline{x}_2,...,\underline{x}_n) \;&=\; \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2}\ |\Sigma|^{1/2}}\ e^{-\frac{1}{2}(\underline{x}_i-\underline{\mu})'\Sigma^{-1}(\underline{x}_i-\underline{\mu})} \\ &=\; \frac{1}{(2\pi)^{np/2}\ |\Sigma|^{n/2}}\ e^{-\frac{1}{2}\sum_{i=1}^{n}(\underline{x}_i-\underline{\mu})'\Sigma^{-1}(\underline{x}_i-\underline{\mu})}\ . \end{aligned} \tag{3.3}$$

Note that the joint distribution of $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$, with the above p.d.f., is the distribution of the matrix X and that

$$\sum_{i=1}^{n}(\underline{x}_i-\underline{\mu})'\Sigma^{-1}(\underline{x}_i-\underline{\mu}) \;=\; tr\left[(\mathcal{X}-E_{n1}\underline{\mu}')\Sigma^{-1}(\mathcal{X}-E_{n1}\underline{\mu}')'\right]\ ,$$

so that (3.3) is the same as (3.2).

Bibliography

Kshirsagar, A. M. (1972). *Multivariate Analysis.* Marcel Dekker. New York.
(Cap. 2 pp. 28-34.)

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* J. Wiley & Sons. New York. (Cap. 3 pp. 79-83.)

Chapter 4

Maximum Likelihood Estimators

4.1 The Likelihood Function

The distribution and density functions, seen as functions of the distribution parameters, may be seen as functions that measure the distance or the similitude of the model they specify to the data.

In this way the p.d.f. in (3.2) or (3.3), seen as a function of the parameters (of the distribution) $\underline{\mu}$ and Σ, may be seen as measuring the similitude or dissimilitude of the model — a random sample of dimension n of a p-multivariate Normal distribution with parameters $\underline{\mu}$ and Σ — to the data, represented by the matrix X. Such similitude or dissimilitude is measured through the differences $|\underline{x}_i - \underline{\mu}|$, which are scaled, through the product $(\underline{x}_i - \underline{\mu})\Sigma^{-1}(\underline{x}_i - \underline{\mu})$. Indeed the term

$$(\underline{x}_i - \underline{\mu})'\Sigma^{-1}(\underline{x}_i - \underline{\mu})$$

represents the so-called Mahalanobis distance (between $\underline{x}_i$ and $\underline{\mu}$).

4.2 MLEs – Maximum Likelihood Estimators

In the present case, the likelihood function may be derived (partially) relatively to each one of the parameters of the distribution. Then, equating such derivatives to zero and solving in order to the parameters we will obtain the Maximum Likelihood Estimators (MLEs) of those parameters.

It is thus time to question ourselves which are the *Maximum Likelihood Estimators* (MLEs) of $\underline{\mu}$ e Σ.

We will denote by L the likelihood function. Then, in our case

$$\begin{aligned} L(\underline{\mu},\Sigma) &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2}\ |\Sigma|^{1/2}}\ e^{-\frac{1}{2}(\underline{x}_i-\underline{\mu})'\Sigma^{-1}(\underline{x}_i-\underline{\mu})} \\ &= \frac{1}{(2\pi)^{np/2}\ |\Sigma|^{n/2}}\ e^{-\frac{1}{2}\,tr\left[(\mathcal{X}-E_{n1}\underline{\mu}')\Sigma^{-1}(\mathcal{X}-E_{n1}\underline{\mu}')'\right]}\ . \end{aligned} \tag{4.1}$$

However, due to a greater easiness in obtaining the derivatives, we often rather use the log-likelihood function, instead of the likelihood function. In our case we will denote by $\mathcal{L}$ the log-likelihood function

$$\begin{aligned} \mathcal{L}(\underline{\mu},\Sigma) &= log\ L(\underline{\mu},\Sigma) \\ &= -\frac{np}{2}log(2\pi) - \frac{n}{2}log|\Sigma| - \frac{1}{2}tr\left[(\mathcal{X}-E_{n1}\underline{\mu}')\Sigma^{-1}(\mathcal{X}-E_{n1}\underline{\mu}')'\right]\ . \end{aligned} \tag{4.2}$$

Let $\overline{\underline{x}}$ be the vector of the means of the n observations $\underline{x}_i$, that is, let

$$\overline{\underline{x}} = [\overline{x}_1\ \overline{x}_2\ \overline{x}_3\ ...\ \overline{x}_p]'$$

where

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \qquad j=1,...,p\ .$$

We are going to add and subtract to each term $(\mathcal{X}-E_{n1}\underline{\mu}')$ of (4.2) the quantity $E_{n1}\overline{\underline{x}}'$. This way we obtain

$$\begin{aligned} & tr\left[(\mathcal{X}-E_{n1}\underline{\mu}')\Sigma^{-1}(\mathcal{X}-E_{n1}\underline{\mu}')'\right] \\ &\quad = tr\left[(\mathcal{X}-E_{n1}\overline{\underline{x}}'+E_{n1}\overline{\underline{x}}'-E_{n1}\underline{\mu}')\Sigma^{-1}(\mathcal{X}-E_{n1}\overline{\underline{x}}'+E_{n1}\overline{\underline{x}}'-E_{n1}\underline{\mu}')'\right] \\ &\quad = tr\left[(\mathcal{X}-E_{n1}\overline{\underline{x}}')\Sigma^{-1}(\mathcal{X}-E_{n1}\overline{\underline{x}}')'+(E_{n1}\overline{\underline{x}}'-E_{n1}\underline{\mu}')\Sigma^{-1}(E_{n1}\overline{\underline{x}}'-E_{n1}\underline{\mu}')'\right] \\ &\qquad\qquad + 2\ tr\left[(\mathcal{X}-E_{n1}\overline{\underline{x}}')\Sigma^{-1}(E_{n1}\overline{\underline{x}}'-E_{n1}\underline{\mu}')'\right] \end{aligned}$$

where

$$\begin{aligned} & tr\left[(\mathcal{X}-E_{n1}\overline{\underline{x}}')\Sigma^{-1}(E_{n1}\overline{\underline{x}}'-E_{n1}\underline{\mu}')'\right] \\ &\qquad = tr\left[\Sigma^{-1}(E_{n1}\overline{\underline{x}}'-E_{n1}\underline{\mu}')'(\mathcal{X}-E_{n1}\overline{\underline{x}}')\right] \\ &\qquad = tr\left[\Sigma^{-1}(\overline{\underline{x}}E_{1n}\mathcal{X}-\overline{\underline{x}}n\overline{\underline{x}}'-\underline{\mu}E_{1n}\mathcal{X}+\underline{\mu}n\overline{\underline{x}}')\right] = 0 \end{aligned}$$

since

$$E_{1n}\mathcal{X} \;=\; n\underline{\overline{x}}' \; . \tag{4.3}$$

We may thus rewrite (4.2) as

$$\begin{aligned}\mathcal{L}(\underline{\mu},\Sigma) \;&=\; K - \tfrac{n}{2}log|\Sigma| - \tfrac{1}{2}tr\left[(\mathcal{X}-E_{n1}\underline{\overline{x}}')\Sigma^{-1}(\mathcal{X}-E_{n1}\underline{\overline{x}}')'\right] \\ &\qquad - \tfrac{1}{2}tr\left[(E_{n1}\underline{\overline{x}}'-E_{n1}\underline{\mu}')\Sigma^{-1}(E_{n1}\underline{\overline{x}}'-E_{n1}\underline{\mu}')'\right]\end{aligned} \tag{4.4}$$

where

$$K \;=\; -\frac{np}{2}log(2\pi) \; .$$

But, we may note that in (4.4) we may write

$$\begin{aligned}&tr\left[(E_{n1}\underline{\overline{x}}'-E_{n1}\underline{\mu}')\Sigma^{-1}(E_{n1}\underline{\overline{x}}'-E_{n1}\underline{\mu}')'\right] \\ &\qquad = tr\left[\Sigma^{-1}(E_{n1}\underline{\overline{x}}'-E_{n1}\underline{\mu}')'(E_{n1}\underline{\overline{x}}'-E_{n1}\underline{\mu}')\right] \\ &\qquad = tr\left[\Sigma^{-1}(\underline{\overline{x}}'-\underline{\mu}')'E_{1n}E_{n1}(\underline{\overline{x}}'-\underline{\mu}')\right] \\ &\qquad = n\;tr\left[\Sigma^{-1}(\underline{\overline{x}}-\underline{\mu})(\underline{\overline{x}}-\underline{\mu})'\right]\end{aligned}$$

and, using (4.3),

$$\begin{aligned}&tr\left[(\mathcal{X}-E_{n1}\underline{\overline{x}}')\Sigma^{-1}(\mathcal{X}-E_{n1}\underline{\overline{x}}')'\right] \\ &\qquad = tr\left[\Sigma^{-1}(\mathcal{X}-E_{n1}\underline{\overline{x}}')'(\mathcal{X}-E_{n1}\underline{\overline{x}}')\right] \\ &\qquad = tr\left[\Sigma^{-1}(\mathcal{X}-\tfrac{1}{n}E_{n1}E_{1n}\mathcal{X})'(\mathcal{X}-\tfrac{1}{n}E_{n1}E_{1n}\mathcal{X})\right] \\ &\qquad = tr\left[\Sigma^{-1}\left((I-\tfrac{1}{n}E_{nn})\mathcal{X}\right)'(I-\tfrac{1}{n}E_{nn})\mathcal{X}\right] \\ &\qquad = tr\left[\Sigma^{-1}\mathcal{X}'(I-\tfrac{1}{n}E_{nn})(I-\tfrac{1}{n}E_{nn})\mathcal{X}\right] \\ &\qquad = tr\left[\Sigma^{-1}\mathcal{X}'(I-\tfrac{1}{n}E_{nn})\mathcal{X}\right] \\ &\qquad = tr\left[\Sigma^{-1}\mathcal{S}\right]\end{aligned}$$

with

$$\begin{aligned}\mathcal{S} \;&=\; (\mathcal{X}-E_{n1}\underline{\overline{x}}')'(\mathcal{X}-E_{n1}\underline{\overline{x}}') = \mathcal{X}'(I-\tfrac{1}{n}E_{nn})\mathcal{X} \\ &=\; \mathcal{X}'\mathcal{X} - \tfrac{1}{n}\mathcal{X}'E_{nn}\mathcal{X} \; .\end{aligned}$$

Taking into account that

$$tr\left[\Sigma^{-1}(\underline{\overline{x}}-\underline{\mu})(\underline{\overline{x}}-\underline{\mu})'\right] \;=\; tr\left[(\underline{\overline{x}}-\underline{\mu})'\Sigma^{-1}(\underline{\overline{x}}-\underline{\mu})\right] \;=\; (\underline{\overline{x}}-\underline{\mu})'\Sigma^{-1}(\underline{\overline{x}}-\underline{\mu}) \; ,$$

we may write (4.4) as

$$\mathcal{L}(\underline{\mu},\Sigma) \;=\; K-\frac{n}{2}log|\Sigma|-\frac{1}{2}tr\left[\Sigma^{-1}\mathcal{S}\right]-\frac{n}{2}(\underline{\overline{x}}-\underline{\mu})'\Sigma^{-1}(\underline{\overline{x}}-\underline{\mu})\;. \qquad (4.5)$$

Now it is easier to obtain

$$\frac{\partial\mathcal{L}}{\partial\underline{\mu}} \;=\; \frac{n}{2}(\underline{\overline{x}}-\underline{\mu})'\Sigma^{-1}\,. \qquad (4.6)$$

By equating (4.6) to zero we obtain $\underline{\overline{x}}$ as the MLe (Maximum Likelihood estimate) of $\underline{\mu}$. The MLE (Maximum Likelihood Estimator) of $\underline{\mu}$ is thus

$$\underline{\hat{\mu}} \;=\; \underline{\overline{X}}\,, \qquad \text{with} \qquad \underline{\overline{X}}=\left[\overline{X}_1,\,\overline{X}_2,\,\ldots,\,\overline{X}_p\right]' \;=\; \frac{1}{n}X'E_{n1}\,. \qquad (4.7)$$

Once the MLE of $\underline{\mu}$ is obtained, we may obtain the MLE of Σ by direct maximization of $\mathcal{L}(\underline{\overline{X}},\Sigma)$. See Muirhead (1982, sec. 3.1).

Once we have obtained $\underline{\overline{X}}$ as the MLE of $\underline{\mu}$, we may write

$$\begin{aligned}
\mathcal{L}(\underline{\overline{X}},\Sigma) \;&=\; K-\frac{n}{2}\log|\Sigma|-\frac{1}{2}\,tr(\Sigma^{-1}\mathcal{S})\\
&=\; K+\frac{n}{2}\log|\Sigma^{-1}|-\frac{1}{2}tr(\Sigma^{-1}\mathcal{S})\\
&=\; K+\frac{n}{2}\log|\Sigma^{-1}|+\frac{n}{2}\log|\mathcal{S}|-\frac{n}{2}\log|\mathcal{S}|-\frac{1}{2}tr(\Sigma^{-1}\mathcal{S})\\
&=\; K+\frac{n}{2}\log|\Sigma^{-1}\mathcal{S}|-\frac{n}{2}\log|\mathcal{S}|-\frac{1}{2}tr(\Sigma^{-1}\mathcal{S})\,,
\end{aligned}$$

with

$$\max\,\mathcal{L}(\underline{\overline{X}},\Sigma)=\mathcal{L}(\underline{\overline{X}},\widehat{\Sigma})=K+\frac{n}{2}\log|\widehat{\Sigma}^{-1}S|-\frac{n}{2}\log|S|-\frac{1}{2}tr(\widehat{\Sigma}^{-1}S)\,.$$

Taking λ_j $(j=1,\ldots,p)$ as the eigenvalues of $\widehat{\Sigma}^{-1}S$, we may then write

$$\begin{aligned}
\mathcal{L}(\underline{\overline{X}},\widehat{\Sigma}) \;&=\; K+\frac{n}{2}\log\prod_{j=1}^{p}\lambda_j-\frac{1}{2}\sum_{j=1}^{p}\lambda_j-\frac{n}{2}\log|S|\\
&=\; K+\frac{1}{2}\sum_{j=1}^{p}(n\,\log\,\lambda_j-\lambda_j)-\frac{n}{2}\log|S|\,,
\end{aligned}$$

where max $(n \log \lambda_j - \lambda_j)$, seen as a function of λ_j, occurs at $\lambda_j = n$, since

$$\frac{d}{d\lambda_j}(n \log \lambda_j - \lambda_j) = n\frac{1}{\lambda_j} - 1\,,$$

so that we have

$$\frac{d}{d\lambda_j} = 0 \Longrightarrow n\frac{1}{\lambda_j} - 1 = 0 \Longrightarrow \lambda_j = n\,,$$

with

$$\frac{d^2}{d\lambda_j^2}(n \log \lambda_j - \lambda_j) = -n\frac{1}{\lambda_j^2} < 0 \Longrightarrow \text{ máximo}\,,$$

what implies $\widehat{\Sigma}^{-1}S = nI_p$, which in turn implies $\widehat{\Sigma}^{-1} = nS^{-1}$, or $\widehat{\Sigma} = \frac{1}{n}S$.

The MLE of Σ is thus,

$$\begin{aligned} \hat{\Sigma} = \frac{1}{n}S &= \frac{1}{n}(X - E_{n1}\underline{\overline{X}}')'(X - E_{n1}\underline{\overline{X}}') = \frac{1}{n}\,X'\left(I - \tfrac{1}{n}E_{nn}\right)X \\ &= \frac{1}{n}\left(X'X - \tfrac{1}{n}X'E_{nn}X\right). \end{aligned} \tag{4.8}$$

We have thus shown that $\underline{\overline{X}}$ and $\frac{1}{n}S$ are the values of $\underline{\mu}$ and Σ that maximize (4.5) or (4.2) and that, equivalently, maximize (4.1).

It may be shown that $\underline{\overline{X}}$ and S are independent (see section 5.2) and also that $\hat{\Sigma} = \frac{1}{n}S$ is not an unbiased estimator of Σ. The unbiased estimator of Σ is $\frac{1}{n-1}S$. The matrix $\frac{1}{n-1}S$ is the usual sample variance-covariance matrix.

Bibliography

Kshirsagar, A. M. (1972). *Multivariate Analysis.* Marcel Dekker. New York.
(Cap. 3 pp. 47-50.)

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* J. Wiley & Sons. New York. (Cap. 3 pp. 83-85.)

Chapter 5

The Wishart Distribution

5.1 The Wishart distribution as the distribution of the matrix of sums of squares and sums of products of differences from the mean

Let us suppose that the matrix X $(n{\times}p)$ is the matrix of a random sample of size n $(\geq p)$ of a $N_p(\underline{\mu},\ \Sigma)$ distribuyion, that is, that X has the distribution (3.1). Then, the matrix

$$A \ = \ (X - E_{n1}\underline{\mu}')'(X - E_{n1}\underline{\mu}') \tag{5.1}$$

has a Wishart distribution, with parameter matrix Σ and n degrees of freedom. We will denote this fact by

$$A \ \sim \ W_p(n,\ \Sigma)\ , \tag{5.2}$$

and the p.d.f. of A is

$$f_A(\mathcal{A}) \ = \ \frac{|\mathcal{A}|^{(n-p-1)/2}}{2^{np/2}\ \Gamma_p\left(\frac{n}{2}\right)\ \ |\Sigma|^{n/2}}\ e^{-\frac{1}{2}tr(\Sigma^{-1}\mathcal{A})} \qquad (\mathcal{A} > 0)\ \ (n \geq p)\ , \tag{5.3}$$

where $\Gamma_p\left(\frac{n}{2}\right)$ represents the p-multivariate Gamma function, with

$$\Gamma_p(a) \ = \ \pi^{p(p-1)/4}\ \prod_{j=1}^{p}\Gamma\left[a - \frac{1}{2}(j-1)\right]\ , \tag{5.4}$$

where $\Gamma(\cdot)$ represents the Gamma function. In the p.d.f. in (5.3) the notation $\mathcal{A}>0$ means 'for $\mathcal{A}$ positive-definite' (a matrix is said to be positive-definite if all of its eigenvalues are positive). That anotation means that the p.d.f. of the Wishart distribution, with n degrees of freedom and parameter matrix Σ, has the value given by (5.3) for $\mathcal{A}$ positive-definite and zero for $\mathcal{A}$ non positive-definite.

The notation $n\geq p$ and the notation $\mathcal{A}>0$ are indeed equivalent since we have $\mathcal{A}>0$ if and only if $n\geq p$, that is, the condition $n\geq p$ is a necessary and sufficient condition for the matrix $\mathcal{A}$ to be positive-definite, or more precisely, the condition $n\geq p$ is a necessary condition in order to have $\mathcal{A}>0$ and it is sufficient, if $\Sigma>0$, which was assumed in the distribution in (3.1). We may also see that once it was assumed in Section 2 that Σ was positive-definite, that is, that there are no dependence relations among the p variables, or, in other words, that none of the p variables is an exact linear combination of the other $p-1$ variables, the condition $n \geq p$ assures that we have enough degrees of freedom (n) in order to be able to represent the p variables. See Appendix 5.A.

The p.d.f. in (5.3) as well as the p.d.f. in (3.2) are joint probability density functions, and as such each one of them is the joint p.d.f. of all the different elements in each of those matrices, that is, the p.d.f. in (3.2) is the joint p.d.f. of the np elements of X, while the p.d.f. in (5.3) is the joint p.d.f. of the $p(p+1)/2$ different elelements of A. The p.d.f. of the Wishart distribution in (5.3) may be obtained in many different ways (see Kshirsagar, 1972, pp. 58-59) among which by direct integration, from (5.1), or otherwise from the p.d.f. of X in (3.1), or rather, from the p.d.f. of $Z=X-E_{n1}\underline{\mu}'$. (Sverdrup, 1947; Muirhead, 1982).

In fact, one of the easiest ways to obtain the p.d.f. of the Wishart distribution, that is, of the matrix $A=\left(X-E_{n1}\underline{\mu}'\right)'\left(X-E_{n1}\underline{\mu}'\right)$ is directly from the p.d.f. of the matrix $Z=X-E_{n1}\underline{\mu}'$, taking into account that the Jacobian of the transformation $Z\to A=Z'Z$ is

$$\frac{dZ}{dA}=|A|^{(n-p-1)/2}\,\frac{\pi^{np/2}}{\Gamma_p\left(\frac{n}{2}\right)}$$

(see Muirhead, 1982, Chap. 2, sec. 2.1, namely Theorems 2.1.14 e 2.1.15). Then, since $Z=X-E_{n1}\underline{\mu}'\sim N_{np}(0_{n\times p},\Sigma\otimes I_n)$, taking into account the p.d.f. of the matrix Z and the Jacobian above for the transformation $Z \to g(Z) = Z'Z$, the p.d.f. of A is given by

$$f_A(\mathcal{A}) \;=\; \frac{|\mathcal{A}|^{(n-p-1)/2}\,\pi^{np/2}}{(2\pi)^{np/2}\,\Gamma_p\left(\frac{n}{2}\right)\;|\Sigma|^{n/2}}\;e^{-\frac{1}{2}tr\left[\Sigma^{-1}\left(\mathcal{X}-E_{n1}\underline{\mu}'\right)'\left(\mathcal{X}-E_{n1}\underline{\mu}'\right)\right]}$$

where $\left(\mathcal{X}-E_{n1}\underline{\mu}'\right)'\left(\mathcal{X}-E_{n1}\underline{\mu}'\right) = \mathcal{Z}'\mathcal{Z} = \mathcal{A}$, being thus the p.d.f. of a matrix with a $W_p(n,\Sigma)$ distribution. ■

If A has the Wishart distribution in (5.2), with p.d.f. given by (5.3), then its two first moments, or more precisely, its first non-centered moment and its second centered moment, are

$$E(A) \;=\; n\Sigma \tag{5.5}$$

and

$$Var(A) \;=\; Var\left(vec(A)\right) \;=\; n\left(I_{p^2}+K_{pp}\right)(\Sigma\otimes\Sigma)$$

where K_{pp} represents the commutation matrix of order $p^2{\times}p^2$. The matrix K_{pp} has this designation because for any matrix C of dimensions $p{\times}p$

$$K_{pp}\;vec(C) \;=\; vec(C')$$

where

$$K_{pp}=\sum_{i,j=1}^{p}(H_{ij}\otimes H'_{ij})\,,\;\;H_{ij}=[h_{\ell m}]\,,\;\;h_{\ell m}=0,(\ell,m\in\{1,...,p\}),h_{ij}=1\,.$$

The matrix A in (5.1) is the matrix of sums of squares and sums of products of the differences of the n observations $\underline{X}_i$ from their true means (population means or expected values) for the p variables, represented in $\underline{\mu}$. If in (5.1) we had used $\underline{\overline{X}}$ instead of $\underline{\mu}$, we would have obtained the matrix

$$S \;=\; \left(X-E_{n1}\underline{\overline{X}}'\right)'\left(X-E_{n1}\underline{\overline{X}}'\right) \tag{5.6}$$

which, using $E_{1n}X = n\underline{\overline{X}}'$, can be shown to be the matrix defined in (4.8). This is the matrix of sums of squares and products of the differences of the n observations $\underline{X}_i$ from the sample means of the p variables, represented in $\underline{\overline{X}}$.

It may be shown that

$$S \;\sim\; W_p\,(n-1,\;\Sigma)\;\;, \tag{5.7}$$

what will be done in the following section.

To the quantities n and $n-1$, respectively in (5.2) or (5.3) and in (5.7), we call 'degrees of freedom' of the corresponding Wishart distribution. Now we are better able to understand the reason for it. In fact when we compute S in a similar manner to that in which we compute A, but using the sample means $\underline{\overline{X}}$ instead of of the population means $\underline{\mu}$, we loose one degree of freedom when we have to compute first $\underline{\overline{X}}$ in order to compute then S.

Using (5.5) above we see that

$$E(S) \;=\; (n-1)\Sigma\;,$$

which is the reason why $\frac{1}{n-1}S$ and not $\frac{1}{n}S$ is the unbiased estimator of Σ, being this last random variable the MLE of Σ. We have thus the main reasons that may go in favor of the use of either $\frac{1}{n-1}S$ or $\frac{1}{n}S$ as estimators of Σ, as well as of their univariate equivalents as estimators of the population variance, being the first one the unbiased estimator and the second one the MLE.

An important result in the construction of matrices with a Wishart distribution is presented in the following Theorem.

Teorema 5.1: Let X be a matrix with dimensions $q{\times}p$, such that

$$X \sim N_{qp}(0_{q\times p},\, \underset{p\times p}{\Sigma} \otimes \underset{q\times q}{B})\,,$$

where B is an idempotent matrix with $rank(B) = tr(B) = r(\leq q)$. Then

$$X'X \sim W_p(r,\Sigma)\,.$$

(And where the distribution may indeed be a pseudo-Wishart (see section 5.3) if $r \leq p$.) ■

The proof of this result is left as an exercise, since it follows similar lines to those used in the obtention of the p.d.f. of the matrix A.

5.2 Obtaining the distribution of $S = (X - E_{n1}\underline{\overline{X}}')'(X - E_{n1}\underline{\overline{X}}')$

In this section we will obtain the distribution of the matrix S, given by (5.6) or (4.8), as well as the distribution of $\underline{\overline{X}}$, and show their relation of independence. This means that we are finally going to obtain the distributions of the MLEs of $\underline{\mu}$ and Σ, as well as show that they are independent. In fact we may even put forward the following Theorem.

Theorem 5.2: Let X $(n{\times}p)$ be the matrix of a andom sample of dimension n from the distribution $N_p(\underline{\mu}, \Sigma)$. Then

i) $\underline{\overline{X}} = \frac{1}{n}X'E_{n1} \sim N_p\left(\underline{\mu}, \frac{1}{n}\Sigma\right)$,

ii) $$\begin{aligned} S &= (X - E_{n1}\underline{\overline{X}}')'(X - E_{n1}\underline{\overline{X}}') \\ &= X'\left(I_n - \tfrac{1}{n}E_{nn}\right)X \sim W_p(n-1, \Sigma)\,, \end{aligned}$$

iii) $\underline{\overline{X}}$ and S are independent,

where $\underline{\overline{X}}$ and $\frac{1}{n-1}\,S$ are respectively the vector of the sample means and the sample variance-covariance matrix of the p random variables with a joint $N_p(\underline{\mu}, \Sigma)$ distribution. We may note that, for $j, k \in \{1, \ldots, p\}$, the matrix $\frac{1}{n-1}\,S$ has as running element

$$\frac{1}{n-1}\,S_{jk} = \widehat{Cov}(X_j, X_k) \;=\; \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ji} - \overline{X}_j\right)\left(X_{ki} - \overline{X}_k\right)\,,$$

which for $j \neq k$ is the sample covariance of the random variables X_j and X_k and for $k = j$ is the sample variance of the random variable X_j.

Proof: We should note that the result in ii) may be directly obtained from the result in Theorem 5.1. However, we want to undertake here a line of proof that will enable us to obtain at the same time a proof for all three items in the body of the Theorem.

We will consider the orthogonal matrix P, with dimensions $n{\times}n$, which last row is

$$\frac{1}{\sqrt{n}}E_{1n}\,. \tag{5.8}$$

Let us then consider the transformation

$$X \longrightarrow Y = PX\,. \tag{5.9}$$

Then the last row of Y will be equal to

$$\frac{1}{\sqrt{n}} E_{1n} X \;=\; \sqrt{n}\,\underline{\overline{X}}' \,, \tag{5.10}$$

and we may then split the matrix Y as

$$Y \;=\; \left[\frac{Z}{\sqrt{n}\,\underline{\overline{X}}'}\right] \,, \tag{5.11}$$

where Z is a matrix of dimensions $(n-1)\times p$. Note also that

$$PE_{n1} \;=\; \left[\; 0 \;\; 0 \;\; \dots \;\; 0 \;\; \sqrt{n} \;\right]' \,,$$

since the first $n-1$ rows of P are orthogonal to its last row, which is defined in (5.8). We then have

$$\begin{aligned} S &= (X - E_{n1}\underline{\overline{X}}')'(X - E_{n1}\underline{\overline{X}}') \\ &= (X - E_{n1}\underline{\overline{X}}')'P'P(X - E_{n1}\underline{\overline{X}}') \\ &= (X' - \underline{\overline{X}}E_{1n})P'P(X - E_{n1}\underline{\overline{X}}') \\ &= X'P'PX \;-\; X'P'PE_{n1}\underline{\overline{X}}' \;-\; \underline{\overline{X}}E_{1n}P'PX \;+\; \underline{\overline{X}}E_{1n}P'PE_{n1}\underline{\overline{X}}' \\ &= Y'Y \;-\; Y'PE_{n1}\underline{\overline{X}}' \;-\; \underline{\overline{X}}E_{1n}P'Y \;+\; \underline{\overline{X}}E_{1n}P'PE_{n1}\underline{\overline{X}}' \end{aligned} \tag{5.12}$$

where

$$Y'Y = \left[\frac{Z}{\sqrt{n}\,\underline{\overline{X}}'}\right]' \left[\frac{Z}{\sqrt{n}\,\underline{\overline{X}}'}\right] = \left[\; Z' \;\middle|\; \sqrt{n}\,\underline{\overline{X}} \;\right] \left[\frac{Z}{\sqrt{n}\,\underline{\overline{X}}'}\right] = Z'Z + n\,\underline{\overline{X}}\,\underline{\overline{X}}', \tag{5.13}$$

and, using (5.10) and (5.11), we have

$$Y'PE_{n1}\underline{\overline{X}}' \;=\; \left[\; Z' \;\middle|\; \sqrt{n}\,\underline{\overline{X}} \;\right] \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \sqrt{n} \end{bmatrix} \underline{\overline{X}}' \;=\; n\,\underline{\overline{X}}\,\underline{\overline{X}}' \,, \tag{5.14}$$

and

$$\underline{\overline{X}}E_{1n}P'Y \;=\; \underline{\overline{X}}\left[\; 0 \;\; 0 \;\; \dots \;\; 0 \;\; \sqrt{n} \;\right] \left[\frac{Z}{\sqrt{n}\,\underline{\overline{X}}'}\right] \;=\; n\,\underline{\overline{X}}\,\underline{\overline{X}}' \,, \tag{5.15}$$

so that, from (5.12), (5.13), (5.14) and (5.15),

$$S \;=\; Z'Z \;. \tag{5.16}$$

Then, from the distribution of X in (3.1), using the rule of derivation of a transformed random variable which says that if $g(\cdot)$ is an invertible function and if

$$Y \;=\; g(X) \tag{5.17}$$

then

$$f(Y) \;=\; f\left(g^{-1}(Y)\right)\left|\frac{dX}{dY}\right| \;,$$

where we call $\left|\frac{dX}{dY}\right|$, namely when we refer to multivariate distributions, the Jacobian of the transformation (5.17).

In the case of the transformation (5.9) we have that the distribution of Y is the joint distribution of Z and $\sqrt{n}\,\underline{\overline{X}}$, being yet the case that

$$\left|\frac{dX}{dY}\right| \;=\; |P^{-1}|^p \;=\; 1\;.$$

Taking into account the definition of the matrix S and to the fact that $E_{np}E_{pq} = pE_{nq}$, from which we have $E_{1n}E_{n1} = n$, we get

$$\begin{aligned}
&\left(X - E_{n1}\underline{\mu}'\right)'\left(X - E_{n1}\underline{\mu}'\right)\\
&\quad= \left(X - E_{n1}\underline{\overline{X}}' + E_{n1}\underline{\overline{X}}' - E_{n1}\underline{\mu}'\right)'\left(X - E_{n1}\underline{\overline{X}}' + E_{n1}\underline{\overline{X}}' - E_{n1}\underline{\mu}'\right)\\
&\quad= \left(X - E_{n1}\underline{\overline{X}}'\right)'\left(X - E_{n1}\underline{\overline{X}}'\right)\\
&\qquad+\underbrace{\left(X-E_{n1}\underline{\overline{X}}'\right)'\left(E_{n1}\underline{\overline{X}}'-E_{n1}\underline{\mu}'\right)}_{=0}+\underbrace{\left(E_{n1}\underline{\overline{X}}'-E_{n1}\underline{\mu}'\right)'\left(X-E_{n1}\underline{\overline{X}}'\right)}_{=0}\\
&\qquad+\left(E_{n1}\underline{\overline{X}}' - E_{n1}\underline{\mu}'\right)'\left(E_{n1}\underline{\overline{X}}' - E_{n1}\underline{\mu}'\right)\\
&\quad= S + \left(\underline{\overline{X}}' - \underline{\mu}'\right)' E_{1n}E_{n1}\left(\underline{\overline{X}}' - \underline{\mu}'\right)\\
&\quad= S + n\left(\underline{\overline{X}} - \underline{\mu}\right)\left(\underline{\overline{X}} - \underline{\mu}\right)',
\end{aligned}$$

since

$$\left(X - E_{n1}\underline{\overline{X}}'\right)'\left(E_{n1}\underline{\overline{X}}' - E_{n1}\underline{\mu}'\right) = X'E_{n1}\underline{\overline{X}}' - \underline{\overline{X}}E_{1n}E_{n1}\underline{\overline{X}}' - X'E_{n1}\underline{\mu}' + \underline{\overline{X}}E_{1n}E_{n1}\underline{\mu}'$$
$$= n\underline{\overline{X}}\,\underline{\overline{X}}' - n\underline{\overline{X}}\,\underline{\overline{X}}' - n\underline{\overline{X}}\underline{\mu}' + n\underline{\overline{X}}\underline{\mu}' = 0$$

and

$$\left(E_{n1}\underline{\overline{X}}' - E_{n1}\underline{\mu}'\right)'\left(X - E_{n1}\underline{\overline{X}}'\right) = \left(\left(X - E_{n1}\underline{\overline{X}}'\right)'\left(E_{n1}\underline{\overline{X}}' - E_{n1}\underline{\mu}'\right)\right)' = 0\,,$$

so that

$$tr\left[\Sigma^{-1}\left(\mathcal{X} - E_{n1}\underline{\mu}'\right)'\left(\mathcal{X} - E_{n1}\underline{\mu}'\right)\right]$$
$$= tr\left[\Sigma^{-1}\left(\mathcal{X} - E_{n1}\underline{\overline{X}}'\right)'\left(\mathcal{X} - E_{n1}\underline{\overline{X}}'\right) + n\Sigma^{-1}\left(\underline{\overline{X}} - \underline{\mu}\right)\left(\underline{\overline{X}} - \underline{\mu}\right)'\right]$$
$$= tr\left[\Sigma^{-1}\mathcal{S}\right] + n\left(\underline{\overline{X}} - \underline{\mu}\right)'\Sigma^{-1}\left(\underline{\overline{X}} - \underline{\mu}\right)$$

and thus we have, from (3.2) and (5.11),

$$f_Y(\mathcal{Y}) = f_{Z,\sqrt{n}\underline{\overline{X}}}(\mathcal{Z}, \sqrt{n}\,\underline{\overline{x}}) = \frac{1}{(2\pi)^{np/2}\,|\Sigma|^{n/2}} \times e^{-\frac{1}{2}tr\left[\Sigma^{-1}\mathcal{Z}'\mathcal{Z} + n(\underline{\overline{x}}-\underline{\mu})'\Sigma^{-1}(\underline{\overline{x}}-\underline{\mu})\right]}\, dZ\ d(\sqrt{n}\,\underline{\overline{X}})$$
$$= \frac{1}{(2\pi)^{p/2}\,|\Sigma|^{1/2}}\, e^{-\frac{1}{2}\sqrt{n}(\underline{\overline{x}}-\underline{\mu})'\Sigma^{-1}\sqrt{n}(\underline{\overline{x}}-\underline{\mu})}\, d(\sqrt{n}\,\underline{\overline{X}}) \times \frac{1}{(2\pi)^{(n-1)p/2}\,|\Sigma|^{(n-1)/2}}\, e^{-\frac{1}{2}tr\left[\Sigma^{-1}\mathcal{Z}'\mathcal{Z}\right]}\, dZ$$
$$= f_{\sqrt{n}\underline{\overline{X}}}(\sqrt{n}\,\underline{\overline{x}})\ f_Z(\mathcal{Z})$$

where

$$\sqrt{n}\,\underline{\overline{X}} \sim N_p\left(\sqrt{n}\,\underline{\mu},\ \Sigma\right) \iff \underline{\overline{X}} \sim N_p\left(\underline{\mu}, \tfrac{1}{n}\Sigma\right)$$

and

$$Z \sim N_{(n-1)p}\left(\underline{0},\ \Sigma \otimes I_{n-1}\right)\,, \tag{5.18}$$

and where $\sqrt{n}\,\underline{\overline{X}}$ and Z are independent since

$$f_{\sqrt{n}\underline{\overline{X}},Z}(\sqrt{n}\,\underline{\overline{x}},\ \mathcal{Z}) = f_{\sqrt{n}\underline{\overline{X}}}(\sqrt{n}\,\underline{\overline{x}})\ f_Z(\mathcal{Z})\ .$$

Then, from (5.18) and Theorem 5.1 we may conclude that

$$S \;=\; Z'Z \;\sim\; W_p\,(n-1,\ \Sigma)\ .$$

But if $\sqrt{n}\,\overline{X}$ and Z are independent, then also $\overline{X}$ and $S = Z'Z$ are independent. ∎

5.3 The non-central Wishart and pseudo-Wishart distributions

The matrix A defined in (5.1) may be written as

$$A \;=\; X^{*'}X^{*}$$

where

$$X^{*} \;=\; X - E_{n1}\underline{\mu}'$$

has the distribution

$$X^{*} \;\sim\; N_{np}\,(0_{n\times p},\ \Sigma\otimes I_n)\ ,$$

given that it is assumed that X has the distribution in (3.1). We say then that the matrix A has a (centered) Wishart distribution, with p.d.f. given by (5.3).

When the matrix A may be written as

$$A \;=\; \widetilde{X}'\widetilde{X} \tag{5.19}$$

where

$$\widetilde{X} \;\sim\; N_{np}\,(M_{n\times p},\ \Sigma\otimes I_n)\ , \tag{5.20}$$

we say that A has a Wishart distribution. If

$$M_{n\times p} \;\neq\; 0_{n\times p}$$

we say that A has a non-central Wishart distribution, with non-centrality parameter

$$\Theta \;=\; M'M\ ,$$

and we denote this fact by

$$A \;\sim\; W_p^{*}\,(n,\ \Sigma,\ \Theta)\ .$$

If

$$M_{n\times p} \;=\; 0_{n\times p}\ ,$$

or, equivalently, if

$$\Theta \;=\; M'M \;=\; 0_{p\times p}$$

then we say that, as we saw above, A has a centered or central Wishart distribution, and we denote this fact as in (5.2).

When we consider the Wishart distribution in 5.1, and namely when we consider its probability density function in (5.3), we consider $n \geq p$ (See Appendix 5.A). In fact only in this case it makes sense to say that A has a Wishart distribution.

Let us suppose that we may write A as in (5.19), where $\widetilde{X}$ still has a distribution of the type in (5.20), but where $n < p$ (note that for distributions of the type of that in (5.20) or (3.1) there is no restriction in terms of the values of n and p). We say that in this case A has a pseudo-Wishart distribution (centered or non-centered, according of having $M_{n\times p} = 0_{n\times p}$ or $M_{n\times p} \neq 0_{n\times p}$). If A has a centered pseudo-Wishart distribution we will denote this fact by

$$A \;\sim\; \widetilde{W}_p\,(n,\ \Sigma) \qquad (n < p)\ ,$$

and if A has a non-centered pseudo-Wishart distribution we denote this fact by

$$A \;\sim\; \widetilde{W}^*\,(n,\ \Sigma,\ \Theta)\ .$$

Usually when we say that a matrix A has a Wishart distribution we want to say that A has a centered Wishart distribution and when we refer that the matrix A has a pseudo-Wishart distribution we want to say that it has a centered pseudo-Wishart distribution.

5.4 The distribution of the MLE of Σ

We showed in 4.2 that the MLE of Σ is

$$V^* \;=\; \frac{1}{n}S\ . \tag{5.21}$$

We may note that if we take into account the definition of S in (5.6) or (4.8), then we may write

$$V^* \;=\; X^{*\prime}DX^*$$

where

$$D \;=\; \frac{1}{n} I_n$$

and

$$X^* \;=\; X - \frac{1}{n}\, E_{n1}\, \underline{\overline{X}}' \;.$$

In Section 5.2 we have shown that the matrix S given by (5.6) or (4.8), has the distribution in (5.7). Then, based on that result, it is easy to show that

$$V^* \;=\; \frac{1}{n} S \;\;\sim\;\; W_p\left(n-1,\; \frac{1}{n}\Sigma\right) \;.$$

If we use wieghts with a unitary sum that are not all equal, that is, if

$$D \;=\; diag(p_i) \qquad\quad i = 1, ..., n$$

with

$$p_i > 0 \;\; (i = 1, \ldots, n) \qquad \text{e} \qquad \sum_{i=1}^{n} p_i \;=\; 1 \;,$$

then the distribution of the matrix V^* will still be approximately Wishart in the sense that the two first moments of its distribution will coincide with the two first moments of a Wishart distribution with parameter matrix $\frac{g}{1-l}\Sigma$ and $\frac{(1-l)^2}{g}$ degrees-of-freedom, where, according to Coelho (1994),

$$l \;=\; E_{1n} D^2 E_{n1} \qquad \text{e} \qquad g \;=\; tr(A^2) \quad \text{com} \quad A = (I_n - E_{nn}D)' D (I_n - E_{nn}D) \;.$$

5.5 Some properties of the Wishart distribution

I) We may note that when $p=1$, the p.d.f. of the Wishart distribution in (5.3) reduces to the p.d.f. of a χ^2 with n degrees-of-freedom multiplied by σ, since then both A and Σ will be scalars, which we will denote by a and σ, allowing us to write

$$f_A(a) \;=\; \frac{a^{(n-2)/2}}{2^{n/2}\,\Gamma\left(\frac{n}{2}\right)\,\sigma^{n/2}}\; e^{-\frac{1}{2}\frac{a}{\sigma}}\; da \;,$$

which shows that if σ=1 then the random variable A has a Chi-square distribution with n degrees-of-freedom.

As such we may see the Wishart distribution as a multivariate generalization of the Chi-square distribution. In fact, random variables with a Wishart distribution have in multivariate methods and models (as the Multivariate Regression or Canonical Analysis and Multivariate Analysis of Variance) the same role that random variables with a chi-square distribution have in the univariate methods and models as the Linear Regression and Analysis of Variance.

The following properties will be in general only enunciated. Their proofs may be found in the references listed at the end of the Chapter.

II) Let the matrices $A_1,\ A_2, ...,\ A_r$ be independent, with

$$A_i \ \sim \ W_p(n_i,\ \Sigma) \quad i = 1, ..., r\ , \quad (n_i > p)$$

then if $A = \sum_{i=1}^{r} A_i$, we have

$$A \ \sim \ W_p(n,\ \Sigma)$$

with

$$n \ = \ \sum_{i=1}^{r} n_i \ .$$

III) Let

$$A \ \sim \ W_p(n,\ \Sigma) \qquad (n > p)$$

and let M be a non-random matrix with dimensions $k{\times}p$, of *rank* k, then

$$MAM' \ \sim \ W_k(n,\ M\Sigma M')$$

and

$$(MA^{-1}M')^{-1} \ \sim \ W_k\left(n - p + k,\ (M\Sigma^{-1}M')^{-1}\right) \ .$$

IV) Let

$$A \ \sim \ W_p(n,\ \Sigma) \qquad (n > p) \ .$$

Let then A and Σ be split in the following way

$$A \;=\; \begin{array}{cc} \begin{array}{cc} p_1 & p_2 \end{array} & \\ \left[\begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array}\right] & \begin{array}{c} p_1 \\ p_2 \end{array} \end{array} , \qquad \Sigma \;=\; \begin{array}{cc} \begin{array}{cc} p_1 & p_2 \end{array} & \\ \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right] & \begin{array}{c} p_1 \\ p_2 \end{array} \end{array} , \qquad p_1 + p_2 = p \ .$$

Then

$$A_{11} \;\sim\; W_{p_1}(n,\ \Sigma_{11}) \quad \text{ e } \quad A_{22} \;\sim\; W_{p_2}(n,\ \Sigma_{22}) \ .$$

[Note that this property is indeed a Corollary of property III).]

V) Let

$$A \;\sim\; W_p(n,\ \Sigma) \qquad (n > p)$$

and let A and Σ be split as in IV) above and further let $\Sigma_{12} = 0$, then

$$A_{11} \;\sim\; W_{p_1}(n,\ \Sigma_{11})$$

$$A_{22} \;\sim\; W_{p_2}(n,\ \Sigma_{22})$$

and they are independent.

VI) Let

$$A \;\sim\; W_p(n,\ \Sigma) \qquad (n > p)$$

and $\underline{Y}$ be a vector of dimensions $p{\times}1$, independent of A, such that $P(\underline{Y}{=}\underline{0}) = 0$, then

$$\frac{\underline{Y}'A\underline{Y}}{\underline{Y}'\Sigma\underline{Y}} \;\sim\; \chi^2_n \qquad \text{independent of} \quad \underline{Y}$$

and

$$\frac{\underline{Y}'\Sigma^{-1}\underline{Y}}{\underline{Y}'A^{-1}\underline{Y}} \;\sim\; \chi^2_{n-p+1} \qquad \text{independent of} \quad \underline{Y}\ ;$$

[in particular, note that we have

$$\frac{\underline{\overline{X}}'S\,\underline{\overline{X}}}{\underline{\overline{X}}'\Sigma\,\underline{\overline{X}}} \;\sim\; \chi^2_{n-1} \qquad \text{independent of} \quad \underline{\overline{X}}$$] .

VII) Let

$$A \;\sim\; W_p(n,\ \Sigma) \qquad (n > p)\ ,$$

with A and Σ subdivided as in IV), and let

$$A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21} \qquad \text{e} \qquad \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\ .$$

Then $A_{11.2}$, $A_{21}|A_{22}$ and A_{22} are independent, with

i) $A_{11.2} \;\sim\; W_{p_1}(n - p_2,\ \Sigma_{11.2})$

ii) $(A_{21}|A_{22}) \;\sim\; N_{p_2p_1}\left(A_{22}\Sigma_{22}^{-1}\Sigma_{21},\ \Sigma_{11.2} \otimes A_{22}\right)$

iii) $A_{22} \;\sim\; W_{p_2}(n,\ \Sigma_{22})$.

Since the proof of this property will use the direct use of the Wishart distribution in an interesting way, we think it to be useful its execution, moreover since the methodoly used is similar to the one used in the proof of similar results of a broader scope.

Proof:

The p.d.f. of A is given by (5.3), so that

$$f_A(\mathcal{A})\,d\mathcal{A} \;=\; \frac{|\mathcal{A}|^{(n-p-1)/2}}{2^{np/2}\ \Gamma_p\left(\frac{n}{2}\right)\ |\Sigma|^{n/2}}\ e^{-\frac{1}{2}tr(\Sigma^{-1}\mathcal{A})}\ d\mathcal{A} \qquad (A > 0) \quad (5.22)$$

where

$$d\mathcal{A} \;=\; d\mathcal{A}_{11}\ d\mathcal{A}_{21}\ d\mathcal{A}_{22}$$

since the distribution in (5.22) is, as already mentioned, the joint distribution of the $p(p+1)/2$ distinct elelements in A.

Let us consider in (5.22) the following changes of variables

$$A_{11} \;\longrightarrow\; A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21} \qquad (5.23)$$

$$A_{21} \;\longrightarrow\; A_{21} \qquad (5.24)$$

$$A_{22} \;\longrightarrow\; A_{22}\ . \qquad (5.25)$$

The Jacobians of the transformations above in (5.23), (5.24) and (5.25) are all equal to 1. Note further that, for a matrix A split as in IV), we have the following relations

$$\begin{aligned}
|A| &= |A_{11}|\,|A_{22} - A_{21}A_{11}^{-1}A_{12}| \qquad \text{se } A_{11} \text{ definida-positiva} \\
&= |A_{11}|\,|A_{22.1}| \\
&= |A_{22}|\,|A_{11} - A_{12}A_{22}^{-1}A_{21}| \qquad \text{se } A_{22} \text{ definida-positiva} \\
&= |A_{22}|\,|A_{11.2}|\,,
\end{aligned}$$

and similar ones for $|\Sigma|$ (and where $A_{22.1}$ is defined in a similar way as that of $A_{11.2}$).

It may be aslo shown that if Σ is split as in IV), and if we define $\Sigma_{22.1}$ in a similar manner to that in which $\Sigma_{11.2}$ was defined in VII), then

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11.2}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1} & \Sigma_{22.1}^{-1} \end{bmatrix}. \tag{5.26}$$

It is also possible to prove that if E and B are two matrices, respectively with dimensions $r{\times}r$ e $s{\times}s$, such that E^{-1} and B^{-1} exist, and if C and D are matrices of dimensions $r{\times}s$ and $s{\times}r$, respectively, then if

$$P = E \pm CBD\,, \tag{5.27}$$

we have

$$\begin{aligned}
P^{-1} &= E^{-1} \mp E^{-1}CB(B \pm BDE^{-1}CB)^{-1}BDE^{-1} \\
&= E^{-1} \mp E^{-1}CB(I_s \pm DE^{-1}CB)^{-1}DE^{-1} \qquad (5.28) \\
&= E^{-1} \mp E^{-1}C(I_s \pm BDE^{-1}C)^{-1}BDE^{-1}\,.
\end{aligned}$$

From (5.27) and (5.28) above, taking into account the definition of $\Sigma_{22.1}$, and taking into account the split of Σ in IV), we have

$$\begin{aligned}
\Sigma_{22.1}^{-1} &= (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \\
&= \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}(\Sigma_{11}^{-1} - \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1})^{-1}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\
&= \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(I - \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\
&= \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\
&= \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}
\end{aligned} \tag{5.29}$$

and, in a similar way,

$$\Sigma_{11.2}^{-1} \;=\; \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} \;. \tag{5.30}$$

Then, using (5.26) and (5.29),

$$\begin{aligned} tr(\Sigma^{-1}\mathcal{A}) &= tr(\Sigma_{11.2}^{-1}\mathcal{A}_{11} - \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}\mathcal{A}_{21}) \\ &\qquad + tr(\Sigma_{22.1}^{-1}\mathcal{A}_{22} - \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\mathcal{A}_{12}) \\ &= tr(\Sigma_{11.2}^{-1}\mathcal{A}_{11.2} + \Sigma_{11.2}^{-1}\mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21} - \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}\mathcal{A}_{21}) \\ &\quad + tr(\Sigma_{22}^{-1}\mathcal{A}_{22} + \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{22} \\ &\qquad - \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\mathcal{A}_{12}) \end{aligned}$$

where, using (5.29) and (5.30),

$$\begin{aligned} tr(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}\mathcal{A}_{21}) &= tr(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{21}) \\ &\qquad + tr(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{21}) \\ &= tr(\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{21}) \\ &\qquad - tr(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{21}) \\ &\qquad + tr(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{21}) \\ &= tr(\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{21}) \end{aligned}$$

so that

$$\begin{aligned} tr(\Sigma^{-1}\mathcal{A}) &= tr(\Sigma_{11.2}^{-1}\mathcal{A}_{11.2}) + tr(\Sigma_{22}^{-1}\mathcal{A}_{22}) + tr(\Sigma_{11.2}^{-1}\mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21}) \\ &\quad - tr(\Sigma_{11.2}^{-1}\mathcal{A}_{12}\Sigma_{22}^{-1}\Sigma_{21}) + tr(\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{22}\Sigma_{22}^{-1}\Sigma_{21}) \\ &\quad - tr(\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{21}) \\ &= tr(\Sigma_{11.2}^{-1}\mathcal{A}_{11.2}) + tr(\Sigma_{22}^{-1}\mathcal{A}_{22}) \\ &\quad + tr\left[\Sigma_{11.2}^{-1}(\mathcal{A}_{12} - \Sigma_{12}\Sigma_{ss}^{-1}\mathcal{A}_{22})\mathcal{A}_{22}^{-1}(\mathcal{A}_{12} - \Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{22})'\right]. \end{aligned}$$

Note that we may yet split $\Gamma_p\left(\frac{n}{2}\right)$ in a convenient way, using (5.4), as

$$\begin{aligned}
\Gamma_p\left(\frac{n}{2}\right) &= \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma\left[\frac{1}{2}(n-j+1)\right] \\
&= \pi^{p_1(p_1-1)/4} \prod_{j=1}^{p_1} \Gamma\left[\frac{1}{2}(n-p_2-j+1)\right] \times \\
&\qquad \times \pi^{p_2(p_2-1)/4} \prod_{j=1}^{p_2} \Gamma\left[\frac{1}{2}(n-j+1)\right] \times \pi^{p_1p_2/2} \\
&= \Gamma_{p_1}\left(\frac{n-p_2}{2}\right) \Gamma_{p_2}\left(\frac{n}{2}\right) \pi^{p_1p_2/2} .
\end{aligned}$$

We are finally able to derive from (5.22) the joint distribution of $A_{11.2}$, A_{21} and A_{22}, since from (5.22) we may write

$$\begin{aligned}
&f(\mathcal{A}_{11.2},\mathcal{A}_{12},\mathcal{A}_{22})\, d\mathcal{A}_{11.2}\, d\mathcal{A}_{12}\, d\mathcal{A}_{22} = \\
&= \frac{|\mathcal{A}_{22}|^{(n-p-1)/2}\ |\mathcal{A}_{11.2}|^{(n-p-1)/2}}{2^{np/2}\ \Gamma_{p_1}\left(\frac{n-p_2}{2}\right)\ \Gamma_{p_2}\left(\frac{n}{2}\right)\ \pi^{p_1p_2/2}\ |\Sigma_{22}|^{n/2}\ |\Sigma_{11.2}|^{n/2}} \\
&\times e^{-\frac{1}{2}\left\{tr\left(\Sigma_{11.2}^{-1}\mathcal{A}_{11.2}\right)+tr\left(\Sigma_{22}^{-1}\mathcal{A}_{22}\right)+tr\left[\Sigma_{11.2}^{-1}(\mathcal{A}_{12}-\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{22})\mathcal{A}_{22}^{-1}(\mathcal{A}_{12}-\Sigma_{12}\Sigma_{22}^{-1}\mathcal{A}_{22})'\right]\right\}} \\
&\qquad\qquad d\mathcal{A}_{11.2}\ d\mathcal{A}_{12}\ d\mathcal{A}_{22} \\
&= \frac{|\mathcal{A}_{11.2}|^{(n-p_2-p_1-1)/2}}{2^{(n-p_2)p_1/2}\ \Gamma_{p_1}\left(\frac{n-p_2}{2}\right)\ |\Sigma_{11.2}|^{(n-p_2)/2}}\ e^{-\frac{1}{2}tr\left(\Sigma_{11.2}^{-1}\mathcal{A}_{11.2}\right)}\ d\mathcal{A}_{11.2} \\
&\quad \times \frac{1}{(2\pi)^{p_2p_1/2}\ |\mathcal{A}_{22}|^{p_1/2}\ |\Sigma_{11.2}|^{p_2/2}} \\
&\qquad \times e^{tr\left[-\frac{1}{2}\Sigma_{11.2}^{-1}(\mathcal{A}_{21}-\mathcal{A}_{22}\Sigma_{22}^{-1}\Sigma_{21})'\mathcal{A}_{22}^{-1}(\mathcal{A}_{21}-\mathcal{A}_{22}\Sigma_{22}^{-1}\Sigma_{21})\right]}\ d\mathcal{A}_{12} \\
&\quad \times \frac{|\mathcal{A}_{22}|^{(n-p_2-1)/2}}{2^{np_2/2}\ \Gamma_{p_2}\left(\frac{n}{2}\right)\ |\Sigma_{22}|^{n/2}}\ e^{-\frac{1}{2}tr(\Sigma_{22}^{-1}\mathcal{A}_{22})}\ d\mathcal{A}_{22}\ ,
\end{aligned}$$

which shows that

i) $A_{11.2} \sim W_{p_1}(n-p_2,\ \Sigma_{11.2})$

ii) $(A_{21}|A_{22}) \sim N_{p_2 p_1}\left(A_{22}\Sigma_{22}^{-1}\Sigma_{21},\ \Sigma_{11.2}\otimes A_{22}\right)$

iii) $A_{22} \sim W_{p_2}(n,\ \Sigma_{22})$

and that $A_{11.2}$, A_{12} and A_{22} are independent since their joint distribution is the product of their marginal distributions;

VIII) Let

$$A \sim W_p(n,\ \Sigma) \qquad (n > p) \ ,$$

be split as in IV) and let

$$A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12} \qquad \text{e} \qquad \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \ .$$

Then

i) $A_{22.1} \sim W_{p_2}(n-p_1,\ \Sigma_{22.1})$

ii) $(A_{12}|A_{11}) \sim N_{p_1 p_2}\left(A_{11}\Sigma_{11}^{-1}\Sigma_{12},\ \Sigma_{22.1}\otimes A_{11}\right)$

iii) $A_{11} \sim W_{p_1}(n,\ \Sigma_{11})$

and they are all independent.

(The proof follows similar lines to those used in VII).)

IX) Let

$$A \sim W_p(n,\ \Sigma) \qquad (n > p) \ ,$$

with A and Σ split as in IV) and let $A_{11.2}$ and $\Sigma_{11.2}$ be defined as in VII) and $A_{22.1}$ and $\Sigma_{22.1}$ as in VIII). Let $\Sigma_{12} = 0$ $(p_1 \times p_2)$, then

if $p_2 > p_1$

i) $A_{12}A_{22}^{-1}A_{21} \sim W_{p_1}(p_2,\ \Sigma_{11})$, independent of $A_{11.2} \sim W_{p_1}(n - p_2,\ \Sigma_{11})$

ii) $A_{21}A_{11}^{-1}A_{12} \sim \widetilde{W}_{p_2}(p_1,\ \Sigma_{22})$, independent of $A_{22.1} \sim W_{p_2}(n - p_1,\ \Sigma_{22})$

and if $p_1 > p_2$

iii) $A_{12}A_{22}^{-1}A_{21} \sim \widetilde{W}_{p_1}(p_2,\ \Sigma_{11})$, independent of $A_{11.2} \sim W_{p_1}(n - p_2,\ \Sigma_{11})$

iv) $A_{21}A_{11}^{-1}A_{12} \;\sim\; W_{p_2}(p_1,\ \Sigma_{22})$, independent of $A_{22.1} \sim W_{p_2}(n - p_1,\ \Sigma_{22})$

Proof:

From ii) in VII), given that $\Sigma_{12} = 0_{(p_1 \times p_2)}$,

$$\begin{aligned}(A_{22}^{-1/2}A_{21}) \;&\sim\; N_{p_2 p_1}\left(A_{22}^{1/2}\Sigma_{22}^{-1}\Sigma_{21},\ \Sigma_{11.2} \otimes I_{p_2}\right)\\ &\equiv\; N_{p_2 p_1}\left(0_{(p_2 \times p_1)},\ \Sigma_{11} \otimes I_{p_2}\right)\end{aligned}$$

(note that if $\Sigma_{12} = 0_{(p_1 \times p_2)}$ then $\Sigma_{11.2} = \Sigma_{11}$).

And then, from section 5.1, if $p_2 > p_1$,

$$(A_{22}^{-1/2}A_{21})'(A_{22}^{-1/2}A_{21}) \;=\; A_{12}A_{22}^{-1}A_{21} \;\sim\; W_{p_1}(p_2,\ \Sigma_{11}) \ .$$

Given that $(A_{21}|A_{22})$ is independent of $A_{11.2}$ (from VII)), also the matrices $A_{22}^{-1/2}A_{21}$ and $A_{12}A_{22}^{-1}A_{21}$ are independent from $A_{11.2}$.

The proof of ii) is similar, taking into account the exposition in section 5.3.

Since if $\Sigma_{12} = 0_{(p_1 \times p_2)}$ we have $\Sigma_{11.2} = \Sigma_{11}$ and $\Sigma_{22.1} = \Sigma_{22}$, we have the parameter matrices in the Wishart distributions of $A_{11.2}$ and $A_{22.1}$ respectively as Σ_{11} and Σ_{22} instead of $\Sigma_{11.2}$ and $\Sigma_{22.1}$.

The proof of iii) and iv) is similar.

Based on the second part of property VI) we have that, if $A \sim W_p(n,\ \Sigma)$, then

$$E(A^{-1}) \;=\; \frac{1}{n-p-1}\Sigma^{-1} \ .$$

It is also possible to easily show the result in the following Theorem, which is very useful in the derivation of the distribution of several likelihood ratio statistics.

Theorem 5.3: If $A \sim W_p(n, \Sigma)$, then

$$E\left(|A|^h\right) \;=\; |\Sigma|^h\, 2^{hp}\, \frac{\Gamma_p\left(\frac{n}{2}+h\right)}{\Gamma_p\left(\frac{n}{2}\right)}\,.$$

Proof: Using the expression of the p.d.f. of A in (5.3), for $h > (p+1-n)/2$ we have

$$\begin{aligned}
E\left(|A|^h\right)& \\
&= \int_{\mathcal{A}>0} |\mathcal{A}|^h\, f_A(\mathcal{A})\, d\mathcal{A} \\
&= \int_{\mathcal{A}>0} |\mathcal{A}|^h\, \frac{e^{-\frac{1}{2}tr(\Sigma^{-1}\mathcal{A})}|\mathcal{A}|^{(n-p-1)/2}}{2^{np/2}\,\Gamma_p\left(\frac{n}{2}\right)|\Sigma|^{n/2}}\, d\mathcal{A} \\
&= \frac{2^{(n+2h)p/2}\,\Gamma_p\left(\frac{n+2h}{2}\right)\;|\Sigma|^{n/2+h}}{2^{np/2}\,\Gamma_p\left(\frac{n}{2}\right)\;|\Sigma|^{n/2}} \underbrace{\int_{\mathcal{A}>0} \underbrace{\frac{e^{-\frac{1}{2}tr(\Sigma^{-1}\mathcal{A})}|\mathcal{A}|^{(n+2h-p-1)/2}}{2^{(n+2h)p/2}\,\Gamma_p\left(\frac{n+2h}{2}\right)|\Sigma|^{(n+2h)/2}}}_{\text{f.d.p. de } W_p(n+2h,\Sigma)}\, d\mathcal{A}}_{=1} \\
&= 2^{hp}\,|\Sigma|^h\,\frac{\Gamma_p\left(\frac{n}{2}+h\right)}{\Gamma_p\left(\frac{n}{2}\right)}
\end{aligned} \tag{5.31}$$

where the notation $\int_{\mathcal{A}>0}$ means that the integral is taken over the space of all symmetric definite-positive matrices.

These results and properties will be used in the construction of statistics used to test hypotheses related with multivariate linear models as well as in the obtention of the distribution of such statistics.

❏❏❏❏❏

Exercises:

5.1 a) Prove property II).

b) Prove property V).

5.2 Let A and B be two independent Wishart matrices, with

$$A \sim W_p(n_1, \Sigma) \quad \text{and} \quad B \sim W_p(n_2, \Sigma)\,.$$

Obtain $E(\Lambda^h)$ and the distribution of Λ, where

$$\Lambda = \frac{|A|}{|A+B|}\,.$$

❑❑❑❑❑

Bibliography

Coelho, C. A. (1994). Brief note on the approximate distribution of the dispersion matrix of a weighted sample from a p-variate normal distribution. *Journal of the Indian Statistical Association*, **32**, 77-81.

Kshirsagar, A. M. (1972). *Multivariate Analysis.* Marcel Dekker, New York. (Cap. 3 pp. 47-82.)

Morrison, D. F. (1976). *Multivariate Statistical Methods.* 2ªed. McGraw-Hill, New York. (Cap.5, Sec. 6)

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* J. Wiley & Sons. New York. (Cap. 3 pp. 85-99.)

Sverdrup, E. (1947). Derivation of the Wishart distribution of the order sample moments by straightforward integration of a multiple integral. *Skandinavisk Aktuarietidskrift*, **30**, 151–166.

Appendix 5.A

It is possible to show that, assuming that Σ is a positive-definite matrix, then for a matrix A defined as in (5.1), or for a matrix S defined as in (5.6) we have respectively that

$$A > 0 \quad \text{if and only if} \quad n \geq p$$

and

$$S > 0 \quad \text{if and only if} \quad n-1 \geq p \; .$$

This fact is shown by Dykstra (1970) [and Muirhead (1982)]. We will, following a procedure in all similar to the one used by those authors, show first (5.77), i.e. that S is positive-definite if and only if $n-1 \geq p$, given that Σ is positive-definite.

We know that from (5.16) we may write $S = Z'Z$, where Z has the distribution in (5.18). Well, it happens that from its definition S is forcedly non-negative-definite, i.e., its eigenvalues will be non-negative. To show that S is positive-definite, if and only if $n-1 \geq p$, all we have to do is to show that S is non-singular if and only if $n-1 \geq p$. More precisely, in this case it is possible to show that S is non-singular, with probability 1. Let us denote by $\underline{z}_1, ..., \underline{z}_p$ the p columns of Z. Then

$$\begin{aligned} P\left(S \text{ being non-singular}\right) &= 1 - P\left(S \text{ being singular}\right) \\ &= 1 - P\left(\underline{z}_1, ..., \underline{z}_p \text{ being linearly dependent}\right) \end{aligned}$$

where

$$\begin{aligned} &P\left(\underline{z}_1, ..., \underline{z}_p \text{ being linearly dependent}\right) \leq \\ &\qquad \leq \textstyle\sum_{j=1}^{p} P\left(\underline{z}_j \text{ to be a linear combination of } \underline{z}_1, ..., \underline{z}_{j-1}, \underline{z}_{j+1}, ..., \underline{z}_p\right) \end{aligned}$$

with

$$\begin{aligned} &P\left(\underline{z}_j \text{ to be a linear combination of } \underline{z}_1, ..., \underline{z}_{j-1}, \underline{z}_{j+1}, ..., \underline{z}_p\right) \\ &\quad = E\Big[P\Big(\underline{z}_j \text{ to be a linear combination of } \underline{z}_1, ..., \underline{z}_{j-1}, \underline{z}_{j+1}, ..., \underline{z}_p \\ &\qquad\qquad\qquad\qquad \mid \underline{z}_1, ..., \underline{z}_{j-1}, \underline{z}_{j+1}, ..., \underline{z}_p\Big)\Big] \end{aligned}$$

where

$$P\big(\underline{z}_j \text{ to be a lin. comb. of } \underline{z}_1, ..., \underline{z}_{j-1}, \underline{z}_{j+1}, ..., \underline{z}_p | \underline{z}_1, ..., \underline{z}_{j-1}, \underline{z}_{j+1}, ..., \underline{z}_p\big) = 0$$

if and only if $n-1 \geq p$, given that $Z_{(n-1)\times p} \sim N_{(n-1)p}(\underline{0}_{(n-1)\times p}, \Sigma \otimes I_{n-1})$ where it is assumed that $\Sigma > 0$ (since if $n-1 < p$ then $rank(Z) = n-1 < p$, and thus the condition $n-1 \geq p$ is then a necessary and sufficient condition for the independence of the columns of Z).

We will thus have

$$\sum_{j=1}^{p} P\left(\underline{z}_j \text{ to be a linear combination of } \underline{z}_1, ..., \underline{z}_{j-1}, \underline{z}_{j+1}, ..., \underline{z}_p\right) \;=\; 0$$

if and only if $n-1 \geq p$, and we will then have

$$P\left(S \text{ to be non-singular}\right) \;\geq\; 1-0 \quad\Longrightarrow\quad P\left(S \text{ to be non-singular}\right) \;=\; 1$$

if and only if $n-1 \geq p$.

Since A, defined in (5.1), may be written as $A = \widetilde{X}'\widetilde{X}$ where $\widetilde{X} \sim N_{np}(\underline{0}_{n\times p}, \Sigma \otimes I_n)$, a proof that follows similar lines to the one above would give A as positive-definite if and only if $n \geq p$.

Bibliografia

Dykstra, R. L. (1970). Establishing the positive definiteness of the sample covariance matrix. *Ann. Math. Statist.*, **41**, 2153-2154.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* J. Wiley & Sons. New York. (Cap. 3 pp. 82-83.)

Chapter 6

Tests for mean vectors based on Hotelling's T^2 statistic

6.1 On the construction of a Hotelling T^2 statistic and its distribution

The results introduced next constitute not only an excellent example of a parallel between a well-known result from the univariate statistic and one coming out of the multivariate statistic as well as they will enable us to test several hypotheses of interest in multivariate statistics.

Let

$$\underline{X} \sim N_p(\underline{\mu}, \Sigma) \quad \text{and} \quad S \sim W_p(f, \Sigma) \tag{6.1}$$

be two independente r.v.'s, with $f > p-1$. Then

$$T^2 = \underline{X}' \left(\frac{1}{f} S\right)^{-1} \underline{X} = f\, \underline{X}' S^{-1} \underline{X} \tag{6.2}$$

is a Hotelling T^2 statistic, with

$$\frac{f-p+1}{p} \frac{T^2}{f} \sim F_{p,f-p+1}(\delta) \tag{6.3}$$

where $\delta = \underline{\mu}' \Sigma^{-1} \underline{\mu}$.

As such, the statistic T^2 in (6.2) will be an adequate statistic for the test of the following hypotheses

$$\begin{array}{c} H_0 \colon \underline{\mu} = \underline{0} \\ \text{vs.} \\ H_1 \colon \underline{\mu} \neq \underline{0} \end{array} \tag{6.4}$$

since under H_0 we have $\underline{X} \sim N_p(\underline{0}, \Sigma)$ and consequently we also have

$$\frac{f-p+1}{p} \frac{T^2}{f} \sim F_{p,f-p+1} \,,$$

with

$$E(T^2) = \frac{(f-p+1)^2}{fp(f-p-1)}$$

while under H_1 we have the distribution in (6.3), with

$$E(T^2) = \frac{(f-p+1)^2}{fp(f-p-1)} \frac{p+\underline{\mu}'\Sigma^{-1}\underline{\mu}}{p} \,,$$

so that an α level test for the hypotheses in (6.4) will reject H_0 if

$$\frac{f-p+1}{p} \frac{T^2_{calc}}{f} > f_{p,f-p+1}(1-\alpha) \,. \tag{6.5}$$

The deduction of the distribution in (6.3) may be easily carried out by writing

$$\frac{T^2}{f} = \underline{X}'S^{-1}\underline{X} = \frac{\underline{X}'\Sigma^{-1}\underline{X}}{\frac{\underline{X}'\Sigma^{-1}\underline{X}}{\underline{X}'S^{-1}\underline{X}}} \tag{6.6}$$

where, from property VI) in subsection 5.5, we know that

$$\frac{\underline{X}'\Sigma^{-1}\underline{X}}{\underline{X}'S^{-1}\underline{X}} \sim \chi^2_{f-p+1} \,,$$

independent of $\underline{X}$, while from property IX) in subsection 2.2, we have

$$\underline{X}'\Sigma^{-1}\underline{X} \sim \chi^2_p(\delta) \tag{6.7}$$

with

$$\delta = \underline{\mu}'\Sigma^{-1}\underline{\mu} \,.$$

But then since the only r.v. that enters in the construction of the r.v. in (6.7) is $\underline{X}$, and since the r.v. in the denominator of (6.6) is independent of $\underline{X}$, the r.v.'s in the numerator and denominator of (6.6) are independent and then

$$\frac{\underline{X}'\Sigma^{-1}\underline{X}\,/\,p}{\dfrac{\underline{X}'\Sigma^{-1}\underline{X}}{\underline{X}'S^{-1}\underline{X}}\Big/(f-p+1)} = \frac{f-p+1}{p}\,\frac{T^2}{f} \sim F_{p,f-p+1}(\delta)\,.$$

If we want to test the hypotheses

$$\begin{array}{c} H_0\colon \underline{\mu} = \underline{\mu}_0 \\ \text{vs.} \\ H_1\colon \underline{\mu} \neq \underline{\mu}_0 \end{array}$$

where $\underline{\mu}_0$ is a given vector in $I\!R^p$, we will then use the test statistic

$$T^2 = f\left(\underline{X}-\underline{\mu}_o\right)' S^{-1}\left(\underline{X}-\underline{\mu}_0\right),$$

which, under H_0, will have an $F_{p,f-p+1}$ distribution, and we will reject H_0 if (6.5) holds.

6.1.1 Confidence Intervals and confidence ellipsoids

A confidence ellipsoid (or confidence region) for $\underline{\mu}$, corresponding to a probability of $1-\alpha$ will then be given by

$$f(\underline{X}-\underline{\mu})'S^{-1}(\underline{X}-\underline{\mu}) \leq T^2_{1-\alpha} = \frac{fp}{f-p+1}\,f_{p,f-p+1}(1-\alpha)$$

and, since for $\underline{u} \in \mathbb{R}^p$ (with $\underline{u} \neq \underline{0}_{p\times 1}$) we have

$$\underline{u}'\underline{X} \sim N(\underline{u}'\underline{\mu}, \underline{u}'\Sigma\underline{u})\,,$$

a Confidence Interval for

$$\underline{u}'\underline{\mu} = \sum_{j=1}^{p} u_j\mu_j\,,$$

linear combination of the mean values μ_j ($j = 1,\ldots,p$), corresponding to a probability of $1-\alpha$, will be given by

$$\left[\underline{u}'\underline{X} - t_f(1-\alpha/2)\sqrt{\frac{\underline{u}'S\underline{u}}{f}},\ \underline{u}'\underline{X} + t_f(1-\alpha/2)\sqrt{\frac{\underline{u}'S\underline{u}}{f}}\right]. \tag{6.8}$$

In fact, from (6.1), we have

$$\underline{u}'\underline{X} \sim N(\underline{u}'\underline{\mu}, \underline{u}'\Sigma\underline{u})$$

or

$$\frac{\underline{u}'\underline{X} - \underline{u}'\underline{\mu}}{\sqrt{\underline{u}'\Sigma\underline{u}}} \sim N(0,1)\,,$$

with the unbiased estimator of $\underline{u}'\Sigma\underline{u}$ being $\frac{1}{f}\underline{u}'S\underline{u}$, since from (6.1) we have, through (5.5), $E(S) = f\Sigma$, where, by property VI) in section 5.5,

$$\frac{\underline{u}'S\underline{u}}{\underline{u}'\Sigma\underline{u}} \sim \chi^2_f\,,$$

independent of $\underline{X}$, through the independence of S and $\underline{X}$, so that

$$T = \frac{\frac{\underline{u}'\underline{X}-\underline{u}'\underline{\mu}}{\sqrt{\underline{u}'\Sigma\underline{u}}}}{\sqrt{\frac{\underline{u}'S\underline{u}}{\underline{u}'\Sigma\underline{u}}\Big/ f}} = \frac{\underline{u}'\underline{X} - \underline{u}'\underline{\mu}}{\sqrt{\frac{\underline{u}'S\underline{u}}{f}}} \sim T_f\,,$$

from which we obtain then (6.8).

6.1.2 Confidence intervals with coverage equal to $1-\alpha$

Although each Confidence Interval (CI) in (6.8) has by itself a coverage with probability $1-\alpha$, we are indeed more interested in obtaining CI's with an overall coverage of $1-\alpha$ for any linear combination $\underline{u}'\underline{\mu}$ and not only for a given linear combination. How should we then define a CI for linear combinations of the μ_j's $(j = 1, \ldots, p)$ so that their global probability of coverage will be exactly equal to $1-\alpha$ for all and any $\underline{u} \in \mathbb{R}^p$?

These CI's should be built on the non-rejection region of H_0, using the "union-intersection principle" (see page 59 for a better explanation of this "principle"), that is, based on the expression

$$\begin{aligned}
1-\alpha &= P\left[\left(\frac{\underline{u}'(\underline{X}-\underline{\mu})}{\sqrt{\frac{\underline{u}'S\underline{u}}{f}}}\right)^2 \le T^2_{1-\alpha}\right] \\
&= P\left[\left(\underline{u}'(\underline{X}-\underline{\mu})\right)^2 \le \frac{1}{f}\,T^2_{1-\alpha}\,\underline{u}'S\underline{u}\right] \\
&= P\left[\underline{u}'\underline{X} - \sqrt{\tfrac{1}{f}\,\underline{u}'S\underline{u}\,T^2_{1-\alpha}} \le \underline{u}'\underline{\mu} \le \underline{u}'\underline{X} + \sqrt{\tfrac{1}{f}\,\underline{u}'S\underline{u}\,T^2_{1-\alpha}}\,\right],
\end{aligned}$$

being this way the set of CI's for a linear combination of the μ_j's of the form $\underline{u}'\underline{\mu}$, with global coverage of $1-\alpha$ given by

$$\left[\underline{u}'\underline{X} - \sqrt{\frac{1}{f}\,\underline{u}'S\underline{u}\,T^2_{1-\alpha}},\ \underline{u}'\underline{X} + \sqrt{\frac{1}{f}\,\underline{u}'S\underline{u}\,T^2_{1-\alpha}} \right],$$

where

$$T^2_{1-\alpha} = \frac{fp}{f-p+1}\, f_{p,f-p+1}(1-\alpha)\,.$$

6.2 A test of hypotheses on $\underline{\mu}$, the vector of expected values of a multivariate Normal vector

Let $\overline{\underline{X}}$ and S^* be respectively the sample mean vector and sample covariance matrix for a sample of size n from a multivariate Normal distribution $N_p(\underline{\mu},\ \Sigma)$, i.e., let $\overline{\underline{X}}$ and S be respectively the vector defined in (4.7) and the matrix of sums of squares and products of the differences to the sample mean defined in (4.8) or (5.6) and let

$$S^* \;=\; \frac{1}{n-1}\, S\,,$$

with

$$\sqrt{n}\,\overline{\underline{X}} \sim N_p(\sqrt{n}\,\underline{\mu}, \Sigma)\,, \qquad S \;\sim\; W_p(n-1, \Sigma)\,.$$

Let then, according to 6.1,

$$T^2 = (n-1)\sqrt{n}\,\overline{\underline{X}}'S^{-1}\sqrt{n}\,\overline{\underline{X}} = n\,(n-1)\,\overline{\underline{X}}'S^{-1}\overline{\underline{X}} = n\,\overline{\underline{X}}'S^{*-1}\overline{\underline{X}}\,,$$

where, according to the exposition in the previous section,

$$\frac{n-p}{p}\,\frac{T^2}{n-1} \;\sim\; F_{p,n-p}(\delta)$$

with

$$\delta \;=\; n\,\underline{\mu}'\Sigma^{-1}\underline{\mu}\;.$$

This way, a test with dimension α for the hypotheses

$$\begin{array}{c} H_0:\ \underline{\mu} = \underline{0} \\ \text{vs} \\ H_1:\ \underline{\mu} \neq \underline{0} \end{array}$$

will be given by the rejection of H_0 if

$$\frac{n-p}{p}\,\frac{T^2_{\text{calc}}}{n-1} \;>\; f_{p,n-p}(1-\alpha)\ ,$$

where $f_{p,n-p}(1-\alpha)$ denots the $1-\alpha$ quantile of the $F_{p,n-p}$ (central) distribution.

To test the hypotheses

$$\begin{array}{rl} & H_0:\ \underline{\mu}=\underline{\mu}_0 \\ \text{vs} & \\ & H_1:\ \underline{\mu}\neq\underline{\mu}_0 \end{array}$$

we will use the statistic

$$T^2 \;=\; n(n-1)\left(\underline{\overline{X}}-\underline{\mu}_0\right)' S^{-1}\left(\underline{\overline{X}}-\underline{\mu}_0\right) \tag{6.9}$$

where under H_0,

$$\frac{n-p}{p}\,\frac{T^2}{n-1} \;\sim\; F_{p,n-p}\,.$$

We will then reject H_0 if

$$\frac{n-p}{p}\,\frac{T^2_{calc}}{n-1} > f_{p,n-p}(1-\alpha)\,.$$

Note, for $p=1$, the analogy with the univariate case, since for $p=1$ the T^2 statistic is exactly the square of the statistic T commonly used to test $H_0:\mu=\mu_0$.

– Exempl of application

Let $p=3$ and let

X_1	–	height of the tree (m)
X_2	–	diameter of the tree (cm)
X_3	–	age of the tree (in years)

the variables from which a sample of size $n=12$ was obtained from a stand of pine trees.

We will assume that

$$\underline{X} = [X_1,\, X_2,\, X_3]' \sim N\left(\underline{\mu},\, \Sigma\right) .$$

We want to test the hypothesis

$$\begin{array}{ll} & H_0 : \underline{\mu} = [20 \;\; 28 \;\; 16]' \\ \text{vs} & \\ & H_1 : \underline{\mu} \neq [20 \;\; 28 \;\; 16]' . \end{array}$$

The sample of size $n = 12$ from X_1, X_2, X_3 has the following values

$$\mathcal{X} = \begin{bmatrix} 21.5 & 25 & 15 \\ 22.6 & 28 & 18 \\ 17.4 & 22 & 17 \\ 18.6 & 21 & 16 \\ 25.6 & 32 & 18 \\ 21.3 & 26 & 17 \\ 18.4 & 20 & 15 \\ 15.4 & 17 & 15 \\ 19.5 & 19 & 17 \\ 20.6 & 22 & 16 \\ 23.4 & 25 & 16 \\ 18.9 & 26 & 17 \end{bmatrix}$$

with

$$\underline{\overline{x}} = [20.2667,\; 23.5833,\; 16.4167]' .$$

Let S be the matrix of sums of squares and products of differences to the sample mean $\underline{\overline{X}}$. Then

$$\mathcal{S} = (\mathcal{X} - E_{1n}\underline{\overline{x}}')'\,(\mathcal{X} - E_{1n}\underline{\overline{x}}')$$

with

$$\mathcal{S} = \begin{bmatrix} 87.0267 & 111.1333 & 16.8667 \\ 111.1333 & 194.9167 & 33.0833 \\ 16.8667 & 33.0833 & 12.9167 \end{bmatrix}$$

and

$$T^2 = 92.8069 .$$

Using the software R, we may program the functions

```
centr<- function(x)
 as.matrix(x)-matrix(1,dim(x)[1],1)%*%
                                t(apply(as.matrix(x),2,mean))

makes<- function(x)
 t(centr(x))%*%centr(x)

maket<- function(x,vec)
 (n<-dim(x)[1])*(n-1)*t(dif<-apply(x,2,mean)-vec)%*%
                                         solve(makes(x))%*%dif
```

or, instead of the function `makes` the function `make2s`

```
make2s<- function(x)
 t(m<-centr(x))%*%m
```

or even the function `make3s`

```
make3s<- function(x)
 (dim(x)[1]-1)*cov(x)
```

which avoids the use of the function `centr`.

After storing the data in a matrix called `matx`, we could use a command like `maket(matx,c(20,28,16))` to obtain the computed value of T^2, to obtain then

$$F_{\text{calc}} = \frac{n-p}{p}\frac{T^2}{n-1} = \frac{9}{3}\frac{92.8069}{11} = 25.311\,,$$

in such a way that for $\alpha = .05$, we have

$$F_{\text{calc}} = 25.311 \; > \; f_{p,n-p}(1-\alpha) = f_{3,9}(.95) = 3.86\,,$$

what would lead us to reject the null hypothesis H_0.

An example of an R function that would give us directly the computed value of the F statistic would be

```
makef<-function(x,vec)
{t<-maket(x,vec)
 np<-dim(x)
 (np[1]-np[2])/np[2]*t/(np[1]-1)}
```

being left as an exercise the construction of another version that would be progammed using a single line of code.

❏❏❏❏❏

Exercises:

6.1. Program an R function which once given the data matrix and the vector $\underline{\mu}_0$, gives as a result the computed value of the statistic T^2 and the corresponding p-value.

6.2. Using the first 11 elements of the random sample being considered, test the hypotheses

$$H_0 : \underline{\mu} = [20 \quad 25 \quad 16]$$

vs

$$H_1 : \underline{\mu} \neq [20 \quad 25 \quad 16]$$

for $\alpha = .01$ and $\alpha = .05$.

6.3. Obtain data sets and implement several tests for expected value vectors on them.

❏❏❏❏❏

6.2.1 Confidence ellipsoid for $\underline{\mu}$ and Confidence Intervals for $\underline{u}'\underline{\mu}$

From the exposition in 6.1.1, a Confidence Ellipsoid for $\underline{\mu}$, corresponding to a probability (of coverage) of $1-\alpha$, will then be given by

$$n(\underline{\overline{X}}-\underline{\mu})'S^{*-1}(\underline{\overline{X}}-\underline{\mu}) \leq \frac{(n-1)p}{n-p}\, f_{p,n-p}(1-\alpha)\,.$$

Note that while the common bilateral Confidence Interval for the expected value of a r.v. with a Normal distribution, used in univariate statistics, is centered on the sample mean, this Confidence Ellipsoid is centered on $\underline{\overline{X}}$.

In Figure 6.1 we have the graphical representation of a set of points $\underline{\mu}$ that fall inside an ellipsoid of this type, for $\alpha = 0.05$ and for a situation where we have a sample of size $n = 10$,

$$\underline{\overline{x}} = [1.3, 1.4, 1.2]' \quad \text{and} \quad S^* = \begin{bmatrix} 5.073 & 0.932 & -1.768 \\ 0.932 & 2.535 & 0.232 \\ -1.768 & 0.232 & 1.566 \end{bmatrix}. \qquad (6.10)$$

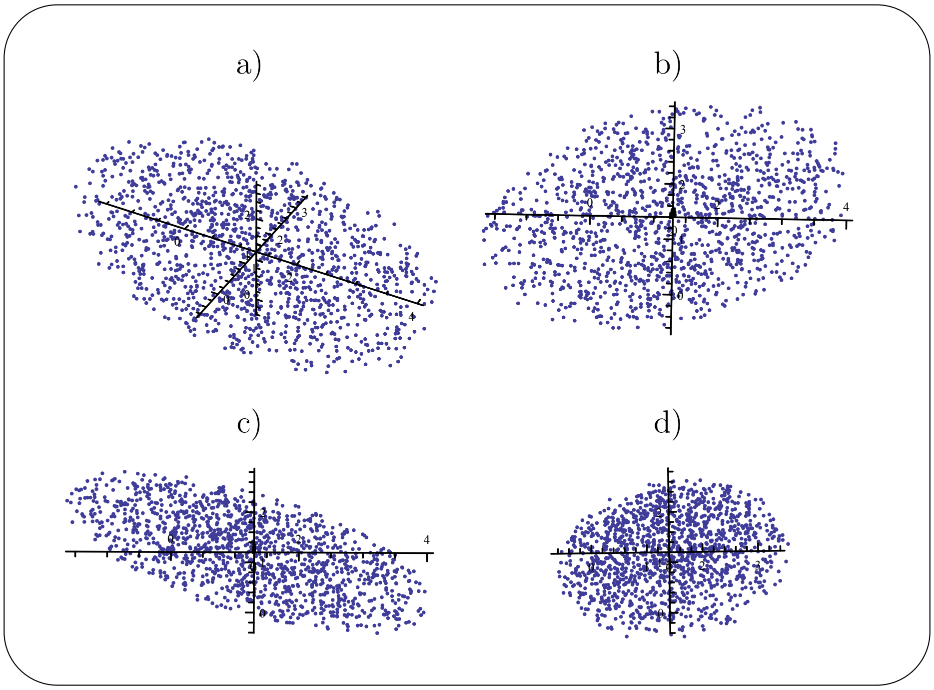

Figure 6.1 – a) Three dimensional graphical representation of a set of points that lie inside the Confidence Ellipsoid for $\alpha = 0.05$ and for a situation where we have $n = 10$ and $\underline{\overline{x}}$ and S^* given in (6.10):
b) frontal perspective of the plan $X_1 - X_2$, c) frontal perspective of the plan $X_1 - X_3$,
d) frontal perspective frontal of the plan $X_2 - X_3$.

In Figure 6.2 we have the graphical representations of a set of points that lie inside the Confidence Ellipsoid for $\alpha = 0.05$ corresponding to the data in the Exercise proposed in section 6.2.

A Confidence Interval for $\underline{u}'\underline{\mu}$ ($\underline{u} \in \mathbb{R}^p$), corresponding to a probability (of coverage) of $1 - \alpha$, will be given by

$$\left[\underline{u}'\underline{\overline{X}} - t_{n-1}(1-\alpha/2)\sqrt{\frac{\underline{u}'S\underline{u}}{n(n-1)}},\ \underline{u}'\underline{\overline{X}} + t_{n-1}(1-\alpha/2)\sqrt{\frac{\underline{u}'S\underline{u}}{n(n-1)}}\right]$$

or

$$\left[\underline{u}'\underline{\overline{X}} - t_{n-1}(1-\alpha/2)\sqrt{\frac{\underline{u}'S^*\underline{u}}{n}},\ \underline{u}'\underline{\overline{X}} + t_{n-1}(1-\alpha/2)\sqrt{\frac{\underline{u}'S^*\underline{u}}{n}}\right].$$

Note the relation with the univariate case.

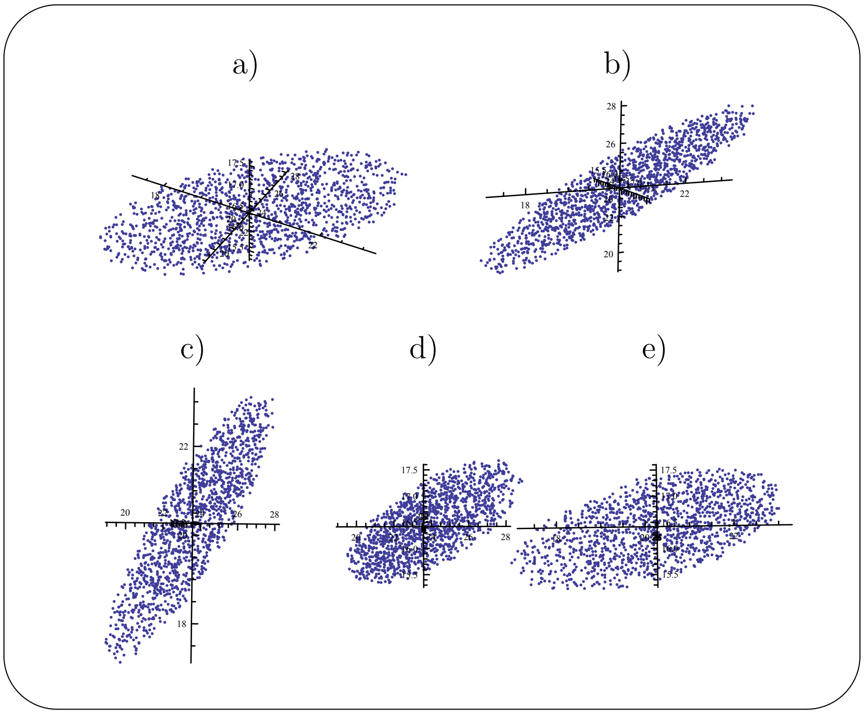

Figure 6.2 – Three dimensional graphical representations of a set of points that lie inside the Confidence Ellipsoid for $\alpha = 0.05$ and for the data in the Exercise proposed in this section:
a) and b) two different perspectives of the set of points,
c), d) and e) frontal perspectives respectively of plans $X_1 - X_2$, $X_1 - X_3$, and $X_2 - X_3$.

While CI's that show an overall coverage probability of $1 - \alpha$ for any linear combination $\underline{u}'\underline{\mu}$ will be given by

$$\left[\underline{u}'\,\underline{\overline{X}} - \sqrt{\frac{\underline{u}'S\underline{u}}{n}\,\frac{p}{n-p}\,f_{p,n-p}(1-\alpha)},\ \underline{u}'\,\underline{\overline{X}} + \sqrt{\frac{\underline{u}'S\underline{u}}{n}\,\frac{p}{n-p}\,f_{p,n-p}(1-\alpha)} \right].$$

6.2.2 An alternative form of building the T^2 statistic for this test

When we want to test

$$H_0 : \underline{\mu} = \underline{\mu}_0 \quad \text{vs.} \quad H_1 : \underline{\mu} \neq \underline{\mu}_0 \tag{6.11}$$

the construction of the Hotelling T^2 statistic may be done using the so-called "union-itersection principle" (Roy, 1953, 1957), what gives us another look over this test.

Once we assume that

$$\underline{X} \sim N_p(\underline{\mu}, \Sigma)\,,$$

then if $\underline{a} \in \mathbb{R}^p$ for any given vector of real values, we have

$$\underline{a}'\underline{X} \sim N(\underline{a}'\underline{\mu}, \underline{a}'\Sigma\underline{a})\,,$$

so that we may test, for a given $\underline{a} \in \mathbb{R}^p$ the hypothesis

$$H_0 : \underline{a}'\underline{\mu} = \underline{a}'\underline{\mu}_0 \quad \text{vs} \quad H_1 : \underline{a}'\underline{\mu} \neq \underline{a}'\underline{\mu}_0\,, \tag{6.12}$$

based on a sample of dimension n from $\underline{X}$, simply by using the Student T statistic

$$T_{\underline{a}} = \frac{\underline{a}'\overline{\underline{X}} - \underline{a}'\underline{\mu}_0}{\sqrt{\underline{a}'S^*\underline{a}/n}} \sim T_{n-1}$$

where S^* is the sample variance-covariance matrix, not rejecting H_0 in (6.12) if $|T_{\underline{a}\,(calc)}| \leq t_{n-1}(1-\alpha/2)$ or $T^2_{\underline{a}\,(calc)} \leq \left(t_{n-1}(1-\alpha/2)\right)^2 = f_{1,n-1}(1-\alpha)$.

However, the null hypothesis in (6.11) holds if and only if the null hypothesis in (6.12) holds for all $\underline{a} \in \mathbb{R}^p$ (see the characterization of the multivariate Normal distribution in X) of section 2.2), so that we will not reject H_0 in (6.11) if we do not reject H_0 in (6.12) for all $\underline{a} \in \mathbb{R}^p$, which is equivalent to say that we will not reject H_0 in (6.11) if

$$\max_{\underline{a}\in\mathbb{R}^p} T_{\underline{a}\,(calc)} \leq t_{n-1}(1-\alpha/2)$$

or, more precisely, in order to keep the α level of the test, if

$$\max_{\underline{a}\in\mathbb{R}^p} T_{\underline{a}\,(calc)} \leq t^*(1-\alpha/2)$$

where $t^*(1-\alpha/2)$ represents the $1-\alpha/2$ quantile of $\max_{\,\underline{a}\in\mathbb{R}^p} T_{\underline{a}}$, or equivalently, if

$$\max_{\underline{a}\in\mathbb{R}^p} T^2_{\underline{a}\,(calc)} \leq \left(t^*(1-\alpha/2)\right)^2 \Big(= f_{p,n-p}(1-\alpha)\Big)\,,$$

thus being the non-rejection zone of H_0 in (6.11) the intersection, for all $\underline{a} \in \mathbb{R}^p$ of the non-rejection zones of each null hypothesis in (6.12), and thus being the rejection zone of H_0 in (6.11) the union of the rejection zones of H_0 in (6.12), which originates the name of "union-intersection principle".

Therefore, we should adopt as test statistic for the test to H_0 in (6.11) the statistic

$$T^2 = \max_{\underline{a} \in \mathbb{R}^p} T^2_{\underline{a}}$$

where $T^2_{\underline{a}}$ is not affected by any changes of scale in $\underline{a}$, what introduces an indetermination when we will try to determine $T^2 = \max_{\underline{a} \in \mathbb{R}^p} T^2_{\underline{a}}$. In order to remove this indetermination we will only consider vectors $\underline{a}$ such that $\underline{a}' S^* \underline{a} = 1$. Then, in order to obtain $\max_{\underline{a} \in \mathbb{R}^p} T^2_{\underline{a}}$ we will build, under the restriction $\underline{a}' S^* \underline{a} = 1$, the Lagrangean

$$\mathcal{L} = n\, \underline{a}'(\overline{\underline{X}} - \underline{\mu}_0)(\overline{\underline{X}} - \underline{\mu}_0)'\underline{a} - \lambda(\underline{a}' S^* \underline{a} - 1)$$

with

$$\frac{\partial \mathcal{L}}{\partial \underline{a}} = 2n\,(\overline{\underline{X}} - \underline{\mu}_0)(\overline{\underline{X}} - \underline{\mu}_0)'\underline{a} - 2\lambda\, S^* \underline{a}$$

which equating to zero and pre-multiplying by S^{*-1} gives

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \underline{a}} = 0 \quad &\Longleftrightarrow \quad n\,(\overline{\underline{X}} - \underline{\mu}_0)(\overline{\underline{X}} - \underline{\mu}_0)'\underline{a} = \lambda\, S^* \underline{a} \\ &\Longleftrightarrow \quad n\, S^{*-1}(\overline{\underline{X}} - \underline{\mu}_0)(\overline{\underline{X}} - \underline{\mu}_0)'\underline{a} = \lambda\, \underline{a} \end{aligned}$$

that is, λ is the only non-null eigenvalue of $n\, S^{*-1}(\overline{\underline{X}} - \underline{\mu}_0)(\overline{\underline{X}} - \underline{\mu}_0)'$, i.e.,

$$\begin{aligned} \lambda &= tr\left[n\, S^{*-1}(\overline{\underline{X}} - \underline{\mu}_0)(\overline{\underline{X}} - \underline{\mu}_0)'\right] \\ &= n\,(\overline{\underline{X}} - \underline{\mu}_0)' S^{*-1}(\overline{\underline{X}} - \underline{\mu}_0) \end{aligned}$$

which is the statistic in (6.9).

6.3 Test to the difference of two mean vectors based on two independent samples

Let us assume that

$$\underline{X}_1 \sim N_p\left(\underline{\mu}_1, \Sigma_1\right)$$

and

$$\underline{X}_2 \sim N_p\left(\underline{\mu}_2, \Sigma_2\right),$$

with

$$\Sigma_1 = \Sigma_2 = \Sigma,$$

and let us suppose that we want to test

$$\begin{array}{ll} & H_0 : \underline{\mu}_1 = \underline{\mu}_2 \\ \text{vs} & \\ & H_1 : \underline{\mu}_1 \neq \underline{\mu}_2 \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} & H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{0} \\ \text{vs} & \\ & H_1 : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{0} \end{array}$$

or, more generally,

$$\begin{array}{ll} & H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{a} \\ \text{vs} & \\ & H_1 : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{a} \end{array} \qquad (\underline{a} \in I\!\!R^p)$$

based on two independent random samples, with sizes n_1 and n_2 respectively.

Then

$$\overline{\underline{X}}_1 \sim N_p\left(\underline{\mu}_1, \frac{1}{n_1}\Sigma\right) \qquad \text{(based on a sample of size } n_1 \text{ from } \underline{X}_1\text{)}$$

and

$$\overline{\underline{X}}_2 \sim N_p\left(\underline{\mu}_2, \frac{1}{n_2}\Sigma\right) \qquad \text{(based on a sample of size } n_2 \text{ from } \underline{X}_2\text{)}$$

will be independent (given the independence of the two samples).

But then we have

$$(\overline{\underline{X}}_1 - \overline{\underline{X}}_2) \sim N_p\left(\underline{\mu}_1 - \underline{\mu}_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\Sigma\right).$$

Let then $S_{\underline{X}_1}$ and $S_{\underline{X}_2}$ be the sample matrices of sums of squares and products of deviations from the sample means, respectively for the samples from $\underline{X}_1$ and $\underline{X}_2$, with

$$\begin{array}{l} S_{\underline{X}_1} \sim W_p(n_1 - 1, \Sigma), \\ S_{\underline{X}_2} \sim W_p(n_2 - 1, \Sigma), \end{array} \qquad \text{(independent)}$$

and let

$$S = S_{\underline{X}_1} + S_{\underline{X}_2} ,$$

with

$$S \sim W_p\left(n_1+n_2-2,\, \Sigma\right) .$$

THen,

$$\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2 - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right) \sim N_p\left(\underline{0},\, \frac{n_1+n_2}{n_1\, n_2}\Sigma\right)$$

that is,

$$\sqrt{\frac{n_1\, n_2}{n_1+n_2}}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2 - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right) \sim N_p\left(\underline{0},\, \Sigma\right) ,$$

independent of S (why?).

We will then have

$$\begin{aligned} T^2 &= \left(\sqrt{\frac{n_1\, n_2}{n_1+n_2}}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2 - \left.\left(\underline{\mu}_1 - \underline{\mu}_2\right)\right|_{H_0}\right)\right)' \left(\frac{1}{n_1+n_2-2}S\right)^{-1} \\ &\qquad \left(\sqrt{\frac{n_1\, n_2}{n_1+n_2}}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2 - \left.\left(\underline{\mu}_1 - \underline{\mu}_2\right)\right|_{H_0}\right)\right) \\ &= \frac{n_1\, n_2}{n_1+n_2}\left(n_1+n_2-2\right)\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2 - \underline{a}\right)' S^{-1}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2 - \underline{a}\right) \end{aligned} \tag{6.13}$$

with

$$\frac{n_1+n_2-2-p+1}{p}\,\frac{T^2}{n_1+n_2-2} = \frac{n_1+n_2-p-1}{p}\,\frac{T^2}{n_1+n_2-2} \sim F_{p,n_1+n_2-p-1} .$$

Example of application:

Let us suppose that two independent random samples were obtained, one with dimension $n_1 = 7$ and the other with dimension $n_2 = 5$ from two stands of pine trees, relative to the variables X_1, X_2 and X_3 referred to before. We want to test if it is plausible the hypothesis that the two stands have the same mean height, diameter and age. That is, we want to test

$$\begin{array}{c} H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{0} \\ \text{vs} \\ H_1 : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{0} \end{array}$$

where

$$\underline{\mu}_1 = [\mu_{11},\, \mu_{12},\, \mu_{13}]'$$

are the population means of X_1, X_2 and X_3 in stand 1 and

$$\underline{\mu}_2 = [\mu_{21},\, \mu_{22},\, \mu_{23}]'$$

are the populationa means of X_1, X_2 and X_3 in stand 2.

Let us suppose that we will use the data shown in the previous section, taking the first 7 observations as pertaining to stand 1 and the last 5 as pertaining to stand 2.

We may program an R function to compute the statistic T^2 necessary to the test, or we may try to use simple R directives to compute the value of the T^2 statistic for this case. Both options are left as an exercise.

❏❏❏❏❏

Exercises:

6.4. Program an R function which once given the two data matrices and the vector $\underline{a} \in I\!R^p$, will give as a result the computed value of the statistic T^2 and the corresponding p-value.

6.5. Use the first 6 and the last 4 observations in the data matrix shown in the previous section as constituting two independent random samples to test the hypotheses

$$\begin{array}{cl} & H_0 : \underline{\mu}_1 - \underline{\mu}_2 = [1 \;\; 5 \;\; 0] \\ \text{vs} & \\ & H_1 : \underline{\mu}_1 - \underline{\mu}_2 \neq [1 \;\; 5 \;\; 0]\ . \end{array}$$

6.6. Show that if there is one only variable (i.e. if $p = 1$) then $\sqrt{T^2}$ has a Student T distribution, and that it is the common Student T statistic used to test the equality of two means.

6.7. Obtain the expressions for the CI's for linear combinations of the type $\underline{u}'(\underline{\mu}_1 - \underline{\mu}_2)$.

6.8. Obtain several data sets and implement on them tests of the type $H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{a}$.

❏❏❏❏❏

6.4 More general tests involving T^2.

6.4.1 Test for $H_0 : C\underline{\mu} = \underline{a}$ $(\underline{a} \in I\!R^k,\ k \leq p)$

Let us suppose that we want to test

$$\begin{array}{ll} & H_0 : C\underline{\mu} \,=\, \underline{a} \\ \text{vs} & \\ & H_1 : C\underline{\mu} \,\neq\, \underline{a}\,, \end{array} \qquad\qquad \left(\underline{a} \in I\!R^k\right)$$

where C is a real matrix of dimensions $k{\times}p$, with $rank(C) = k \leq p$.

Since, if

$$\underline{\overline{X}} \,\sim\, N_p\left(\underline{\mu},\, \frac{1}{n}\Sigma\right)$$

then

$$C\underline{\overline{X}} \,\sim\, N_k\left(C\underline{\mu},\, \frac{1}{n}C\Sigma C'\right)$$

and

$$C\left(\underline{\overline{X}} - \underline{\mu}\right) \,\sim\, N_k\left(\underline{0},\, \frac{1}{n}C\Sigma C'\right)\,,$$

$$\sqrt{n}C\left(\underline{\overline{X}} - \underline{\mu}\right) \,\sim\, N_k\left(\underline{0},\, C\Sigma C'\right)\,.$$

Since then

$$S \,\sim\, W_p\left(n-1,\, \Sigma\right)\,,$$

and thus

$$CSC' \,\sim\, W_k\left(n-1,\, C\Sigma C'\right)\,,$$

are independent of $C\left(\underline{\overline{X}} - \underline{\mu}\right)$,

$$\begin{aligned} T^2 \;&=\; \left(\sqrt{n}\,C\left(\underline{\overline{X}} - \underline{\mu}\right)\right)'\left(\frac{1}{n-1}\,CSC'\right)^{-1}\left(\sqrt{n}\,C\left(\underline{\overline{X}} - \underline{\mu}\right)\right) \\ &=\; n\,(n-1)\,\left(\underline{\overline{X}} - \underline{\mu}\right)'C'\left(CSC'\right)^{-1}C\left(\underline{\overline{X}} - \underline{\mu}\right) \end{aligned}$$

or rather,

$$T^2 \,=\, n\,(n-1)\left(C\underline{\overline{X}} - C\underline{\mu}\right)'\left(CSC'\right)^{-1}\left(C\underline{\overline{X}} - C\underline{\mu}\right)$$

where, under $H_0 : C\underline{\mu} = \underline{a}$, T^2 has a Hotelling T^2 distribution, with

$$\frac{n-k}{k}\,\frac{T^2}{n-1} \,\sim\, F_{k,n-k}\,.$$

❑❑❑❑❑

Exercises:

6.9. Program an R function that implements the test in this subsection and which will give as a result the computed value of the statistic T^2 and the corresponding p-value.

6.10. Obtain data sets and implement the test in this subsection on them.

❑❑❑❑❑

6.4.2 Test for $H_0 : C\underline{\mu}_1 - C\underline{\mu}_2 = \underline{a}$ $\;(\underline{a} \in I\!R^k,\ k \leq p)$

If we want to test

$$\begin{array}{ll} & H_0 :\ C\underline{\mu}_1 - C\underline{\mu}_2 \ = \ \underline{a} \\ \text{vs} & \\ & H_1 :\ C\underline{\mu}_1 - C\underline{\mu}_2 \ \neq \ \underline{a}\,, \end{array} \qquad \left(\underline{a} \in I\!R^k\right)$$

where C is a real matrix of dimensions $k{\times}p$ and $rank(C) = k$, $(k \leq p)$, based on two independent random samples with dimensions n_1 and n_2, respectively, we will use

$$T^2 = \frac{n_1\, n_2}{n_1 + n_2}\,(n_1{+}n_2{-}2)\left(C\left(\underline{\overline{X}}_1{-}\underline{\overline{X}}_2\right){-}\underline{a}\right)'\left(CSC'\right)^{-1}\left(C\left(\underline{\overline{X}}_1{-}\underline{\overline{X}}_2\right){-}\underline{a}\right),$$

since in this case we will have, under H_0,

$$C\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2 - \underline{a}\right) \;\sim\; N_k\left(\underline{0},\, \frac{n_1 + n_2}{n_1 n_2} C\Sigma C'\right)$$

and

$$CSC' \;\sim\; W_k\left(n_1 + n_2 - 2,\, C\Sigma C'\right)$$

where

$$S \;=\; S_1 + S_2\,.$$

The construction of the statistic, from the assumptions taken, following the lines in 6.1 and 6.4.1, is left as an exercise.

Under $H_0 : C\left(\underline{\mu}_1 - \underline{\mu}_2\right) = \underline{a}$, we will have

$$\frac{n_1 + n_2 - k - 1}{k}\,\frac{T^2}{n_1 + n_2 - 2} \;\sim\; F_{k, n_1 + n_2 - k - 1}\,.$$

❏❏❏❏❏

Exercises:

6.11. Program an R function that implements the test in this subsection and that gives as a result the computed value of the T^2 statistic and the corresponding p-value.

6.12. Obtain data sets and implement the test on this subsection on them.

❏❏❏❏❏

6.4.3 Test to the equality of the mean values in $\underline{\mu}$ ($H_0 : \underline{\mu} = \mu_0 E_{p1}$)

Let us suppose that $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ and let us suppose that we want to test the equality of the p mean values $\mu_1, \ldots, \mu_p$. We may write this hypothesis as

$$H_0 : \underline{\mu} = \mu_0 E_{p1} \quad (\mu_0 \in \mathbb{R}) .$$

The question is how we will be able to encompass this test in a test using a Hotelling T^2 statistic.

Well, if we consider a matrix C of dimensions $(p-1)\times p$ with

$$\underset{(p-1)\times p}{C} = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{bmatrix}$$

then

$$H_0 : C\underline{\mu} = \underset{(p-1)\times 1}{\underline{0}} \iff H_0 : \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \\ \vdots \\ \mu_{p-1} - \mu_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \iff H_0 : \underline{\mu} = \mu_0 E_{p1} . \tag{6.14}$$

But then, according to what was exposed in subsection 6.4.1, the test statistic will be

$$T^2 = n(n-1)(C\overline{\underline{X}})'(CSC')^{-1}C\overline{\underline{X}}$$

with

$$\frac{n-p+1}{p-1}\,\frac{T^2}{n-1} \sim F_{p-1,n-p+1}$$

and, for an α level test, we will reject H_0 in (6.14) if $\frac{n-p+1}{p-1}\frac{T^2_{calc}}{n-1} > f_{p-1,n-p+1}(1-\alpha)$.

This test is sometimes referred to as a test for a repeated measures model or design, since the n observation units or "individuals" are the same for the p variables observed or measured, which may be the values of a "same" variable measured on different times or after the application of different treatments or different doses of a given treatment to the n individuals, as for example the values of the diastolic or systolic blood pressure on n individuals 1 day after the administration of each one of p different doses of a given medication.

In case of the rejection of the null hypothesis we may be interested in using other inferential procedures in order to evaluate which of the p variables are responsible for this rejection.

Note that the matrix C is indeed implementing particular contrasts among the μ_j, where a contrast is any linear combination of the μ_j's of the type

$$\underline{u}'\underline{\mu} = \sum_{j=1}^{p} u_j \mu_j$$

where

$$\sum_{j=1}^{p} u_j = 0\,.$$

Note that indeed, given the structure of the matrix C, if we take any $\underline{a} \in \mathbb{R}^{p-1}$, any linear combination $\underline{a}'C\underline{\mu}$ will always be a contrast, given that the elements of the vector

$$\underline{u} = C'\underline{a} \ \in \mathbb{R}^p$$

will always add up tp 0 (zero), since

$$\underline{u} = C'\underline{a} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ 0 & 0 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 \\ 0 & 0 & 0 & \dots & 0 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{p-2} \\ a_{p-1} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 - a_1 \\ a_3 - a_2 \\ a_4 - a_3 \\ \vdots \\ a_{p-2} - a_{p-3} \\ a_{p-1} - a_{p-2} \\ -a_{p-1} \end{bmatrix}.$$

But then, with the aim of testing several of these contrasts, that is, hypotheses of the type

$$H_0 : \underline{a}' C \underline{\mu} = \sum_{j=1}^{p} u_j \mu_j = 0 \,,$$

we may use a building technique for the CI's with a joint coverage probability of $1-\alpha$ similar to the one used in 6.1.2, being the case that now we have

$$\frac{\sqrt{n}\, \underline{a}'(C\underline{\overline{X}} - C\underline{\mu})}{\sqrt{\frac{\underline{a}'CSC'\underline{a}}{n-1}}} \sim T_{n-1}$$

so that we take

$$\begin{aligned} 1-\alpha = P &\left[\left(\frac{\sqrt{n}\, \underline{a}'(C\underline{\overline{X}} - C\underline{\mu})}{\sqrt{\frac{\underline{a}'CSC'\underline{a}}{n-1}}} \right)^2 \leq T^2_{1-\alpha} \right] \\ = P &\Big[\underline{a}'C\underline{\overline{X}} - \sqrt{\tfrac{1}{n}\, \underline{a}'CS^*C'\underline{a}\, T^2_{1-\alpha}} \leq \underline{a}'C\underline{\mu} \\ &\qquad\qquad \leq \underline{a}'C\underline{\overline{X}} + \sqrt{\tfrac{1}{n}\, \underline{a}'CS^*C'\underline{a}\, T^2_{1-\alpha}} \Big] \end{aligned}$$

and we obtain the set of CI's for $\underline{a}'C\underline{\mu}$, with a global coverage of $1-\alpha$ given by

$$\left[\underline{a}'C\underline{\overline{X}} - \sqrt{\frac{1}{n}\, \underline{a}'CS^*C'\underline{a}\, T^2_{1-\alpha}},\ \underline{a}'C\underline{\overline{X}} + \sqrt{\frac{1}{n}\, \underline{a}'CS^*C'\underline{a}\, T^2_{1-\alpha}} \right],$$

where

$$T^2_{1-\alpha} = \frac{(n-1)(p-1)}{n-p+1}\, f_{p-1,n-p+1}(1-\alpha)\,.$$

Note how the present test reduces himself to the common T test for two paired samples for $p=2$.

❑❑❑❑❑

Exercises:

6.13. Show that indeed for $p=2$ the test in this subsection reduces to the common T test for two paired samples.

6.14. Program an R function that implements the test in this subsection and which will give as a result the computed value of the T^2 statistic and the corresponding p-value.

6.15. Obtain data sets on which you will implement the test in this subsection.

6.16. Program an R function to obtain the CI at the end of the present subsection and implement it on several data sets.

❑❑❑❑❑

6.5 Profile analysis (two profiles)

Let us suppose that in two populations, for example two stands of *Pinus pinaster*, we measured the same three variables X_1, X_2 and X_3 referred to before, obtaining a sample of size n_1 from the first population and a sample of size n_2 from the second population. Let us suppose that, given the way the samples were taken, these may be considered independent.

We may represent the results obtained, in terms of sample means, in the for of a plot as the one presented in Figure 6.1, where the first index refers to the population and the second to the variable to which the sample mean corresponds. To each broken line that passes through the sample means for a given popultion we call 'profile'.

Relative to the corresponding population mean values, besides the question of the possible coincidence of the two profiles, that is, the test to the null hypothesis $H_0: \underline{\mu}_1 = \underline{\mu}_2$, we may be interested in testing oter hypotheses of interest, as for example:

I – Are the profiles parallel?

II – Assuming that the profiles are parallel; are the profiles coincident?

III – Assuming that the profiles are coincident; are the profiles horizontal?

where this last test may sometimes loose its meaning, since it will mostly have a meaning namely when all variables are measured in the same units. We may note that the non-rejection of both hypotheses in I and II will correspond to the non-rejection of the hypothesis of equality of $\underline{\mu}_1$ and $\underline{\mu}_2$.

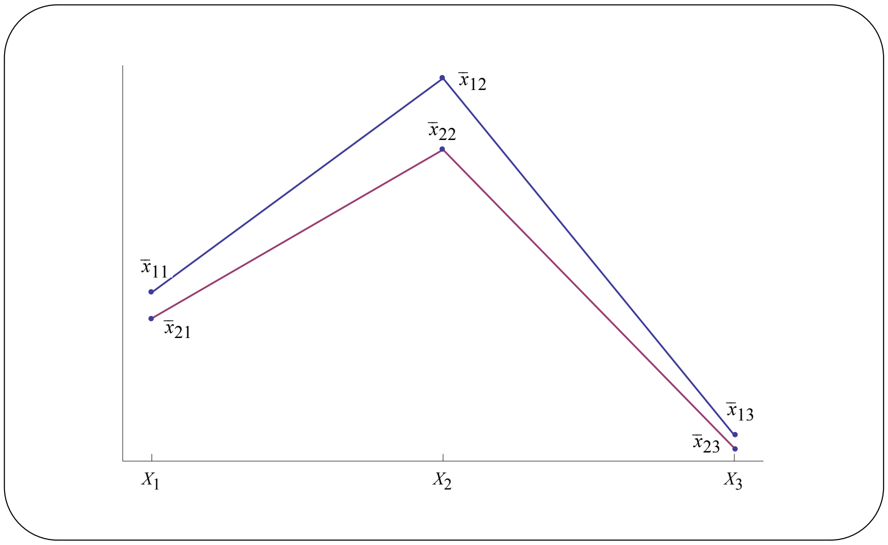

Figura 6.1. – Example of a plot of two profiles, corresponding to 3 random variables

Let us consider the tests to be carried out in order to give an answer to each of the three questions above.

I – Test to the parallelism of the profiles.

We want to test the hypothesis

$$H_{0(1)} : \mu_{1i} - \mu_{1,i+1} = \mu_{2i} - \mu_{2,i+1} , \quad i = 1, 2, 3, ..., p-1$$

or, if we consider the matrix

$$\underset{(p-1)\times p}{C} = \begin{bmatrix} 1 & -1 & 0 & ... & 0 & 0 \\ 0 & 1 & -1 & ... & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & ... & 1 & -1 \end{bmatrix} \tag{6.15}$$

we may then write the hypothesis $H_{0(1)}$ above in an equivalent form, which will be much easier to implement with the use of a T^2 statistic, and which is

$$\begin{array}{ll} & H_{0(1)} : C\underline{\mu}_1 = C\underline{\mu}_2 \iff H_{0(1)} : C\underline{\mu}_1 - C\underline{\mu}_2 = \underline{0} \\ \text{vs} & \\ & H_{1(1)} : C\underline{\mu}_1 \neq C\underline{\mu}_2 \iff H_{1(1)} : C\underline{\mu}_1 - C\underline{\mu}_2 \neq \underline{0} \, , \end{array}$$

where

$$\underline{\mu}_1 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{bmatrix}, \qquad \underline{\mu}_2 = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{bmatrix},$$

so that

$$C\underline{\mu}_1 = \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1,p-1} - \mu_{1p} \end{bmatrix}, \qquad C\underline{\mu}_2 = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2,p-1} - \mu_{2p} \end{bmatrix},$$

and thus

$$C\underline{\mu}_1 = C\underline{\mu}_2 \iff \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1,p-1} - \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2,p-1} - \mu_{2p} \end{bmatrix}.$$

We will then use a statistic similar to the one defined in 6.4.1, that is, the statistic

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(n_1 + n_2 - 2)\left(C\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)\right)'\left(CSC'\right)^{-1}\left(C\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)\right),$$

with

$$S = S_1 + S_2\,,$$

where S_1 and S_2 represent respectively the matrices of sums of squares and products of deviations from the sample means of each profile, rejecting $H_{0(1)}$ if

$$F_{calc} = \frac{n_1 + n_2 - p}{p - 1}\,\frac{T^2_{calc}}{n_1 + n_2 - 2} > f_{p-1,n_1+n_2-p}(1-\alpha)\,.$$

In case we do not reject $H_{0(1)}$, then we may pursue to **II**.

II – Test to the coincidence of the profiles.

We wnat to test the coincidence of the profiles, assuming that we did not reject $H_{0(1)}$ above, that is, assuming that the profiles are paraççeç. We wnat then to test

$$\begin{array}{ll} & H_{0(2)}: E_{1p}\underline{\mu}_1 = E_{1p}\underline{\mu}_2 \iff H_{0(2)}: E_{1p}\underline{\mu}_1 - E_{1p}\underline{\mu}_2 = 0 \\ \text{vs} & \\ & H_{1(2)}: E_{1p}\underline{\mu}_1 \neq E_{1p}\underline{\mu}_2 \iff H_{1(2)}: E_{1p}\underline{\mu}_1 - E_{1p}\underline{\mu}_2 \neq 0\,, \end{array}$$

where $E_{1p}\underline{\mu}_j$ $(j = 1, 2)$, represents the sum of the expected values of each one of the profiles.

We will use then a statistic of the type studied in 6.4.2, which inthis case will be the statistic

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(n_1+n_2-2)\left(E_{1p}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)\right)'\left(E_{1p}SE_{p1}\right)^{-1}\left(E_{1p}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)\right)\,,$$

rejecting $H_{0(2)}$ if

$$F_{calc} = \frac{n_1 + n_2 - 2}{1}\,\frac{T^2_{calc}}{n_1 + n_2 - 2} = T^2_{calc} > f_{1,n_1+n_2-2}(1-\alpha)\,.$$

If we do not reject $H_{0(2)}$, then we may pursue to **III**.

Note that the null hypotheses tested in **I** and **II**, combined, correspond to the hypothesis $H_0 : \underline{\mu}_1 = \underline{\mu}_2$.

III – Test to the horizontality of the profiles.

We want to test the horizontality of the profiles, assuming that we did not reject $H_{0(2)}$ above, that is, assuming that the teo profiles are coincident. We want thus to test

$$\begin{array}{ll} & H_{0(3)}: C\underline{\mu} = \underline{0} \\ \text{vs} & \\ & H_{1(3)}: C\underline{\mu} \neq \underline{0}\,, \end{array}$$

where C is the matrix defined in **I** and

$$\underline{\mu} = \frac{n_1}{n_1+n_2}\underline{\mu}_1 + \frac{n_2}{n_1+n_2}\underline{\mu}_2 \left(= \underline{\mu}_1 = \underline{\mu}_2\right) \tag{6.16}$$

is the ‘pooled’ mean of the two profiles, since the joint non-rejection of $H_{0(1)}$ and $H_{0(2)}$ is equivalent to the non-rejection of $H_0 : \underline{\mu}_1 = \underline{\mu}_2$.

We will then use a statistic of the type defined in 6.4.1,

$$
\begin{aligned}
T^2 &= \left(C\underline{\overline{X}} - C\underline{\mu}\big|_{H_0}\right)' \left(\frac{1}{n_1+n_2-2}\,\frac{1}{n_1+n_2}\,CSC'\right)^{-1} \left(C\underline{\overline{X}} - C\underline{\mu}\big|_{H_0}\right) \\
&= (n_1+n_2-2)(n_1+n_2)\left(C\underline{\overline{X}}\right)'\left(CSC'\right)^{-1} C\underline{\overline{X}}
\end{aligned}
\tag{6.17}
$$

with

$$
\frac{n_1+n_2-p}{p-1}\,\frac{T^2}{n_1+n_2-2} \sim F_{p-1,n_1+n_2-p} \;,
$$

where

$$
\underline{\overline{X}} = \frac{n_1}{n_1+n_2}\,\underline{\overline{X}}_1 + \frac{n_2}{n_1+n_2}\,\underline{\overline{X}}_2 \sim N_p\Bigg(\underbrace{\frac{n_1}{n_1+n_2}\,\underline{\mu}_1 + \frac{n_2}{n_1+n_2}\,\underline{\mu}_2}_{=\underline{\mu}},\; \frac{n_1+n_2}{(n_1+n_2)^2}\,\Sigma\Bigg)
\tag{6.18}
$$

is the pooled estimator of $\underline{\mu}$ in (6.16). Note that from (6.18) we have

$$
C\underline{\overline{X}} \sim N_{p-1}\left(C\left(\frac{n_1}{n_1+n_2}\,\underline{\mu}_1 + \frac{n_2}{n_1+n_2}\,\underline{\mu}_2\right),\; \frac{1}{n_1+n_2}C\Sigma C'\right),
$$

or, taking into account the definition of $\underline{\mu}$ in (6.16),

$$
\left(C\underline{\overline{X}} - C\underline{\mu}\right) \sim N_{p-1}\left(\underline{0},\; \frac{1}{n_1+n_2}\,C\Sigma C'\right),
$$

while,

$$
\frac{1}{n_1+n_2}\,CSC' \sim W_{p-1}\left(n_1+n_2-2,\; \frac{1}{n_1+n_2}\,C\Sigma C'\right),
$$

where $S = S_1 + S_2$ is the pooled estimator of the variance-covariance matrix of the p variables.

To test $H_{0(3)}$ we obtain the computed value of the statistic T^2 in (6.17), replacing $C\underline{\mu}$ by $\underline{0}$ and we will reject $H_{0(3)}$ if, for a gicen α

$$
\frac{n_1+n_2-p}{p-1}\,\frac{T^2_{calc}}{n_1+n_2-2} > f_{p-1,n_1+n_2-p}(1-\alpha)\,.
$$

❑❑❑❑❑

Exercises:

6.17. Program R function which implement the analysis of two profiles.

6.18. Keeping in mind the possible problems associated with data snooping, use the data set relative to the 3 species of *Iris* to carry out a profile analysis using the profiles relative to the two species for which you think the hypothesis of parallelism of the profiles is most likely to be not rejected.

6.19. Obtain data sets on which you will implement the profile analysis in this section.

❑❑❑❑❑

Chapter 7

Likelihood ratio test for the equality of several mean vectors

7.1 Derivation of the likelihood ratio statistic

Let us suppose that $\underline{X}_k \sim N_p(\underline{\mu}_k, \Sigma_k)$, $k = 1, \ldots, q$, with $\Sigma_1 = \ldots = \Sigma_q (= \Sigma)$. Let us further suppose that we have a sample of dimesion n_k from $\underline{X}_k$ and that the q samples are independent, with $n = \sum_{k=1}^{q} n_k$.

Let us suppose that we are interested in testing the hypotheses

$$\begin{aligned} & H_0 : \underline{\mu}_1 = \ldots = \underline{\mu}_q (= \underline{\mu}) \\ \text{vs.} \quad & \\ & H_1 : \exists k, k' \in \{1, \ldots, q\} : \underline{\mu}_k \neq \underline{\mu}_{k'} \ (k \neq k') \,, \end{aligned} \tag{7.1}$$

assuming $\Sigma_1 = \ldots = \Sigma_q (= \Sigma)$.

We may note that we are indeed making an hypothesis test in a model which is commonly called "one-way MANOVA", that is, a Multivariate Analysis of Variance model, where we have several p response variables, and a design with q treatments or one factor with q levels.

Let X_k $(k = 1, \ldots, q)$ be the $n_k \times p$ sample matrices for the q random samples. Both under H_0, as well as under H_1, the likelihood function will be the joint distribution of these q matrices.

Let then L_0 and L_1 be respectively the likelihood functions under H_0 and under H_1 (see Appendix 7.A for a brief note on likelihood ratio tests). Under

H_1, from (3.2), we have

$$\begin{aligned} L_1 &= \prod_{k=1}^{q} (2\pi)^{-n_k p/2} |\Sigma|^{-n_k/2} \; e^{-\frac{1}{2} tr\left[\Sigma^{-1}(X_k - E_{n_k 1}\underline{\mu}_k')'(X_k - E_{n_k 1}\underline{\mu}_k')\right]} \\ &= (2\pi)^{-np/2} \, |\Sigma|^{-n/2} \; e^{-\frac{1}{2}\sum_{k=1}^{q} tr\left[\Sigma^{-1}(X_k - E_{n_k 1}\underline{\mu}_k')'(X_k - E_{n_k 1}\underline{\mu}_k')\right]} \end{aligned}$$

where, under H_1,

$$\widehat{\underline{\mu}}_k = \overline{\underline{X}}_k = \frac{1}{n_k} X_k' E_{n_k 1} \quad \text{e} \quad \widehat{\Sigma}_{H_1} = \frac{1}{n} S \,,$$

with

$$S = \sum_{k=1}^{q} S_k = \sum_{k=1}^{q} \left(X_k - E_{n_k 1} \overline{\underline{X}}_k' \right)' \left(X_k - E_{n_k 1} \overline{\underline{X}}_k' \right) , \tag{7.2}$$

so that

$$\begin{aligned} \max \, L_1 &= L_1(\widehat{\underline{\mu}}_1, \ldots, \widehat{\underline{\mu}}_q, \widehat{\Sigma}) \\ &= (2\pi)^{-np/2} \left(\frac{1}{n}\right)^{-np/2} |S|^{-n/2} \; e^{-\frac{1}{2} tr\left[\sum_{k=1}^{q}(nS_k S^{-1})\right]} \\ &= (2\pi)^{-np/2} \left(\frac{1}{n}\right)^{-np/2} |S|^{-n/2} \; e^{-\frac{1}{2} tr(nSS^{-1})} \\ &= (2\pi)^{-np/2} \left(\frac{1}{n}\right)^{-np/2} |S|^{-n/2} \; e^{-\frac{1}{2} np} \,. \end{aligned}$$

Under H_0 we have

$$\begin{aligned} L_0 &= \prod_{k=1}^{q} (2\pi)^{-n_k p/2} \, |\Sigma|^{-n_k/2} \; e^{-\frac{1}{2} tr\left[\Sigma^{-1}(X_k - E_{n_k 1}\underline{\mu}')'(X_k - E_{n_k 1}\underline{\mu}')\right]} \\ &= (2\pi)^{-np/2} \, |\Sigma|^{-n/2} \; e^{-\frac{1}{2} tr \sum_{k=1}^{q}\left[\Sigma^{-1}(X_k - E_{n_k 1}\underline{\mu}')'(X_k - E_{n_k 1}\underline{\mu}')\right]} \end{aligned}$$

where

$$\widehat{\underline{\mu}} = \overline{\underline{X}} = \frac{1}{n} \sum_{k=1}^{q} n_k \overline{\underline{X}}_k \tag{7.3}$$

and

$$\widehat{\Sigma}_{H_0} = \frac{1}{n} \sum_{k=1}^{q} \left(X_k - E_{n_k 1} \overline{\underline{X}}' \right)' \left(X_k - E_{n_k 1} \overline{\underline{X}}' \right) = \frac{1}{n} S^* ,$$

with

$$S^* = \sum_{k=1}^{q} \left(X_k - E_{n_k1}\underline{\overline{X}}'\right)' \left(X_k - E_{n_k1}\underline{\overline{X}}'\right).$$

Note that

$$\begin{aligned}
&\left(X_k - E_{n_k1}\underline{\overline{X}}'\right)' \left(X_k - E_{n_k1}\underline{\overline{X}}'\right) \\
&\quad = \left(X_k - E_{n_k1}\underline{\overline{X}}'_k + E_{n_k1}\underline{\overline{X}}'_k - E_{n_k1}\underline{\overline{X}}'\right)' \\
&\qquad\qquad\qquad \times \left(X_k - E_{n_k1}\underline{\overline{X}}'_k + E_{n_k1}\underline{\overline{X}}'_k - E_{n_k1}\underline{\overline{X}}'\right) \\
&\quad = \left(X_k - E_{n_k1}\underline{\overline{X}}'_k\right)' \left(X_k - E_{n_k1}\underline{\overline{X}}'_k\right) + \left(\underline{\overline{X}}'_k - \underline{\overline{X}}'\right)' E_{1n_k}E_{n_k1}\left(\underline{\overline{X}}'_k - \underline{\overline{X}}'\right) \\
&\qquad + \underbrace{\left(X_k - E_{n_k1}\underline{\overline{X}}'_k\right)' \left(E_{n_k1}\underline{\overline{X}}'_k - E_{n_k1}\underline{\overline{X}}'\right)}_{=0} \\
&\qquad + \underbrace{\left(E_{n_k1}\underline{\overline{X}}'_k - E_{n_k1}\underline{\overline{X}}'\right)' \left(X_k - E_{n_k1}\underline{\overline{X}}'_k\right)}_{=0} \\
&\quad = S_k + \underbrace{n_k(\underline{\overline{X}}_k - \underline{\overline{X}})(\underline{\overline{X}}_k - \underline{\overline{X}})'}_{=B_k},
\end{aligned}$$

since in fact

$$\underline{\overline{X}}_k = \frac{1}{n_k} X'_k E_{n_k1}\,,$$

so that

$$\begin{aligned}
&\left(X_k - E_{n_k1}\underline{\overline{X}}'_k\right)' \left(E_{n_k1}\underline{\overline{X}}'_k - E_{n_k1}\underline{\overline{X}}'\right) \\
&\quad = X'_k E_{n_k1}\underline{\overline{X}}'_k - X'_k E_{n_k1}\underline{\overline{X}}' - \underbrace{\underline{\overline{X}}_k \underbrace{E_{1n_k}E_{n_k1}}_{n_k} \underline{\overline{X}}'_k}_{=X'_k E_{n_k1}\underline{\overline{X}}'_k} + \underbrace{\underline{\overline{X}}_k \underbrace{E_{1n_k}E_{n_k1}}_{n_k} \underline{\overline{X}}'}_{=X'_k E_{n_k1}\underline{\overline{X}}'} = 0
\end{aligned}$$

and as such

$$S^* = S + B\,, \quad \text{onde} \quad B = \sum_{k=1}^{q} B_k\,. \tag{7.4}$$

But then

$$\begin{aligned} \max L_0 &= (2\pi)^{-np/2}\, n^{np/2}\, |S+B|^{-n/2}\; e^{-\frac{1}{2}tr[nS^* S^{*-1}]} \\ &= (2\pi)^{-np/2}\, n^{np/2}\, |S+B|^{-n/2}\; e^{-\frac{1}{2}np} \end{aligned}$$

and thus

$$\begin{aligned} \Lambda = \frac{\max L_0}{\max L_1} &= \frac{(2\pi)^{-np/2}\, n^{np/2}\, |S+B|^{-n/2}\; e^{-\frac{1}{2}np}}{(2\pi)^{-np/2}\, n^{np/2}\, |S|^{-n/2}\; e^{-\frac{1}{2}np}} \\ &= \left(\frac{|S|}{|S+B|}\right)^{n/2} . \end{aligned}$$

Instead of the statistic Λ, is often used the statistic

$$\Lambda^* = \Lambda^{2/n} = \frac{|S|}{|S+B|}\,, \tag{7.5}$$

obtaining an equivalent test, and rejecting the null hypothesis in (7.1) if $\Lambda^*_{calc} < \Lambda^*_\alpha$ or if $\Lambda_{calc} < \Lambda_\alpha$.

Both Λ and Λ^* are commonly called Wilks Λ statistics since it was Samuel Stanley Wilks who first derived and studied their distribution (Wilks, 1932). See Appendix 7.B for a brief reference to the Wilks Λ statistic. However, in order to be able to make use of these statistics in the implementation of the test to the above hypothesis we need to obtain their distribution, at least under H_0.

7.2 Obtaining the distribution of the likelihood ratio statistic

In order to obtain the distribution of the statistic Λ^* we will first obtain the expression for its h-th moment.

But, in order to obtain the expression for the h-th moment of Λ^* we need to obtain first the distributions of the matrices S and B. Since

$$S = \sum_{k=1}^{q} S_k\,,$$

where the q matrices S_k are independent, with

$$S_k \sim W_p(n_k - 1, \Sigma)\,,$$

we have that

$$S \sim W_p(n - q, \Sigma)\,. \tag{7.6}$$

Concerning the matrix B, we may note that we may write

$$B = \sum_{k=1}^{q} B_k = \sum_{k=1}^{q} n_k(\underline{\overline{X}}_k - \underline{\overline{X}})'(\underline{\overline{X}}_k - \underline{\overline{X}}) = n\overline{X}^{*\prime}(I_q - E_{qq}D)'D(I_q - E_{qq}D)\,\overline{X}^{*}$$

where $\overline{X}^{*}$ is the matrix of dimensions $q{\times}p$ which k-th row is equal to $\underline{\overline{X}}_k'$ multiplied by $\sqrt{n_k}$, that is,

$$\underset{q\times p}{\overline{X}^{*}} = \left[\begin{array}{c} \underline{\overline{X}}_1' \\ \hline \underline{\overline{X}}_2' \\ \hline \vdots \\ \hline \underline{\overline{X}}_q' \end{array}\right]$$

and

$$D = diag(n_1/n, n_2/n, \ldots, n_q/n)\,, \qquad \text{com} \qquad tr(D) = 1\,.$$

Note that, under H_0,

$$\overline{X}^{*} \sim N_{qp}\left(E_{q1}\underline{\mu}', \Sigma \otimes \frac{1}{n}D^{-1}\right),$$

we may write

$$B = Y'Y$$

where

$$Y = \sqrt{n}D^{1/2}(I_q - E_{qq}D)\overline{X}^{*}\,,$$

so that by Result 3.1 we have

$$\begin{aligned} Y &\sim N_{qp}\left(\sqrt{n}D^{1/2}(I_q - E_{qq}D)E_{q1}\underline{\mu}', \Sigma \otimes D^{1/2}(I_q - E_{qq}D)D^{-1}(I_q - DE_{qq})D^{1/2}\right) \\ &\equiv N_{qp}\left(0_{q\times p}, \Sigma \otimes (I_q - D^{1/2}E_{qq}D^{1/2})\right), \end{aligned}$$

where $(I_q - D^{1/2}E_{qq}D^{1/2})$ is a symmetric and idempotent matrix, since i)

$$D^{1/2}(I_q - E_{qq}D)E_{q1}\underline{\mu}' = D^{1/2}E_{q1}\underline{\mu}' - D^{1/2}\underbrace{E_{qq}DE_{q1}}_{=E_{q1}}\underline{\mu}' = 0_{q\times 1};$$

ii)

$$\begin{aligned}
&D^{1/2}(I_q - E_{qq}D)D^{-1}(I_q - DE_{qq})D^{1/2}\\
&\qquad = D^{1/2}(I_q - E_{qq}D)D^{-1/2}D^{-1/2}(I_q - DE_{qq})D^{1/2}\\
&\qquad = (I_q - D^{1/2}E_{qq}D^{1/2})(I_q - D^{1/2}E_{qq}D^{1/2})\\
&\qquad = I_q - D^{1/2}E_{qq}D^{1/2} - D^{1/2}E_{qq}D^{1/2}\\
&\qquad\qquad\qquad + D^{1/2}\underbrace{E_{qq}DE_{qq}}_{E_{qq}}D^{1/2}\\
&\qquad = I_q - D^{1/2}E_{qq}D^{1/2}\,,
\end{aligned}$$

which also shows that $I_q - D^{1/2}E_{qq}D^{1/2}$ is an idempotent matrix.

But then, given the idempotency of $I_q - D^{1/2}E_{qq}D^{1/2}$, we have

$$\begin{aligned}
rank(I_q - D^{1/2}E_{qq}D^{1/2}) &= tr(I_q - D^{1/2}E_{qq}D^{1/2})\\
&= tr(I_q) - tr(D^{1/2}E_{qq}D^{1/2})\\
&= q - tr(E_{qq}D) = q - 1\,,
\end{aligned}$$

so that from Theorem 5.1 we have that, under H_0,

$$B \sim W_p(q-1, \Sigma)\,, \tag{7.7}$$

where the distribution will be a pseudo-Wishart distribution if $q - 1 < p$.

But then, from the distributions of S and B we have

$$
\begin{aligned}
E\left[\left(\frac{|S|}{|S+B|}\right)^h\right] &= E\left(\frac{|S|^h}{|S+B|^h}\right) \\
&= \int_{\mathcal{S}>0}\int_{\mathcal{B}>0} \frac{|\mathcal{S}|^h}{|\mathcal{S}+\mathcal{B}|^h}\, \frac{e^{-\frac{1}{2}tr(\Sigma^{-1}\mathcal{S})}|\mathcal{S}|^{(n-q-p-1)/2}}{2^{(n-q)p/2}\,\Gamma_p\left(\frac{n-q}{2}\right)|\Sigma|^{(n-q)/2}} \\
&\qquad\qquad \times \frac{e^{-\frac{1}{2}tr(\Sigma^{-1}\mathcal{B})}|\mathcal{B}|^{(q-p-2)/2}}{2^{(q-1)p/2}\,\Gamma_p\left(\frac{q-1}{2}\right)|\Sigma|^{(q-1)/2}}\, d\mathcal{B}\, d\mathcal{S} \\
&= 2^{hp}\,|\Sigma|^h\, \frac{\Gamma_p\left(\frac{n-q}{2}+h\right)}{\Gamma_p\left(\frac{n-q}{2}\right)} \\
&\qquad \int_{\mathcal{S}>0}\int_{\mathcal{B}>0} \frac{1}{|\mathcal{S}+\mathcal{B}|^h}\, \underbrace{\frac{e^{-\frac{1}{2}tr(\Sigma^{-1}\mathcal{S})}|\mathcal{S}|^{(n-q+2h-p-1)/2}}{2^{(n-q+2h)p/2}\,\Gamma_p\left(\frac{n-q+2h}{2}\right)|\Sigma|^{(n-q+2h)/2}}}_{\text{p.d.f. of } W_p(n-q+2h,\Sigma)} \\
&\qquad\qquad \times \underbrace{\frac{e^{-\frac{1}{2}tr(\Sigma^{-1}\mathcal{B})}|\mathcal{B}|^{(q-p-2)/2}}{2^{(q-1)p/2}\,\Gamma_p\left(\frac{q-1}{2}\right)|\Sigma|^{(q-1)/2}}}_{\text{p.d.f. of } W_p(q-1,\Sigma)}\, d\mathcal{B}\, d\mathcal{S} \\
&= 2^{hp}\,|\Sigma|^h\, \frac{\Gamma_p\left(\frac{n-q}{2}+h\right)}{\Gamma_p\left(\frac{n-q}{2}\right)}\, \underbrace{E\left(|S+B|^{-h}\right)}_{\text{with } S+B\,\sim\, W_p(n-1+2h,\Sigma)}\,,
\end{aligned}
$$

so that, using the result in Theorem 5.3, we may write

$$
\begin{aligned}
E\left[\left(\frac{|S|}{|S+B|}\right)^h\right] &= 2^{hp}\,|\Sigma|^h\, \frac{\Gamma_p\left(\frac{n-q}{2}+h\right)}{\Gamma_p\left(\frac{n-q}{2}\right)}\, 2^{-hp}\,|\Sigma|^{-h}\, \frac{\Gamma_p\left(\frac{n-1}{2}+h-h\right)}{\Gamma_p\left(\frac{n-1}{2}+h\right)} \\
&= \frac{\Gamma_p\left(\frac{n-q}{2}+h\right)}{\Gamma_p\left(\frac{n-q}{2}\right)}\, \frac{\Gamma_p\left(\frac{n-1}{2}\right)}{\Gamma_p\left(\frac{n-1}{2}+h\right)}\,,
\end{aligned}
$$

which, using (5.4), may be written, for $h>(p+q-n-1)/2$, as

$$
E\left[\left(\frac{|S|}{|S+B|}\right)^h\right] = \prod_{j=1}^{p} \frac{\Gamma\left(\frac{n-q}{2}+h-\frac{j-1}{2}\right)}{\Gamma\left(\frac{n-q}{2}-\frac{j-1}{2}\right)}\, \frac{\Gamma\left(\frac{n-1}{2}-\frac{j-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}+h-\frac{j-1}{2}\right)} = \prod_{j=1}^{p} E\left(Y_j^h\right) \tag{7.8}
$$

where

$$Y_j \sim Beta\left(\frac{n-q+1-j}{2}, \frac{q-1}{2}\right), \qquad j=1,\ldots,p \tag{7.9}$$

form a set of p independent r.v.'s.

But then, since the support of Λ is delimited, since $0 < \Lambda < 1$, the distribution of Λ is completely determined by its moments, so that we may write

$$\Lambda^* \stackrel{st}{\sim} \prod_{j=1}^{p} Y_j$$

where the r.v.'s Y_j have the distributions in (7.9), and where '$\stackrel{st}{\sim}$' is to be read as 'stochastically equivalent', or 'has the same distribution as'.

In fact it is also possible to show that

$$\Lambda^* \stackrel{st}{\sim} \prod_{k=1}^{q-1} Y_k^* \tag{7.10}$$

where

$$Y_k^* \sim Beta\left(\frac{n-p-k}{2}, \frac{p}{2}\right), \qquad k=1,\ldots,q-1 \tag{7.11}$$

form a set of $q-1$ independent r.v.'s, since it is indeed possible to show that, without any restrictions on either p or q,

$$\prod_{j=1}^{p} Y_j \stackrel{d}{\equiv} \prod_{k=1}^{q-1} Y_k^* \,.$$

In fact, from (7.8) we may write

$$\begin{aligned}
E\left((\Lambda^*)^h\right) &= \prod_{j=1}^{p} \frac{\Gamma\left(\frac{n-q+1-j}{2}+h\right)\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-q+1-j}{2}\right)\Gamma\left(\frac{n-j}{2}+h\right)} \\
&= \left\{\prod_{j=q}^{p+q-1} \frac{\Gamma\left(\frac{n-j}{2}+h\right)}{\Gamma\left(\frac{n-j}{2}\right)}\right\}\left\{\prod_{j=1}^{p} \frac{\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-j}{2}+h\right)}\right\} \\
&= \left\{\prod_{j=1}^{p+q-1} \frac{\Gamma\left(\frac{n-j}{2}+h\right)}{\Gamma\left(\frac{n-j}{2}\right)}\right\}\left\{\prod_{j=1}^{q-1} \frac{\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-j}{2}+h\right)}\right\}\left\{\prod_{j=1}^{p} \frac{\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-j}{2}+h\right)}\right\}
\end{aligned}$$

$$= \left\{\prod_{j=p+1}^{p+q-1} \frac{\Gamma\left(\frac{n-j}{2}+h\right)}{\Gamma\left(\frac{n-j}{2}\right)}\right\}\left\{\prod_{j=1}^{q-1} \frac{\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-j}{2}+h\right)}\right\}$$

$$= \prod_{j=1}^{q-1} \frac{\Gamma\left(\frac{n-p-j}{2}+h\right)\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-p-j}{2}\right)\Gamma\left(\frac{n-j}{2}+h\right)} = \prod_{k=1}^{q-1} E\left(Y_k^*\right)^h , \tag{7.12}$$

where Y_k^* $(k = 1, \ldots, q-1)$ are the r.v.'s in (7.11).

From (7.8) and (7.12) we may see the intermutability of p and $q-1$ in the definition of the distribution of Λ^* and Λ.

An important result is also the one on the independence of Λ^* and $S+B$. This fact is not too hard to prove, since, being the determinant of a matrix A a function of all its elements, and being the distribution of the matrix A nothing else than the joint distribution of all its elements, to prove the independence of Λ^* and $S+B$ it is enough to prove that Λ^* is independent of $|S+B|$.

Let then $S \sim W_p(n-q, \Sigma)$ and $S+B \sim W_p(n-1, \Sigma)$. Then, from (5.31) we have

$$E\left(|S|^h\right) = 2^{hp}|\Sigma|^h \frac{\Gamma_p\left(\frac{n-q}{2}+h\right)}{\Gamma_p\left(\frac{n-q}{2}\right)} \quad \text{e} \quad E\left(|S+B|^h\right) = 2^{hp}|\Sigma|^h \frac{\Gamma_p\left(\frac{n-1}{2}+h\right)}{\Gamma_p\left(\frac{n-1}{2}\right)} ,$$

expressions which are indeed valid for any $h \in \mathbb{C}$, so that, taking

$$W_1 = -\log|S| \quad \text{and} \quad W_2 = -\log|S+B| \tag{7.13}$$

we have the characteristic functions of W_1 and W_2 respectively given by

$$\Phi_{W_1}(t) = E\left(e^{\mathrm{i}tW_1}\right) = E\left(e^{-\mathrm{i}t\log|S|}\right) = E\left(|S|^{-\mathrm{i}t}\right) = 2^{-\mathrm{i}t}|\Sigma|^{-\mathrm{i}t} \frac{\Gamma_p\left(\frac{n-q}{2}-\mathrm{i}t\right)}{\Gamma_p\left(\frac{n-q}{2}\right)} \tag{7.14}$$

and

$$\begin{aligned}\Phi_{W_2}(t) = E\left(e^{\mathrm{i}tW_2}\right) = E\left(e^{-\mathrm{i}t\log|S+B|}\right) &= E\left(|S+B|^{-\mathrm{i}t}\right)\\ &= = 2^{-\mathrm{i}t}|\Sigma|^{-\mathrm{i}t} \frac{\Gamma_p\left(\frac{n-1}{2}-\mathrm{i}t\right)}{\Gamma_p\left(\frac{n-1}{2}\right)} .\end{aligned} \tag{7.15}$$

But then, taking

$$W = -\log\Lambda^* , \tag{7.16}$$

we have, from (7.8),

$$\Phi_W(t) = E\left(e^{\mathrm{i}tW}\right) = E\left(e^{-\mathrm{i}t\log\Lambda}\right) = E\left(\Lambda^{-\mathrm{i}t}\right) = \frac{\Gamma_p\left(\frac{n-q}{2}-\mathrm{i}t\right)}{\Gamma_p\left(\frac{n-q}{2}\right)}\,\frac{\Gamma_p\left(\frac{n-1}{2}\right)}{\Gamma_p\left(\frac{n-1}{2}-\mathrm{i}t\right)}\,, \tag{7.17}$$

so that in fact we have, from (7.14), (7.15) and (7.17),

$$\phi_{W_1}(t) = \Phi_W(t)\,\Phi_{W_2}(t)\,, \tag{7.18}$$

while from the definitions of W_1, W_2 and W in (7.13) and (7.16) we have

$$W_1 = W + W_2\,. \tag{7.19}$$

But then (7.18) and (7.19) show that W and W_2 are two independent r.v.'s, and thus also $\Lambda^* = e^{-W}$ and $|S+B| = e^{-W_2}$ are independent.

In fact this independence relation between Λ in (7.5) and $|S+B|$ has its origin in the well-known fact that each Beta r.v. Y_k^* in (7.11) may be written as

$$Y_k^* = \frac{W_k}{W_k+Z_k}\,,$$

where

$$W_k \sim \chi^2_{n-p-k} \quad \text{and} \quad Z_k \sim \chi^2_p$$

are two independent r.v.'s and where Y_k^* is independent of W_k+Z_k.

7.3 Relation, for $q=2$, with the Hotelling T^2 statistic in section 6.3

For the case $q=2$, the use of the test and of the statistic in section 6.3 or the present test results in fact in two absolutely equivalent approaches, since it is possible to show that

$$\frac{T^2}{n-2} = \frac{1-\Lambda^*}{\Lambda^*}\,.$$

In fact, since for $q=2$, $\underline{\overline{X}}$ in (7.3) may be written as

$$\underline{\overline{X}} = \tfrac{1}{n}\left(n_1\underline{\overline{X}}_1 + n_2\underline{\overline{X}}_2\right) = \tfrac{n_1}{n}\underline{\overline{X}}_1 + \tfrac{n_2}{n}\underline{\overline{X}}_2\,,$$

so that, for $n = n_1 + n_2$,

$$\begin{aligned}
& n_1 \left(\underline{\overline{X}}_1 - \underline{\overline{X}}\right) \left(\underline{\overline{X}}_1 - \underline{\overline{X}}\right)' + n_2 \left(\underline{\overline{X}}_2 - \underline{\overline{X}}\right) \left(\underline{\overline{X}}_2 - \underline{\overline{X}}\right)' \\
& \quad = n_1 \left(\underline{\overline{X}}_1 - \tfrac{n_1}{n}\underline{\overline{X}}_1 - \tfrac{n_2}{n}\underline{\overline{X}}_2\right) \left(\underline{\overline{X}}_1 - \tfrac{n_1}{n}\underline{\overline{X}}_1 - \tfrac{n_2}{n}\underline{\overline{X}}_2\right)' \\
& \qquad\qquad + n_2 \left(\underline{\overline{X}}_2 - \tfrac{n_1}{n}\underline{\overline{X}}_1 - \tfrac{n_2}{n}\underline{\overline{X}}_2\right) \left(\underline{\overline{X}}_2 - \tfrac{n_1}{n}\underline{\overline{X}}_1 - \tfrac{n_2}{n}\underline{\overline{X}}_2\right)' \\
& \quad = n_1 \left(\tfrac{n-n_1}{n}\underline{\overline{X}}_1 - \tfrac{n_2}{n}\underline{\overline{X}}_2\right) \left(\tfrac{n-n_1}{n}\underline{\overline{X}}_1 - \tfrac{n_2}{n}\underline{\overline{X}}_2\right)' \\
& \qquad\qquad + n_2 \left(\tfrac{n-n_2}{n}\underline{\overline{X}}_2 - \tfrac{n_1}{n}\underline{\overline{X}}_1\right) \left(\tfrac{n-n_2}{n}\underline{\overline{X}}_2 - \tfrac{n_1}{n}\underline{\overline{X}}_1\right)' \\
& \quad = n_1 \left(\tfrac{n_2}{n}\right)^2 \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' + n_2 \left(\tfrac{n_1}{n}\right)^2 \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' \\
& \quad = \tfrac{n_1 n_2^2 + n_2 n_1^2}{n^2} \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' = \tfrac{n_1 n_2}{n_1 + n_2} \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)'.
\end{aligned} \tag{7.20}$$

Then, since if the matrix H is partitioned as

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

we have

$$|H| = |H_{11}|\,|H_{22.1}| \;=\; |H_{22}|\,|H_{11.2}|\,,$$

were

$$H_{11.2} = H_{11} - H_{12}H_{22}^{-1}H_{21}\,, \qquad \text{e} \qquad H_{22.1} = H_{22} - H_{21}H_{11}^{-1}H_{12}\,,$$

following a procedure similar to the one used by Morrison (2005, sec. 4.2), considering the matrix

$$H = \begin{bmatrix} -1 & \sqrt{\frac{n_1 n_2}{n_1+n_2}} \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' \\ \sqrt{\frac{n_1 n_2}{n_1+n_2}} \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) & S \end{bmatrix}$$

where S is the matrix in (7.2), for $q = 2$.

Then we have

$$\begin{aligned}
|H| &= -\left| S + \frac{n_1 n_2}{n_1 + n_2} \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' \right| && (7.21) \\
&= |S| \left(-1 - \frac{n_1 n_2}{n_1 + n_2} \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' S^{-1} \left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) \right), && (7.22)
\end{aligned}$$

so that equating (7.21) and (7.22) we have

$$\left|S+\frac{n_1n_2}{n_1+n_2}\left(\underline{\overline{X}}_1-\underline{\overline{X}}_2\right)\left(\underline{\overline{X}}_1-\underline{\overline{X}}_2\right)'\right| = |S|\left(1+\underbrace{\frac{n_1n_2}{n_1+n_2}\left(\underline{\overline{X}}_1-\underline{\overline{X}}_2\right)'S^{-1}\left(\underline{\overline{X}}_1-\underline{\overline{X}}_2\right)}_{T^2/(n-2)}\right)$$

or, taking into account the definition of the statistic T^2 in (6.13), the equality in (7.20) and the definition of the matrix B in (7.4) for $q=2$,

$$\begin{aligned} 1+\frac{T^2}{n-2} &= \frac{\left|S+\frac{n_1n_2}{n_1+n_2}\left(\underline{\overline{X}}_1-\underline{\overline{X}}_2\right)\left(\underline{\overline{X}}_1-\underline{\overline{X}}_2\right)'\right|}{|S|} \\ &= \frac{\left|S+n_1\left(\underline{\overline{X}}_1-\underline{\overline{X}}\right)\left(\underline{\overline{X}}_1-\underline{\overline{X}}\right)'+n_2\left(\underline{\overline{X}}_2-\underline{\overline{X}}\right)\left(\underline{\overline{X}}_2-\underline{\overline{X}}\right)'\right|}{|S|} \\ &= \frac{|S+B|}{|S|}=\frac{1}{\Lambda^*} \end{aligned}$$

or

$$\frac{T^2}{n-2}=\frac{1}{\Lambda^*}-1=\frac{1-\Lambda^*}{\Lambda^*}\,,$$

where, from (7.10) and (7.11),

$$\Lambda^* \sim Beta\left(\frac{n-p-1}{2},\frac{p}{2}\right), \tag{7.23}$$

so that in fact we have, with $n=n_1+n_2$,

$$\frac{n-p-1}{p}\,\frac{T^2}{n-2}\sim F_{p,n-p-1}\,,$$

so that the two tests are in fact equivalent, since to reject H_0 if Λ^*_{calc} is smaller than the quantile α of a r.v. with the Beta distribution in (7.23) is in fact equivalent to reject H_0 if

$$\frac{n-p-1}{p}\,\frac{T^2_{calc}}{n-2}\sim F_{p,n-p-1}>f_{p,n-p-1}(1-\alpha)\,.$$

7.4 Relation with the univariate case

Note that for $p=1$ we would have the univariate case, that is, a univariate Analysis of Variance model, with a singlr factor with q levels, situation in which we know that we might use an $F_{q-1,n-q}$ statistic to implement the test to the equality of the means of the response variable being analyzed, for the q levels of the factor.

In fact the approach we followed reduces exactly to this test for the case $p=1$, which shows also that the test obtained from a linear regression model, and that as such uses mean square estimators, coincides with the likelihood ratio test. Note that for $p=1$ we have, from the results in section 7.2,

$$\Lambda^* \overset{st}{\sim} Beta\left(\frac{n-q}{2}, \frac{q-1}{2}\right) \quad \Longleftrightarrow \quad \frac{1-\Lambda^*}{\Lambda^*}\,\frac{n-q}{q-1} \sim F_{q-1,n-q}\,,$$

given the well-known relation between r.v.'s with a Beta and an F distribution, being in fact the statistic

$$F = \frac{1-\Lambda^*}{\Lambda^*}\,\frac{n-q}{q-1}\,,$$

the statistic commonly used in a univariate Analysis of Variance model.

7.5 The exact distribution of Λ^* for even p or odd q

For even p we have, from (7.12), taking $W=-\log\Lambda^*$, and using the relation

$$\frac{\Gamma(a+n)}{\Gamma(a)} = \prod_{\ell=0}^{n-1}(a+\ell)\,, \tag{7.24}$$

which is valid for any $a \in \mathbb{C}$ and $n \in \mathbb{N}_0$,

$$\begin{aligned}
\Phi_W(t) &= E\left(e^{\mathrm{i}tW}\right) = E\left(e^{-\mathrm{i}t\log\Lambda^*}\right) = E\left((\Lambda^*)^{-\mathrm{i}t}\right) \\
&= \prod_{j=1}^{q-1} \frac{\Gamma\left(\frac{n-j}{2}\right)\,\Gamma\left(\frac{n-j}{2}-\frac{p}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-j}{2}-\frac{p}{2}\right)\,\Gamma\left(\frac{n-j}{2}-\mathrm{i}t\right)} \\
&= \prod_{j=1}^{q-1}\prod_{\ell=0}^{\frac{p}{2}-1}\left(\frac{n-j}{2}-\frac{p}{2}+\ell\right)\left(\frac{n-j}{2}-\frac{p}{2}+\ell-\mathrm{i}t\right)^{-1} \qquad (7.25)\\
&= \prod_{j=1}^{p+q-3}\left(\frac{n-2-j}{2}\right)^{r_j}\left(\frac{n-2-j}{2}-\mathrm{i}t\right)^{-r_j}
\end{aligned}$$

with

$$r_j = \begin{cases} h_j\,, & j=1,2 \\ h_j + r_{j-2}\,, & j=3,\dots,p+q-3 \end{cases} \qquad (7.26)$$

where

$$h_j = (\#\text{ of elements in } \{p, q-1\} \geq j) - 1\,, \quad j=1,\dots,p+q-3 \qquad (7.27)$$

and for odd q, from (7.8), and using again (7.24),

$$\begin{aligned}
\Phi_W(t) &= E\left((\Lambda^*)^{-\mathrm{i}t}\right) \\
&= \prod_{j=1}^{p} \frac{\Gamma\left(\frac{n-j}{2}\right)\,\Gamma\left(\frac{n-j}{2}-\frac{q-1}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-j}{2}-\frac{q-1}{2}\right)\,\Gamma\left(\frac{n-j}{2}-\mathrm{i}t\right)} \\
&= \prod_{j=1}^{p}\prod_{\ell=0}^{\frac{q-1}{2}-1}\left(\frac{n-j}{2}-\frac{q-1}{2}+\ell\right)\left(\frac{n-j}{2}-\frac{q-1}{2}+\ell-\mathrm{i}t\right)^{-1} \\
&= \prod_{j=1}^{p+q-3}\left(\frac{n-2-j}{2}\right)^{r_j}\left(\frac{n-2-j}{2}-\mathrm{i}t\right)^{-r_j}
\end{aligned} \qquad (7.28)$$

with r_j given by (7.26) and (7.27), which is in fact the same expression that was obtained in (7.25).

The expressions in (7.25) and (7.28) show that the exact distribution of W is in these cases a GIG (Generalized Integer Gamma) distribution of depth $p+q-3$ (the GIG distribution of depth p is the distribution of the sum of p

independent r.v.'s with Gamma distributions with integer shape parameters and different scale parameters (Coelho, 1998; Coelho and Arnold, 2019, Chap. 2 and Appendix 2.A)), with shape parameters r_j and rate parameters $\frac{n-2-j}{2}$ $(j=1,\dots,p+q-3)$.

The p.d.f. of W is thus given by (Coelho, 1998; Coelho and Arnold, 2019)

$$f_W(w) = K \sum_{j=1}^{p+q-3} P_j(w)\, e^{-\lambda_j w}\,, \qquad (w>0)$$

and the c.d.f. by

$$F_W(w) = 1 - K \sum_{j=1}^{p+q-3} P_j^*(w)\, e^{-\lambda_j w}\,, \qquad (w>0)$$

where

$$K = \prod_{j=1}^{p+q-3} \lambda_j^{r_j}\,, \qquad P_j(w) = \sum_{k=1}^{r_j} c_{j,k}\, w^{k-1}$$

and

$$P_j^*(w) = \sum_{k=1}^{r_j} c_{j,k}(k-1)! \sum_{i=0}^{k-1} \frac{w^i}{i!\,\lambda_j^{k-i}}\,,$$

with

$$c_{j,r_j} = \frac{1}{(r_j-1)!} \prod_{\substack{i=1\\ i\neq j}}^{p+q-3} (\lambda_i - \lambda_j)^{-r_i}\,, \qquad j=1,\dots,p+q-3\,,$$

and, for $k=1,\dots,r_j-1$ e $j=1,\dots,p+q-3$,

$$c_{j,r_j-k} = \frac{1}{k} \sum_{i=1}^{k} \frac{(r_j-k+i-1)!}{(r_j-k-1)!}\, R(i,j,p) c_{j,r_j-(k-i)}\,,$$

where

$$R(i,j,p) = \sum_{\substack{k=1\\ k\neq j}}^{p+q-3} r_k(\lambda_j-\lambda_k)^{-i} \qquad (i=1,\dots,r_j-1)\,,$$

being thus the exact distribution of Λ^* an EGIG (Exponentiated Generalized Integer Gamma) distribution (Arnold et al., 2013; Coelho and Arnold, 2019, Chap. 2 and Appendix 2.A) with p.d.f.

$$f_{\Lambda^*}(\ell) = K \sum_{j=1}^{p+q-3} P_j(-\log\,\ell)\,\ell^{\lambda_j - 1}\,, \qquad (0<\ell<1)$$

and c.d.f.

$$F_{\Lambda^*}(\ell) = 1 - K \sum_{j=1}^{p+q-3} P_j^*(-\log\,\ell)\,\ell^{\lambda_j}\,, \qquad (0<\ell<1)\,,$$

and where the intermutability between p and $q-1$ is clear.

7.6 The distribution of Λ^* for odd p and even q

When p is odd and q is even, the exact distributions of Λ^* and Λ have p.d.f.'s and c.d.f.'s with expressions too complicated to be used in practice, while the existing asymptotic distributions do not perform well neither for higher values of p nor for small samples, where for 'small samples' we understand the samples where n exceeds $p+q-1$ by small values (note that from (7.9), (7.11), (7.8) and (7.12) it is clear that we need to have $n > p+q-1$).

In these cases we should use the near-exact distributions (Coelho, 2004) which are asymptotic distributions developed using a different approach based on a factorization of the characteristic function of W. These near-exact distributions are manageable distributions that lie extremely close to the exact distribution even for very small samples and high values of p, being, opposite to the common asymptotic distributions, also asymptotic for increasing values of p and q, and as such extremely adequate for 'big-data' situations where $p\to\infty$ with $n/p\searrow 1$, that is, with $n/p>1$ but tending to 1.

These distributions are developed by factorizing the characteristic function (c.f.) of W in such a way that we will be able to collect the major part of the factors in a c.f. that we identify as corresponding to a known manageable distribution and the remainder factors in another c.f. which we will be able to approximate asymptotically (for increasing sample sizes) by another c.f., in such a way that the product of these two c.f.'s corresponds to a c.f. of a known and manageable distribution, from which we will be able to compute quantiles and p-values. This distribution should show an asymptotic behavior for increasing values of all the parameters in the original distribution, namely the ones pertaining to the sample size and to the number of variables involved.

When p is odd and q is even we may write the c.f. of W as

$$
\begin{aligned}
\Phi_W(t) &= \left\{\prod_{j=1}^{q-1}\frac{\Gamma\left(\frac{n-j-1}{2}\right)\,\Gamma\left(\frac{n-j-p}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-j-p}{2}\right)\,\Gamma\left(\frac{n-j-1}{2}-\mathrm{i}t\right)}\right\}\left\{\prod_{j=1}^{q-1}\frac{\Gamma\left(\frac{n-j}{2}\right)\,\Gamma\left(\frac{n-j-1}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-j-1}{2}\right)\,\Gamma\left(\frac{n-j}{2}-\mathrm{i}t\right)}\right\}\\
&= \left\{\prod_{j=1}^{q-1}\prod_{\ell=0}^{\frac{p-1}{2}-1}\left(\frac{n-j-p}{2}+\ell\right)\left(\frac{n-j-p}{2}+\ell-\mathrm{i}t\right)^{-1}\right\}\\
&\qquad\times\left\{\prod_{j=2}^{q-1}\frac{\Gamma\left(\frac{n-j}{2}\right)\,\Gamma\left(\frac{n-j-1}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-j-1}{2}\right)\,\Gamma\left(\frac{n-j}{2}-\mathrm{i}t\right)}\right\}\frac{\Gamma\left(\frac{n-1}{2}\right)\,\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\,\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)}\\
&= \left\{\prod_{j=1}^{p+q-4}\left(\frac{n-3-j}{2}\right)^{r_j^*}\left(\frac{n-3-j}{2}-\mathrm{i}t\right)^{-r_j^*}\right\}\\
&\qquad\times\left\{\prod_{\substack{j=2\\ \text{step } 2}}^{q-1}\frac{\Gamma\left(\frac{n-j}{2}\right)\,\cancel{\Gamma\left(\frac{n-j-1}{2}-\mathrm{i}t\right)}}{\cancel{\Gamma\left(\frac{n-j-1}{2}\right)}\,\Gamma\left(\frac{n-j}{2}-\mathrm{i}t\right)}\;\frac{\cancel{\Gamma\left(\frac{n-j-1}{2}\right)}\,\Gamma\left(\frac{n-j-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-j-2}{2}\right)\,\cancel{\Gamma\left(\frac{n-j-1}{2}-\mathrm{i}t\right)}}\right\}\\
&\qquad\times\frac{\Gamma\left(\frac{n-1}{2}\right)\,\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\,\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)}\\
&= \left\{\prod_{j=1}^{p+q-4}\left(\frac{n-p-q+j}{2}\right)^{r_j^*}\left(\frac{n-p-q+j}{2}-\mathrm{i}t\right)^{-r_j^*}\right\}\\
&\qquad\times\left\{\prod_{\substack{j=2\\ \text{step } 2}}^{q-1}\left(\frac{n-j-2}{2}\right)\left(\frac{n-j-2}{2}-\mathrm{i}t\right)^{-1}\right\}\frac{\Gamma\left(\frac{n-1}{2}\right)\,\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\,\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)}\\
&= \underbrace{\left\{\prod_{j=1}^{p+q-4}\left(\frac{n-p-q+j}{2}\right)^{r_j}\left(\frac{n-p-q+j}{2}-\mathrm{i}t\right)^{-r_j}\right\}}_{\Phi_{W,1}(t)}\underbrace{\frac{\Gamma\left(\frac{n-1}{2}\right)\,\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\,\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)}}_{\Phi_{W,2}(t)}
\end{aligned}
$$

where

$$
r_j^* = \begin{cases} h_j^*\,, & j=1,2\\ h_j^*+r_{j-2}^*\,, & j=3,\dots,p+q-4\end{cases}
$$

with

$$
h_j^* = (\#\text{ of elements in } \{p-1,q-1\}\geq j)-1 \qquad j=1,\dots,p+q-4\,,
$$

and

$$r_j = \begin{cases} h_j\,, & j = 1,2 \\ h_j + r_{j-2}\,, & j = 3,\dots,p+q-4 \end{cases}$$

with

$$h_j = (\#\text{ of elements in } \{p, q-1\} \geq j) - 1 \qquad j = 1,\dots,p+q-4\,.$$

Note that the r_j^* are symmetrical relative to their indexation, that is, $r_j^* = r_{p+q-3-j}^*$ for $j = 1,\dots,p+q-4$, but the same does not happen with the r_j.

In order to build a near-exact distribution we will then keep the c.f. $\Phi_{W,1}(t)$ untouched, since it corresponds to a GIG distribution, of depth $p+q-4$, with shape parameters r_j and rates $\frac{n-p-q+j}{2}$ $(j = 1,\dots,p+q-4)$, and approximate asymptotically $\Phi_{W,2}(t)$ which is the c.f. of a *Logbeta* $\left(\frac{n-2}{2}, \frac{1}{2}\right)$ distribution.

In fact from the two first expressions in section 5 of Tricomi and Erdélyi (1951) and also from expressions (11) and (14) in the same paper, we may write

$$\frac{\Gamma(a-\mathrm{i}t)}{\Gamma(a+b-\mathrm{i}t)} \simeq \sum_{k=0}^{\infty} p_k(b)(a-\mathrm{i}t)^{-b-k}$$

where $p_0(b) = 1$ and for $k = 1, 2, \dots,$

$$p_k(b) = \frac{1}{k}\sum_{m=0}^{k-1}\left(\frac{\Gamma(1-b-m)}{\Gamma(-b-k)(k-m+1)!} + (-1)^{k+m}b^{k-m+1}\right)p_m(b)\,,$$

so that, since the h-th moment of a r.v. X with a *Beta*(a, b) distribution is given by

$$E(X^h) = \frac{\Gamma(a+b)}{\Gamma(a)}\,\frac{\Gamma(a+h)}{\Gamma(a+b+h)}\,,$$

expression that remains valid for any $h \in \mathbb{C}$, we may write the c.f. of the r.v. $Y = -\log X$, which will be a r.v. with a *Logbeta*(a, b) distribution, as

$$\begin{aligned}
\Phi_Y(t) &= E\left(e^{\mathrm{i}tY}\right) = E\left(e^{-\mathrm{i}t\log X}\right) = E\left(X^{-\mathrm{i}t}\right) \\
&= \frac{\Gamma(a+b)}{\Gamma(a)}\,\frac{\Gamma(a-\mathrm{i}t)}{\Gamma(a+b-\mathrm{i}t)} \simeq \frac{\Gamma(a+b)}{\Gamma(a)}\sum_{k=0}^{\infty} p_k(b)\,(a-\mathrm{i}t)^{-(b+k)} \\
&= \sum_{k=0}^{\infty} \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)}\,\frac{p_k(b)}{a^{b+k}}}_{p_k^*(a,b)}\, a^{b+k}(a-\mathrm{i}t)^{-(b+k)}\,,
\end{aligned}$$

which is the c.f. of an infinite mixture of $\Gamma(b+k,a)$ distributions, with weights $p_k^*(a,b)$.

This way, using a somewhat heuristic approach, we will approximate $\Phi_{W,2}(t)$ by the c.f. of a finite mixture of $\Gamma\left(\frac{1}{2}+k, \frac{n-2}{2}\right)$ distributions. More precisely, we will, for a given $m^* \in \mathbb{N}$, approximate $\Phi_{W,2}(t)$ by

$$\Phi_2^*(t) = \sum_{k=0}^{m^*} \pi_k \left(\frac{n-2}{2}\right)^{\frac{1}{2}+k} \left(\frac{n-2}{2} - \mathrm{i}t\right)^{-\left(\frac{1}{2}+k\right)}$$

where the weights π_k, $k=0,\ldots,m^*-1$ will be determined in such a way that the first m^* derivatives of $\Phi_{W,2}(t)$ and $\Phi_2^*(t)$ relative to t, at $t=0$, are equal, i.e., in such a way that

$$\left.\frac{\partial^h}{\partial t^h}\Phi_{W,2}(t)\right|_{t=0} = \left.\frac{\partial^h}{\partial t^h}\Phi_2^*(t)\right|_{t=0}, \qquad h=1,\ldots,m^*, \tag{7.29}$$

and with $\pi_{m^*} = 1 - \sum_{k=0}^{m^*} \pi_k$.

This way we will have as near-exact c.f. for W

$$\Phi_W^*(t) = \Phi_{W,1}(t)\,\Phi_2^*(t) = \sum_{k=0}^{m^*} \pi_k \left(\frac{n-2}{2}\right)^{\frac{1}{2}+k} \left(\frac{n-2}{2} - \mathrm{i}t\right)^{-\left(\frac{1}{2}+k\right)} \times \prod_{j=1}^{p+q-4} \left(\frac{n-p-q+j}{2}\right)^{r_j} \left(\frac{n-p-q+j}{2} - \mathrm{i}t\right)^{-r_j} \tag{7.30}$$

which is the c.f. of a mixture with m^*+1 components, each of which is a GNIG (Generalized Near-Integer Gamma) distribution (the GNIG distribution of depth p is the distribution of the sum of p independent r.v.'s with Gamma distribution with different scale parameters, $p-1$ of which have integer shape parameters and the remaining one a non-integer shape parameter (Coelho, 2004); the GNIG distribution of depth p reduces to a GIG distribution of depth p when the r.v. that is supposed to have a non-integer shape parameter has in fact an integer shape parameter) with depth $p+q-3$ and with weights π_k $(k=0,\ldots,m^*)$.

To these near-exact distributions correspond p.d.f.'s for W given by

$$\sum_{k=0}^{m^*} \pi_k f^{GNIG}\left(w \,\middle|\, \{r_j\}_{j=1:p+q-4}, \tfrac{1}{2}+k; \left\{\tfrac{n-p-q+j}{2}\right\}_{j=1:p+q-4}, \tfrac{n-2}{2}\right)$$

and c.d.f.'s given by

$$\sum_{k=0}^{m^*}\pi_k F^{GNIG}\left(w\,\middle|\,\{r_j\}_{j=1:p+q-4},\tfrac{1}{2}+k;\left\{\tfrac{n-p-q+j}{2}\right\}_{j=1:p+q-4},\tfrac{n-2}{2}\right)$$

and p.d.f.'s for Λ^* given by

$$\sum_{k=0}^{m^*}\pi_k f^{GNIG}\left(-\log\,\ell\,\middle|\,\{r_j\}_{j=1:p+q-4},\tfrac{1}{2}+k;\left\{\tfrac{n-p-q+j}{2}\right\}_{j=1:p+q-4},\tfrac{n-2}{2}\right)\tfrac{1}{\ell}$$

and c.d.f.'s given by

$$1-\sum_{k=0}^{m^*}\pi_k F^{GNIG}\left(-\log\,\ell\,\middle|\,\{r_j\}_{j=1:p+q-4},\tfrac{1}{2}+k;\left\{\tfrac{n-p-q+j}{2}\right\}_{j=1:p+q-4},\tfrac{n-2}{2}\right),$$

where

$$f^{GNIG}\left(w\,\middle|\,\{r_j\}_{j=1:p},r;\{\lambda_j\}_{j=1:p},\lambda\right)=K\lambda^r e^{-\lambda w}\sum_{j=1}^{p}P_j(w)$$

and

$$\begin{aligned}F^{GNIG}\left(w\,\middle|\,\{r_j\}_{j=1:p},r;\{\lambda_j\}_{j=1:p},\lambda\right)=\lambda^r\tfrac{w^r}{\Gamma(r+1)}\,{}_1F_1(r,r+1,-\lambda w)\\ -K\lambda^r e^{-\lambda w}\sum_{j=1}^{p}P_j^*(w)\end{aligned}$$

are respectively the p.d.f. and the c.d.f. of a GNIG distribution, with

$$P_j(w)=\sum_{k=1}^{r_j}c_{j,k}\frac{\Gamma(k)}{\Gamma(k+r)}w^{k+r-1}\,{}_1F_1(k,k+r,(\lambda-\lambda_j)w)$$

and

$$P_j^*(w)=\sum_{k=1}^{r_j}c_{j,k}^*\sum_{i=0}^{k-1}\frac{w^{r+i}}{\Gamma(r+1+i)}\,{}_1F_1(i+1,r+1+i,(\lambda-\lambda_j)w)$$

with

$$c_{j,k}^*=\frac{c_{j,k}}{\lambda_j^k}\,\Gamma(k)$$

and with

$$ {}_1F_1(a,b,w) = \sum_{i=0}^{\infty} \frac{\Gamma(a+i)\,\Gamma(b)}{\Gamma(a)\,\Gamma(b+i)}\,\frac{w^i}{i!} $$

representing the so-called Kummer confluent hypergeometric function.

Note that solving the system of equations in (7.29), using an adequate software, is very simple, since this is a linear system of equations.

7.6.1 Measuring the distance between the near-exact and the exact distributions

Although we do not have at hand the expression for the exact c.d.f. of Λ or Λ^*, we may measure the distance between the near-exact c.d.f.'s and the exact c.d.f., by using a measure based on c.f.'s, which is a very sharp upper-bound on the distance between these c.d.f.'s, the near-exact and the exact ones.

This measure is

$$ \Delta = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| \frac{\Phi_W(t) - \Phi_W^+(t)}{t} \right| dt $$

with

$$ \Delta \geq \sup_{w>0} \left| F_W(w) - F_W^+(w) \right| = \sup_{0<\ell<1} \left| F_\Lambda(\ell) - F_\Lambda^+(\ell) \right| \,, $$

where $\Phi_W(t)$ represents the exact c.f. of W and $\Phi_W^+(t)$ the approximate, near-exact or asymptotic, and where $F_W(\,\cdot\,)$ and $F_\Lambda(\,\cdot\,)$ are, respectively, the exact c.d.f.'s of $W = -\log\Lambda$ and Λ, and $F_W^+(\,\cdot\,)$ is the c.d.f. that corresponds to $\Phi_W^+(\,\cdot\,)$, with $F_\Lambda^+(\ell) = 1 - F_W^+(-\log\ell)$.

In Table 7.1 we may see values of the measure Δ for the near-exact distributions with c.f. given by $\Phi_W^*(t)$ in (7.30), for different values of p and q and for (overall) samples with dimensions $n = (p+1)q + 0, 50, 100$, and where the measure Δ is computed using $\Phi_W(t)$ on page 91 as the exact c.f. of W and $\Phi_W^*(t)$ in (7.30) as the approximate (near-exact) c.f. of W.

We may see as the near-exact distributions exhibit a clear asymptotic character not only for increasing values of the sample size but also for increasing values of p and q, which is not the case with the usual asymptotic distributions, as well as the very good performance for small samples.

A pertinent question is the one about how small really are the values in Table 7.1. To have a more precise idea, we may note that for $p = 25$, $q = 4$

Table 7.1. – Values of the measure Δ for near-exact distributions that match m^* exact moments
of W, for different values of p and q and samples of size $n = (p+1)q + 0, 50, 100$

			m^*				
p	q	n	0	2	4	6	10
3	4	16	2.99×10^{-3}	3.53×10^{-7}	6.70×10^{-10}	2.74×10^{-12}	1.31×10^{-15}
		66	7.51×10^{-4}	5.06×10^{-9}	6.01×10^{-13}	2.29×10^{-16}	2.39×10^{-22}
		116	4.28×10^{-4}	9.21×10^{-10}	3.49×10^{-14}	4.30×10^{-18}	5.08×10^{-25}
7	4	32	8.97×10^{-4}	1.07×10^{-8}	2.65×10^{-12}	2.23×10^{-15}	7.12×10^{-21}
		82	3.73×10^{-4}	7.34×10^{-10}	2.99×10^{-14}	4.36×10^{-18}	8.55×10^{-25}
		132	2.35×10^{-4}	1.80×10^{-10}	2.85×10^{-15}	1.62×10^{-19}	5.05×10^{-27}
15	4	64	2.95×10^{-4}	3.94×10^{-10}	1.19×10^{-14}	1.38×10^{-18}	1.94×10^{-25}
		114	1.76×10^{-4}	8.15×10^{-11}	8.47×10^{-16}	3.39×10^{-20}	6.09×10^{-28}
		164	1.24×10^{-4}	2.89×10^{-11}	1.49×10^{-16}	2.96×10^{-21}	1.32×10^{-29}
25	4	104	1.39×10^{-4}	4.13×10^{-11}	2.85×10^{-16}	7.83×10^{-21}	7.16×10^{-29}
		154	9.81×10^{-5}	1.45×10^{-11}	4.96×10^{-17}	6.70×10^{-22}	1.50×10^{-30}
		204	7.57×10^{-5}	6.64×10^{-12}	1.34×10^{-17}	1.07×10^{-22}	8.31×10^{-32}
55	4	224	4.29×10^{-5}	1.23×10^{-12}	8.36×10^{-19}	2.30×10^{-24}	2.31×10^{-34}
		274	3.61×10^{-5}	7.31×10^{-13}	3.50×10^{-19}	6.77×10^{-25}	3.35×10^{-35}
		324	3.11×10^{-5}	4.67×10^{-13}	1.65×10^{-19}	2.37×10^{-25}	6.39×10^{-36}
15	6	96	1.57×10^{-4}	5.92×10^{-11}	5.16×10^{-16}	1.77×10^{-20}	2.47×10^{-28}
		146	1.06×10^{-4}	1.84×10^{-11}	7.33×10^{-17}	1.15×10^{-21}	3.42×10^{-30}
		196	8.04×10^{-5}	7.94×10^{-12}	1.79×10^{-17}	1.60×10^{-22}	1.54×10^{-31}
15	12	192	5.41×10^{-5}	2.47×10^{-12}	2.64×10^{-18}	1.14×10^{-23}	2.74×10^{-33}
		242	4.35×10^{-5}	1.28×10^{-12}	8.80×10^{-19}	2.45×10^{-24}	2.45×10^{-34}
		292	3.63×10^{-5}	7.45×10^{-13}	3.57×10^{-19}	6.92×10^{-25}	3.37×10^{-35}
15	20	320	2.49×10^{-5}	2.42×10^{-13}	5.57×10^{-20}	5.20×10^{-26}	6.06×10^{-37}
		370	2.17×10^{-5}	1.60×10^{-13}	2.78×10^{-20}	1.97×10^{-26}	1.32×10^{-37}
		420	1.92×10^{-5}	1.11×10^{-13}	1.51×10^{-20}	8.39×10^{-27}	3.44×10^{-38}

and $n = 204$ the near-exact distribution that matches the first 10 exact moments of W has a value of Δ equal to 8.31×10^{-32}. It happens that if we make an error of this magnitude in measuring the diameter of the Milky Way, which including the so-called 'tidal streams', is about 120 000 light years (being one light year about $9.46\times10^{12} Km$), we will make an error of about 0.094 nanometers, which would be less than half of the diameter of a carbon atom, which is about 0.22 nanometers.

However, it is possible to obtain even better results if we use a slightly different heuristic approach, where the rate parameter in $\Phi_2^*(t)$ instead of being fixed at $\frac{n-2}{2}$ is determined as the rate parameter λ^* in the c.f.

$$\Phi^*(t) = \theta(\lambda^*)^{r_1}(\lambda^* - \mathrm{i}t)^{-r_1} + (1-\theta)(\lambda^*)^{r_2}(\lambda^* - \mathrm{i}t)^{-r_2}$$

where θ, λ^*, r_1 and r_2 are jointly determined, in such a way that

$$\left.\frac{\partial^h}{\partial t^h}\Phi_{W,2}(t)\right|_{t=0} = \left.\frac{\partial^h}{\partial t^h}\Phi^*(t)\right|_{t=0}, \qquad h=1,\ldots,4\,.$$

All the rest is done in exactly the same way as before. The values of the measure Δ for these new near-exact distributions are shown in Table 7.2, where we may see that, in general, they assume smaller values than those in Table 7.1, thus showing that the near-exact distributions developed using this new procedure will lie closer to the exact distribution. This improvement is more notorious for smaller values of m^*, namely for $m^*=0$, and also somewhat more notorious for larger values of p and q.

In Table 7.3 we have the values Δ for three asymptotic distributions for W, together with the values of Δ for the near-exact distribution that matches 2 exact moments of W. These asymptotic distributions are:

as.1: the asymptotic distribution in Muirhead (2005, sec. 10.5.3), which corresponds to a mixture of two chi-square distributions, with c.f., for $W=-\log\Lambda^*$, given by

$$\Phi_W^{as.1}(t) = (1-w_2)\left(\frac{n}{2}\rho\right)^{f/2}\left(\frac{n}{2}\rho-\mathrm{i}t\right)^{-f/2} + w_2\left(\frac{n}{2}\rho\right)^{\frac{f+4}{2}}\left(\frac{n}{2}\rho-\mathrm{i}t\right)^{-\frac{f+4}{2}}$$

where $f=p(q-1)$ and

$$w_2=\frac{p(q-1)(p^2+(q-1)^2-5)}{48(\rho n)^2}\,,\quad \rho = 1-\frac{1}{n}\left[1+\frac{1}{2}(p+q)\right]\,.$$

We may note that this is also the asymptotic distribution in Anderson (1958, sec. 8.6.2; 2003, sec. 8.5.2) if we consider only the 2 first terms.

as.2: the asymptotic distribution above, where we consider only the first term, giving rise to an asymptotic distribution with c.f.

$$\Phi_W^{as.2}(t) = \left(\frac{n}{2}\rho\right)^{f/2}\left(\frac{n}{2}\rho-\mathrm{i}t\right)^{-f/2}\,.$$

Table 7.2. – Values of the measure Δ for the new near-exact distributions that match m^* exact moments of W, for different values of p and q and samples of dimension $n = (p+1)q + 0, 50, 100$

			quase-exactas				
					m^*		
p	q	n	0	2	4	6	10
3	4	16	1.25×10^{-4}	1.18×10^{-7}	6.11×10^{-10}	6.86×10^{-12}	1.28×10^{-15}
		66	5.59×10^{-6}	3.12×10^{-10}	9.12×10^{-14}	6.56×10^{-17}	1.82×10^{-22}
		116	1.76×10^{-6}	3.14×10^{-11}	2.91×10^{-15}	6.65×10^{-19}	1.93×10^{-25}
7	4	32	1.56×10^{-5}	1.57×10^{-9}	9.80×10^{-13}	1.67×10^{-15}	2.78×10^{-20}
		82	2.31×10^{-6}	3.79×10^{-11}	3.75×10^{-15}	1.02×10^{-18}	5.09×10^{-25}
		132	8.84×10^{-7}	5.65×10^{-12}	2.16×10^{-16}	2.27×10^{-20}	1.70×10^{-27}
15	4	64	2.42×10^{-6}	2.73×10^{-11}	2.02×10^{-15}	4.42×10^{-19}	1.69×10^{-25}
		114	7.84×10^{-7}	3.06×10^{-12}	7.71×10^{-17}	5.69×10^{-21}	2.50×10^{-28}
		164	3.82×10^{-7}	7.42×10^{-13}	9.26×10^{-18}	3.38×10^{-22}	3.60×10^{-30}
25	4	104	6.86×10^{-7}	1.72×10^{-12}	2.90×10^{-17}	1.47×10^{-21}	3.33×10^{-29}
		154	3.24×10^{-7}	4.03×10^{-13}	3.33×10^{-18}	8.26×10^{-23}	4.43×10^{-31}
		204	1.87×10^{-7}	1.38×10^{-13}	6.71×10^{-19}	9.80×10^{-24}	1.81×10^{-32}
55	4	224	9.69×10^{-8}	2.35×10^{-14}	3.84×10^{-20}	1.94×10^{-25}	4.60×10^{-35}
		274	6.64×10^{-8}	1.13×10^{-14}	1.31×10^{-20}	4.63×10^{-26}	5.39×10^{-36}
		324	4.83×10^{-8}	6.10×10^{-15}	5.21×10^{-21}	1.36×10^{-26}	8.61×10^{-37}
15	6	96	8.41×10^{-7}	2.69×10^{-12}	5.70×10^{-17}	3.63×10^{-21}	1.27×10^{-28}
		146	3.70×10^{-7}	5.40×10^{-13}	5.20×10^{-18}	1.50×10^{-22}	1.07×10^{-30}
		196	2.07×10^{-7}	1.72×10^{-13}	9.36×10^{-19}	1.53×10^{-23}	3.49×10^{-32}
15	12	192	1.43×10^{-7}	5.49×10^{-14}	1.42×10^{-19}	1.12×10^{-24}	6.43×10^{-34}
		242	9.08×10^{-8}	2.25×10^{-14}	3.73×10^{-20}	1.90×10^{-25}	4.48×10^{-35}
		292	6.27×10^{-8}	1.08×10^{-14}	1.25×10^{-20}	4.43×10^{-26}	5.06×10^{-36}
15	20	320	3.93×10^{-8}	3.21×10^{-15}	1.78×10^{-21}	3.04×10^{-27}	8.30×10^{-38}
		370	2.96×10^{-8}	1.83×10^{-15}	7.68×10^{-22}	9.92×10^{-28}	1.55×10^{-38}
		420	2.30×10^{-8}	1.12×10^{-15}	3.67×10^{-22}	3.71×10^{-28}	3.55×10^{-39}

as.3: the asymptotic distribution above, that is, in *as.2*, for $\rho = 1$; note that when $n \to \infty$ we have $\rho \to 1$; this is in fact the asymptotic distribution based on the fact that $-2\log\Lambda \xrightarrow{d} \chi^2_f$, where f is the difference between the number of parameters in H_1 and the number of parameters in H_0, with $f = p(q-1)$; to this asymptotic distribution corresponds thus a c.f., for $W = -\log\Lambda^*$, given by

$$\Phi_W^{as.3}(t) = \left(\frac{n}{2}\right)^{f/2}\left(\frac{n}{2} - \mathrm{i}t\right)^{-f/2} .$$

Table 7.3. – Values of the measure Δ for the near-exact distribution that matches 2 exact moments of W and for the asym+totic distributions, for the same values of p, q and n in Table 7.2

			quase-exacta m^*	assimptóticas		
p	q	n	2	$as.1$	$as.2$	$as.3$
3	4	16	1.18×10^{-7}	9.40×10^{-5}	6.84×10^{-3}	2.97×10^{-1}
		66	3.12×10^{-10}	1.14×10^{-7}	2.37×10^{-4}	6.33×10^{-2}
		116	3.14×10^{-11}	1.05×10^{-8}	8.76×10^{-3}	3.03×10^{-1}
7	4	32	1.57×10^{-9}	1.25×10^{-4}	8.76×10^{-3}	3.03×10^{-1}
		82	3.79×10^{-11}	1.61×10^{-6}	9.92×10^{-4}	1.10×10^{-1}
		132	5.65×10^{-12}	2.11×10^{-7}	3.59×10^{-4}	6.71×10^{-2}
15	4	64	2.73×10^{-11}	2.16×10^{-4}	1.26×10^{-2}	3.45×10^{-1}
		114	3.06×10^{-12}	1.53×10^{-5}	3.36×10^{-3}	1.87×10^{-1}
		164	7.42×10^{-13}	3.16×10^{-6}	1.53×10^{-3}	1.28×10^{-1}
25	4	104	1.72×10^{-12}	3.35×10^{-4}	1.64×10^{-2}	3.96×10^{-1}
		154	4.03×10^{-13}	5.55×10^{-5}	6.67×10^{-3}	2.63×10^{-1}
		204	1.38×10^{-13}	1.61×10^{-5}	3.60×10^{-3}	1.96×10^{-1}
55	4	224	2.35×10^{-14}	6.84×10^{-4}	2.46×10^{-2}	5.15×10^{-1}
		274	1.13×10^{-14}	2.72×10^{-4}	1.55×10^{-2}	4.23×10^{-1}
		324	6.10×10^{-15}	1.29×10^{-4}	1.06×10^{-2}	3.57×10^{-1}
15	6	96	2.69×10^{-12}	6.28×10^{-5}	6.99×10^{-3}	3.13×10^{-1}
		146	5.40×10^{-13}	9.75×10^{-6}	2.75×10^{-3}	2.02×10^{-1}
		196	1.72×10^{-13}	2.75×10^{-6}	1.46×10^{-3}	1.49×10^{-1}
15	12	192	5.49×10^{-14}	1.30×10^{-5}	3.27×10^{-3}	2.82×10^{-1}
		242	2.25×10^{-14}	4.81×10^{-6}	1.99×10^{-3}	2.24×10^{-1}
		292	1.08×10^{-14}	2.17×10^{-6}	1.34×10^{-3}	1.85×10^{-1}
15	20	320	3.21×10^{-15}	7.39×10^{-6}	2.54×10^{-3}	2.81×10^{-1}
		370	1.83×10^{-15}	4.00×10^{-6}	1.87×10^{-3}	2.43×10^{-1}
		420	1.12×10^{-15}	2.35×10^{-6}	1.43×10^{-3}	2.14×10^{-1}

We may see how the asymptotic distribution *as.1* shows a better performance than the distribution *as.2* and this one a better performance than the distribution *as.3*. While the distribution *as.1* is even able to outperform the near-exact distribution for $m^* = 0$ for smaller values of p, we may see that although all the asymptotic distributions exhibit a clear asymptotic behavior for increasing sample sizes, and also for increasing values of q, that is not the case for increasing values of p. In fact, the asymptotic distribution *as.1* is only able to outperform the near-exact distribution with $m^* = 0$ for $p = 3$, while for $p = 7$ that does not happen anymore.

Table 7.4. – Values of the measure Δ for the near-exact distributions that match m^* exact moments of W, for different values of p and q and very small samples of dimension $n = p + q + 1$

			m^*				
p	q	n	0	2	4	6	10
3	4	8	4.67×10^{-3}	1.62×10^{-6}	5.44×10^{-9}	1.61×10^{-10}	1.72×10^{-13}
			5.12×10^{-3}	1.51×10^{-6}	6.61×10^{-9}	1.12×10^{-10}	1.57×10^{-13}
7	4	12	1.41×10^{-3}	5.19×10^{-8}	3.79×10^{-11}	2.79×10^{-14}	3.56×10^{-17}
			1.02×10^{-4}	3.02×10^{-8}	6.87×10^{-11}	3.86×10^{-13}	4.39×10^{-17}
15	4	20	4.02×10^{-4}	1.23×10^{-9}	9.82×10^{-14}	2.86×10^{-17}	5.98×10^{-23}
			1.30×10^{-5}	3.41×10^{-10}	7.39×10^{-14}	5.28×10^{-17}	1.75×10^{-22}
25	4	30	1.61×10^{-4}	7.92×10^{-11}	1.08×10^{-15}	6.39×10^{-20}	1.32×10^{-27}
			3.12×10^{-6}	1.32×10^{-11}	4.63×10^{-16}	5.74×10^{-20}	9.80×10^{-27}
55	4	60	3.66×10^{-5}	9.15×10^{-13}	6.51×10^{-19}	2.19×10^{-24}	4.44×10^{-34}
			3.26×10^{-7}	6.96×10^{-14}	1.23×10^{-19}	7.86×10^{-25}	4.50×10^{-34}
15	6	22	2.83×10^{-4}	4.08×10^{-10}	1.45×10^{-14}	1.90×10^{-18}	3.96×10^{-25}
			8.05×10^{-6}	9.97×10^{-11}	9.54×10^{-15}	2.92×10^{-18}	1.92×10^{-24}
15	12	28	1.45×10^{-4}	5.26×10^{-11}	4.67×10^{-16}	1.61×10^{-20}	4.36×10^{-29}
			3.05×10^{-6}	9.54×10^{-12}	2.20×10^{-16}	1.62×10^{-20}	7.19×10^{-28}
15	20	36	7.97×10^{-5}	8.58×10^{-12}	2.28×10^{-17}	2.45×10^{-22}	2.38×10^{-31}
			1.25×10^{-6}	1.15×10^{-12}	7.81×10^{-18}	1.69×10^{-22}	7.39×10^{-31}
25	6	32	1.22×10^{-4}	3.21×10^{-11}	2.20×10^{-16}	6.28×10^{-21}	3.54×10^{-29}
			2.18×10^{-6}	4.96×10^{-12}	8.72×10^{-17}	5.14×10^{-21}	1.79×10^{-28}
25	12	38	7.25×10^{-5}	6.49×10^{-12}	1.45×10^{-17}	1.33×10^{-22}	1.05×10^{-31}
			1.07×10^{-6}	8.21×10^{-13}	4.65×10^{-18}	8.50×10^{-23}	2.74×10^{-31}
25	20	46	4.52×10^{-5}	1.55×10^{-12}	1.31×10^{-18}	4.60×10^{-24}	6.42×10^{-34}
			5.37×10^{-7}	1.58×10^{-13}	3.35×10^{-19}	2.29×10^{-24}	1.05×10^{-33}
55	6	62	2.96×10^{-5}	4.54×10^{-13}	1.86×10^{-19}	3.39×10^{-25}	1.75×10^{-35}
			2.54×10^{-7}	3.33×10^{-14}	3.37×10^{-20}	1.17×10^{-25}	1.69×10^{-35}
55	12	68	2.07×10^{-5}	1.50×10^{-13}	2.74×10^{-20}	2.15×10^{-26}	1.86×10^{-37}
			1.61×10^{-7}	9.96×10^{-15}	4.50×10^{-21}	6.65×10^{-27}	1.56×10^{-37}
55	20	76	1.52×10^{-5}	5.81×10^{-14}	5.55×10^{-21}	2.24×10^{-27}	5.09×10^{-39}
			1.05×10^{-7}	3.43×10^{-15}	8.04×10^{-22}	6.09×10^{-28}	3.65×10^{-39}

But our attention also goes to situations where $p \to \infty$ and at the same time $n/p \searrow 1$. Although some of the sample sizes taken into account so far were somewhat small, even for the smaller ones we have $n/p \to q$, so that we want to analyze smaller sample sizes, where in fact $n/p \searrow 1$. We decided thus to analyze samples of size $n = p + q + 1$. The values for the measure Δ for cases with sample sizes of these dimensions are shown in Table 7.4, where, for each case, we have in the first row the values for the near-exact distributions derived in the previous section and in the second row the values for the near-exact distributions obtained with the procedure described in the present section. In Table 7.5 we have the values of Δ for the asymptotic distributions.

We may see from the values in Table 7.5 that the asymptotic distributions in these cases do not exhibit anymore an asymptotic behavior for increasing values of q, and in general show too high values ofthe measure Δ, even for the cases where p has quite low values.

Table 7.5. – Values of the measure Δ for the asymptotic distributions, for the same values of p, q and n used in Table 7.4

p	q	n	$as.1$	$as.2$	$as.3$
3	4	8	1.23×10^{-2}	8.19×10^{-2}	7.35×10^{-1}
7	4	12	7.64×10^{-2}	2.29×10^{-1}	9.71×10^{-1}
15	4	20	3.15×10^{-1}	5.17×10^{-1}	1.20×10^{0}
25	4	30	6.57×10^{-1}	7.75×10^{-1}	1.34×10^{0}
55	4	60	1.40×10^{0}	1.13×10^{0}	1.56×10^{0}
15	6	22	3.77×10^{-1}	5.74×10^{-1}	1.28×10^{0}
15	12	28	5.49×10^{-1}	7.05×10^{-1}	1.42×10^{0}
15	20	36	8.24×10^{-1}	8.66×10^{-1}	1.53×10^{0}
25	6	32	7.67×10^{-1}	8.36×10^{-1}	1.42×10^{0}
25	12	38	9.43×10^{-1}	9.21×10^{-1}	1.54×10^{0}
25	20	46	1.15×10^{0}	1.01×10^{0}	1.63×10^{0}
55	6	62	1.63×10^{0}	1.19×10^{0}	1.64×10^{0}
55	12	68	1.94×10^{0}	1.25×10^{0}	1.75×10^{0}
55	20	76	2.13×10^{0}	1.28×10^{0}	1.82×10^{0}

Some of the values of the measure Δ for the asymptotic distributions, in Table 7.5, even exceed the value of 1, which actually, at first sight, does not make much sense. However, these values are correct because they occur in situations where the asymptotic distribution *as.1* is no more a legitimate distribution, with a p.d.f. and c.d.f. which assume negative values for some values of W. Situations where the values of Δ are close to 1 occur when the c.d.f.'s of the asymptotic distributions *as.2* or *as.3* attain the value 1, while the c.d.f. of the exact distribution would still be near zero. This fact may be analyzed in Figure 7.1, where plots of p.d.f.'s and c.d.f.'s are shown for cases of very small sample sizes, namely for $p = 25$, $q = 20$ and $n = 46$.

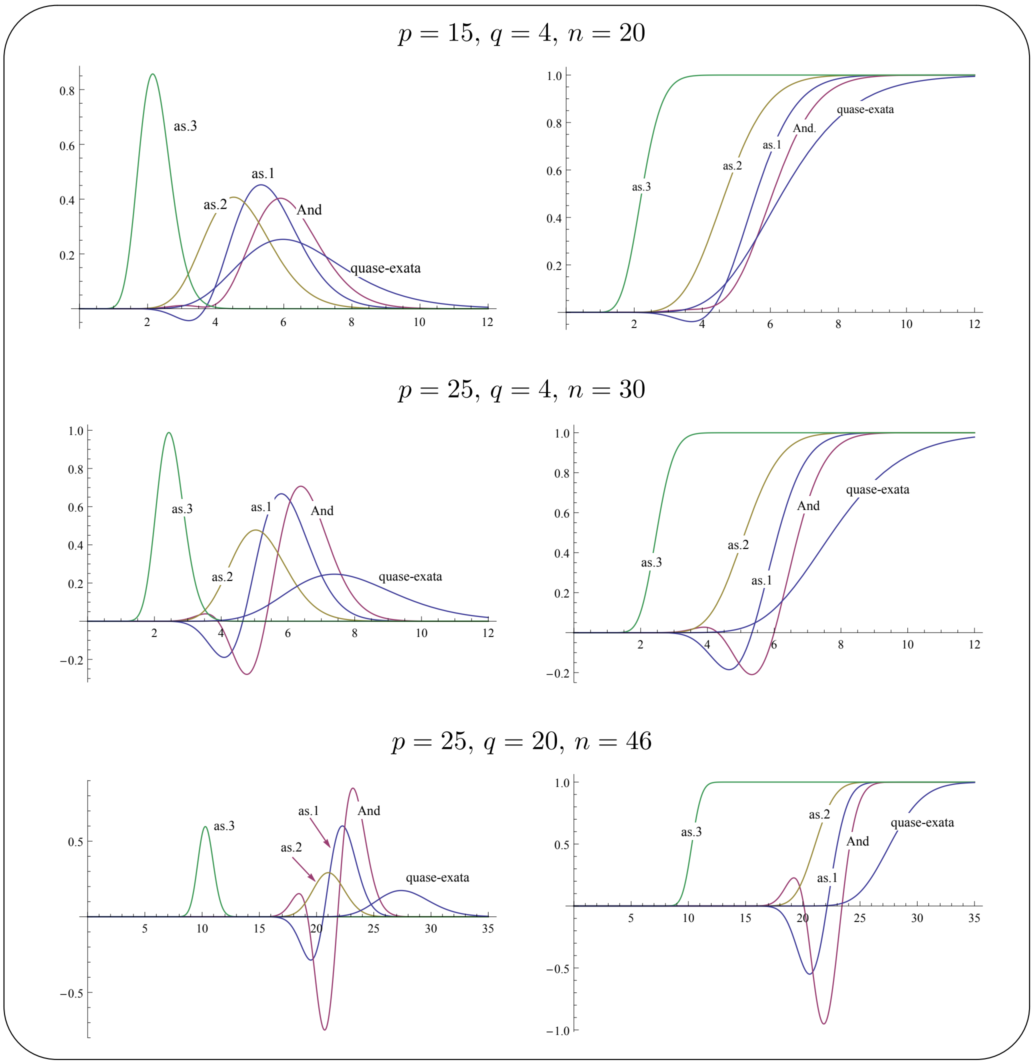

Figure 7.1. – Asymptotic and near-exact p.d.f.'s and c.d.f.'s for some of the cases of very small samples

In Figure 7.1 are also shown plots of the p.d.f.'s and c.d.f.'s of the asymptotic distribution in Anderson (1958, sec. 8.6.2; 2003, sec. 8.5.2), denotaded by *And*, which has as c.f. for $W = -\log\Lambda^*$ the c.f.

$$\Phi_W^{And}(t) = \left(1 - w_2 - \frac{\gamma_2}{(\rho n)^4} + w_2^2\right)\left(\frac{n}{2}\rho\right)^{f/2}\left(\frac{n}{2}\rho - \mathrm{it}\right)^{-f/2} + \left(w_2 - w_2^2\right)\left(\frac{n}{2}\rho\right)^{\frac{f+4}{2}}\left(\frac{n}{2}\rho - \mathrm{it}\right)^{-\frac{f+4}{2}} + \frac{\gamma_2}{(\rho n)^4}\left(\frac{n}{2}\rho\right)^{\frac{f+8}{2}}\left(\frac{n}{2}\rho - \mathrm{it}\right)^{-\frac{f+8}{2}},$$

where w_2, ρ and f are defined above and

$$\gamma_2 = \frac{\gamma^2}{2} + \frac{p(q-1)}{1920}\left[3p^4 + 3(q-1)^4 + 10p^2(q-1)^2 - 50(p^2 + (q-1)^2) + 159\right].$$

This asymptotic distribution is based on the results of Box (1949), and it has as a particular case the asymptotic distribution *as.1*, which although for the samples of a somewhat larger size it shows good values of the measure Δ, namely for the smaller values of p, and although it also shows for these samples an asymptotic behavior for increasing values of q, as it may be seen from the values in Table 7.6, for very small samples it shows, for the larger values of p an even worse behavior than that of the asymptotic distribution *as.1*, as it may be seen from the values in Tables 7.6 and 7.7 and also from the plots in Figure 7.1.

In fact, from the analysis of the plots in Figure 7.1 we may see how for $p = 15$, $q = 4$ and $n = 20$ the asymptotic 'distribution' *as.1* is indeed no more a legitimate distribution, while the asymptotic distribution *And* still keeps a quite good behavior. However, this latter distribution starts to have a worse behavior than the asymptotic distribution *as-1* for larger values of p and q.

In Table 7.7 we may analyze values for the quantiles 0.95 and 0.99 of $W = -\log\Lambda^*$ and the corresponding quantiles 0.05 and 0.01 of Λ^*, for the cases of very small samples considered in Figure 7.1.

From the analysis of the values in Table 7.7 we may once again see that the asymptotic distributions cannot be used in situations where $p \to \infty$ and $n/p \searrow 1$, with the near-exact quantiles of Λ^* being sometimes approximately equal to 10^{-3} times the asymptotic 'quantiles' (we should recall here the fact that we should reject the null hypothesis, for a level α test, if the computed value of the statistic Λ^* is less than the α quantile of Λ^*, or, equivalently, if the computed value of W exceeds the $1 - \alpha$ quantile of W).

Table 7.6. – Values of the measure Δ for the asymptotic distribution of W in Anderson (2003, sec. 8.5.2), for different values of p and q and samples of small dimensions

		n			
p	q	$p+q+1$	$(p+1)q$	$(p+1)q+50$	$(p+1)q+100$
3	4	1.97×10^{-3}	1.36×10^{-6}	5.75×10^{-11}	1.62×10^{-12}
7	4	2.70×10^{-2}	1.96×10^{-6}	2.88×10^{-9}	1.36×10^{-10}
15	4	1.89×10^{-1}	3.82×10^{-6}	7.23×10^{-8}	6.78×10^{-9}
25	4	5.41×10^{-1}	6.91×10^{-6}	4.68×10^{-7}	7.34×10^{-8}
55	4	1.73×10^{0}	1.91×10^{-5}	4.81×10^{-6}	1.56×10^{-6}
15	6	2.43×10^{-1}	5.80×10^{-7}	3.55×10^{-8}	5.32×10^{-9}
15	12	4.18×10^{-1}	5.28×10^{-8}	1.19×10^{-8}	3.61×10^{-9}
15	20	7.70×10^{-1}	2.18×10^{-8}	8.67×10^{-9}	3.90×10^{-9}
25	6	6.84×10^{-1}	9.32×10^{-7}	1.48×10^{-7}	3.63×10^{-8}
25	12	9.53×10^{-1}	4.92×10^{-8}	1.91×10^{-8}	8.42×10^{-9}
25	20	1.32×10^{0}	1.04×10^{-8}	5.82×10^{-9}	3.44×10^{-9}
55	6	2.28×10^{0}	2.55×10^{-6}	1.02×10^{-6}	4.63×10^{-7}
55	12	3.26×10^{0}	1.01×10^{-7}	6.42×10^{-8}	4.22×10^{-8}
55	20	3.98×10^{0}	1.17×10^{-8}	8.94×10^{-9}	6.90×10^{-9}

Tabela 7.7. – Quantiles of order $\alpha = 0.95$ and $\alpha = 0.99$ of W and order $\alpha = 0.05$ and $\alpha = 0.01$ of Λ^*, for the asymptotic distributions and for the near-exact distribution that matches the first 4 exact moments of W, for some of the samples of very small size

	W		Λ^*	
	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.05$	$\alpha = 0.01$
	$p = 15,\ q = 4,\ n = 20$		$p = 15,\ q = 4,\ n = 20$	
as.3	3.0828116688	3.4978416033	$4.5830215935\times10^{-2}$	$3.0262631746\times10^{-2}$
as.2	6.4901298291	7.3638770596	$1.5183518690\times10^{-3}$	$6.3373665558\times10^{-4}$
as.1	7.3481465516	8.2217481154	$6.4378447658\times10^{-4}$	$2.6874485701\times10^{-4}$
And	7.9347163374	8.8388674726	$3.5809353405\times10^{-4}$	$1.4498685532\times10^{-4}$
quase-exata	9.6051662374	11.3927585260	$6.7379736046\times10^{-5}$	$1.1276850904\times10^{-5}$
	$p = 25,\ q = 4,\ n = 30$		$p = 25,\ q = 4,\ n = 30$	
as.3	3.2072223585	3.5464307643	$4.0468865304\times10^{-2}$	$2.8827347849\times10^{-2}$
as.2	6.6356324658	7.3374429607	$1.3127482143\times10^{-3}$	$6.5071229236\times10^{-4}$
as.1	7.4555483871	8.1236188024	$5.7822448226\times10^{-4}$	$2.9645390885\times10^{-4}$
And	8.0399075614	8.7050258003	$3.2233874529\times10^{-4}$	$1.6575068432\times10^{-4}$
quase-exata	11.0475591743	12.8440941292	$1.5925974356\times10^{-5}$	$2.6416835672\times10^{-6}$
	$p = 25,\ q = 20,\ n = 46$		$p = 25,\ q = 20,\ n = 46$	
as.3	11.4523718827	11.9484717167	$1.0624245231\times10^{-5}$	$6.4691119416\times10^{-6}$
as.2	23.4137380713	24.4279866209	$6.7848891226\times10^{-11}$	$2.4607087103\times10^{-11}$
as.1	24.7594588331	25.6410024133	$1.7664593746\times10^{-11}$	$7.3156705832\times10^{-12}$
And	25.6782332481	26.5029586876	$7.0483099721\times10^{-12}$	$3.0896642508\times10^{-12}$
quase-exata	31.8285620334	33.8564429535	$1.5032497559\times10^{-14}$	$1.9784890383\times10^{-15}$

That in fact the near-exact distributions lie very close to the exact distribution and that the common asymptotic distributions are in fact not legitimate distributions for many of the cases of very small samples may be easily seen from the analysis of the plots of the absolute value of the c.f.'s in Figure 7.2. In this Figure we may see that the c.f.'s for the asymptotic distributions *as.1* and *And* are in fact not c.f.'s since their absolute value exceeds 1, for some values of t, so that the 'distributions' that correspond to them are in fact not legitimate distributions, while from the plots in Figure 7.1 we may easily see that for the large majority of cases of small samples the asymptotic distributions *as.2* and *as.1* are approximations that do not exhibit the necessary precision.

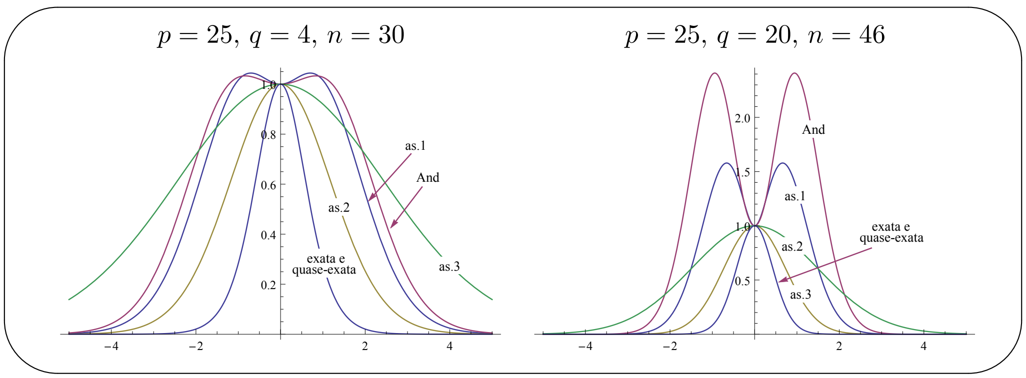

Figure 7.2. – Plots of the absolute values of the characteristic functions for the asymptotic, near-exact and exact distributions, for cases of small samples

The near-exact distributions considered in Figures 7.1 and 7.2 are those that use λ^* as rate parameter and that match the first $m^* = 4$ exact moments of W. That these near-exact distributions lie very close to the exact distribution may be seen from the fact that in Figure 7.2 the plots of the exact and near-exact c.f.'s are indistinguishable.

7.7 Profile Analysis (for several profiles)

Similar to what happened in section 6.5, we want to test the hypotheses of

I) – profile parallelism

II) – profile coincidence (assuming that they are parallel)

III) – profile horizontality (assuming they are coincident)

now for a set of q profiles, each of which defined by the mean values of p variables.

7.7.1 Test to the profile parallelism

Let us consider a matrix C analogous to the matrix in (6.15). The parallelism hypothesis of the q profiles may be then written as

$$H_0 : C\underline{\mu}_1 = \ldots = C\underline{\mu}_q \tag{7.31}$$

where $C\underline{\mu}_k$ $(k = 1, \ldots, q)$ is the expected value of $C\underline{X}_k \sim N_{p-1}(C\underline{\mu}_k, C\Sigma_k C')$, where we assume $\Sigma_1 = \ldots = \Sigma_q = \Sigma$.

This way the test to the null hypothesis in (7.31) will be just a tst of equality of the q vectors of expected values of $(p-1)$-multivariate random vectors with multivariate Normal distributions.

Thus, based on the exposition in section 7.1, the $n/2$-th power of the likelihood ratio statistic to test H_0 in (7.31) will be

$$\Lambda^* = \frac{|CSC'|}{|CSC' + CBC'|} = \frac{|CSC'|}{|C(S+B)C'|} \tag{7.32}$$

where S and B are the matrices in (7.2) and (7.4).

Based on the exposition in section 7.2 we have

$$\Lambda^* \overset{st}{\sim} \prod_{j=1}^{p-1} Y_j \overset{d}{\equiv} \prod_{k=1}^{q-1} Y_k^*$$

where

$$Y_j \sim Beta\left(\frac{n-q+1-j}{2}, \frac{q-1}{2}\right), \qquad j = 1, \ldots, p-1$$

and

$$Y_k^* \sim Beta\left(\frac{n-p+1-k}{2}, \frac{p-1}{2}\right), \qquad k = 1, \ldots, q-1$$

form two sets of independent r.v.'s.

This way, all that was established about the distribution of the statistic Λ^* in (7.5) may be transposed to the distribution of the statistic Λ^* in (7.32), replacing p by $p-1$, now with p and q being interchangeable.

In the context of a "one-way MANOVA" model the parallelism test is seen by some authors as a test to the non-existence of interaction between the p response variables and the q treatments or levels of the factor.

7.7.2 Test to the profile coincidence

If the null hypothesis in (7.31) is not rejected, we may then pursue to test the hypothesis of profile coincidence, assuming their parallelism. Similar to what was said in section 6.5, once the parallelism of the profiles is assumed, the hypothesis of profile coincidence may be written as

$$H_0 : E_{1p}\underline{\mu}_1 = \ldots = E_{1p}\underline{\mu}_q \tag{7.33}$$

which is in fact the null hypothesis of a test of equality of q univariate mean values, that is, the test in a univariate one-way ANOVA model, with a factor with q levels, which as such may be carried out using an F statistic with an $F_{q-1,n-q}$ distribution.

In fact, similar to what happens in the previous subsection, we may use to test the null hypothesis in (7.33) the statistic

$$\Lambda^* = \frac{|E_{1p}SE_{p1}|}{|E_{1p}(S+B)E_{p1}|} = \frac{E_{1p}SE_{p1}}{E_{1p}(S+B)E_{p1}},$$

which, according to the exposition in section 7.2,

$$\Lambda^* \overset{st}{\sim} Beta\left(\frac{n-q}{2}, \frac{q-1}{2}\right),$$

so that

$$\frac{1-\Lambda^*}{\Lambda^*}\,\frac{n-q}{q-1} \sim F_{q-1,n-q}\,.$$

7.7.3 Test to profile horizontality

Once the null hypothesis of profile coincidence is not rejected, we may pursue to test the profile horizontality, assuming that they are coincident.

Note that the hypotheses in (7.31) and (7.33) are jintly equivalent to the null hypothesis of equality of the expected value vectors in (7.1), which we will assume in this subsection as holding.

The hypothesis of profile horizontality, once assumed their coincidence, will be, similar to what happens in section 6.5, a test to the null hypothesis

$$H_0 : C\underline{\mu} = \underline{0}_{(p-1)\times 1}\,, \tag{7.34}$$

where C is the same matrix used in 7.7.1.

Since the estimator of $C\underline{\mu}$ is $C\underline{\overline{X}}$, where

$$\underline{\overline{X}} = \frac{1}{n}\sum_{k=1}^{q} n_k\,\underline{\overline{X}}_k$$

with

$$\underline{\overline{X}} \sim N_p\left(\underline{\mu}, \tfrac{1}{n}\Sigma\right),$$

that is, with

$$\sqrt{n}\,C\underline{\overline{X}} \sim N_{p-1}\left(\sqrt{n}\,C\underline{\mu}, C\Sigma C'\right)\,,$$

being m.l.e. of $C\Sigma C'$,

$$CSC' \sim W_{p-1}(n-q, C\Sigma C')\,,$$

the test to the null hypothesis in (7.34) will be carried out using the T^2 statistic

$$T^2 = (n-q)n(C\underline{\overline{X}})'(CSC')^{-1}C\underline{\overline{X}}\,,$$

built using the technique in section 6.1, with

$$\frac{n-q-p+2}{p-1}\,\frac{T^2}{n-q} \sim F_{p-1,n-q-p+2}\,.$$

❏❏❏❏❏

Exercises:

7.1. Program a function or module that allows to obtain the computed value of the statistic Λ^*, once given the q matrices with the q random samples.

7.2. Program a function or module, for the case of even p or odd q, which uses the function in 7.1 and which allows to obtain the p-value for the test of equality of q vectors of expected values.

7.3. Program a function or module, for the case of odd p and even q, which uses the function in 7.1 and which allows to obtain the p-value for the test of equality of q vectors of expected values.

7.4. Using the data corresponding to the 3 species of *Iris* carry out a test to the equality of their ean vectors.

- 7.5.] Program R functions to implement the 4 asymptotic distributions in subsection 7.6.1.

7.5. Program R functions to implement the Analysis of more than 2 Profiles.

7.6. Using the data for the 3 species of *Iris*

   a) carry out a profile analysis using the profiles for the 3 species;

   b) being aware of the possible problems arising from '*data snooping*', carry out a profile analysis using only the profiles corresponding to the 2 species for which it seems to be most likely to not reject their parallelism.

7.7. Obtain data sets onto which you perform tests of equality of expected value vectors and at least one profile analysis.

❏❏❏❏❏

Bibliography

Anderson, T. W. (1958, 2003). *An Introduction to Multivariate Statistical Analysis*, 1.ª, 3.ª ed., J. Wiley & Sons Inc., New York (1.ª ed.), Hoboken (3.ª ed.).

Arnold, B. C., Coelho, C. A., Marques, F. J. (2013). The distribution of the product of powers of independent Uniform random variables, *Journal of Multivariate Analysis*, **113**, 19-36.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**, 317–346.

Coelho, C. A. (1998). The Generalized Integer Gamma distribution – a basis for distributions in multivariate statistics. *Journal of Multivariate Analysis*, **64**, 86–102.

Coelho, C. A. (2004). The Generalized Near-Integer Gamma distribution: a basis for 'near-exact' approximations to the distribution of statistics which are the product of an odd number of independent Beta random variables, *Journal of Multivariate Analysis*, **89**, 191–218.

Coelho, C. A., Arnold, B. C. (2019). Finite form representations for instances of Meijer G and Fox H functions – Application to the implementation of likelihood ratio tests in Multivariate Analysis. Lecture Notes in Statistics, Springer.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*, 2.ª ed., J. Wiley & Sons, Hoboken.

Tricomi, F. G., Erdélyi, A. (1951). The asymptotic expansion of a ratio of Gamma functions, *Pacific Journal of Mathematics*, **1**, 133–142.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471–494.

Appendix 7.A

Brief note on likelihood ratio tests

It is called a likelihood ratio test a test that uses a statistic of the form

$$\Lambda \;=\; \frac{sup\, L_0}{sup\, L_1} \qquad \text{or} \qquad \Lambda \;=\; \frac{max\, L_0}{max\, L_1}$$

where L_0 and L_1 represent respectively the likelihood functions for the set of parameters being considered, with their values conditioned respectively by H_0 and by H_1. The values of the parameters which respectively under H_0 and H_1 maximize respectively L_0 and L_1 are the Maximum Likelihood Estimates of these parameters, respectively under H_0 and H_1, and the corresponding expressions in terms of random variables are the MLE's (Maximum Likelihood Estimators) of these parameters, respectively under H_0 and H_1.

This way, in order to obtain Λ we need to maximize the two likelihood functions L_0 and L_1, obtaining the expressions for the parameter estimators that maximize these two functions, and then, by replacing in L_0 e L_1 those parameters by such expressions, obtain respectively $sup\, L_0$ or $max\, L_0$ and $sup\, L_1$ or $max\, L_1$.

It may be shown that $0 \leq \Lambda \leq 1$, and from the formulation for Λ it is clear that under H_0 we have $\Lambda = 1$, so that in order to obtain a test of level α we will reject H_0 if $\Lambda_{calc} < \Lambda_\alpha$, where Λ_α represents the α quantile of Λ.

Appendix 7.B

Brief note on the Wilks Λ statistic

Let A and B be two independent matrices, with

$$A \;\sim\; W_p(f-q,\; \Sigma)$$

and

$$B \;\sim\; W_p(q,\; \Sigma) \qquad \text{or} \qquad B \;\sim\; \widetilde{W}_p(q,\; \Sigma)\;.$$

Then,

$$\Lambda \;=\; \frac{|A|}{|A+B|}$$

is a random variable with a Wilks Lambda distribution, with parameters f, p and q, which is denoted by several authors by

$$\Lambda \;\sim\; \Lambda(f,p,q)\ .$$

A and B are usually called the 'error matrix' or matrix of sums of squares and products of the error, and 'hypothesis matrix' or matrix of sums of squares and products of the hypothesis, respectively. These designations are related with the role that these two matrices play in the tests of several hypotheses. When these hypotheses are tested, the matrix A has always a (central) Wishart distribution, but the matrix B will only have a central Wishart (or pseudo-Wishart) distribution under the null hypothesis. If such null hypothesis is not verified, then B will have a non-central distribution and will not be independent of A.

It may be shown that

$$\Lambda(f,p,q) \;\equiv\; \Lambda(f,q,p)\ ,$$

since, as shown in Section 7.2, we have

$$\Lambda \sim \prod_{j=1}^{p} Y_j \stackrel{d}{\equiv} \prod_{k=1}^{q} Y_k^*$$

where, for $j=1,\dots,p$ and $k=1,\dots,q$,

$$Y_j \sim Beta\left(\frac{f-q-j}{2},\frac{q}{2}\right) \quad \text{and} \quad Y_k^* \sim Beta\left(\frac{f-p-k}{2},\frac{p}{2}\right)$$

form two sets of independent r.v.'s.

Based on these results, it is possible to easily obtain an expressions for the h-th moment of Λ, and since Λ has a delimited support, with $0 \leq \Lambda \leq 1$, these moments will determine and identify the distribution of Λ. As such, a random varialbe Λ has a Wilks Lambda distribution, with parameters f, p and q if its h-th moment is

$$\begin{aligned} E(\Lambda^h) &= \prod_{j=1}^{p} \frac{\Gamma\left(\frac{f-j}{2}\right)\ \Gamma\left(\frac{f-q-j}{2}+h\right)}{\Gamma\left(\frac{f-j}{2}+h\right)\ \Gamma\left(\frac{f-q-j}{2}\right)} = \frac{\Gamma_p\left(\frac{f-1}{2}\right)\ \Gamma_p\left(\frac{f-q-1}{2}+h\right)}{\Gamma_p\left(\frac{f-1}{2}+h\right)\ \Gamma_p\left(\frac{f-q-1}{2}\right)} \\ &= \prod_{k=1}^{q} \frac{\Gamma\left(\frac{f-k}{2}\right)\ \Gamma\left(\frac{f-p-k}{2}+h\right)}{\Gamma\left(\frac{f-k}{2}+h\right)\ \Gamma\left(\frac{f-p-k}{2}\right)} = \frac{\Gamma_q\left(\frac{f-1}{2}\right)\ \Gamma_q\left(\frac{f-p-1}{2}+h\right)}{\Gamma_q\left(\frac{f-1}{2}+h\right)\ \Gamma_q\left(\frac{f-p-1}{2}\right)}\ . \end{aligned}$$

For a number of tests in Multivariate Analysis the likelihood ratio test statistic is a Wilks Lambda statistic, or it is closely related with such a statistic.

Chapter 8

The test of independence of two random vectors

8.1 The setup

Let

$$\underline{X} = \left[\, \underline{X}_1' \,\middle|\, \underline{X}_2' \,\right]' \sim N_p(\underline{\mu},\ \Sigma)\,, \tag{8.1}$$

where $\underline{X}_1$ is $p_1{\times}1$ and $\underline{X}_2$ is $p_2{\times}1$. To this split of the vector $\underline{X}$ corresponds a split of the vector $\underline{\mu}$ and a split of the matrix Σ similar to the one in IV) in section 5.3, with

$$\underline{\mu} = \left[\, \underline{\mu}_1' \,\middle|\, \underline{\mu}_2' \,\right]' \qquad \text{and} \qquad \Sigma = \begin{array}{c} \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right] \begin{array}{c} p_1 \\ p_2 \end{array} \\ \begin{array}{cc} p_1 & \quad p_2 \end{array} \end{array} .$$

We thus have

$$\underline{X}_1 \sim N_{p_1}(\underline{\mu}_1, \Sigma_{11}) \quad \text{and} \quad \underline{X}_2 \sim N_{p_2}(\underline{\mu}_2, \Sigma_{22})$$

and

$$Cov(\underline{X}_1, \underline{X}_2) = \Sigma_{12} = \Sigma_{21}' = (Cov(\underline{X}_2, \underline{X}_1))'\ .$$

According to property II) in section 2.2, the independence of $\underline{X}_1$ and $\underline{X}_2$ is equivalent to have $\Sigma_{12} = 0_{(p_1\times p_2)}$. We may thus write the hypothesis of independence of $\underline{X}_1$ and $\underline{X}_2$ as

$$H_0 : \Sigma_{12} = 0_{(p_1\times p_2)} \iff H_0 : \Sigma = bdiag(\Sigma_{11}, \Sigma_{22})\,. \tag{8.2}$$

8.2 The likelihood ratio test

In this section we will obtain the likelihood ratio statistic to test the null hypothesis in (8.2) versus the general alternative

$$H_1 : \Sigma_{12} \neq 0_{(p_1 \times p_2)} , \tag{8.3}$$

and we will show that this statistic is a power of a Wilks Lambda statistic.

Let

$$\underset{n\times p}{X} = \begin{array}{c} \left[\begin{array}{c|c} X_1 & X_2 \end{array}\right] n \\ \;\; p_1 \quad p_2 \quad\;\; \end{array}$$

be the matrix of a random sample of dimension n of $\underline{X}$, where X_1 and X_2 are respectively the matrices of dimensions $n{\times}p_1$ and $n{\times}p_2$ of the random samples of dimension n of $\underline{X}_1$ and $\underline{X}_2$.

Let S be the matrix defined in (5.6), and $V^* = \frac{1}{n}S$ the matrix defined in (5.21), and let S_{11}, S_{22}, V^*_{11} and V^*_{22} the matrices defined respectively by

$$S_{11} = (X_1 - E_{n1}\overline{\underline{X}}'_1)'(X_1 - E_{n1}\overline{\underline{X}}'_1) ,$$

$$S_{22} = (X_2 - E_{n1}\overline{\underline{X}}'_2)'(X_2 - E_{n1}\overline{\underline{X}}'_2) ,$$

and

$$\begin{aligned} V^*_{11} &= (X_1 - E_{n1}\overline{\underline{X}}'_1)'D(X_1 - E_{n1}\overline{\underline{X}}'_1) , \\ V^*_{22} &= (X_2 - E_{n1}\overline{\underline{X}}'_2)'D(X_2 - E_{n1}\overline{\underline{X}}'_2) , \end{aligned}$$

where

$$D = \frac{1}{n}I_n = diag\left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right) .$$

We may note that we then have

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

and

$$V^* = \begin{bmatrix} V^*_{11} & V^*_{12} \\ V^*_{21} & V^*_{22} \end{bmatrix}$$

where

$$S_{12} = (X_1 - E_{n1}\overline{\underline{X}}'_1)'(X_2 - E_{n1}\overline{\underline{X}}'_2) = S'_{21} ,$$

and

$$V_{12}^* \;=\; (X_1 - E_{n1}\underline{\overline{X}}_1')'D(X_2 - E_{n1}\underline{\overline{X}}_2') \;=\; V_{21}^{*\prime}\;.$$

Then, under H_1, the likelihood function is the function in (4.1) and thus the MLEs of $\underline{\mu}$ and Σ, under H_1 are given by (4.7) and (4.8), respectively. Under H_0, given the independence of $\underline{X}_1$ and $\underline{X}_2$, the likelihoods of $\underline{X}_1$ and $\underline{X}_2$ factorize to give rise to the likelihood of $\underline{X}$, i.e., under H_0

$$\begin{aligned}
L_0(\underline{\mu},\Sigma) &= L_1(\underline{\mu}_1,\underline{\mu}_2,\Sigma_{11},\Sigma_{22})\\
&= \prod_{j=1}^{2}\prod_{i=1}^{n}\frac{1}{(2\pi)^{p_j/2}\;|\Sigma_{jj}|^{1/2}}\;e^{-\frac{1}{2}(\underline{x}_{ji}-\underline{\mu}_j)'\Sigma_{jj}^{-1}(\underline{x}_{ji}-\underline{\mu}_j)}\\
&= \frac{1}{(2\pi)^{np_1/2}\;|\Sigma_{11}|^{n/2}}\;e^{tr\left[-\frac{1}{2}(X_1-E_{n1}\underline{\mu}_1')\Sigma_{11}^{-1}(X_1-E_{n1}\underline{\mu}_1')'\right]}\times\\
&\qquad\times\frac{1}{(2\pi)^{np_2/2}\;|\Sigma_{22}|^{n/2}}\;e^{tr\left[-\frac{1}{2}(X_2-E_{n1}\underline{\mu}_2')\Sigma_{22}^{-1}(X_2-E_{n1}\underline{\mu}_2')'\right]}\,,
\end{aligned}\tag{8.4}$$

so that, following a line of thought and deduction similar to the ones followed in section 4.2 we may obtain from the first term in (8.4) the MLEs of $\underline{\mu}_1$ and Σ_{11} and from the second term the MLEs of $\underline{\mu}_2$ and Σ_{22}. It is not hard to show that

$$\hat{\underline{\mu}}_1 \;=\; \underline{\overline{X}}_1\;,\quad \hat{\underline{\mu}}_2 \;=\; \underline{\overline{X}}_2$$

and

$$\hat{\Sigma}_{11} \;=\; \frac{1}{n}S_{11} \;=\; V_{11}^*\;,\quad \hat{\Sigma}_{22} \;=\; \frac{1}{n}S_{22} \;=\; V_{22}^*\;.$$

If we denote by $\hat{\Sigma}_{H_0}$ and $\hat{\Sigma}_{H_1}$ the MLEs of Σ under H_0 and H_1, respectively, we have

$$\hat{\Sigma}_{H_0} \;=\; \frac{1}{n}\begin{bmatrix} S_{11} & 0\\ 0 & S_{22}\end{bmatrix} \;=\; \begin{bmatrix} V_{11}^* & 0\\ 0 & V_{22}^*\end{bmatrix}$$

and

$$\hat{\Sigma}_{H_1} \;=\; \frac{1}{n}S \;=\; \frac{1}{n}\begin{bmatrix} S_{11} & S_{12}\\ S_{21} & S_{22}\end{bmatrix} = V^* = \begin{bmatrix} V_{11}^* & V_{12}^*\\ V_{21}^* & V_{22}^*\end{bmatrix}\;.$$

Taking into account (4.1) and (8.4) it is then easy to show that we will have as lielihood ratio statistic the statistic

$$\Lambda = \frac{\sup L_0}{\sup L_1} = \frac{\underset{\underline{\mu},\Sigma}{sup}\ L}{\underset{\underline{\mu}_1,\underline{\mu}_2,\Sigma_{11}\Sigma_{22}}{sup}\ L}$$

$$= \frac{\dfrac{e^{tr\left[-\frac{1}{2}(X_1-E_{n1}\underline{\overline{x}}_1')V_{11}^{*-1}(X_1-E_{n1}\underline{\overline{x}}_1')'\right]+tr\left[-\frac{1}{2}(X_2-E_{n1}\underline{\overline{x}}_2')V_{22}^{*-1}(X_2-E_{n1}\underline{\overline{x}}_2')'\right]}}{(2\pi)^{np/2}\,|V_{11}^*|^{n/2}\,|V_{22}^*|^{n/2}}}{\dfrac{1}{(2\pi)^{np/2}\,|V^*|^{n/2}}\,e^{tr\left[-\frac{1}{2}(X-E_{n1}\underline{\overline{x}}')\,V^{*-1}(X-E_{n1}\underline{\overline{x}}')'\right]}}$$

$$= \frac{|V^*|^{n/2}}{|V_{11}^*|^{n/2}\,|V_{22}^*|^{n/2}}\,e^{-\frac{1}{2}\left\{tr\left[\left(nS_{11}^{-1}S_{11}\right)\right]+tr\left[\left(nS_{22}^{-1}S_{22}\right)\right]-tr\left[\left(nS^{-1}S\right)\right]\right\}}$$

$$= \frac{|V^*|^{n/2}}{|V_{11}^*|^{n/2}\,|V_{22}^*|^{n/2}} = \frac{|S|^{n/2}}{|S_{11}|^{n/2}\,|S_{22}|^{n/2}} = (\Lambda^*)^{n/2}\,,$$

where

$$\Lambda^* = \frac{|S|}{|S_{11}|\,|S_{22}|}\,, \tag{8.5}$$

since if V is $p{\times}p$ we have $|aV| = a^p|V|$, and thus

$$|V^*| = \left(\frac{1}{n}\right)^p |S|\,, \quad |V_{11}^*| = \left(\frac{1}{n}\right)^{p_1} |S_{11}| \quad \text{e} \quad |V_{22}^*| = \left(\frac{1}{n}\right)^{p_2} |S_{22}|\,.$$

We may note that since

$$|S| = |S_{11}|\,|S_{22.1}| = |S_{22}|\,|S_{11.2}|$$

with

$$\begin{aligned} S_{11.2} &= S_{11} - S_{12}S_{22}^{-1}S_{21} \\ S_{22.1} &= S_{22} - S_{21}S_{11}^{-1}S_{12} \end{aligned}$$

we may write

$$\begin{aligned} \Lambda^* = \frac{|S|}{|S_{11}|\,|S_{22}|} &= \frac{|S_{11}|\,|S_{22.1}|}{|S_{11}|\,|S_{22}|} = \frac{|S_{22.1}|}{|S_{22}|} = \frac{|S_{22.1}|}{|S_{22.1} + S_{21}S_{11}^{-1}S_{12}|} \\ &= \frac{|S_{22}|\,|S_{11.2}|}{|S_{11}|\,|S_{22}|} = \frac{|S_{11.2}|}{|S_{11}|} = \frac{|S_{11.2}|}{|S_{11.2} + S_{12}S_{22}^{-1}S_{21}|} \end{aligned} \tag{8.6}$$

where, under H_0, by property VII) in section 5.5,

$$\begin{aligned} S_{22.1} &\sim W_{p_2}(n-1-p_1,\, \Sigma_{22}) \\ S_{21}S_{11}^{-1}S_{12} &\sim W_{p_2}(p_1,\, \Sigma_{22}) \end{aligned} \tag{8.7}$$

are independent, and, by the same property,

$$\begin{aligned} S_{11.2} &\sim W_{p_1}(n-1-p_2,\, \Sigma_{11}) \\ S_{12}S_{22}^{-1}S_{21} &\sim W_{p_1}(p_2,\, \Sigma_{11}) \end{aligned} \tag{8.8}$$

are also independent, being $S_{22.1}$ or $S_{11.2}$ commonly called the error matrices and $S_{21}S_{11}^{-1}S_{12}$ or $S_{12}S_{22}^{-1}S_{21}$ the hypothesis matrices, since in the definition of these latter ones is involved the matrix S_{12}, estimator of Σ_{12}, matrix that under H_0 is null. We should note that under $H_0 : \Sigma_{12} = 0$, we have $\Sigma_{22.1} = \Sigma_{22}$ and $\Sigma_{11.2} = \Sigma_{11}$.

Through the analysis of the Wishart distributions of the matrices in (8.7) and (8.8) we may see how p_1 and p_2 are interchangeable in the distribution of Λ^* in (8.5), as well as, by establishing a parallel between these distributions and the distributions of the matrices S and B in (7.6) and (7.7) we may also establish a parallel between the distribution of the statistic Λ^* in (8.5) and the distribution of the statistic Λ^* in (7.5), showing that for Λ^* in (8.5) we have

$$\Lambda^* \overset{st}{\sim} \prod_{j=1}^{p_1} Y_j \overset{d}{\equiv} \prod_{k=1}^{p_2} Y_k^*$$

where

$$Y_j \sim Beta\left(\frac{n-p_2-j}{2}, \frac{p_2}{2}\right), \qquad j = 1, \ldots, p_1$$

and

$$Y_k^* \sim Beta\left(\frac{n-p_1-j}{2}, \frac{p_1}{2}\right), \qquad k = 1, \ldots, p_2$$

form two sets of independent r.v.'s.

This way the distribution of the statistic Λ^* in (8.5) is in fact identical to the distribution of the statistic Λ^* in Chapter 7, by replacing p by p_1 and $q-1$ by p_2, or vice-versa. We have this way the exact distribution of Λ^* in (8.5), when p_1 is even or p_2 is even, given by the results in section 7.5, once the replacement of parameters referred to above is made, that is the replacement of p by p_1 and $q-1$ by p_2, or vice-versa, i.e., with the distribution of $W = -\log \Lambda^*$ being in this case a GIG or that of Λ^* an EGIG of depth p_1+p_2-2 with shape parameters

$$r_j = \begin{cases} h_j\,, & j=1,2 \\ h_j + r_{j-2}\,, & j=3,\dots,p_1+p_2-2 \end{cases}$$

where

$$h_j = (\#\text{ of elements in } \{p_1,p_2\} \geq j) - 1\,, \quad j=1,\dots,p_1+p_2-2\,,$$

and corresponding rate parameters

$$\frac{n-2-j}{2} \quad \text{ou} \quad \frac{n-p_1-p_2-1+j}{2}\,, \quad j=1,\dots,p_1+p_2-2\,.$$

In case both p_1 and p_2 are odd, we will then have to resort to the use of near-exact distributions.

8.3 The test of independence of $\underline{X}_1$ e $\underline{X}_2$ as the test of fit of the Multivariate Regression model

The test to the null hypothesis in (8.2) is indeed also the test to

$$H_0: \underset{(p_2\times p_1)}{\beta} = 0 \tag{8.9}$$

in the model

$$\underset{(n\times p_1)}{X_1^*} = \underset{(n\times p_2)}{X_2^*}\ \underset{(p_2\times p_1)}{\beta} + \underset{(n\times p_1)}{E} \tag{8.10}$$

where

$$\begin{array}{l} X_1^* = X_1 - E_{n1}\overline{\underline{X}}_1' = (I_n - \frac{1}{n}E_{nn})X_1 \\ X_2^* = X_2 - E_{n1}\overline{\underline{X}}_2' = (I_n - \frac{1}{n}E_{nn})X_2 \end{array}\,, \quad \text{with} \quad \overline{\underline{X}}_k = \tfrac{1}{n}X_k'E_{n1}\,,$$

i.e., it is the test of fit to the model in (8.10), with the rejection of H_0 in (8.2) being equivalent to assuming that the model (8.10) fits, that is, that the p_2 variables in $\underline{X}_2$ model 'significantly well' the p_1 variables in $\underline{X}_1$, the model in (8.10) being a Multivariate Regression model with p_1

response variables (the variables $X_1, \ldots, X_{p1}$) and p_2 explanatory variables (the variables $X_{p_1+1}, \ldots, X_p$, with $p = p_1 + p_2$).

We may note that the choice of the r.v.'s in $\underline{X}_1$ as response variables and that of the r.v.'s in $\underline{X}_2$ as the explanatory variables is in fact arbitrary, being possible to make the reverse choice, with the test to the null hypothesis in (8.2) remaining equivalent to the test to the nullity of the parameter matrix of this other Multivariate Regression model.

This model is also called a Canonical Correlation model since it is possible to show that (Coelho, 1986, 1992; Sengupta, 2004, 2014)

$$\Lambda^* = \frac{|S|}{|S_{11}|\,|S_{22}|} = \prod_{j=1}^{p} \lambda_j$$

where

$$SQ\underline{u}_j = \lambda_j \underline{u}_j\,, \qquad j = 1, \ldots, p$$

that is, where λ_j $(j = 1, \ldots, p)$ are the eigenvalues of SQ, with

$$Q = \begin{bmatrix} S_{11}^{-1} & 0 \\ 0 & S_{22}^{-1} \end{bmatrix},$$

associated with the eigenvectors $\underline{u}_j \in \mathbb{R}^p$.

It is also possible to show that

$$0 < \lambda_j < 2$$

and that, for $q = \min(p_1, p_2)$ and $s = \max\,(p_1, p_2)$

$$\lambda_j = \begin{cases} 2 - \lambda_{p-j+1}\,, & j = 1, \ldots, q \\ 1\,, & j = q+1, \ldots, s \\ 2 - \lambda_{p-j+1}\,, & j = s+1, \ldots, p\,. \end{cases}$$

being common to call the quantities

$$\rho_j = \lambda_j - 1 = 1 - \lambda_{p-j+1}\,, \qquad j = 1, \ldots, q$$

the canonical correlations, since

$$\rho_j = corr^2(u_{1j}, u_{2j}) = cos^2(\underline{u}_{1j}, \underline{u}_{2j})$$

where

$$u_{1j} = \underline{u}'_{1j}\underline{X}_1 \quad \text{e} \quad u_{2j} = \underline{u}'_{2j}\underline{X}_2 \quad \text{com} \quad \underline{u}_j = \begin{bmatrix} \underline{u}_{1j} \\ \hline \underline{u}_{2j} \end{bmatrix} \begin{matrix} p_1 \\ p_2 \end{matrix}$$

represent linear combinations of the variables in $\underline{X}_1$ and $\underline{X}_2$ subject to the conditions that u_{11} and u_{21} have the maximum correlation among all possible linear combinations of variables in $\underline{X}_1$ and variables in $\underline{X}_2$, that is, among all vectors $\underline{u}_{11} \in \mathbb{R}^{p_1}$ and $\underline{u}_{21} \in \mathbb{R}^{p_2}$ and that $corr(u_{12}, u_{22})$ is maximum amomng all linear combinations of variables in $\underline{X}_1$ and variables in $\underline{X}_2$ with $corr(u_{11}, u_{12}) = 0$ and $corr(u_{21}, u_{22}) = 0$, being then $corr(u_{13}, u_{23})$ maximum among all linear combinations of variables in $\underline{X}_1$ and variables in $\underline{X}_2$, with $corr(u_{11}, u_{13}) = corr(u_{12}, u_{13}) = 0$ and $corr(u_{21}, u_{23}) = corr(u_{22}, u_{23}) = 0$, and so on.

8.3.1 Testing between a Multivariate Linear Model and one of its Submodels

Once the null hypothesis in (8.9) is rejected we may be interested in testing if a given subgroup of explanatory variables in $\underline{X}_2$ is or not important in modeling the variables in $\underline{X}_1$, in a model where the other variables in $\underline{X}_2$ remain as explanatory variables. The statistic to be used in such a test will be a statistic to which we may call partial Wilks Λ, given its parallel with the F statistic used in the partial F test.

In as much the same way as the statistic used in the partial F test in Multiple Regression, ends up itself to have an F distribution, also the partial Wilks Λ statistic is itself a Wilks Λ statistic in its own right.

In order to simplify the exposition and the formulation, and without any loss of generality, given that the Wilks Λ statistics used in this test are not sensible to the order of the variables in each of the subvectors $\underline{X}_1$ and $\underline{X}_2$, let us suppose that the variables which importance in the model we want ot test are the variables

$$X_{p_1+p_{21}+1}, \ldots, X_p\,, \tag{8.11}$$

in the presence of the variables

$$X_{p_1+1}, \ldots, X_{p_1+p_{21}}\,. \tag{8.12}$$

Let us consider then the subvectors $\underline{X}_1$ and $\underline{X}_2$ already defined in (8.1), now with $\underline{X}_2$ subdivided into $\underline{X}_{21}$, of dimensions $p_{21}{\times}1$ and $\underline{X}_{22}$ of dimensions

$p_{22}\times 1$, where $p_{22} = p_2 - p_{21}$, being these the subvectors whose components are the variables in (8.12) and (8.11), respectively.

Let us consider then the submodel

$$\underset{(n\times p_1)}{X_1^*} = \underset{(n\times p_{21})}{X_{21}^*} \underset{(p_{21}\times p_1)}{\beta_1} + \underset{(n\times p_1)}{E} \tag{8.13}$$

where X_{21}^* represents the sub-matrix of X_2^* formed by its first p_{21} columns, that is, with

$$\underset{(n\times p_2)}{X_2^*} = \left[\underset{(n\times p_{21})}{X_{21}^*} \;\middle|\; \underset{(n\times p_{22})}{X_{22}^*} \right], \tag{8.14}$$

and where β_1 represents the matrix formed by the first p_{21} rows of β, that is, with

$$\beta = \begin{array}{c} p_1 \\ \left[\begin{array}{c} \beta_1 \\ \hline \beta_2 \end{array}\right] \end{array} \begin{array}{l} \\ p_{21} \\ p_{22} \end{array} ,$$

being this a submodel of the model in (8.10). We want then to test between the model in (8.10) and the submodel (8.13), with

$$\begin{array}{ll} & H_0^*: \;\; X_1^* = X_{21}^*\beta_1 + E \\ \text{vs.} & \\ & H_1: \;\; X_1^* = X_2^*\beta + E\,, \end{array} \tag{8.15}$$

leading the rejection of H_0^* to the choice for the original model (8.10), that is, the model in H_1 in (8.15), since we conclude that the explanatory variables in (8.12) are important in the model (that is, in a model where the explanatory variables in (8.11) are present), while the non-rejection of H_0^* will lead us to choose the submodel in (8.13). We should note that in case we do not reject H_0^* in (8.15), that is, in case we choose the submodel instead of the original model, we will be assuming that the submatrix of β formed by its last p_{22} rows is a submatrix of zeros. However, this test only makes sense once the null hypothesis in (8.2) or (8.9) was rejected.

This way, the test to the hypotheses in (8.15) is in fact equivalent to the test to the hypotheses

$$\begin{array}{ll} & H_0^*: \;\; \beta_2 = 0 \,|\, \beta_1 \neq 0 \\ \text{vs.} & \\ & H_1: \;\; \beta_2 \neq 0 \,|\, \beta_1 \neq 0\,. \end{array}$$

Using the notation $\Lambda_{H_0^*,H_1}$ to denote the statistic used in the test to the hypotheses H_0^* and H_1, given the nested form of the null hypotheses H_0 and H_0^*, we have

$$\Lambda_{H_0^*,H_1} = \frac{\Lambda_{H_0,H_1}}{\Lambda_{H_0,H_0^*}}$$

where Λ_{H_0,H_1}, being the statistic used in the test of the null hypothesis in (8.9) or (8.2) versus the alternative hypothesis in (8.3), equivalent to the hypothesis $H_1 : \beta \neq 0$, is indeed the statistic Λ^* in (8.5), used in the test of fit of the model (8.10), or equivalently yet, in the test to the hypotheses

$$\begin{array}{ll} & H_0: \;\; X_1^* = E \\ \text{vs.} & \\ & H_1: \;\; X_1^* = X_2^*\beta + E \end{array}$$

where $X_1^* = E$ represents the null model, and where Λ_{H_0,H_0^*} is the statistic used in the test of fit of the submodel (8.13), that is, the statistic used in the test to the hypotheses

$$\begin{array}{ll} & H_0: \;\; X_1^* = E \\ \text{vs.} & \\ & H_0^*: \;\; X_1^* = X_{21}^*\beta_1 + E \end{array} \iff \begin{array}{ll} & H_0: \;\; \beta_1 = 0 \\ \text{vs.} & \\ & H_0^*: \;\; \beta_1 \neq 0\,. \end{array} \tag{8.16}$$

We will have then, in a similar way the statistic Λ_{H_0,H_0^*} defined as

$$\Lambda_{H_0,H_0^*} = \frac{|S_{11.2(1)}|}{|S_{11}|} \tag{8.17}$$

where

$$S_{11.2(1)} = S_{11} - S_{12(1)}S_{22(11)}^{-1}S_{21(1)}\,,$$

for the subdivision of S induced by the subdivision of $\underline{X}_2$ into $\underline{X}_{21}$ and $\underline{X}_{22}$, or, equivalently, induced by the subdivision of the matrix X_2^* in (8.14), giving raise to

$$S = \left[\begin{array}{c:c} S_{11} & S_{12} \\ \hdashline S_{21} & S_{22} \end{array}\right] = \left[\begin{array}{c:cc} S_{11} & S_{12(1)} & S_{12(2)} \\ \hdashline S_{21(1)} & S_{22(11)} & S_{22(12)} \\ S_{21(2)} & S_{22(21)} & S_{22(22)} \end{array}\right], \tag{8.18}$$

with

$$S_{12(i)} = X_1'(I_n - \tfrac{1}{n}E_{nn})X_{2i} = X_1^{*\prime}X_{2i}^* \qquad (i = 1, 2)$$

and

$$S_{22(ij)} = X'_{2i}(I_n - \tfrac{1}{n}E_{nn})X_{2j} = X^{*\prime}_{2i}X^*_{2j}\,, \qquad i,j \in \{1,2\}$$

for a subdivision of the matrix X_2, similar to the subdivision of the matrix X_2^* in (8.14),

$$\underset{(n\times p_2)}{X_2} = \left[\; \underset{(n\times p_{21})}{X_{21}} \;\middle|\; \underset{(n\times p_{22})}{X_{22}} \;\right].$$

While under H_0^* in (8.16) we have

$$S_{11.2(1)} \sim W_{p_1}(n-1-p_{21}, \Sigma_{11} - \Sigma_{12(1)}\Sigma_{22(11)}^{-1}\Sigma_{21(1)})\,,$$

under H_0 in (8.16) we have

$$S_{11.2(1)} \sim W_{p_1}(n-1-p_{21}, \Sigma_{11})\,,$$

so that the test to the hypotheses in (8.16) is in fact also a test to the hypotheses

$$\begin{array}{rl} & H_0: \;\; \Sigma_{12(1)} = 0 \\ \text{vs.} & \\ & H_0^*: \;\; \Sigma_{12(1)} \neq 0\,, \end{array}$$

for a subdivision of Σ similar to that of S in (8.18).

But then, from (8.6) and (8.17) we have

$$\Lambda_{H_0^*,H_1} = \frac{|S_{11.2}|}{|S_{11.2(1)}|}$$

with

$$S_{11.2(1)} = S_{11,2} + \hat{\beta}_2' S_{22(2.1)}\hat{\beta}_2$$

where

$$\hat{\beta}_2 = \left(S_{12(2)} - S_{12(1)}S_{22(11)}^{-1}S_{22(12)}\right)S_{22(2.1)}^{-1}\,,$$

and

$$S_{22(2.1)} = S_{22(22)} - S_{22(21)}S_{22(11)}^{-1}S_{22(12)}\,,$$

(see Coelho and Marques (2014) and Coelho and Arnold (2017, sec. 5.1.7)) so that the test to the hypotheses in (8.15) is in fact a test also to the hypotheses

$$\begin{array}{rl} & H_0^*: \;\; \Sigma_{12(2)} - \Sigma_{12(1)}\Sigma_{22(11)}^{-1}\Sigma_{22(12)} = 0 \\ \text{vs.} & \\ & H_1: \;\; \Sigma_{12(2)} - \Sigma_{12(1)}\Sigma_{22(11)}^{-1}\Sigma_{22(12)} \neq 0\,, \end{array} \tag{8.19}$$

and not to the hypotheses

$$\begin{array}{ll} & H_0^*: \ \ \Sigma_{12(2)} = 0 \\ \text{vs.} & \\ & H_1: \ \ \Sigma_{12(2)} \neq 0\,. \end{array}$$

Under H_0^* in (8.19) we have

$$\hat{\beta}_2' S_{22(2.1)} \hat{\beta}_2 \sim W_{p_1}(p_{22}, \Sigma_{11.2|H_0^*})$$

and

$$S_{11.2} \sim W_{p_1}(n-1-p_2, \Sigma_{11.2|H_0^*})\,,$$

so that we have

$$\Lambda_{H_0^*,H_1} \stackrel{st}{\sim} \prod_{j=1}^{p_1} Y_j \stackrel{d}{\equiv} \prod_{k=1}^{p_{22}} Y_k^*$$

where

$$Y_j \sim Beta\left(\frac{n-p_2-j}{2}, \frac{p_{22}}{2}\right), \qquad j = 1, \ldots, p_1$$

and

$$Y_k^* \sim Beta\left(\frac{n-p_1-p_{21}-k}{2}, \frac{p_1}{2}\right), \qquad k = 1, \ldots, p_{22}$$

form two sets of independent r.v.'s.

We have thus the exact distribution of $\Lambda_{H_0^*,H_1}$ for even p_1 or even p_{22} as an EGIG, while when p_1 and p_{22} are both odd we will have to resort to the use of a near-exact distribution. Note that the distribution of $\Lambda_{H_0^*,H_1}$ is similar to that of Λ^* in (8.5) or (8.6), with the due adaptations, which are left as an exercise.

8.4 The test of equality of mean vectors as a test of independence – the MANOVA model

In this section the test of equality of mean vectors, or expected value vectors, will be presented as a test of independence between the random vector $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, from which we have an overall sample of size n, formed by the piling of the q independent samples used in Chap. 7, the k-th of which has dimension n_k $(k = 1, \ldots, q)$, and a non-random vector, which we will call

$\underline{M}$, which is the vector of the indicator variables for the sub-populations, the k-th of which will assume the value 1 or 0 (zero), according to the fact that the corresponding observation belongs or not to the k-th sub-population $(k = 1, \ldots, q)$.

Let X be the $n{\times}p$ matrix of the overal random sample of dimension n from $\underline{X}$, that is, the $n{\times}p$ matrix formed by the piling of the q samples used in Chap. 7, the k-th of which of dimension n_k, with $n = \sum_{k=1}^{q} n_k$,

$$\underset{(p\times n)}{X'} = \left[\begin{array}{cccccccccc} \overbrace{X_{111} \ldots X_{11n_1}}^{n_1} & \overbrace{X_{211} \ldots X_{21n_2}}^{n_2} & \ldots & \overbrace{X_{k11} \ldots X_{k1n_k}}^{n_k} & \ldots & \overbrace{X_{q11} \ldots X_{q1n_q}}^{n_q} \\ \vdots \quad\quad \vdots & \vdots \quad\quad \vdots & & \vdots \quad\quad \vdots & & \vdots \quad\quad \vdots \\ X_{1j1} \ldots X_{1jn_1} & X_{2j1} \ldots X_{2jn_2} & \ldots & X_{kj1} \ldots X_{kjn_k} & \ldots & X_{qj1} \ldots X_{qjn_q} \\ \vdots \quad\quad \vdots & \vdots \quad\quad \vdots & & \vdots \quad\quad \vdots & & \vdots \quad\quad \vdots \\ X_{1p1} \ldots X_{1pn_1} & X_{2p1} \ldots X_{2pn_2} & \ldots & X_{kp1} \ldots X_{kpn_k} & \ldots & X_{qp1} \ldots X_{qpn_q} \end{array}\right] \tag{8.20}$$

where X_{kji} represents the i-th observation on the j-th variable in the k-th sub-population, for $k = 1, \ldots, q$, $j = 1, \ldots, p$ and $i = 1, \ldots, n_k$, and let M be the matrix of dimensions $n{\times}q$ of the q indicator variables, that is, the so-called 'design matrix'

$$\underset{(q\times n)}{M'} = \left[\begin{array}{cccccccccccccc} \overbrace{1 \;\; 1 \;\; \ldots \;\; 1}^{n_1} & \overbrace{0 \;\; 0 \;\; \ldots \;\; 0}^{n_2} & \ldots & \overbrace{0 \;\; 0 \;\; \ldots \;\; 0}^{n_k} & \ldots & \overbrace{0 \;\; 0 \;\; \ldots \;\; 0}^{n_q} \\ 0 \;\; 0 \;\; \ldots \;\; 0 & 1 \;\; 1 \;\; \ldots \;\; 1 & \ldots & 0 \;\; 0 \;\; \ldots \;\; 0 & \ldots & 0 \;\; 0 \;\; \ldots \;\; 0 \\ \vdots \;\; \vdots \quad\quad \vdots & \vdots \;\; \vdots \quad\quad \vdots & & \vdots \;\; \vdots \quad\quad \vdots & & \vdots \;\; \vdots \quad\quad \vdots \\ 0 \;\; 0 \;\; \ldots \;\; 0 & 0 \;\; 0 \;\; \ldots \;\; 0 & \ldots & 1 \;\; 1 \;\; \ldots \;\; 1 & \ldots & 0 \;\; 0 \;\; \ldots \;\; 0 \\ \vdots \;\; \vdots \quad\quad \vdots & \vdots \;\; \vdots \quad\quad \vdots & & \vdots \;\; \vdots \quad\quad \vdots & & \vdots \;\; \vdots \quad\quad \vdots \\ 0 \;\; 0 \;\; \ldots \;\; 0 & 0 \;\; 0 \;\; \ldots \;\; 0 & \ldots & 0 \;\; 0 \;\; \ldots \;\; 0 & \ldots & 1 \;\; 1 \;\; \ldots \;\; 1 \end{array}\right]. \tag{8.21}$$

Then, to test the null hypothesis of equality of the q mean vectors, or expected value vectors, $\underline{\mu}_k$ in Chap. 7, will be the same as to test the null hypothesis of independence of the vectors $\underline{X}$ and $\underline{M}$. We will prove this fact by showing that the likelihood ratio statistic for these two tests is indeed the same statistic.

The test to the hypothesis of independence between $\underline{X}$ and $\underline{M}$ is in fact the test of fit of a 'one-way MANOVA' model, that is, of a MANOVA model with a single factor with q levels, corresponding the rejection of the null hypothesis of independence between $\underline{X}$ and $\underline{M}$ to the assumption of existence of a significant effect of the factor upon the mean value of the variables in $\underline{X}$, that is, to the assumption of differences among the q expected value vectors $\underline{\mu}_k$ $(k = 1, \ldots, q)$.

With the setup that we are considering, the MLE's of Σ_{11}, Σ_{22}, and Σ_{12} are respectively

$$V_{11} = \tfrac{1}{n} X(I_n - \tfrac{1}{n} E_{nn})X' \,, \qquad V_{22} = \tfrac{1}{n} M(I_n - \tfrac{1}{n} E_{nn})M' \tag{8.22}$$

and

$$V_{12} = \tfrac{1}{n} X(I_n - \tfrac{1}{n} E_{nn})M' \,, \tag{8.23}$$

so that the likelihood ratio statistic would be, in principle, given by (8.6).

However, it happens that the matrix V_{22} in (8.22) is not full rank, and as such it is non-invertible. In fact the q indicator variables in $\underline{M}$ form a set of 'one more than necessary' indicator variables, given the fact that if we know the values of $q-1$ of them, then we will also know the values for the indicator variable that is missing. This shows that there is an exact linear relation among these q indicator variables. For any given observation, they always add up to 1.

For this reason, we should in fact rather consider the vector $\underline{M}^*$ formed by any $q-1$ of the q components of $\underline{M}$, and consider then the matrix M^*, obtained from the matrix M by deleting the column that corresponds to the indicator variable that was droped from $\underline{M}$ to obtain $\underline{M}^*$. To make the exposition more precise and simple, both in terms of notation and of mental setup, we will consider that $\underline{M}^*$ is the vector obtained from $\underline{M}$ by deleting the first indicator variable, and that, accordingly, M^* is the matrix obtained from M by deleting its first column. If we denote this column by M_1, we will then have

$$\underset{(n\times q)}{M} = \left[\begin{array}{c|c} \underset{(n\times 1)}{M_1} & \underset{(n\times(q-1))}{M^*} \end{array} \right] \tag{8.24}$$

We will then reformulate the matrices V_{22} and V_{12} as

$$V_{22} = \frac{1}{n} M^*(I_n - \tfrac{1}{n} E_{nn})M^{*\prime} \qquad \text{e} \qquad V_{12} = \frac{1}{n} X(I_n - \tfrac{1}{n} E_{nn})M^{*\prime}$$

so that the inverse V_{22}^{-1} exists and $V_{11.2}$ may then be defined as $V_{11} - V_{12}V_{22}^{-1}V_{21}$.

Thus, in order to show that the likelihood ratio statistic for the test of independence of $\underline{X}$ and $\underline{M}$ is the same as the one in (8.6) we will have to show that the matrix

$$S_{11.2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$$

is equal to the matrix S in (7.5) in Chap. 7 and that the matrix

$$S_{11.2} + S_{12}S_{22}^{-1}S_{21} = S_{11}$$

is the same as the matrix $S+B$, for S and B in (7.5).

We will first show that the matrix

$$S_{12}S_{22}^{-1}S_{21} = X' \underbrace{\left(I_n - \tfrac{1}{n}E_{nn}\right) M^* \left(M^{*\prime}\left(I_n - \tfrac{1}{n}E_{nn}\right) M^*\right)^{-1} M^{*\prime}\left(I_n - \tfrac{1}{n}E_{nn}\right)}_{G} X$$

is the same as the matrix B in the expression after (7.6), where

$$B = n\overline{X}^{*\prime}\left(I_q - E_{qq}D\right)' D \left(I_q - E_{qq}D\right)\overline{X}^*$$

with

$$D = diag(n_1/n, n_2/n, \ldots, n_q/n) = \tfrac{1}{n}M'M$$

and

$$\overline{X}^* = \tfrac{1}{n}D^{-1}M'X = (M'M)^{-1}M'X$$

so that we may write

$$\begin{aligned}
B &= nX'M(M'M)^{-1}\left(I_q - \tfrac{1}{n}E_{qq}M'M\right)' \tfrac{1}{n}M'M\left(I_q - \tfrac{1}{n}E_{qq}M'M\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (M'M)^{-1}M'X \\
&= X'M\left((M'M)^{-1} - \tfrac{1}{n}E_{qq}\right)M'M\left((M'M)^{-1} - \tfrac{1}{n}E_{qq}\right)M'X \\
&= X'M\left((M'M)^{-1} - \tfrac{1}{n}E_{qq}\right)\left(I_q - \tfrac{1}{n}M'ME_{qq}\right)M'X \\
&= X'M\Big((M'M)^{-1} - \tfrac{1}{n}E_{qq} - \tfrac{1}{n}E_{qq} + \tfrac{1}{n^2}\underbrace{E_{qq}M'ME_{qq}}_{nE_{qq}}\Big)M'X \\
&= X'\underbrace{M\left((M'M)^{-1} - \frac{1}{n}E_{qq}\right)M'}_{F}X
\end{aligned}$$

where

$$F = M(M'M)^{-1}M' - \tfrac{1}{n}ME_{qq}M'$$

with

$$(M'M)^{-1} = diag\left(\tfrac{1}{n_k}, k = 1, \ldots, q\right),$$

in such a way that

$$M(M'M)^{-1}M' = bdiag\left(\tfrac{1}{n_k}E_{n_k n_k}, k = 1, \ldots, q\right)$$

while

$$\tfrac{1}{n}\underbrace{ME_{qq}}_{E_{nq}}M' = \tfrac{1}{n}E_{nn}\,,$$

so that we may write

$$F = bdiag\left(\tfrac{1}{n_k}E_{n_k n_k}, k = 1, \ldots, q\right) - \tfrac{1}{n}E_{nn}\,.$$

Thus, we have to show that $F = G$. In what matters G, we have to obtain an expression for $\left(M^{*\prime}\left(I_n - \frac{1}{n}E_{nn}\right)M^*\right)^{-1}$.

Since

$$\begin{aligned} M^{*\prime}\left(I_n - \tfrac{1}{n}E_{nn}\right)M^* &= M^{*\prime}M^* - \tfrac{1}{n}M^{*\prime}E_{nn}M^* \\ &= diag(n_2, \ldots, n_q) - \tfrac{1}{n}M^{*\prime}E_{n1}E_{1n}M^* \end{aligned}$$

where

$$M^{*\prime}E_{n1} = \begin{bmatrix} n_2 \\ \vdots \\ n_q \end{bmatrix} = diag(n_2, \ldots, n_q)E_{q-1,1}$$

and

$$E_{1n}M^* = [n_2, \ldots, n_q] = E_{1,q-1}diag(n_2, \ldots, n_q)\,,$$

we may write

$$\begin{aligned} M^{*\prime}\left(I_n - \tfrac{1}{n}E_{nn}\right)M^* &= diag(n_2, \ldots, n_q) - \tfrac{1}{n}\,diag(n_2, \ldots, n_q)E_{q-1,1}E_{1,q-1} \\ &\qquad \times diag(n_2, \ldots, n_q) \\ &= diag(n_2, \ldots, n_q)\underbrace{\left(I_{q-1} - \tfrac{1}{n}E_{q-1,q-1}diag(n_2, \ldots, n_q)\right)}_{H}, \end{aligned}$$

or,

$$\left(M^{*\prime}\left(I_n - \frac{1}{n}E_{nn}\right)M^*\right)^{-1} = H^{-1}\,diag\left(\tfrac{1}{n_2}, \ldots, \tfrac{1}{n_q}\right)\,,$$

where, using a slight generalization of expression (9) in section 2.11 of Morrison (1976), which shows that if

$$H = A - c\,\underline{b}\,\underline{d}'$$

then

$$H^{-1} = A^{-1} + \frac{c}{1 - c\,\underline{d}' A^{-1}\underline{b}} A^{-1}\underline{b}\,\underline{d}' A^{-1}\,,$$

so that for

$$\begin{aligned} H &= I_{q-1} - \tfrac{1}{n} E_{q-1,q-1}\, diag(n_2, \ldots, n_q) \\ &= \underbrace{I_{q-1}}_{A} - \underbrace{\tfrac{1}{n}}_{c}\, \underbrace{E_{q-1,1}}_{\underline{b}}\, \underbrace{E_{1,q-1}\, diag(n_2, \ldots, n_q)}_{\underline{d}'} \end{aligned}$$

we have

$$\begin{aligned} H^{-1} &= I_{q-1} + \frac{\frac{1}{n}}{1 - \frac{1}{n} \underbrace{E_{1,q-1}\, diag(n_2, ..., n_q)\, E_{q-1,1}}_{= n - n_1}} E_{q-1,1} E_{1,q-1}\, diag(n_2, ..., n_q) \\ &= I_{q-1} + \frac{\frac{1}{n}}{1 - \frac{1}{n}(n - n_1)} E_{q-1,q-1}\, diag(n_2, \ldots, n_q) \\ &= I_{q-1} + \tfrac{1}{n_1} E_{q-1,q-1}\, diag(n_2, \ldots, n_q) \end{aligned}$$

we may finally write

$$\left(M^{*\prime} \left(I_n - \frac{1}{n} E_{nn} \right) M^* \right)^{-1} = diag\left(\tfrac{1}{n_2}, \ldots, \tfrac{1}{n_q} \right) + \tfrac{1}{n_1} E_{q-1,q-1}\,.$$

We thus have

$$G = (I_n - \tfrac{1}{n} E_{nn}) M^* \left(diag\left(\tfrac{1}{n_2}, \ldots, \tfrac{1}{n_q} \right) + \tfrac{1}{n_1} E_{q-1,q-1} \right) M^{*\prime} (I_n - \tfrac{1}{n} E_{nn})$$

where

$$\underbrace{M^* diag\left(\tfrac{1}{n_2}, \ldots, \tfrac{1}{n_q} \right) M^{*\prime}}_{Z} = bdiag\left(0_{n_1 \times n_1}, \tfrac{1}{n_2} E_{n_2 n_2}, \ldots, \tfrac{1}{n_q} E_{n_q n_q} \right)$$

and

$$\begin{aligned} \underbrace{\tfrac{1}{n_1} M^* E_{q-1,q-1} M^{*\prime}}_{W} &= \tfrac{1}{n_1} \underbrace{M^* E_{q-1,1}}\, \underbrace{E_{1,q-1} M^{*\prime}} \\ &= \tfrac{1}{n_1} \left[\frac{\underline{0}_{n_1 \times 1}}{E_{(n-n_1),1}} \right] \left[\underline{0}_{1 \times n_1} \;\middle|\; E_{1,(n-n_1)} \right] \\ &= bdiag\left(0_{n_1 \times n_1}, \tfrac{1}{n_1} E_{(n-n_1)(n-n_1)} \right), \end{aligned}$$

so that we may write

$$G = (I_n - \tfrac{1}{n}E_{nn})\,(Z+W)\,(I_n - \tfrac{1}{n}E_{nn})$$

where

$$\begin{aligned}(I_n - \tfrac{1}{n}E_{nn})Z(I_n - \tfrac{1}{n}E_{nn}) &= \left(Z - \tfrac{1}{n}E_{nn}Z\right)(I_n - \tfrac{1}{n}E_{nn})\\ &= Z - \tfrac{1}{n}ZE_{nn} - \tfrac{1}{n}E_{nn}Z + \tfrac{1}{n^2}\,\underbrace{E_{nn}ZE_{nn}}_{=(n-n_1)E_{nn}}\\ &= Z - \tfrac{1}{n}\left[\begin{array}{c} 0_{n_1\times n} \\ \hline E_{(n-n_1)n}\end{array}\right] - \tfrac{1}{n}\left[\,0_{n\times n_1}\;\middle|\;E_{n_1(n-n_1)}\right] + \tfrac{n-n_1}{n^2}\,E_{nn}\end{aligned}$$

and

$$\begin{aligned}&(I_n - \tfrac{1}{n}E_{nn})W(I_n - \tfrac{1}{n}E_{nn})\\ &\quad= \left(W - \tfrac{1}{n}E_{nn}W\right)(I_n - \tfrac{1}{n}E_{nn})\\ &\quad= W - \tfrac{1}{n}WE_{nn} - \tfrac{1}{n}E_{nn}W + \tfrac{1}{n^2}\,\underbrace{E_{nn}WE_{nn}}_{=\frac{(n-n_1)^2}{n_1}E_{nn}}\\ &\quad= W - \tfrac{1}{nn_1}\left[\begin{array}{c} 0_{n_1\times n} \\ \hline (n-n_1)E_{(n-n_1)n}\end{array}\right] - \tfrac{1}{nn_1}\left[\,0_{n\times n_1}\;\middle|\;(n-n_1)E_{n_1(n-n_1)}\right]\\ &\qquad\qquad + \tfrac{n-n_1}{n^2n_1}\,E_{nn}\,,\end{aligned}$$

so that in the first diagonal block of G of dimension $n_1\times n_1$ we have

$$\frac{n-n_1}{n^2}E_{nn} + \frac{(n-n_1)^2}{n^2n_1}E_{nn} = \frac{n^2+n_1^2-2nn_1+nn_1-n_1^2}{n^2n_1}\,E_{nn} = \frac{n-n_1}{nn_1}\,E_{nn}$$

which is clearly equal to

$$\left(\tfrac{1}{n_1} - \tfrac{1}{n}\right)E_{n_1n_1}$$

and as such equal to the first diagonal block of dimensions $n_1\times n_1$ of F, being the remaining of G equal to the remaining of Z added with

$$-\tfrac{1}{n}ZE_{nn} - \tfrac{1}{n}E_{nn}Z - \tfrac{1}{n}WE_{nn} - \tfrac{1}{n}E_{nn}W + \tfrac{n-n_1}{nn_1}E_{nn} = -\tfrac{1}{n}\,E_{nn}\,,$$

so that we have indeed

$$F = G$$

and as such, for B in (7.5)

$$B = S_{12}S_{22}^{-1}S_{21}\,,$$

remaining now to show that we have, for S in (7.5)

$$S = S_{11.2}\,.$$

We may note that in Chap. 7 we have

$$\begin{aligned} S^* &= \sum_{k=1}^{q}\left(X_k - E_{n_k1}\underline{\overline{X}}'\right)'\left(X_k - E_{n_k1}\underline{\overline{X}}'\right) \\ &= \sum_{k=1}^{q} S_k^* = \sum_{k=1}^{q}(S_k + B_k) = \underbrace{\sum_{k=1}^{q} S_k}_{S} + \underbrace{\sum_{k=1}^{q} B_k}_{B} = S + B\,, \end{aligned}$$

or

$$\begin{aligned} S &= S^* - B \\ &= \sum_{k=1}^{q}\left(X_k - E_{n_k1}\underline{\overline{X}}'\right)'\left(X_k - E_{n_k1}\underline{\overline{X}}'\right) - B \end{aligned}$$

while now in Chap. 8 we have

$$\begin{aligned} S &= S_{11.2} = S_{11} - S_{12}S_{22}^{-1}S_{21} \\ &= X'(I_n - \tfrac{1}{n}E_{nn})X - B \end{aligned}$$

being enough to show that

$$\sum_{k=1}^{q}\left(X_k - E_{n_k1}\underline{\overline{X}}'\right)'\left(X_k - E_{n_k1}\underline{\overline{X}}'\right) = X'(I_n - \tfrac{1}{n}E_{nn})X\,,$$

where

$$\underset{n\times p}{X} = \begin{array}{c} \left[\begin{array}{c} X_1 \\ \hline \vdots \\ \hline X_k \\ \hline \vdots \\ \hline X_q \end{array}\right] \end{array} \begin{array}{c} n_1 \\ \\ n_k \\ \\ n_q \end{array}$$

and

$$\underline{\overline{X}} = \frac{1}{n}\sum_{k=1}^{q} n_k\,\underline{\overline{X}}_k \qquad \text{com} \qquad \underline{\overline{X}}_k = \frac{1}{n_k}\,X_k'\,E_{n_k1}\,,$$

so that

$$\underline{\overline{X}} = \frac{1}{n}\sum_{k=1}^{q} X_k'\,E_{n_k1}$$

and

$$E_{n_k1}\underline{\overline{X}}' = \frac{1}{n}\sum_{k=1}^{q} E_{n_k1}E_{1n_k}X_k = \frac{1}{n}\sum_{k=1}^{q} E_{n_kn_k}X_k$$

and thus

$$\begin{aligned}\sum_{k=1}^{q}\left(X_k - E_{n_k1}\underline{\overline{X}}'\right)'\left(X_k - E_{n_k1}\underline{\overline{X}}'\right) &= \sum_{k=1}^{q}\left(X_k - \tfrac{1}{n}\sum_{k=1}^{q}E_{n_kn_k}X_k\right)' \\ &\qquad\times\left(X_k - \tfrac{1}{n}\sum_{k=1}^{q}E_{n_kn_k}X_k\right) \\ &= \left(X - \tfrac{1}{n}E_{nn}X\right)'\left(X - \tfrac{1}{n}E_{nn}X\right) \\ &= \left((I_n - \tfrac{1}{n}E_{nn})X\right)'(I_n - \tfrac{1}{n}E_{nn})X \\ &= X'(I_n - \tfrac{1}{n}E_{nn})(I_n - \tfrac{1}{n}E_{nn})X \\ &= X'(I_n - \tfrac{1}{n}E_{nn})X\,,\end{aligned}$$

which shows that we have in fact, for S and B in Chap. 7,

$$S = S_{11.2} \qquad \text{and} \qquad B = S_{12}S_{22}^{-1}S_{21}\,,$$

being thus the likelihood ratio statistic used in Chap. 7 and the one used in this section the same statistic.

8.5 'Multi-way MANOVA' models

Considering together the results in subsection 8.3.1 and in section 8.4, we may easily develop a complete methodology for the analysis of a multi-way MANOVA model, in terms of significance tests for factors and their interactions, that is, in a Multivariate Analysis of Variance model with more than one factor, with the corresponding interactions, keeping in mind that when testing a given interaction all the indicator variables corresponding to all factors and interactions involved in that interaction.

Implement the study of a MANOVA model with 2 factors and interaction, using observations 1 through 150 in the file `body.dat` and using as response variables the variables 6 through 9, taking as factors the age class, with 3 levels: (i) ≤ 35, (ii) 35–≤ 50 and (iii) >50, and the height class, with 2 levels: (i) ≤ 70 e (ii) >70. Implement then other MANOVA models where you may consider as response variables any of the last 10 variables in the file, using all or only some of the observations in the file `body.dat`, considering as factors the class age, with the same levels as above and the height class now with 4 levels: (i) ≤ 67, (ii) 67–≤ 70, (iii) 70–≤ 72 e (iv) >72.

Implement other MANOVA models with datasets of your choice, with at least 3 factors.

Bibliografia

Coelho, C. A. (1986). Métodos Factoriais de Análise de Dados, 253 pp., Monografia apresentada no âmbito das Provas de Aptidão Científica e Capacidade Pedagógica, Instituto Sup. de Agronomia, Universidade Técnica de Lisboa.

Coelho, C. A. (1992). *Generalized Canonical Analysis*, Ph.D. Thesis, The University of Michigan, Ann Arbor, MI, U.S.A.

Coelho, C. A., Arnold, B. C. (2019) *Finite Form Representations for Meijer G and Fox H Functions – Applied to Multivariate Likelihood Ratio Tests using Mathematica®, Maxima and R*, Lecture Notes in Statistics, Springer, Cham, Switzerland, 515+xviii pp.

Coelho, C. A., Marques, F. J. (2014). A double decomposition of the test of independence of sets of variables that allows for a modeling view of this test, in *Mathematical Methods in Science and Mechanics*, Mastorakis, N. E., Mainardi, F. and Milanova, M. (eds.), 13–19 (ISBN: 978-960-474-396-4).

Morrison, D. F. (1976). *Multivariate Statistical Methods.* 2ªed. McGraw-Hill, New York. (Cap.5, Sec. 6)

SenGupta, A. (2004, 2014). Generalized Canonical Variables, in *Wiley StatsRef: Statistical Reference Online*, J. Wiley & Sons, Inc. (2014) (http://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat02670/abstract)