Lecture Notes

on

Multivariate Analysis

Preface

We would like to give an idea of some of the special feature of this monograph on Multivariate Analysis before we go further; let us tell you that it has been prepared according to new syllabus prescribed by UGC. We have written this notes in such a simple style that even the weak student will be able to understand very easily.

We are sure you will agree with us that the facts and formula of multivariate analysis is just the same in all the books, the difference lies in the method of presenting these facts to the students in such a simple way that while going through this notes, a student will feel as if a teacher is sitting by his side and explaining various things to him. We are sure that after reading this lecture notes, the student will develop a special interest in this field and would like to help to analyze such type of data in other discipline as well.

We think that the real judges of this monograph are the teachers concern and the student for whom it is meant. So, we request our teacher friends as well as students to point out our mistakes, if any, and send their comments and suggestions for the further improvement of this monograph. Wishes you a great success.

Yours sincerely

Dr. Rafiqullah Khan
aruke@rediffmail.com
Department of Statistics and Operations Research
Aligarh Muslim University, Aligarh.

CONTENTS

BASIC CONCEPT OF MULTIVARIATE ANALYSIS

Multivariate analysis is the analysis of observations on several correlated random variables for a number of individuals in one or more samples simultaneously, this analysis, has been used in almost all scientific studies.

For example, the data may be the nutritional anthropometrical measurements like height, weight, arm circumference, chest circumference, etc. taken from randomly selected students to assess their nutritional studies. Since here we are considering more than one variable this is called multivariate analysis.

The sample data may be a collection of measurements such as lengths and breadths of petals and sepals of different flowers from two different species to identify their group characteristics.

The data may be information such as annual income, saving, assets and number of children and so on collected from a randomly selected families.

As in the univariate case, we shall assume that a random sample of multivariate observations has been collected from an infinite or finite population. Also the multivariate techniques like their univariate counterpart allow one to evaluate hypothesis or results regarding the population by means of sample observations.

For this purpose, we consider the multiple measurements together as a system of measurements. Thus $x_{i\alpha}$, denotes the $i-$th measurement of the $\alpha-$th individual. Normally, we denote the number of variables by p and number of individuals by n. The n measurements on p variables can be arranged as follows:

Variable/Characteristic	Individuals					
	1	2	...	α	...	n
X_1	x_{11}	x_{12}	...	$x_{1\alpha}$	...	x_{1n}
X_2	x_{21}	x_{22}	...	$x_{2\alpha}$	...	x_{2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_i	x_{i1}	x_{i2}	...	$x_{i\alpha}$	...	x_{in}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_p	x_{p1}	x_{p2}	...	$x_{p\alpha}$	...	x_{pn}

This can be written in the matrix form as

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1\alpha} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2\alpha} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{i\alpha} & \cdots & x_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{p\alpha} & \cdots & x_{pn} \end{pmatrix} = (x_{i\alpha})\,;\; i=1,2,\cdots p,\; \alpha=1,2,\cdots n.$$

This is also known as data matrix.

The main aim of the most of the multivariate technique is to summarize a large body of data by means of relatively few parameters, or to reduce the number of variables by employing

suitable linear transformations and to choose a very limited number of the resulting linear combination in some optimum manner, disregarding the remaining linear combination in the hope that they do not contain much information. Among the multiple techniques there are two main groups. Some methods concerned with relationship between variables called **variable-oriented.** The important variable-oriented techniques are total, partial, multiple and canonical correlations, principal component analysis and so on. While the other group concerned with the relationships between individuals called **individual-oriented**. In such cases problem of classification arises to assign an individual of unknown origin. One of the two populations we make use of discriminant analysis.

Basic theory of matrices

- Given a $p\times p$ square matrix $A=(a_{ij})$, the elements $a_{11}\cdots a_{pp}$ form the principal diagonal or simply the diagonal of A and are called its diagonal elements, whereas the elements $a_{1p}\,a_{2\,p-1}\cdots a_{p1}$ constitute the secondary diagonal of A. For $i\neq j$, a_{ij} are the off diagonal elements of A.
- A matrix of order p, say $A=(a_{ij})$ is said to be symmetric, if

  $a_{ij}=a_{ji}$, $\forall\;\; i, j=1,2,\cdots,p$, or $A=A'$.
- And skew-symmetric, if $a_{ij}=-a_{ji}$, $\forall\;\; i, j=1,2,\cdots,p$, or $A=-A'$.
- A matrix $A=(a_{ij})$ is called a diagonal matrix, if

  $a_{ij}=0$, $\forall\;\; i\neq j$, and we write $A=diag\,(a_{11},\cdots,a_{pp})$.
- A nonzero diagonal matrix whose diagonal elements are all equal is called a scalar matrix and a scalar matrix whose diagonal elements are all unities is called a unit matrix or an identity matrix. A unit matrix is generally denoted by I, in which case a scalar matrix is of the form cI, where c is some nonzero scalar. It may be easily seen that $AI=A$, $IB=B$, $A(cI)=cA$, $(aI)B=aB$, provided the products are all defined.
- A square matrix $A=(a_{ij})$ is said to be upper triangular if $a_{ij}=0$, for all $i>j$ and lower triangular if $a_{ij}=0$, for all $i<j$, triangular if A is either upper or lower triangular.
- A square matrix A is said to be idempotent if $A^2=A$.
- A square matrix A is said to be orthogonal if $AA'=A'A=I$.
- Let $\underline{a}_i'$ denote the $i-$th row of matrix $A=(a_{ij})$ of order p and $\underline{b}_j$ its $j-$th column, then $(i, j)-$th element of AA' and $A'A$ are $\underline{a}_i'\,\underline{a}_j$ and $\underline{b}_i'\,\underline{b}_j$ respectively. Now if A is orthogonal, we have $AA'=A'A=I$, and hence

  $\underline{a}_i'\,\underline{a}_j=\underline{b}_i'\,\underline{b}_j=1$, whenever $i=j$

  $=0$, whenever $i\neq j$

  Thus, both $(\underline{a}_1,\cdots,\underline{a}_p)$ and $(\underline{b}_1,\cdots,\underline{b}_p)$ are orthonormal.
- The trace of a matrix A, written '$tr\,A$', is the sum of the diagonal elements of A.
- $tr\,(A+B)=tr\,A+tr\,B$, whenever A and B are square matrices of same order.

- $tr\,AB = tr\,BA$, whenever both AB and BA are defined.
- $tr\,ABC = tr\,BCA = tr\,CAB$, provided all the three products are defined.
- A square matrix A is said to be non-singular, it there exists a square matrix B of same order, called the inverse of A, such that $AB = BA = I$.
- The inverse of a non-singular matrix A is unique, and it is because of this uniqueness that the inverse of A, if it exists, is generally denoted by A^{-1}.
- $(A^{-1})^{-1} = A$ and $(A^{-1})' = (A')^{-1}$, whenever A is non-singular.
- $(AB)^{-1} = B^{-1}A^{-1}$, whenever A and B are non-singular matrices of the same order.
- The determinant of a diagonal or triangular matrix equals the product of its diagonal elements.
- $|A'| = |A|$, $|AB| = |BA| = |A||B| = |B||A|$.
- $|A| \neq 0$, if A is non-singular, since $1 = |AA^{-1}| = |A||A^{-1}|$.
- $|A| = 1$ or -1, if A is orthogonal.
- $|aA| = a^p |A|$, where p is the order of A.
- $|A| = 0$ whenever A has a zero-row or a zero-column or two identical rows (columns).
- The determinant changes sign if any two rows (columns) are interchanged.
- The determinant remains unaltered when a row (column) is replaced with the sum of that row (column) and a scalar multiple of another row (column).
- If $A = diag\,(A_{11} \cdots A_{qq})$, where $A_{11} \cdots A_{qq}$ are all square matrices, then $|A| = |A_{11}| \cdots |A_{qq}|$.
- If A is of the form
$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ 0 & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{qq} \end{pmatrix} \text{ or } \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ A_{21} & A_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ A_{1q} & A_{q2} & \cdots & A_{qq} \end{pmatrix}, \text{ then } |A| = |A_{11}| \cdots |A_{qq}|.$$
- If $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, where both A_{11} and A_{22} are non-singular, then
$$A^{-1} = \begin{pmatrix} A_{11.2}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22.1}^{-1} \end{pmatrix}.$$
- If $A = diag\,(A_{11} \cdots A_{qq})$, then $A^{-1} = diag\,(A_{11}^{-1} \cdots A_{qq}^{-1})$.
- Given a square matrix A of order p, the matrix $(A - \lambda I)$, where λ is a scalar variable, is called the latent (or characteristic) matrix of A, and the polynomial, $|A - \lambda I|$, which is of degree p, is called the latent (or characteristic) polynomial of A and the equation $|A - \lambda I| = 0$ its characteristic equation.

- If we write $|A - \lambda I| = a_0 + a_1\lambda + \cdots + a_p\lambda^p$ it can be easily seen that $a_0 = |A|$, $a_p = (-1)^p$.
- Every square matrix satisfies its own characteristic equation.
- Given a square matrix A, the roots of its characteristic equation, $|A - \lambda I| = 0$, are called its latent (or characteristic or secular) roots or eigen (or proper) values. Clearly if A is $p \times p$, then A has p latent roots.
- The latent roots of a diagonal and triangular matrix are their diagonal elements.
- The determinant of a square matrix is the product of its latent roots.
- A square matrix is non-singular if and only if none of its latent roots is zero.
- Given a latent roots λ of a square matrix A, $\underline{x}$ is called a latent vector of A corresponding to the root λ if $(A - \lambda I)\underline{x} = \underline{0}$, i.e. $A\underline{x} = \lambda\,\underline{x}$.
- If λ is a latent root of a non-singular matrix A, then λ^{-1} is a latent root of A^{-1}.
- If λ is a latent root of an orthogonal matrix A, so is λ^{-1} a latent roots of A.
- Latent roots of an idempotent matrix are zeros and unities only.
- Latent roots of a real symmetric matrix are all real.
- Two square matrices A and B of the same order are said to be similar if there exists a non-singular matrix C such that $A = C^{-1}BC$.
- Similar matrices have the same characteristic polynomials.
- Similar matrices have the same latent roots.
- If the latent roots of A are all distinct, then A is similar to a diagonal matrix whose diagonal elements are the latent roots of A.
- If A is a real symmetric matrix, then there exits an orthogonal matrix C such that $C'AC$ is a diagonal matrix.
- Let $A = (a_{ij})$ be a real symmetric matrix of order p and $\underline{x}' = (x_1, \cdots, x_p)$ a $p-$tuple of real variables. The polynomial
$\sum_i \sum_j a_{ij}\, x_i\, x_j = \underline{x}' A \underline{x}$ is called a $p-$ary quadratic form.
- The quadratic form $\underline{x}' A \underline{x}$ (or equivalently the real symmetric matrix A) is said to be

  i) positive definite, if $\underline{x}' A \underline{x} > 0$ for all $\underline{x} \neq 0$.

  ii) negative definite, if $\underline{x}' A \underline{x} < 0$ for all $\underline{x} \neq 0$.

  iii) positive semi-definite or nonnegative definite, if $\underline{x}' A \underline{x} \geq 0$ for all $\underline{x} \neq 0$.

iv) negative semi-definite or no positive definite, if $\underline{x}' A\underline{x} \le 0$ for all $\underline{x} \ne 0$.

v) indefinite if it takes on both positive and negative values.

- The diagonal elements of a positive definite matrix are all positive.
- The diagonal elements of a negative definite matrix are all negative.
- An $p-$ary quadratic form $\underline{x}' A\underline{x}$ can be equivalently reduced by an orthogonal transformation to the diagonal form $\lambda_1 y_1^2 + \cdots + \lambda_p y_p^2$, where $\lambda_1, \cdots, \lambda_p$ are the latent roots of A.
- An $p-$ary quadratic form $\underline{x}' A\underline{x}$ can be equivalently reduced by a non-singular transformation to the canonical form $d_1 y_1^2 + \cdots + d_p y_p^2$.
- An $p-$ary quadratic form $\underline{x}' A\underline{x}$ is positive definite (negative definite) if and only if the talent roots of A are all positive (negative).
- If A is positive definite (negative definite), $-A$ is negative definite (positive definite).
- If A is positive definite, $|A| > 0$.
- If Σ is $p \times p$ and positive definite and if A is a $q \times p$ matrix of rank q, then there exists an $(p-q) \times p$ matrix B of rank $p-q$ such that $A\Sigma B' = 0$, in which case $D = \begin{pmatrix} A \\ B \end{pmatrix}$ is non-singular.
- Gamma integral

  i) $\Gamma(m) = \int_0^\infty x^{m-1} e^{-x}\, dx$, ii) $\alpha^{-m}\Gamma(m) = \int_0^\infty x^{m-1} e^{-\alpha x}\, dx$
- Gamma duplication formula

  $2^{2m-1}\Gamma(m)\Gamma(m+1/2) = \sqrt{\pi}\,\Gamma(2m)$.

THE MULTIVARAITE NORMAL DISTRIBUTION

Notations of multivariate distribution

If X is a random variable (rv) then the cumulative distribution function (abbreviated as cdf) of X is given by

$F(x) = \Pr(X \le x)$, where x is a real number.

If $F(x)$ is absolutely continuous, then $\frac{d}{dx} F(x) = f(x)$ is called the density function of X

and

$$F(x) = \int_{-\infty}^{x} f(u)\, du .$$

Joint distribution

First consider the case of two (real) random variables say X and Y.

Let X and Y be two rv. The joint cdf of X and Y is given by

$F(x, y) = \Pr(X \le x,\ Y \le y)$, defined for every pair of real numbers $(x,\ y)$.

We are interested in cases where $F(x,\ y)$ is absolutely continuous. If $F(x,\ y)$ is absolutely continuous, then

$\frac{\partial^2 F(x, y)}{\partial x\, \partial y} = f(x, y)$, and is called the joint density function of X and Y, and

$$F(x,\ y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v)\, du\, dv .$$

Now we consider the case of p random variables $X_1, X_2, \cdots, X_p$. The cdf is

$F(x_1, x_2, \cdots, x_p) = \Pr(X_1 \le x_1, X_2 \le x_2, \cdots, X_p \le x_p)$,

defined for every set of real numbers $x_1, x_2, \cdots, x_p$, and the density function, if $F(x_1, x_2, \cdots, x_p)$ is absolutely continuous, is

$\frac{\partial^p F(x_1, x_2, \cdots, x_p)}{\partial x_1\, \partial x_2 \cdots \partial x_p} = f(x_1, x_2, \cdots, x_p)$, and

$$F(x_1, x_2, \cdots, x_p) = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(u_1, u_2, \cdots, u_p)\, du_1 \cdots du_p .$$

Marginal distribution

Let $F(x,\ y)$ be the cdf of two $rv's$ X and Y, the marginal cdf of X is

$\Pr(X \le x) = \Pr\{X \le x,\ Y \le \infty\} = F(x, \infty)$

Let this be $F(x)$. Clearly

$$F(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, v)\, dv\, du . \tag{2.1}$$

Now

$$\int_{-\infty}^{\infty} f(u, v)\, dv = f(u)$$

where $f(u)$ is called the marginal density function of X and from (2.1)

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

In a similar fashion we define $G(y)$, the marginal cdf of Y, and $g(y)$, the marginal density function of Y.

Now we turn to the general case

Let $F(x_1, x_2, \cdots, x_p)$ be the cdf of the $rv's$ $X_1, X_2, \cdots, X_p$. The marginal cdf of some of $X_1, X_2, \cdots, X_p$, say $X_1, X_2, \cdots, X_r$ $(r < p)$ is

$$\begin{aligned}\Pr(X_1 \le x_1, \cdots, X_r \le x_r) &= \Pr(X_1 \le x_1, \cdots, X_r \le x_r, X_{r+1} \le \infty, \cdots, X_p \le \infty)\\ &= F(x_1, x_2, \cdots, x_r, \infty, \cdots, \infty)\\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_r} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \cdots u_p)\, du_1 \cdots du_p\end{aligned}$$

and the marginal density function of $X_1, X_2, \cdots, X_r$ is

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \cdots u_p)\, du_{r+1} \cdots du_p .$$

The marginal distribution and density of any other subset of $X_1, X_2, \cdots, X_p$ are obtained in the obvious similar fashion.

Statistical Independence

Two random variables X, Y with cdf $F(x, y)$ are said to be independent if

$F(x, y) = F(x)\, G(y)$,

where $F(x)$ is the marginal cdf of X and $G(y)$ be the marginal cdf of Y, then the density function of X and Y is

$f(x\ y) = \frac{\partial^2 F(x, y)}{\partial x\, \partial y} = \frac{\partial^2 F(x)\, G(y)}{\partial x\, \partial y} = \frac{d\, F(x)}{\partial x} \frac{d\, G(y)}{\partial y} = f(x)\, g(y)$. Conversely, if

$f(x, y) = f(x)\, g(y)$, then

$$\begin{aligned}F(x, y) &= \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v)\, du\, dv = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u)\, g(v)\, du\, dv\\ &= \int_{-\infty}^{x} f(u\, du \int_{-\infty}^{y} g(v)\, dv = F(x)\, G(y) .\end{aligned}$$

General case,

Given $F(x_1, x_2, \cdots, x_p)$ be the cdf of $X_1, X_2, \cdots, X_p$, the variables $X_1, X_2, \cdots, X_p$ are called mutually independent if

$F(x_1, x_2, \cdots, x_p) = F_1(x_1) \cdots F_p(x_p)$,

where $F_i(x_i)$ is the marginal cdf of X_i, $i = 1, 2, \cdots, p$.

The set $X_1, X_2, \cdots, X_r$ is independent of $X_{r+1}, X_{r+2}, \cdots, X_p$ if

$$F(x_1, x_2, \cdots, x_p) = F(x_1, x_2, \cdots, x_r, \infty \cdots \infty)\, F(\infty \cdots \infty, x_{r+1}, x_{r+2}, \cdots, x_p).$$

Conditional distribution

The conditional density function of Y, given X, for two $rv's$ X and Y is defined as

$$f(y \mid x) = \frac{f(x, y)}{f(x)}, \text{ provided that } f(x) > 0.$$

In the general case of $X_1, X_2, \cdots, X_p$ with *cdf* $F(x_1, x_2, \cdots, x_p)$, the conditional density function of $X_1, X_2, \cdots, X_r$, given $X_{r+1} = x_{r+1}, X_{r+2} = x_{r+2}, \cdots, X_p = x_p$ is

$$\frac{f(x_1, x_2, \cdots, x_p)}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \cdots u_r, u_{r+1}, \cdots, u_p)\, du_1 \cdots du_r}.$$

Transformation of variables

Let $X_1, X_2, \cdots, X_p$ have the joint density function $f(x_1, x_2, \cdots, x_p)$. Consider p real-valued functions $y_i = y_i(x_1, x_2, \cdots, x_p)$, $i = 1, 2, \cdots, p$. We assumed that the transformation of Y to X be one-to-one, the inverse transformation is $x_i = x_i(y_1, y_2, \cdots, y_p)$, $i = 1, 2, \cdots, p$. Let the random variable $Y_1, Y_2, \cdots, Y_p$ be defined by

$$Y_i = y_i(X_1, X_2, \cdots, X_p)$$

Then the joint density function of $Y_1, Y_2, \cdots, Y_p$ is

$$g(y_1, y_2, ..., y_p) = f[x_1(y_1, y_2, ..., y_p), ..., x_p(y_1, y_2, ..., y_p)] \mid J \mid, \text{ where}$$

$$J = \text{mod} \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_p} \\ \vdots & \cdots & \vdots \\ \frac{\partial x_p}{\partial y_1} & \cdots & \frac{\partial x_p}{\partial y_p} \end{vmatrix} = \text{Jacobian of transformation.}$$

Form of normal density function

1) The univariate normal density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{(2\pi)^{1/2} \mid \Sigma \mid^{1/2}}\, e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}, \quad -\infty < x < \infty,$$

$$-\infty < \mu < \infty, \quad \Sigma = \sigma^2 > 0.$$

To show that

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Consider,

$$\int_{-\infty}^{\infty} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx. \text{ Put } \frac{x-\mu}{\sigma} = y, \text{ then } dx = \sigma\, dy.$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} \sigma\, dy = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}y^2} dy.$$

Again put

$$\frac{1}{2}y^2 = t \text{ or } t^2 = 2y \text{ or } 2t\, dt = 2\, dy$$

$$\text{or} \quad dt = \frac{1}{t} dy \quad \text{or} \quad dt = \frac{1}{\sqrt{2y}} dy = \frac{1}{\sqrt{2}} y^{-\frac{1}{2}} dy, \text{ then}$$

$$\int_{-\infty}^{\infty} f(x)\, dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-y} \left(\frac{1}{\sqrt{2}} y^{-1/2}\right) dy = \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-y} y^{\frac{1}{2}-1} dy = \frac{1}{\sqrt{\pi}} \Gamma 1/2 = 1.$$

Since gamma function $\int_0^{\infty} e^{-\alpha} \alpha^{p-1} d\alpha = \Gamma p$, and $\Gamma 1/2 = \sqrt{\pi}$.

Moment generating function of univariate normal distribution

If $X \sim N(\mu, \sigma^2)$, then $Y = \frac{X-\mu}{\sigma} \sim N(0,1)$.

By definition,

$$M_Y(t) = E[e^{ty}] = \int e^{ty} f(y)\, dy = \int_{-\infty}^{\infty} e^{ty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y^2 - 2ty)} dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y^2 - 2ty + t^2 - t^2)} dy = \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t)^2} dy.$$

Let

$$y - t = z, \text{ and } dy = dz, \text{ then}$$

$$M_Y(t) = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = e^{t^2/2}.$$

Now,

$$Y = \frac{X-\mu}{\sigma} \quad \Rightarrow \quad X = \mu + \sigma Y.$$

Thus,

$$M_X(t) = E\, e^{tx} = E\, e^{t(\mu + \sigma y)} = E\, e^{t\mu} e^{t\sigma y} = e^{t\mu} E[e^{t(\sigma y)}]$$

$$= e^{t\mu} M_Y(t\sigma) = e^{t\mu + t^2\sigma^2/2}.$$

2) The bivariate normal density function

$$f(x_1,x_2)=\frac{1}{2\pi\,\sigma_1\sigma_2\sqrt{1-\rho^2}}\,e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}$$

or $f(\underline{x})=\dfrac{1}{2\pi|\Sigma|^{1/2}}\,e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})}$.

Proof: Let $\underline{X}=\begin{pmatrix}X_1\\X_2\end{pmatrix}$, corresponding $\underline{\mu}=\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix}$, and

$$\Sigma=\begin{pmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{21}&\sigma_{22}\end{pmatrix}=\begin{pmatrix}\sigma_1^2&\rho\,\sigma_1\sigma_2\\\rho\,\sigma_1\sigma_2&\sigma_2^2\end{pmatrix},\ \text{where }\rho=\frac{\sigma_{ij}}{\sigma_i\sigma_j}.$$

Now

$$|\Sigma|=\sigma_1^2\sigma_2^2-\rho^2\sigma_1^2\sigma_2^2=\sigma_1^2\sigma_2^2\,(1-\rho^2)\,.$$

Thus,

$$\Sigma^{-1}=\frac{1}{(1-\rho^2)\,\sigma_1^2\sigma_2^2}\begin{pmatrix}\sigma_2^2&-\rho\,\sigma_1\sigma_2\\-\rho\,\sigma_1\sigma_2&\sigma_1^2\end{pmatrix}=\frac{1}{(1-\rho^2)}\begin{pmatrix}\frac{1}{\sigma_1^2}&-\frac{\rho}{\sigma_1\sigma_2}\\-\frac{\rho}{\sigma_1\sigma_2}&\frac{1}{\sigma_2^2}\end{pmatrix}$$

Consider

$$(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})=(x_1-\mu_1,x_2-\mu_2)\frac{1}{(1-\rho^2)}\begin{pmatrix}\frac{1}{\sigma_1^2}&-\frac{\rho}{\sigma_1\sigma_2}\\-\frac{\rho}{\sigma_1\sigma_2}&\frac{1}{\sigma_2^2}\end{pmatrix}\begin{pmatrix}x_1-\mu_1\\x_2-\mu_2\end{pmatrix}$$

$$=\frac{1}{1-\rho^2}\left(\frac{x_1-\mu_1}{\sigma_1^2}-\frac{\rho\,(x_2-\mu_2)}{\sigma_1\sigma_2},\ -\frac{\rho\,(x_1-\mu_1)}{\sigma_1\sigma_2}+\frac{x_2-\mu_2}{\sigma_2^2}\right)\begin{pmatrix}x_1-\mu_1\\x_2-\mu_2\end{pmatrix}$$

$$=\frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-\frac{\rho\,(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}-\frac{\rho\,(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$$

$$=\frac{1}{1-\rho^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right].$$

Therefore,

$$f(\underline{x})=\frac{1}{(2\pi)^{2/2}\,|\Sigma|^{1/2}}\,e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})}\,.$$

This is the *pdf* of bivariate normal distribution with mean $\underline{\mu}$ and variance covariance matrix Σ, thus $\underline{X}\sim N_2\,(\underline{\mu},\Sigma)$.

Exercise: Find the marginal of X_1 and X_2, if $f(x_1,x_2)=\dfrac{1}{(2\pi)\,\sigma_1\sigma_2\sqrt{1-\rho^2}}\,e^{-Q/2}$,

where

$$Q=\frac{1}{(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right].$$

Solution: For, marginal of X_1, we have

$$Q=\frac{1}{(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]$$

$$=\frac{1}{(1-\rho^2)}\left[\rho^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2+(1-\rho^2)\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right.$$

$$\left.-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]$$

$$=\frac{1}{(1-\rho^2)}\left[\left(\frac{x_2-\mu_2}{\sigma_2}\right)-\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right]^2+\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2$$

$$=\frac{1}{\sigma_2^2(1-\rho^2)}\left[x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)\right]^2+\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2$$

$$=\left(\frac{x_2-\mu_2^*}{\sigma_2^*}\right)^2+\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2,$$

where $\mu_2^*=\mu_2+\rho\dfrac{\sigma_2}{\sigma_1}(x_1-\mu_1)$, and $\sigma_2^{*2}=\sigma_2^2(1-\rho^2)$.

Therefore,

$$f(x_1,x_2)=\frac{1}{(2\pi)\,\sigma_1\sigma_2^*}\exp\left[-\frac{1}{2}\left\{\left(\frac{x_2-\mu_2^*}{\sigma_2^*}\right)^2+\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\}\right]$$

$$=\left[\frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\}\right]\left[\frac{1}{\sqrt{2\pi}\,\sigma_2^*}\exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2\,^*}{\sigma_2\,^*}\right)^2\right\}\right]$$

$$=f(x_1)\,f(x_2\,|\,x_1),\ \text{since}\ \int_{x_2}f(x_1,x_2)\,dx_2=f(x_1)\,.$$

Hence, the marginal distribution of X_1 is $N(\mu_1,\sigma_1^2)$, i.e. $X_1\sim N(\mu_1,\sigma_1^2)$.

For, marginal of X_2. We have

$$Q=\frac{1}{(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]$$

$$=\frac{1}{(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\rho^2\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right.$$

$$\left.+(1-\rho^2)\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]$$

$$=\frac{1}{(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)-\rho\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2$$

$$=\frac{1}{\sigma_1^2(1-\rho^2)}\left[x_1-\mu_1-\rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right]^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2$$

$$=\left(\frac{x_1-\mu_1^*}{\sigma_1^*}\right)^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2,$$

where, $\mu_1^*=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)$, and $\sigma_1^{*2}=\sigma_1^2(1-\rho^2)$. Therefore, the *pdf* of bi-variate normal distribution is

$$f(x_1,x_2)=\frac{1}{(2\pi)\sigma_1^*\sigma_2}\exp\left[-\frac{1}{2}\left\{\left(\frac{x_1-\mu_1^*}{\sigma_1^*}\right)^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right]$$

$$=\left[\frac{1}{\sqrt{2\pi}\,\sigma_2}\exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right]\left[\frac{1}{\sqrt{2\pi}\,\sigma_1^*}\exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1^*}{\sigma_1^*}\right)^2\right\}\right]$$

$$=f(x_2)\ f(x_1\mid x_2)\text{, where, } \int_{x_1}f(x_1,x_2)\,dx_1=f(x_2).$$

Hence, the marginal distribution of X_2 is $N(\mu_2,\sigma_2^2)$, i.e. $X_2\sim N(\mu_2,\sigma_2^2)$.

Moment generating function of bi-variate normal distribution

It is defined as

$$M_{\underline{X}}(\underline{t})=E[e^{\underline{t}'\underline{x}}]\text{, where } \underline{t}=\begin{pmatrix}t_1\\t_2\end{pmatrix}\text{, and } \underline{x}=\begin{pmatrix}x_1\\x_2\end{pmatrix}.$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{t_1x_1+t_2x_2}f(x_1,x_2)\,dx_1\,dx_2$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{t_1x_1}\,e^{t_2x_2}f(x_1)\,f(x_2\mid x_1)\,dx_1\,dx_2$$

$$=\int_{-\infty}^{\infty}e^{t_1x_1}f(x_1)\left(\int_{-\infty}^{\infty}e^{t_2x_2}f(x_2\mid x_1)\,dx_2\right)dx_1.$$

The integral within the brackets is the *mgf* of the conditional *pdf* $f(x_2|x_1)$. Since $f(x_2|x_1)$ is *pdf* of a normal distribution with mean $\mu_2^*=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)$, and variance $\sigma_2^{*2}=\sigma_2{}^2(1-\rho^2)$.

Thus,

$$\int_{-\infty}^{\infty}e^{t_2x_2}f(x_2|x_1)\,dx_2=e^{t_2\mu_2^*+\frac{1}{2}\sigma_2^{*2}t_2^2}\text{, as } M_X(t)=e^{t\mu+\frac{1}{2}t^2\sigma^2}\text{ for } N(\mu,\sigma^2).$$

Therefore,

$$M_{\underline{X}}(\underline{t})=\int_{-\infty}^{\infty}e^{t_1x_1}\,e^{t_2\left[\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)\right]+\frac{1}{2}t_2^2\sigma_2^2(1-\rho^2)}f(x_1)\,dx_1$$

$$=\exp\left[t_2\mu_2-t_2\,\rho\frac{\sigma_2}{\sigma_1}\mu_1+\frac{1}{2}t_2^2\,\sigma_2^2\,(1-\rho^2)\right]\int_{-\infty}^{\infty}e^{\left(t_1+t_2\,\rho\frac{\sigma_2}{\sigma_1}\right)x_1}f(x_1)\,dx_1$$

$$=\exp\left[t_2\mu_2-t_2\,\rho\frac{\sigma_2}{\sigma_1}\mu_1+\frac{1}{2}t_2^2\,\sigma_2^2\,(1-\rho^2)\right]$$

$$\times\exp\left[\left(t_1+t_2\,\rho\frac{\sigma_2}{\sigma_1}\right)\mu_1+\frac{1}{2}\left(t_1+t_2\,\rho\frac{\sigma_2}{\sigma_1}\right)^2\sigma_1^2\right]$$

$$=\exp\left[t_2\mu_2-t_2\,\rho\frac{\sigma_2}{\sigma_1}\mu_1+\frac{1}{2}t_2^2\,\sigma_2^2\,(1-\rho^2)+t_1\,\mu_1+t_2\,\mu_1\,\rho\frac{\sigma_2}{\sigma_1}\right.$$

$$\left.+\frac{\sigma_1^2}{2}\left(t_1^2+t_2^2\,\rho^2\,\frac{\sigma_2^2}{\sigma_1^2}+2t_1t_2\rho\frac{\sigma_2}{\sigma_1}\right)\right]$$

$$=\exp\left(t_1\,\mu_1+t_2\,\mu_2+\frac{1}{2}t_2^2\,\sigma_2^2-\frac{1}{2}t_2^2\,\rho^2\,\sigma_2^2+\frac{1}{2}t_1^2\,\sigma_1^2+\frac{1}{2}t_2^2\rho^2\sigma_2^2+t_1t_2\rho\,\sigma_1\sigma_2\right)$$

$$=\exp\left[t_1\,\mu_1+t_2\,\mu_2+\frac{1}{2}(t_1^2\,\sigma_1^2+2\,t_1t_2\,\rho\sigma_1\sigma_2+t_2^2\,\sigma_2^2)\right].$$

$$M_{X_1,X_2}(t_1,0)=E[e^{t_1x_1+0}]=E[e^{t_1x_1}]=M_{X_1}(t_1)=\exp\left(t_1\,\mu_1+\frac{1}{2}t_1^2\,\sigma_1^2\right)$$

$$M_{X_1,X_2}(0,t_2)=\exp\left(t_2\,\mu_2+\frac{1}{2}t_2^2\,\sigma_2^2\right).$$

Note: If $\rho=0$, $M_{X_1X_2}(t_1,t_2)=M_{X_1X_2}(t_1,0)\ M_{X_1X_2}(0,t_2)$, i.e. X_1 and X_2 are independently distributed, if and only if $M(t_1,t_2)=M(t_1,0)\ M(0,t_2)$.

Define,

$$K(t_1,t_2)=\log M(t_1,t_2)$$

i) $\dfrac{\partial K(t_1,t_2)}{\partial t_2}\Big|_{t_1=0,t_2=0}=E(X_2)=\mu_2$.

ii) $\dfrac{\partial K(t_1,t_2)}{\partial t_1}\Big|_{t_1=0,t_2=0}=E(X_1)=\mu_1$.

iii) $\dfrac{\partial^2 K(t_1,t_2)}{\partial t_1^2}\Big|_{t_1,t_2=0}=V(X_1)=\sigma_1^2$.

iv) $\dfrac{\partial^2 K(t_1,t_2)}{\partial t_2^2}\Big|_{t_1,t_2=0}=V(X_2)=\sigma_2^2$.

v) $\dfrac{\partial^2 K(t_1,t_2)}{\partial t_1\,\partial t_2}\Big|_{t_1,t_2=0}=Cov(X_1,X_2)=\rho\sigma_1\sigma_2$.

Exercise: Given $(X,Y)\sim N_2(3,1,16,25,3/5)$. Find i) $P(3<Y<8)$,

ii) $P(3<Y<8\,|\,X=7)$, iii) $P(-3<X<3)$, and iv) $P(-3<X<3\,|\,Y=-4)$.

Solution: We are given $\mu_1=3,\ \mu_2=1,\ \sigma_1^2=16,\ \sigma_2^2=25$, and $\rho=\dfrac{3}{5}$, i.e. $X\sim N(3,16)$, and $Y\sim N(1,25)$.

i) $P(3<Y<8)=P\left(\dfrac{3-1}{5}<\dfrac{Y-\mu_2}{\sigma_2}<\dfrac{8-1}{5}\right)=P(0.4<Z<1.4)=0.4192-0.1554\cong 0.264$.

ii) Since $(Y\,|\,X=x)\sim N(\mu_2^*,\sigma_2^{*2})$, where

$$\mu_2^*=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)=1+\frac{3}{5}\times\frac{5}{4}(7-3)=4, \text{ and}$$

$$\sigma_2^{*2}=\sigma_2^2(1-\rho^2)=25\left(1-\frac{9}{25}\right)=16.$$

Therefore,

$$P(3<Y<8\,|\,X=7)=P\left(\frac{3-4}{4}<\frac{Y-\mu_2^*}{\sigma_2^*}<\frac{8-4}{4}\right)=P(-\frac{1}{4}<Z<1)$$

$$=P(-0.25<Z<1)=P(-0.25<Z<0)+P(0<Z<1)=0.440.$$

iii) Given $X\sim N(3,16)$

$$P(-3<X<3)=P\left(\frac{-3-3}{4}<\frac{X-\mu_1}{\sigma_1}<\frac{3-3}{4}\right)$$

$$=P\left(-\frac{6}{4}<Z<0\right)=P(-1.5<Z<0)=0.4332.$$

iv) Since $(X\,|\,Y=-4)\sim N(\mu_1^*,\sigma_1^*)$,

where $\mu_1^*=\mu_1+\rho\dfrac{\sigma_1}{\sigma_2}(y-\mu_2)=3\times\dfrac{3}{5}\times\dfrac{4}{5}(-4-1)=0.6$,

and

$$\sigma_1^{*2}=\sigma_1^2(1-\rho^2)=16\left(1-\frac{9}{25}\right)=\frac{16}{25}\times 16,\ \Rightarrow\ \sigma_1^*=\frac{16}{5}.$$

Therefore,

$$P(-3<X<3\,|\,Y=-4)=P\left(\frac{-3-\frac{3}{5}}{\frac{16}{5}}<Z<\frac{3-\frac{3}{5}}{\frac{16}{5}}\right)=P\left(-\frac{18}{16}<Z<\frac{12}{16}\right)=0.642.$$

Exercise: Given $(X,Y)\sim N_2(5,10,1,25,\rho)$, and $P(4<Y<16\,|\,X=5)=0.954$. Find ρ.

Solution: Since $(Y\,|\,X=5)\sim N(\mu_2^*,\sigma_2^{*2})$, where

$$\mu_2^*=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)=10+\rho\frac{5}{1}(5-5)=10, \text{ and } \sigma_2^{*2}=\sigma_2^2(1-\rho^2)=25(1-\rho^2).$$

Thus,

$$P(4<Y<16\,|\,X=5)=P\left(\frac{4-10}{5\sqrt{1-\rho^2}}<Z<\frac{16-10}{5\sqrt{1-\rho^2}}\right)=P\left(-\frac{6}{5\sqrt{1-\rho^2}}<Z<\frac{6}{5\sqrt{1-\rho^2}}\right)$$

$$=P\left(\frac{-1.2}{\sqrt{1-\rho^2}}<Z<\frac{1.2}{\sqrt{1-\rho^2}}\right)=0.954.$$

or $2P\left(0<Z<\dfrac{1.2}{\sqrt{1-\rho^2}}\right)=0.954$ or $P\left(0<Z<\dfrac{1.2}{\sqrt{1-\rho^2}}\right)=0.477$.

$\Rightarrow\ \dfrac{1.2}{\sqrt{1-\rho^2}}=2$, or $2\sqrt{1-\rho^2}=1.4$ or $4(1-\rho^2)=1.2$

or $1-\rho^2=\dfrac{1.44}{4}=0.36\ \Rightarrow\ \rho^2=1-0.36=0.64$ or $\rho=\pm 0.8$.

Exercise: Find $\underset{\sim}{\mu}$ mean vector and Σ variance covariance matrix or dispersion matrix of the following density functions:

i) $f(x,y)=\dfrac{1}{2\pi}e^{-\frac{1}{2}[(x-1)^2+(y-2)^2]}$,

ii) $f(x,y)=\dfrac{1}{2.4\pi}e^{-\frac{1}{0.72}\left[\frac{x^2}{4}-1.6\frac{xy}{2}+y^2\right]}$,

iii) $f(x,y)=\frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2+4x-6y+13)}$, and

iv) $f(x,y)=\frac{1}{2\pi}e^{-\frac{1}{2}(2x^2+y^2+2xy-22x-14y+65)}$.

Solution: If $(X,Y)\sim N_2(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, then

$$f(x,y)=\frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-Q/2},$$

where

$$Q=\frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right].$$

i) We have $f(x,y)=\frac{1}{2\pi}e^{-\frac{1}{2}[(x-1)^2+(y-2)^2]}$,

$\Rightarrow\ \mu_1=1,\ \mu_2=2,\ \sigma_1=1,\ \sigma_2=1$, and $\rho=0$.

i.e. $\underline{\mu}=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}=\begin{pmatrix}1\\ 2\end{pmatrix}$, and $\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$.

ii) We have

$$f(x,y)=\frac{1}{2.4\pi}\exp\left[-\frac{1}{0.72}\left(\frac{x^2}{4}-1.6\frac{xy}{2}+y^2\right)\right]$$

$$=\frac{1}{(2\pi)(2)(1)(0.6)}\exp\left[-\frac{1}{2\times 0.36}\left\{\left(\frac{x}{2}\right)^2-2\times 0.8\left(\frac{x}{2}\right)\left(\frac{y}{1}\right)+\left(\frac{y}{1}\right)^2\right\}\right]$$

$\Rightarrow\ 1-\rho^2=0.36$ or $\rho^2=0.64\ \Rightarrow\ \rho=\pm 0.8$.

Hence,

$\mu_1=0,\ \mu_2=0,\ \sigma_1=2,\ \sigma_2=1$, and $\rho=0.8$, i.e. $\underline{\mu}=\begin{pmatrix}0\\ 0\end{pmatrix}$, and

$$\Sigma=\begin{pmatrix}4 & (0.8)(2)(1)\\ (0.8)(2)(1) & 1\end{pmatrix}=\begin{pmatrix}4 & 1.6\\ 1.6 & 1\end{pmatrix}.$$

iii) We have

$$f(x,y)=\frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2+4x-6y+13)}=\frac{1}{2\pi}e^{-\frac{1}{2}[(x+2)^2+(y-3)^2]}.$$

Here,

$\mu_1=-2,\ \mu_2=3,\ \sigma_1=1,\ \sigma_2=1$, and $\rho=0$. $\underline{\mu}=\begin{pmatrix}-2\\ 3\end{pmatrix}$, and $\Sigma=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$.

Alternative Solution

Since $Q=x^2+y^2+4x-6y+13$, $\frac{\partial Q}{\partial x}=2x+4=0\ \Rightarrow\ x=-2$, i.e. $\mu_1=-2$.

$\frac{\partial Q}{\partial y}=2y-6=0\ \Rightarrow\ y=3$, i.e. $\mu_2=3$ or $\underline{\mu}=\begin{pmatrix}-2\\ 3\end{pmatrix}$, and

$$\Sigma^{-1}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}=\Sigma,\ \text{where } \Sigma^{-1}=\begin{Bmatrix}coeff.\ of\ x^2 & \frac{1}{2}coeff.\ of\ x\,y\\ \frac{1}{2}coeff.\ of\ x\,y & coeff.\ of\ y^2\end{Bmatrix}$$

iv) We have

$$f(x,y)=\frac{1}{2\pi}e^{-\frac{1}{2}(2x^2+y^2+2xy-22x-14y+65)},$$

where $Q=2x^2+y^2+2xy-22x-14y+65$.

$\frac{\partial Q}{\partial x}=4x+2y-22=0$, and $\frac{\partial Q}{\partial y}=2y+2x-14=0$.

Solving these two equations, we get

$x=4$, i.e. $\mu_1=4$, and $y=3$, i.e. $\mu_2=3$, or $\underline{\mu}=\begin{pmatrix}4\\ 3\end{pmatrix}$.

Now, $\Sigma^{-1}=\begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}$, then, $\Sigma=\begin{pmatrix}1 & -1\\ -1 & 2\end{pmatrix}$.

Therefore,

$\sigma_1^2=1,\ \sigma_2^2=2,\ \rho\sigma_1\sigma_2=-1\ \Rightarrow\ \rho=\frac{-1}{\sqrt{2}}$.

Exercise: Let $f(\underline{x})=C\,e^{-Q/2}$, where

i) $Q=3x^2+2y^2-2xy-32x+4y+92$, and

ii) $Q=2x_1^2+3x_2^2+4x_3^2+2x_1x_2-2x_1x_3-4x_2x_3-6x_1-2x_2+10x_3+8$.

Find $\underline{\mu}$ and Σ.

3) By the same analogy we have that the multivariate normal density function may be written in the form

$$f(\underline{x})=k\,e^{-\frac{1}{2}(\underline{x}-\underline{b})'A(\underline{x}-\underline{b})}.$$

where, $k>0$ is a constant to be determined such that the integral over the entire $p-$dimensional Euclidean space of $x_1,x_2,\cdots,x_p$ is unity.

We assume that the matrix A is positive definite (if the quadratic form $\underline{x}'A\underline{x}\ge 0$) and symmetric.

Since A is positive definite, we have $(\underline{x}-\underline{b})' A (\underline{x}-\underline{b}) \geq 0$, so that $f(\underline{x})$ is bounded above also clearly $f(\underline{x}) \geq 0$.

Since A is positive definite matrix, then by matrix algebra, there exist a nonsingular matrix C (a matrix whose determinant is not zero), such that

$$C' A C = I \tag{2.2}$$

where, I denotes the identity matrix and C' the transpose of C.

Make the nonsingular transformation

$$\underline{X} - \underline{b} = C \underline{Y} \tag{2.3}$$

The Jacobian of the transformation is $|J| = \dfrac{d(\underline{x}-\underline{b})}{d\underline{y}} = \dfrac{d\underline{x}}{d\underline{y}} = |C|$, where, $|C|$ indicates the absolute value of the determinant of C.

herefore, the density function of $\underline{y}$ will be

$$g(\underline{y}) = k\, e^{-\frac{1}{2}\underline{y}' C' A C \underline{y}} |C| = k\, e^{-\frac{1}{2}\underline{y}' I \underline{y}} |C|, \text{ since } C' A C = I.$$

$$= k|C| e^{-\frac{1}{2}\underline{y}'\underline{y}} \tag{2.4}$$

From equation (2.3) we know that the range of the $y's$ are also from $-\infty$ to ∞. Integrating over the entire range

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} k|C| e^{-\frac{1}{2}\underline{y}'\underline{y}}\, dy_1 \ldots dy_p = 1$$

or $\displaystyle \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} k|C| e^{-\frac{1}{2}(y_1^2+y_2^2+\ldots+y_p^2)}\, dy_1 \ldots dy_p = 1$

or $\displaystyle k|C| \left[\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}y_1^2} dy_1\right) \ldots \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}y_p^2} dy_p\right)\right] = 1$

or $\displaystyle k|C|(\sqrt{2\pi})\ldots(\sqrt{2\pi}) = 1$, as $\displaystyle \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = 1$ or $\displaystyle k = \frac{1}{|C|(2\pi)^{p/2}}$ (2.5)

Substituting from (2.5) in (2.4), we get

$$g(\underline{y}) = \frac{|C| e^{-\frac{1}{2}\underline{y}'\underline{y}}}{|C|(2\pi)^{p/2}} = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_1^2}\right) \cdots \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_p^2}\right).$$

Thus, $y_1, y_2, \ldots, y_p$ are independently normally distributed each with mean zero and variance unity.

Therefore,

$$E(\underline{x}-\underline{b}) = E(C\underline{y}) = C E(\underline{y}) = 0$$

and

$$\underline{b} = E(\underline{x}) = \underline{\mu}_x \text{ (the mean of } \underline{x}\text{, denoted by } \underline{\mu}) \tag{2.6}$$

Let Σ_y, variance covariance matrix of $\underline{y}$. Since $y_1, y_2, \ldots, y_p$ are independent $N(0,1)$, we have

$$\Sigma_y = E[\underline{y}-0][\underline{y}-0]' = E\,\underline{y}\,\underline{y}' = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I$$

Hence,

$$I = E\,\underline{y}\,\underline{y}' = E[C^{-1}(\underline{x}-\underline{b})(\underline{x}-\underline{b})' C^{-1'}], \text{from equation (2.3)}$$

$$= C^{-1}[E(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})'] C^{-1'} = C^{-1}\Sigma_x C^{-1'}, \text{from equation (2.6)}$$

Multiply by C on the left and by C' on the right, we get

$$C I C' = \Sigma_x \quad \text{or} \quad C C' = \Sigma_x \tag{2.7}$$

Also, we have

$$C' A C = I \text{ or } A = C'^{-1} I C^{-1} = C'^{-1} C^{-1}$$

$$\text{or } A^{-1} = C C', \text{ as } (AC)^{-1} = C^{-1} A^{-1} \tag{2.8}$$

From equation (2.7) and (2.8), we get, $A = \Sigma_x^{-1}$.

Also

$$\left|A^{-1}\right| = |CC'| = |C||C'| = |C|^2, \text{ as } |C| = |C'|, \text{ so that}$$

$$|C| = |\Sigma_x|^{1/2}.$$

Hence,

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma_x|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_x)' \Sigma_x^{-1} (\underline{x}-\underline{\mu}_x)\right].$$

Definition: A random vector $\underline{X} = (X_1, X_2, \ldots, X_p)'$ taking values $\underline{x} = (x_1, x_2, \ldots, x_p)'$ in E^n (Euclidean space of dimension p) is said to have a p-variate normal distribution if its probability density function can be written as

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma_x|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_x)' \Sigma_x^{-1} (\underline{x}-\underline{\mu}_x)\right],$$

where

$\underline{\mu} = (\mu_1, \cdots, \mu_p)'$, and Σ is a positive definite symmetric matrix of order p.

Properties of multivariate normal distribution

Theorem: If the variance covariance matrix of $p-$variate normal random vector $\underline{X}=(X_1, X_2, ..., X_p)'$ is diagonal matrix, then the components of $\underline{X}$ are independently normally distributed random variables.

Proof: The pdf of $p-$variate normal random vector is

$$f(\underline{x})=\frac{1}{(2\pi)^{p/2}|\Sigma_x|^{1/2}}\exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_x)'\,\Sigma_x^{-1}(\underline{x}-\underline{\mu}_x)\right].$$

Given

$$\Sigma=\begin{pmatrix}\sigma_1^2 & \cdots & 0\\ \vdots & \cdots & \vdots\\ 0 & \cdots & \sigma_p^2\end{pmatrix}, \text{ then the quadratic form } (\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}) \text{ will be}$$

$$[(x_1-\mu_1),\cdots,(x_p-\mu_p)]_{1\times p}\begin{pmatrix}1/\sigma_1^2 & \cdots & 0\\ \vdots & \cdots & \vdots\\ 0 & \cdots & 1/\sigma_p^2\end{pmatrix}_{p\times p}\begin{pmatrix}(x_1-\mu_1)\\ \vdots\\ (x_p-\mu_p)\end{pmatrix}_{p\times 1}=\sum_{i=1}^{p}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$

and

$|\Sigma|=\prod_{i=1}^{p}\sigma_i^2$ (a square matrix A is said to be diagonal, if $a_{ij}=0, i\neq j$, then $|A|=\prod_{i=1}^{p}a_{ii}$).

Thus,

$$|\Sigma|^{1/2}=\prod_{i=1}^{p}\sigma_i .$$

Hence,

$$f(\underline{x})=\frac{1}{(2\pi)^{p/2}\prod_{i=1}^{p}\sigma_i}\exp\left[-\frac{1}{2}\sum_{i=1}^{p}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right]=\prod_{i=1}^{p}\frac{1}{(2\pi)^{1/2}\sigma_i}\exp\left[-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right]$$

$$=f(x_1)\,f(x_2)\cdots f(x_p).$$

Therefore, $X_1, X_2,\cdots, X_p$ are independently normally distributed random variable with mean μ_i, and variance σ_i^2.

Theorem: If $\underline{X}$ (with p components) be distributed according to $N(\underline{\mu},\Sigma)$. Then $\underline{Y}=C\,\underline{X}$ (nonsingular transformation) is distributed according to $N(C\,\underline{\mu}, C\Sigma C')$ for C nonsingular.

Proof: We have

$$f(\underline{x})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})'\,\Sigma^{-1}(\underline{x}-\underline{\mu}\right]$$

Consider the transformation

$\underline{Y}=C\underline{X}$ or $\underline{X}=C^{-1}\underline{Y}$. The Jacobian of the transformation is $|C^{-1}|$, therefore,

$$g(\underline{y})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(C^{-1}\underline{y}-\underline{\mu})'\,\Sigma^{-1}(C^{-1}\underline{y}-\underline{\mu})\right]|C^{-1}|$$

The quadratic form in the exponent of $g(\underline{y})$ is

$$(C^{-1}\underline{y}-\underline{\mu})'\,\Sigma^{-1}(C^{-1}\underline{y}-\underline{\mu})=(C^{-1}\underline{y}-C^{-1}C\,\underline{\mu})'\Sigma^{-1}(C^{-1}\underline{y}-C^{-1}C\,\underline{\mu})$$

$$=[C^{-1}(\underline{y}-C\,\underline{\mu})]'\,\Sigma^{-1}[C^{-1}(\underline{y}-C\,\underline{\mu})]$$

$$=(\underline{y}-C\,\underline{\mu})'C^{-1'}\Sigma^{-1}C^{-1}(\underline{y}-C\,\underline{\mu}), \text{ since } (AB)'=B'A'$$

$$=(\underline{y}-C\,\underline{\mu})'C'^{-1}\Sigma^{-1}C^{-1}(\underline{y}-C\,\underline{\mu}), \text{ since } (C^{-1})'=(C')^{-1}$$

$$=(\underline{y}-C\,\underline{\mu})'(C\,\Sigma\,C')^{-1}(\underline{y}-C\,\underline{\mu}), \text{ since } (A\,B)^{-1}=B^{-1}A^{-1}.$$

And the Jacobian of the transformation, which is

$$|C^{-1}|=\frac{1}{|C|}=\sqrt{\frac{1}{|C|^2}}=\sqrt{\frac{1}{|C||C'|}}=\sqrt{\frac{|\Sigma|}{|C||\Sigma||C'|}}=\frac{|\Sigma|^{1/2}}{|C\Sigma C'|^{1/2}}, \text{ since } |AB|=|A||B|$$

Thus, the density function of $\underline{Y}$ is

$$g(\underline{y})=\frac{1}{(2\pi)^{p/2}|C\Sigma C'|^{1/2}}\exp\left[-\frac{1}{2}(\underline{y}-C\,\underline{\mu})'(C\,\Sigma\,C')^{-1}(\underline{y}-C\,\underline{\mu})\right]$$

Therefore,

$\underline{Y}\sim N(C\,\mu, C\,\Sigma\,C')$.

Exercise: Given $(X_1, X_2)\sim N_2(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$. Find the joint density function of $Y_1=X_1+X_2$, and $Y_2=X_1-X_2$.

Solution: The transformation is $\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}=\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}X_1\\ X_2\end{pmatrix}$, then, the matrix of transformation $C=\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$ is nonsingular, therefore,

$$\underline{Y}\sim N_2(C\,\mu, C\,\Sigma\,C'), \text{ where, } C\,\underline{\mu}=\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}=\begin{pmatrix}\mu_1+\mu_2\\ \mu_1-\mu_2\end{pmatrix}=\underline{\mu}_y$$

and

$$C\,\Sigma\,C' = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\,\sigma_2 \\ \rho\,\sigma_1\,\sigma_2 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 + \rho\,\sigma_1\,\sigma_2 & \rho\,\sigma_1\,\sigma_2 + \sigma_2^2 \\ \sigma_1^2 - \rho\sigma_1\,\sigma_2 & \rho\,\sigma_1\,\sigma_2 - \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 + 2\,\rho\,\sigma_1\,\sigma_2 + \sigma_2^2 & \sigma_1^2 - \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 - 2\,\rho\,\sigma_1\,\sigma_2 + \sigma_2^2 \end{pmatrix} = \Sigma_y .$$

Hence, $\underline{Y} \sim N_2\,(\mu_y, \Sigma_y)$.

Exercise: Given $\underline{X}$ be distributed according to $N_3\,(\underline{0}, I_3)$ and if $Y_1 = X_2 + X_3$, $Y_2 = 2X_3 - X_1$, and $Y_3 = X_2 - X_1$, then find the covariance matrix of $\underline{Y}$.

Solution: The transformation is

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 2 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}.$$ Here the matrix of transformation $C = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 2 \\ -1 & 1 & 0 \end{pmatrix}$ is nonsingular. Therefore,

$$\underline{Y} \sim N_3\,(\underline{0},\, C\,I\,C') \text{ or } \underline{Y} \sim N_3\,(\underline{0},\, CC'), \text{ where, } C\,C' = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Theorem: Let $\underline{X}' = (X_1, X_2, ..., X_p)$ have a joint normal distribution, a necessary and sufficient condition that a subset $\underline{X}^{(1)}$ of the components of $\underline{X}$ be independent of the subset $\underline{X}^{(2)}$ consisting of the remaining components of $\underline{X}$ is that the covariance between each component of $\underline{X}^{(1)}$ with a component of $\underline{X}^{(2)}$ is zero.

Proof: Let us partition

$$\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{pmatrix}, \text{ where, } \underline{X}^{(1)} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_q \end{pmatrix}, \text{ and } \underline{X}^{(2)} = \begin{pmatrix} X_{q+1} \\ X_{q+2} \\ \vdots \\ X_p \end{pmatrix},$$

the corresponding partition of $\underline{\mu}$ and Σ will be

$$\underline{\mu} = \begin{pmatrix} \underline{\mu}^{(1)} \\ \underline{\mu}^{(2)} \end{pmatrix}, \text{ and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \text{ where, } \Sigma_{11} = E\,(\underline{X}^{(1)} - \underline{\mu}^{(1)})\,(\underline{X}^{(1)} - \underline{\mu}^{(1)})'$$

$$\Sigma_{22} = E\,(\underline{X}^{(2)} - \underline{\mu}^{(2)})\,(\underline{X}^{(2)} - \underline{\mu}^{(2)})', \; \Sigma_{12} = E\,(\underline{X}^{(1)} - \underline{\mu}^{(1)})\,(\underline{X}^{(2)} - \underline{\mu}^{(2)})', \text{ and}$$

$$\Sigma_{21} = \Sigma_{12}'.$$

Necessary part: Let $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ be independent

$\Sigma_{12} = E\,(\underline{X}^{(1)} - \underline{\mu}^{(1)})\,(\underline{X}^{(2)} - \underline{\mu}^{(2)})' = 0$, since $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are independent.

Thus if $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ be independent, then the components are uncorrelated i.e. covariances are zero.

Sufficient part: From the necessary part, we have

$\Sigma = \begin{pmatrix} \Sigma_{11} & \underline{0} \\ \underline{0} & \Sigma_{22} \end{pmatrix}$, we have to prove that $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are independent.

The joint density function of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ is

$$f\,(\underline{x}) = \frac{1}{(2\pi)^{p/2}\,|\,\Sigma\,|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu})'\,\Sigma^{-1}\,(\underline{x} - \underline{\mu})\right]$$

$$= \frac{1}{(2\pi)^{q/2}\,(2\pi)^{(p-q)/2}\,|\,\Sigma\,|^{1/2}} \exp\left(-\frac{1}{2}Q\right),$$

where,

$$Q = (\underline{x} - \underline{\mu})'\,\Sigma^{-1}\,(\underline{x} - \underline{\mu}) = [(\underline{x}^{(1)} - \underline{\mu}^{(1)})',\, (\underline{x}^{(2)} - \underline{\mu}^{(2)})'] \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} \underline{x}^{(1)} - \underline{\mu}^{(1)} \\ \underline{x}^{(2)} - \underline{\mu}^{(2)} \end{pmatrix}$$

$$= [(\underline{x}^{(1)} - \underline{\mu}^{(1)})'\,\Sigma_{11}^{-1},\, (\underline{x}^{(2)} - \underline{\mu}^{(2)})'\,\Sigma_{22}^{-1}] \begin{pmatrix} \underline{x}^{(1)} - \underline{\mu}^{(1)} \\ \underline{x}^{(2)} - \underline{\mu}^{(2)} \end{pmatrix}$$

$$= (\underline{x}^{(1)} - \underline{\mu}^{(1)})'\,\Sigma_{11}^{-1}\,(\underline{x}^{(1)} - \underline{\mu}^{(1)}) + (\underline{x}^{(2)} - \underline{\mu}^{(2)})'\,\Sigma_{22}^{-1}\,(\underline{x}^{(2)} - \underline{\mu}^{(2)})$$

$$= Q_1 + Q_2 .$$

Also we note that

$$|\,\Sigma\,| = |\,\Sigma_{11}\,|\,|\,\Sigma_{22}\,|, \text{ because } \begin{vmatrix} A_{11} & 0 \\ 0 & A_{22} \end{vmatrix} = \begin{vmatrix} A_{11} & 0 \\ 0 & I \end{vmatrix} \times \begin{vmatrix} I & 0 \\ 0 & A_{22} \end{vmatrix}$$

The density of $\underline{X}$ can be written

$$f\,(\underline{x}) = \frac{1}{(2\pi)^{q/2}\,(2\pi)^{(p-q)/2}\,|\,\Sigma_{11}\,|^{1/2}\,|\,\Sigma_{22}\,|^{1/2}} \exp\left[-\frac{1}{2}(Q_1 + Q_2)\right]$$

$$= \frac{1}{(2\pi)^{q/2}\,|\,\Sigma_{11}\,|^{1/2}} \exp\left(-\frac{1}{2}Q_1\right) \times \frac{1}{(2\pi)^{(p-q)/2}\,|\,\Sigma_{22}\,|^{1/2}} \exp\left(-\frac{1}{2}Q_2\right)$$

$$= f\,(\underline{x}^{(1)})\; f\,(\underline{x}^{(2)}) .$$

Therefore, $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are independent.

Further, the marginal density of $\underline{X}^{(1)}$ is given by the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(\underline{x}^{(1)})\, f(\underline{x}^{(2)})\, dx_{q+1}\, dx_{q+2}\cdots dx_p$$

$$= f(\underline{x}^{(1)})\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(\underline{x}^{(2)})\, dx_{q+1}\, dx_{q+2}\cdots dx_p = f(\underline{x}^{(1)})$$

Hence, the marginal distribution of $\underline{X}^{(1)}$ is $N_q(\underline{\mu}^{(1)},\Sigma_{11})$, similarly the marginal distribution of $\underline{X}^{(2)}$ is $N_{p-q}(\underline{\mu}^{(2)},\Sigma_{22})$.

Theorem: If $\underline{X}$ is distributed according to $N(\underline{\mu},\Sigma)$, the marginal distribution of any set of components of $\underline{X}$ is multivariate normal with means, variances and covariances obtained by taking corresponding components of $\underline{\mu}$ and Σ, respectively.

Proof: We partition $\underline{X}$, $\underline{\mu}$ and Σ as

$$\underline{X}=\begin{pmatrix} X_1\\ X_2\\ \vdots\\ X_p\end{pmatrix}=\begin{pmatrix}\underline{X}^{(1)}\\ \underline{X}^{(2)}\end{pmatrix},\text{ where } \underline{X}^{(1)}=\begin{pmatrix} X_1\\ X_2\\ \vdots\\ X_q\end{pmatrix},\text{ and } \underline{X}^{(2)}=\begin{pmatrix} X_{q+1}\\ X_{q+2}\\ \vdots\\ X_p\end{pmatrix}$$

and the corresponding partition of $\underline{\mu}$ and Σ will be

$$\underline{\mu}=\begin{pmatrix}\underline{\mu}^{(1)}\\ \underline{\mu}^{(2)}\end{pmatrix},\text{ and }\quad \Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}.$$

We shall make a nonsingular linear transformation to sub vectors

$$\underline{Y}^{(1)}=\underline{X}^{(1)}+B\,\underline{X}^{(2)},\text{ and }\quad \underline{Y}^{(2)}=\underline{X}^{(2)}.$$

where B is the matrix chosen such that $\underline{Y}^{(1)}$ and $\underline{Y}^{(2)}$ are uncorrelated. i.e. B must satisfy the equation

$$E(\underline{Y}^{(1)}-E\underline{Y}^{(1)})(\underline{Y}^{(2)}-E\underline{Y}^{(2)})'=0$$

or $E(\underline{X}^{(1)}+B\underline{X}^{(2)}-\underline{\mu}^{(1)}-B\underline{\mu}^{(2)})(\underline{X}^{(2)}-\underline{\mu}^{(2)})'=0$

or $E(\underline{X}^{(1)}-\underline{\mu}^{(1)})(\underline{X}^{(2)}-\underline{\mu}^{(2)})'+B\,E(\underline{X}^{(2)}-\underline{\mu}^{(2)})(\underline{X}^{(2)}-\underline{\mu}^{(2)})'=0$

or $\Sigma_{12}+B\Sigma_{22}=0$.

Thus,

$$B=-\Sigma_{12}\Sigma_{22}^{-1},\text{ and }\quad \underline{Y}^{(1)}=\underline{X}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\underline{X}^{(2)}.$$

Therefore, the transformation is

$$\begin{pmatrix}\underline{Y}^{(1)}\\ \underline{Y}^{(2)}\end{pmatrix}=\begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1}\\ \underline{0} & I\end{pmatrix}\begin{pmatrix}\underline{X}^{(1)}\\ \underline{X}^{(2)}\end{pmatrix}$$

i.e. $\underline{Y}=C\underline{X}$.

Since the transformation is nonsingular, therefore, the distribution $\underline{Y}$ is also $p-$variate normal with

$$E\underline{Y}=E\begin{pmatrix}\underline{Y}^{(1)}\\ \underline{Y}^{(2)}\end{pmatrix}=E\begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1}\\ \underline{0} & I\end{pmatrix}\begin{pmatrix}\underline{X}^{(1)}\\ \underline{X}^{(2)}\end{pmatrix}=E\begin{pmatrix}\underline{X}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\underline{X}^{(2)}\\ \underline{X}^{(2)}\end{pmatrix}$$

$$=\begin{pmatrix}\underline{\mu}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}^{(2)}\\ \underline{\mu}^{(2)}\end{pmatrix}$$

and

$$\Sigma_{\underline{Y}}=\begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1}\\ \underline{0} & I\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\begin{pmatrix} I & \underline{0}\\ -\Sigma_{12}\Sigma_{22}^{-1} & I\end{pmatrix},\text{ since } \underline{Y}=C\underline{X}.$$

$$=\begin{pmatrix}\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{12}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\begin{pmatrix} I & \underline{0}\\ -\Sigma_{12}\Sigma_{22}^{-1} & I\end{pmatrix}$$

$$=\begin{pmatrix}\Sigma_{11.2} & \underline{0}\\ \Sigma_{21}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} & \Sigma_{22}\end{pmatrix}=\begin{pmatrix}\Sigma_{11.2} & \underline{0}\\ \underline{0} & \Sigma_{22}\end{pmatrix}.$$

The joint density function of $\underline{Y}^{(1)}$ and $\underline{Y}^{(2)}$ is

$$f(\underline{y})=\frac{1}{(2\pi)^{q/2}\,|\Sigma_{11.2}|^{1/2}}\exp[-\frac{1}{2}\{\underline{y}^{(1)}-(\underline{\mu}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}^{(2)})\}'$$

$$\Sigma_{11.2}^{-1}\{\underline{y}^{(1)}-(\underline{\mu}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}^{(2)})\}]$$

$$\times\frac{1}{(2\pi)^{(p-q)/2}\,|\Sigma_{22}|^{1/2}}\exp[-\frac{1}{2}(\underline{y}^{(2)}-\underline{\mu}^{(2)})'\Sigma_{22}^{-1}(\underline{y}^{(2)}-\underline{\mu}^{(2)})]$$

Since $\underline{Y}^{(2)}=\underline{X}^{(2)}$ and $\underline{Y}^{(1)}$ are independently distributed, so

$f(\underline{y})=f(\underline{y}^{(1)})\,f(\underline{x}^{(2)})$. Therefore, the marginal density function of $\underline{X}^{(2)}=\underline{Y}^{(2)}$ is given by the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(\underline{y}^{(1)})\,f(\underline{x}^{(2)})\,d\underline{y}^{(1)}=f(\underline{x}^{(2)})\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} f(\underline{y}^{(1)})\,d\underline{y}^{(1)}=f(\underline{x}^{(2)})$$

Hence, the marginal distribution of $\underline{X}^{(2)}$ is $N_{p-q}(\underline{\mu}^{(2)},\ \Sigma_{22})$.

Conditional distribution

The joint density function of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ can be obtained by substituting $\underline{x}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\underline{x}^{(2)}$ for $\underline{y}^{(1)}$, $\underline{x}^{(2)}$ for $\underline{y}^{(2)}$ and multiplying by the Jacobian of the transformation. The Jacobian of the transformation is $|J|=1$, since $\underline{y}=C\underline{x}$ or $\underline{x}=C^{-1}\underline{y}$ and $\dfrac{d\underline{x}}{d\underline{y}}=\left|C^{-1}\right|=\dfrac{1}{|C|}$, where $|C|=\begin{vmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1}\\ \underline{0} & I\end{vmatrix}=1$.

The resulting density of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ is

$$f(\underline{x}^{(1)},\underline{x}^{(2)})=\frac{1}{(2\pi)^{q/2}|\Sigma_{11.2}|^{1/2}}\exp[-\frac{1}{2}(\underline{x}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\underline{x}^{(2)}-\underline{\mu}^{(1)}+\Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}^{(2)})'$$

$$\Sigma_{11.2}^{-1}(\underline{x}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\underline{x}^{(2)}-\underline{\mu}^{(1)}+\Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}^{(2)})]$$

$$\times\frac{1}{(2\pi)^{(p-q)/2}|\Sigma_{22}|^{1/2}}\exp[-\frac{1}{2}(\underline{x}^{(2)}-\underline{\mu}^{(2)})'\Sigma_{22}^{-1}(\underline{x}^{(2)}-\underline{\mu}^{(2)})]$$

$$=\frac{1}{(2\pi)^{q/2}|\Sigma_{11.2}|^{1/2}}\exp[-\frac{1}{2}\{\underline{x}^{(1)}-\underline{\mu}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)}-\underline{\mu}^{(2)})\}'$$

$$\Sigma_{11.2}^{-1}\{\underline{x}^{(1)}-\underline{\mu}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)}-\underline{\mu}^{(2)})\}]$$

$$\times\frac{1}{(2\pi)^{(p-q)/2}|\Sigma_{22}|^{1/2}}\exp[-\frac{1}{2}(\underline{x}^{(2)}-\underline{\mu}^{(2)})'\Sigma_{22}^{-1}(\underline{x}^{(2)}-\underline{\mu}^{(2)})].$$

Now

$$\frac{f(\underline{x}^{(1)},\underline{x}^{(2)})}{f(\underline{x}^{(2)})}=\frac{1}{(2\pi)^{q/2}|\Sigma_{11.2}|^{1/2}}\exp[-\frac{1}{2}\{\underline{x}^{(1)}-\underline{\mu}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)}-\underline{\mu}^{(2)})\}'$$

$$\Sigma_{11.2}^{-1}\{\underline{x}^{(1)}-\underline{\mu}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)}-\underline{\mu}^{(2)})\}]$$

$$=f(\underline{x}^{(1)}\mid\underline{x}^{(2)}).$$

Therefore, the density function $f(\underline{x}^{(1)}\mid\underline{x}^{(2)})$ is a $q-$variate normal with mean $\underline{\mu}^{(1)}+\Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)}-\underline{\mu}^{(2)})=\underline{\mu}^{(1)*}$, and covariance matrix $\Sigma_{11.2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, i.e. $(\underline{X}^{(1)}\mid\underline{X}^{(2)}=\underline{x}^{(2)})\sim N_q(\underline{\mu}^{(1)*},\Sigma_{11.2})$.

Note: The mean vector in the distribution of $(\underline{X}^{(1)}\mid\underline{X}^{(2)}=\underline{x}^{(2)})$ is linear function of $\underline{X}^{(2)}$ while the covariance matrix is independent of $\underline{X}^{(2)}$.

Exercise: In a similar fashion we can prove the following results

a) $\underline{X}^{(1)}$, $\underline{X}^{(2)}-\Sigma_{21}\Sigma_{11}^{-1}\underline{X}^{(1)}$ are independently normally distributed.

Hint: non singular linear transformation $\underline{Y}^{(1)}=\underline{X}^{(1)}$, $\underline{Y}^{(2)}=\underline{X}^{(2)}+B\,\underline{X}^{(1)}$

b) The marginal distribution of $\underline{X}^{(1)}\sim N_q(\underline{\mu}^{(1)},\ \Sigma_{11})$.

c) The conditional distribution of $\underline{X}^{(2)}$ given $\underline{X}^{(1)}=\underline{x}^{(1)}$ is normal with mean $\underline{\mu}^{(2)}+\Sigma_{21}\Sigma_{11}^{-1}(\underline{x}^{(1)}-\underline{\mu}^{(1)})=\underline{\mu}^{(2)*}$ and covariance matrix $\Sigma_{22.1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

Exercise: If $\underline{X}\sim N_3(0,\Sigma)$, where $\Sigma=\begin{pmatrix}1.0 & 0.8 & -0.4\\ 0.8 & 1.0 & -0.56\\ -0.4 & -0.56 & 1.0\end{pmatrix}$. Find the conditional distribution of i) X_1,X_2 given X_3, ii) X_1,X_3 given X_2, and iii) X_2,X_3 given X_1.

Solution:

i) We partition $\underline{X}$ for X_1,X_2 given X_3 as

$$\underline{X}=\begin{pmatrix}X_1\\ X_2\\ \cdots\\ X_3\end{pmatrix}=\begin{pmatrix}\underline{X}^{(1)}\\ X^{(2)}\end{pmatrix},\text{ where }\underline{X}^{(1)}=\begin{pmatrix}X_1\\ X_2\end{pmatrix},\text{ and }X^{(2)}=X_3$$

$$\Sigma=\begin{pmatrix}1.0 & 0.8 & \vdots & -0.4\\ 0.8 & 1.0 & \vdots & -0.56\\ \cdots & \cdots & \cdots & \cdots\\ -0.4 & -0.56 & \vdots & 1.0\end{pmatrix}=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}.$$

The conditional distribution of X_1,X_2 given X_3 is $N_2(\underline{\mu}^{(1)*},\Sigma_{11.2})$, where

$$\underline{\mu}^{(1)*}=\underline{\mu}^{(1)}+\Sigma_{12}\Sigma_{22}^{-1}(x^{(2)}-\mu^{(2)})=0+\begin{pmatrix}-0.4\\ -0.56\end{pmatrix}(1)(x_3-0)=\begin{pmatrix}-0.4\,x_3\\ -0.56\,x_3\end{pmatrix},\text{ and}$$

$$\Sigma_{11.2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}=\begin{pmatrix}1.0 & 0.8\\ 0.8 & 1.0\end{pmatrix}-\begin{pmatrix}-0.4\\ -0.56\end{pmatrix}(1)(-0.4\quad -0.56)$$

$$=\begin{pmatrix}0.8400 & 0.5760\\ 0.5760 & 0.6864\end{pmatrix}.$$

ii) For X_1,X_3 given X_2, we partition $\underline{X}$ as

$$\underline{X}=\begin{pmatrix}X_1\\ X_3\\ \cdots\\ X_2\end{pmatrix}=\begin{pmatrix}\underline{X}^{(1)}\\ X^{(2)}\end{pmatrix},\text{ where }\underline{X}^{(1)}=\begin{pmatrix}X_1\\ X_3\end{pmatrix},\text{ and }X^{(2)}=X_2$$

$$\Sigma=\begin{pmatrix}1.0 & -0.4 & \vdots & 0.8\\ -0.4 & 1.0 & \vdots & -0.56\\ \cdots & \cdots & \cdots & \cdots\\ 0.8 & -0.56 & \vdots & 1.0\end{pmatrix}=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}.$$

The conditional distribution of X_1,X_3 given X_2 is $N_2(\underline{\mu}^{(1)*},\Sigma_{11.2})$,

where

$$\underline{\mu}^{(1)*}=\underline{\mu}^{(1)}+\Sigma_{12}\Sigma_{22}^{-1}(x^{(2)}-\mu^{(2)})=0+\begin{pmatrix}0.8\\ -0.56\end{pmatrix}(1)(x_2-0)=\begin{pmatrix}0.8\,x_2\\ -0.56\,x_2\end{pmatrix}$$

and

$$\Sigma_{11.2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}=\begin{pmatrix}1.0 & -0.4\\ -0.4 & 1.0\end{pmatrix}-\begin{pmatrix}0.8\\ -0.56\end{pmatrix}(1)(0.8 \quad -0.56)$$

$$=\begin{pmatrix}1.0 & -0.4\\ -0.4 & 1.0\end{pmatrix}-\begin{pmatrix}0.6400 & -0.4480\\ -0.4480 & 0.3136\end{pmatrix}=\begin{pmatrix}0.36 & 0.048\\ 0.048 & 0.6864\end{pmatrix}.$$

iii) For X_2, X_3 given X_1 we partition $\underline{X}$ as

$$\underline{X}=\begin{pmatrix}X_1\\ \cdots\\ X_2\\ X_3\end{pmatrix}=\begin{pmatrix}X^{(1)}\\ \underline{X}^{(2)}\end{pmatrix}, \text{ where } \underline{X}^{(2)}=\begin{pmatrix}X_2\\ X_3\end{pmatrix}, \text{ and } X^{(1)}=X_1$$

$$\Sigma=\begin{pmatrix}1.0 & \vdots & 0.8 & -0.4\\ \cdots & \cdots & \cdots & \cdots\\ 0.8 & \vdots & 1.0 & -0.56\\ -0.4 & \vdots & -0.56 & 1.0\end{pmatrix}=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}.$$

The conditional distribution of X_2, X_3 given X_1 is $N_2(\underline{\mu}^{(2)*}, \Sigma_{22.1})$, where

$$\underline{\mu}^{(2)*}=\underline{\mu}^{(2)}+\Sigma_{21}\Sigma_{11}^{-1}(x^{(1)}-\mu^{(1)})=0+\begin{pmatrix}0.8\\ -0.4\end{pmatrix}(1)(x_1-0)=\begin{pmatrix}0.8\,x_1\\ -0.4\,x_1\end{pmatrix}, \text{ and}$$

$$\Sigma_{22.1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}=\begin{pmatrix}1.0 & -0.56\\ -0.56 & 1.0\end{pmatrix}-\begin{pmatrix}0.8\\ -0.4\end{pmatrix}(1)(0.8 \quad 0.4)$$

$$=\begin{pmatrix}0.36 & -0.24\\ -0.24 & 0.84\end{pmatrix}.$$

Exercise: Let X_1 and X_2 are jointly distributed as normal with $E(X_i)=0$ and Var.$(X_i)=1$, for $i=1,2$. If the distribution of X_2 given X_1 is normal $N_1(0, 1-\rho^2)$ with $|\rho|<1$, find the covariance matrix of X_1 and X_2.

Solution: Given $\underline{X}=\begin{pmatrix}X_1\\ X_2\end{pmatrix}$, $\quad \underline{\mu}=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$, and $\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}=\begin{pmatrix}1 & \Sigma_{12}\\ \Sigma_{21} & 1\end{pmatrix}$

The conditional distribution of X_2 given X_1 is $N_1(\mu_2^*, \Sigma_{22.1})$

$$\Sigma_{22.1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}=1-\rho^2 \text{ or } 1-\Sigma_{21}\Sigma_{12}=1-\rho^2$$

or $\Sigma_{21}\Sigma_{12}=\rho^2$ or $\Sigma_{21}=\Sigma_{12}=\rho$.

Therefore, the variance covariance matrix of X_1 and X_2 will be $\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$.

Theorem: If $\underline{X}$ is distributed according to $N_p(\underline{\mu}, \Sigma)$ and $\underline{Y}=D\,\underline{X}$, where D is $q\times p$ $(q<p)$ matrix of rank q, then $\underline{Y}$ is distributed according to $N_q(D\,\underline{\mu}, D\Sigma D')$.

Note: This result has been proved when D is $p\times p$ and nonsingular.

Proof: The transformation is $\underline{Y}=D\,\underline{X}$, where $\underline{Y}$ has q components and D is $q\times p$ real matrix. The expected value of $\underline{Y}$ is

$E(\underline{Y})=D\,E(\underline{X})=D\,\underline{\mu}$, and the covariance matrix is

$$\Sigma_y=E[\underline{Y}-E(\underline{Y})]\,[\underline{Y}-E(\underline{Y})]'=E[D\,\underline{X}-D\,\underline{\mu}]\,[D\,\underline{X}-D\,\underline{\mu}]'=D\Sigma D'.$$

Since $R(D)=q$ the q rows of D are independent. We know that a set of q independent vectors can be extended to form a basis of the $p-$dimensional vector space by adding to it $p-q$ vectors.

Let $C=\begin{pmatrix}D_{q\times p}\\ E_{(p-q)\times p}\end{pmatrix}$, then C is nonsingular. Make the transformation

$\underline{Z}=C\,\underline{X}$. Since C nonsingular, therefore, $\quad \underline{Z} \sim N_p(C\,\underline{\mu}, C\Sigma C')$, i.e.

$$\underline{Z}=\begin{pmatrix}D\\ E\end{pmatrix}\underline{X}=\begin{pmatrix}D\underline{X}\\ E\underline{X}\end{pmatrix}.$$

But, $D\,\underline{X}$ being the partition vector of $\underline{Z}$ has a marginal $q-$variate normal distribution, therefore,

$$\underline{Y}=D\,\underline{X} \sim N_q(D\,\underline{\mu}, D\Sigma D').$$

Note: This theorem tells us if $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, then every linear combination of the components of $\underline{X}$ has a univariate normal distribution.

Proof: Let $Y=D_{1\times p}\,\underline{X}=(l_1, l_2, \ldots, l_p)\begin{pmatrix}X_1\\ X_2\\ \vdots\\ X_p\end{pmatrix}=\underline{l}'\,\underline{X}$, then $E(Y)=\underline{l}'\,\underline{\mu}$, and $\Sigma_y=\underline{l}'\,\Sigma\,\underline{l}$,

therefore,

$$Y=D_{1\times p}\,\underline{X}=\underline{l}'\,\underline{X} \sim N_1(\underline{l}'\,\underline{\mu}, \underline{l}'\,\Sigma\,\underline{l}).$$

Exercise: Let $\underline{X} \sim N(\underline{\mu}, \Sigma)$, where $\underline{X}'=(X_1, X_2, X_3)$, $\underline{\mu}'=(0,0,0)$, and

$\Sigma=\begin{pmatrix}7 & 3 & 2\\ 3 & 4 & 1\\ 2 & 1 & 2\end{pmatrix}$. Obtain the distribution of $X_1+2X_2-3X_3$.

Solution: We know that

$$Y=D_{1\times p}\,\underline{X}=(l_1, l_2, \ldots, l_p)\begin{pmatrix}X_1\\ X_2\\ \vdots\\ X_p\end{pmatrix}=\underline{l}'\,\underline{X}, \text{ then } Y \sim N_1(\underline{l}'\,\underline{\mu}, \underline{l}'\,\Sigma\,\underline{l}).$$

Therefore, in our case

$$Y = D_{1\times 3}\underline{X} = (1,\ 2,\ -3)\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \text{ then } \quad \underline{l}'\underline{\mu} = (1,\ 2,\ -3)\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0$$

and

$$\underline{l}'\Sigma\underline{l} = (1,\ 2,\ -3)\begin{pmatrix} 7 & 3 & 2 \\ 3 & 4 & 1 \\ 2 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} = (7,\ 8,\ -1)\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} = 29.$$

Hence $Y \sim N(0, 29)$.

Exercise: Let $\underline{X}_\alpha \sim N_p(\underline{\mu}_\alpha, \Sigma_\alpha)$, $\alpha = 1, 2, \cdots, n$. Let $\underline{Y} = a_1\underline{X}_1 + \cdots + a_n\underline{X}_n$, then show that $\underline{Y} \sim N_p\left(\sum_\alpha a_\alpha \underline{\mu}_\alpha,\ \sum_\alpha a_\alpha^2 \Sigma_\alpha\right)$, where a_α's are scalar.

Proof: Consider a linear combination of $\underline{Y}$

$$\underline{l}'\underline{Y} = a_1\underline{l}'\underline{X}_1 + \cdots + a_n\underline{l}'\underline{X}_n, \text{ then } E(\underline{l}'\underline{Y}) = a_1\underline{l}'\underline{\mu}_1 + \cdots + a_n\underline{l}'\underline{\mu}_n = \underline{l}'\left(\sum_\alpha a_\alpha \underline{\mu}_\alpha\right), \text{ and}$$

$$\Sigma_{\underline{l}'\underline{Y}} = Cov(a_1\underline{l}'\underline{X}_1 + \cdots + a_n\underline{l}'\underline{X}_n) = a_1^2\underline{l}'\Sigma_1\underline{l} + \cdots + a_n^2\underline{l}'\Sigma_n\underline{l} = \underline{l}'\left(\sum_\alpha a_\alpha^2\Sigma_\alpha\right)\underline{l}.$$

Thus,

$$\underline{l}'\underline{Y} \sim N_1\left(\underline{l}'\sum_\alpha a_\alpha\underline{\mu}_\alpha,\ \underline{l}'\sum_\alpha a_\alpha^2\Sigma_\alpha\underline{l}\right) \quad \Rightarrow \underline{Y} \sim N_p\left(\sum_\alpha a_\alpha\underline{\mu}_\alpha,\ \sum_\alpha a_\alpha^2\Sigma_\alpha\right).$$

Exercise: Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, let $Y_1 = \underline{a}'\underline{X}$ and $Y_2 = \underline{b}'\underline{X}$, where $\underline{a}$ and $\underline{b}$ are given vectors of scalars. Obtain the joint distribution of Y_1 and Y_2, and derive the condition for the independence of Y_1 and Y_2.

Solution: The transformation is $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \underline{a}' \\ \underline{b}' \end{pmatrix}\underline{X}$ i.e. $\underline{Y} = D\underline{X}$, where $\underline{Y}$ has two components and D is $2\times p$ real matrix. The expected value of $\underline{Y}$ is

$E(\underline{Y}) = D\ E(\underline{X}) = D\underline{\mu}$, and the covariance matrix is

$$\Sigma_y = E[\underline{Y} - E(\underline{Y})][\underline{Y} - E(\underline{Y})]' = E[D\underline{X} - D\underline{\mu}][D\underline{X} - D\underline{\mu}]' = D\Sigma D'$$

We know that, if

$\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, and $\underline{Y} = D\underline{X}$, where, D is a matrix of rank q $(q < p)$, then

$\underline{Y} \sim N_q(D\underline{\mu}, D\Sigma D')$.

Therefore, in our case $\underline{Y} \sim N_2(D\underline{\mu}, D\Sigma D')$,

where,

$$D\underline{\mu} = \begin{pmatrix} \underline{a}' \\ \underline{b}' \end{pmatrix}\underline{\mu} = \begin{pmatrix} \underline{a}'\underline{\mu} \\ \underline{b}'\underline{\mu} \end{pmatrix}, \text{ and } \quad D\Sigma D' = \begin{pmatrix} \underline{a}' \\ \underline{b}' \end{pmatrix}\Sigma(\underline{a},\ \underline{b}) = \begin{pmatrix} \underline{a}'\Sigma \\ \underline{b}'\Sigma \end{pmatrix}(\underline{a},\ \underline{b}) = \begin{pmatrix} \underline{a}'\Sigma\underline{a} & \underline{a}'\Sigma\underline{b} \\ \underline{b}'\Sigma\underline{a} & \underline{b}'\Sigma\underline{b} \end{pmatrix}.$$

The off-diagonal term $\underline{a}'\Sigma\underline{b}$ is the covariance matrix for Y_1 and Y_2. Consequently when $\underline{a}'\underline{b} = 0$, so that $\underline{a}'\Sigma\underline{b} = 0$, Y_1 and Y_2 are independent.

Exercise: Let $\underline{X}$ be $N_3(\underline{\mu}, \Sigma)$ with $\underline{\mu}' = (-3, 1, 4)$ and $\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

Which of the following random variables are independent? Explain

i) X_1 and X_2

ii) X_2 and X_3

iii) (X_1, X_2) and X_3

iv) $\dfrac{X_1 + X_2}{2}$ and X_3

v) X_2 and $X_2 - \dfrac{5}{2}X_1 - X_3$

vi) Conditional distribution of X_2 given $X_1 = x_1$ and $X_3 = x_3$.

Solution:

i) Since X_1 and X_2 have covariance $\sigma_{12} = -2$, they are not independent.

ii) Since X_2 and X_3 have covariance $\sigma_{23} = 0$, they are independent.

iii) Partitioning $\underline{X}$ and Σ as

$$\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \cdots \\ X_3 \end{pmatrix} = \begin{pmatrix} \underline{X}^{(1)} \\ X^{(2)} \end{pmatrix}, \text{ and } \Sigma = \begin{pmatrix} 1 & -2 & \vdots & 0 \\ -2 & 5 & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & 2 \end{pmatrix}.$$

We see that $\underline{X}^{(1)} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and X_3 have covariance matrix $\Sigma_{12} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, therefore, (X_1, X_2) and X_3 are independent.

iv) The transformation is

$$\underline{Y} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} \dfrac{X_1 + X_2}{2} \\ X_3 \end{pmatrix}, \text{ i.e. } \underline{Y} = D\underline{X}.$$

Therefore,

$$\underline{Y} \sim N_2(D\underline{\mu},\, D\Sigma D'), \text{ where } D\underline{\mu} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} -3 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \end{pmatrix} \text{ and}$$

$$D\Sigma D' = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} 1/2 & 0 \\ 1/2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}.$$

We see that $\dfrac{X_1 + X_2}{2}$ and X_3 have covariance $\sigma_{12} = 0$, they are independent.

v) The transformation is

$$\underline{Y} = \begin{pmatrix} 0 & 1 & 0 \\ -5/2 & 1 & -1 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} X_2 \\ X_2 - (5/2)X_1 - X_3 \end{pmatrix}, \text{ i.e. } \underline{Y} = D\,\underline{X}$$

Therefore,

$$\underline{Y} \sim N_2(D\underline{\mu},\, D\Sigma D'), \text{ where } D\underline{\mu} = \begin{pmatrix} 0 & 1 & 0 \\ -5/2 & 1 & -1 \end{pmatrix}\begin{pmatrix} -3 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 9/2 \end{pmatrix}$$

and

$$D\Sigma D' = \begin{pmatrix} 0 & 1 & 0 \\ -5/2 & 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} 0 & -5/2 \\ 1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 5 & 10 \\ 10 & 93/4 \end{pmatrix}.$$

We see that X_2 and $X_2 - \dfrac{5}{2}X_1 - X_3$ have covariance $\sigma_{12} = 10$, they are not independent.

Exercise: Let $\underline{X}$ be $N_3(\underline{\mu}, \Sigma)$ with $\underline{\mu}' = (2, -3, 1)$ and $\Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}$

i) Obtain the distribution of $3X_1 - 2X_2 + X_3$.

ii) Find a vector $\underline{l}$ such that X_2 and $X_2 - \underline{l}'\begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$ are independent.

iii) Conditional distribution of X_3 given $X_1 = x_1$ and $X_2 = x_2$.

Solution: Consider the transformation

$$\underline{Y} = \begin{pmatrix} 0 & 1 & 0 \\ -l_1 & 1 & -l_2 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} X_2 \\ X_2 - \underline{l}'\,\underline{X}^{(2)} \end{pmatrix}, \text{ where } \underline{X}^{(2)} = \begin{pmatrix} X_1 \\ X_3 \end{pmatrix},$$

i.e. $\underline{Y}_{2\times 1} = D_{2\times 3}\,\underline{X}_{3\times 1}$.

The variance covariance of this transformation is

$$D\Sigma D' = \begin{pmatrix} 0 & 1 & 0 \\ -l_1 & 1 & -l_2 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}\begin{pmatrix} 0 & -l_1 \\ 1 & 1 \\ 0 & -l_2 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & -l_1 + 3 - 2l_2 \\ -l_1 + 3 - 2l_2 & l_1^2 + 2l_1l_2 - 2l_1 - 4l_2 + 3 + 2l_2^2 \end{pmatrix}.$$

The off-diagonal term $-l_1 + 3 - 2l_2$ is the covariance for X_2 and $X_2 - \underline{l}'\begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$.

For independent $-l_1 + 3 - 2l_2 = 0$, so that $\underline{l} = \begin{pmatrix} l_1 \\ -\dfrac{l_1 + 3}{2} \end{pmatrix}$.

Exercise: X_1 and X_2 are two $rv's$ with respective expectations μ_1 and μ_2 and variances σ_{11} and σ_{22} and correlation ρ such that

$$X_1 = \mu_1 + \sqrt{\sigma_{11}}\,Y_1, \text{ and } X_2 = \mu_2 + \sqrt{\sigma_{22}(1-\rho^2)}\,Y_1 + \sqrt{\sigma_{22}}\,Y_2,$$

where Y_1 and Y_2 are two independent normal $N(0,1)$ variates. Find the joint distribution of X_1 and X_2.

Solution: The transformation is

$$\begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix} = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 \\ \sqrt{\sigma_{22}(1-\rho^2)} & \sqrt{\sigma_{22}} \end{pmatrix}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \text{ i.e. } \underline{X} - \underline{\mu} = C\underline{Y} \text{ or } \underline{X} = C\underline{Y} + \underline{\mu}, \text{ so that}$$

$$E\underline{X} = C\,E(Y) + \underline{\mu} = \underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

and

$$\Sigma_{\underline{X}} = E[\underline{X} - \underline{\mu}][\underline{X} - \underline{\mu}]' = E[C\underline{Y} + \underline{\mu} - \underline{\mu}][C\underline{Y} + \underline{\mu} - \underline{\mu}]' = E[C\underline{Y}\,\underline{Y}'C'] = CC'$$

$$= \begin{pmatrix} \sqrt{\sigma_{11}} & 0 \\ \sqrt{\sigma_{22}(1-\rho^2)} & \sqrt{\sigma_{22}} \end{pmatrix}\begin{pmatrix} \sqrt{\sigma_{11}} & \sqrt{\sigma_{22}(1-\rho^2)} \\ 0 & \sqrt{\sigma_{22}} \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{11} & \sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)} \\ \sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)} & \sigma_{22}(1-\rho^2) + \sigma_{22} \end{pmatrix}.$$

Exercise: Let $A = \Sigma^{-1}$ and Σ be partitioned similarly as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \text{ by solving } A\Sigma = I \text{ according to partitioning prove}$$

i) $\Sigma_{12}\,\Sigma_{22}^{-1} = -A_{11}^{-1}\,A_{12}$

ii) $\Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21} = A_{11}^{-1}$.

Solution: Consider the matrix equation $A\Sigma = I$

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} I & \underset{\sim}{0} \\ \underset{\sim}{0} & I \end{pmatrix}$$

or $\begin{pmatrix} A_{11}\Sigma_{11} + A_{12}\Sigma_{21} & A_{11}\Sigma_{12} + A_{12}\Sigma_{22} \\ A_{21}\Sigma_{11} + A_{22}\Sigma_{21} & A_{21}\Sigma_{12} + A_{22}\Sigma_{22} \end{pmatrix} = \begin{pmatrix} I & \underset{\sim}{0} \\ \underset{\sim}{0} & I \end{pmatrix}$.

Consider

$A_{11}\Sigma_{12} + A_{12}\Sigma_{22} = \underset{\sim}{0}$, then $A_{11}\Sigma_{12} = -A_{12}\Sigma_{22}$

or $\Sigma_{12} = -A_{11}^{-1} A_{12}\Sigma_{22}$ or $\Sigma_{12}\Sigma_{22}^{-1} = -A_{11}^{-1} A_{12}$.

Also

$A_{11}\Sigma_{11} + A_{12}\Sigma_{21} = I$, then

$A_{11}\Sigma_{11} = I - A_{12}\Sigma_{21}$ or $\Sigma_{11} = A_{11}^{-1} - A_{11}^{-1} A_{12}\Sigma_{21}$

or $A_{11}^{-1} = \Sigma_{11} + A_{11}^{-1} A_{12}\Sigma_{21}$

or $A_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Exercise: Prove explicitly that Σ is positive definite, then

$$|\Sigma| = |\Sigma_{22}|\,|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}| = |\Sigma_{22}|\,|\Sigma_{11.2}|.$$

Solution: Let Σ be partitioned as

$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, and define a matrix $C_{p\times p}$ as $C = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ \underset{\sim}{0} & I \end{pmatrix}$.

Obviously C is a non-singular matrix because determinant of C is 1.

Now consider the matrix

$$C\Sigma C' = \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \underset{\sim}{0} \\ \underset{\sim}{0} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11.2} & \underset{\sim}{0} \\ \underset{\sim}{0} & \Sigma_{22} \end{pmatrix}.$$

Therefore,

$|C\Sigma C'| = |\Sigma_{11.2}|\ |\Sigma_{22}| \Rightarrow |C|\,|\Sigma|\,|C'| = |\Sigma_{11.2}|\ |\Sigma_{22}|$, because $|C| = 1$.

Similarly, we can prove

$$|\Sigma| = |\Sigma_{11}|\,|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}| = |\Sigma_{11}|\,|\Sigma_{22.1}|.$$

Exercise: Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, let A be a symmetric matrix of order p, show that

i) $E(\underset{\sim}{X}\,\underset{\sim}{X}') = \Sigma + \underset{\sim}{\mu}\,\underset{\sim}{\mu}'$.

ii) $E(\underset{\sim}{X}' A\,\underset{\sim}{X}) = tr\, A\Sigma + \underset{\sim}{\mu}' A\,\underset{\sim}{\mu}$.

Solution:

i) We know that

$$\Sigma = E(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})' = E(\underset{\sim}{X}\,\underset{\sim}{X}' - \underset{\sim}{X}\,\underset{\sim}{\mu}' - \underset{\sim}{\mu}\,\underset{\sim}{X}' + \underset{\sim}{\mu}\,\underset{\sim}{\mu}')$$

$$= E(\underset{\sim}{X}\,\underset{\sim}{X}') - \underset{\sim}{\mu}\,\underset{\sim}{\mu}' - \underset{\sim}{\mu}\,\underset{\sim}{\mu}' + \underset{\sim}{\mu}\,\underset{\sim}{\mu}'$$

$\Rightarrow\ E(\underset{\sim}{X}\,\underset{\sim}{X}') = \Sigma + \underset{\sim}{\mu}\,\underset{\sim}{\mu}'$.

ii) $E(\underset{\sim}{X}' A\,\underset{\sim}{X}) = E[tr\,\underset{\sim}{X}' A\,\underset{\sim}{X}]$, since $tr\, a = a$, a is a scalar.

$= E[tr\, A\,(\underset{\sim}{X}\,\underset{\sim}{X}')]$, since $tr\,(CD) = tr\,(DC)$

$= tr\, A\, E(\underset{\sim}{X}\,\underset{\sim}{X}') = tr\, A(\Sigma + \underset{\sim}{\mu}\,\underset{\sim}{\mu}')$

$= tr\, A\Sigma + tr\, A\,\underset{\sim}{\mu}\,\underset{\sim}{\mu}' = tr\, A\Sigma + \underset{\sim}{\mu}' A\,\underset{\sim}{\mu}$.

Exercise: Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{0}, I)$ and if A and B are real symmetric matrices of order p, then

i) $E(\underset{\sim}{X}' A\underset{\sim}{X}) = tr\, A$

ii) $V(\underset{\sim}{X}' A\underset{\sim}{X}) = 2\,tr\, A^2$

iii) $\mathrm{Cov}(\underset{\sim}{X}' A\underset{\sim}{X}, \underset{\sim}{X}' B\underset{\sim}{X}) = 2\,tr\, AB$

Solution: Since A is a real symmetric matrix of order p, there exists an orthogonal matrix C such that

$C'AC = \mathrm{diag}\ (\lambda_1, \lambda_2, \cdots, \lambda_p) = \Delta$, where $\lambda_1, \lambda_2, \cdots, \lambda_p$ are latent roots of A.

Consider an orthogonal transformation

$\underset{\sim}{Y} = C\underset{\sim}{X}$, then $E(\underset{\sim}{Y}) = C\underset{\sim}{\mu} = \underset{\sim}{0}$, and $\Sigma_{\underset{\sim}{Y}} = E(\underset{\sim}{Y}\,\underset{\sim}{Y}') = C\,E(\underset{\sim}{X}\,\underset{\sim}{X}')\,C' = CC' = I$

This shows that $\underset{\sim}{Y} \sim N_p(\underset{\sim}{0}, I)$, i.e. $Y_i \sim N(0, 1)$.

Now

i) $\underset{\sim}{X}' A\underset{\sim}{X} = \underset{\sim}{Y}' C^{-1'} A C^{-1}\underset{\sim}{Y} = \underset{\sim}{Y}'\Delta\underset{\sim}{Y}$, since C being orthogonal, $C^{-1} = C'$.

$= \lambda_1 Y_1^2 + \lambda_2 Y_2^2 + \cdots \lambda_p Y_p^2$.

Therefore,

$E(\underset{\sim}{X}' A\underset{\sim}{X}) = \lambda_1 E(Y_1^2) + \lambda_2 E(Y_2^2) + \cdots \lambda_p E(Y_p^2)$

$= \lambda_1 + \lambda_2 + \cdots \lambda_p$, since $Y_i^2 \sim \chi_1^2$ for each i

$= tr\, C'AC = tr\, ACC' = tr\, A$.

ii) $V(\underset{\sim}{X}' A\underset{\sim}{X}) = \lambda_1^2 V(Y_1^2) + \lambda_2^2 V(Y_2^2) + \cdots \lambda_p^2 V(Y_p^2)$

$= 2(\lambda_1^2 + \lambda_2^2 + \cdots \lambda_p^2) = 2\,tr\, A^2$. Since $\lambda_1^2, \cdots, \lambda_p^2$ are the latent roots of A^2.

iii) Since A and B are real symmetric matrices of order p, so that, $A+B$ is real symmetric matrix of order p. Hence,

$$2\,tr\,(A+B)^2 = V[\underline{X}'(A+B)\underline{X}] = V(\underline{X}'A\underline{X} + \underline{X}'B\underline{X})$$

$$= V(\underline{X}'A\underline{X}) + V(\underline{X}'B\underline{X}) + 2\,\text{Cov}(\underline{X}'A\underline{X},\ \underline{X}'B\underline{X})$$

$$\Rightarrow \text{Cov}(\underline{X}'A\underline{X},\ \underline{X}'B\underline{X}) = \frac{1}{2}[2\,tr\,(A+B)^2 - V(\underline{X}'A\underline{X}) - V(\underline{X}'B\underline{X})]$$

$$= tr\,(A+B)^2 - tr\,A^2 - tr\,B^2$$

$$= tr\,(A^2 + B^2 + AB + BA) - tr\,A^2 - tr\,B^2$$

$$= tr\,AB + tr\,BA = 2\,tr\,AB, \text{ since } tr\,AB = tr\,BA.$$

Characteristic function

The characteristic function of a random vector $\underline{X}$ is defined as

$$\phi_{\underline{X}}(\underline{t}) = E[e^{i\underline{t}'\underline{X}}], \text{ where } \underline{t} \text{ is a vector of reals, } i = \sqrt{-1}.$$

Theorem: Let $\underline{X} = (X_1, X_2, \cdots, X_p)'$ be normally distributed random vector with mean $\underline{\mu}$ and positive definite covariance matrix Σ, then the characteristic function of $\underline{X}$ is given by

$$\phi_{\underline{X}}(\underline{t}) = e^{i\underline{t}'\underline{\mu} - \frac{1}{2}\underline{t}'\Sigma\underline{t}}, \text{ where } \underline{t} = (t_1, t_2, \cdots, t_p)' \text{ is a real vector of order } p \times 1.$$

Proof: We have

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right]$$

Since Σ is a symmetric and positive definite, there exits a non-singular matrix C such that

$$C'\Sigma^{-1}C = I, \text{ and } \Sigma = CC'.$$

Let $\underline{X} - \underline{\mu} = C\underline{Y}$, so that $\underline{Y} = C^{-1}(\underline{X} - \underline{\mu})$ a nonsingular transformation and the Jacobian of the transformation is $|J| = |C|$, therefore, the density function of $\underline{Y}$ is

$$f(\underline{y}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(C\underline{y} + \underline{\mu} - \underline{\mu})'\Sigma^{-1}(C\underline{y} + \underline{\mu} - \underline{\mu})\right]|C|$$

$$= \frac{1}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\underline{y}'C'\Sigma^{-1}C\underline{y})\right], \text{ since } |C| = |\Sigma|^{1/2}$$

$$= \frac{1}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}\underline{y}'\underline{y}\right] = \left(\frac{1}{(2\pi)^{1/2}}\exp-\frac{1}{2}y_1^2\right)\cdots\left(\frac{1}{(2\pi)^{1/2}}\exp-\frac{1}{2}y_p^2\right).$$

It shows that $Y_1, Y_2, \cdots, Y_p$ are independently normally distributed each with mean zero and variance one.

Now the characteristic function of $\underline{Y}$ is

$$\phi_{\underline{Y}}(\underline{u}) = E[e^{i\underline{u}'\underline{Y}}] = E\,e^{i(u_1Y_1 + \cdots + u_pY_p)} = E\,e^{iu_1Y_1} \cdots E\,e^{iu_pY_p} \tag{2.9}$$

Since $Y_1, Y_2, \cdots, Y_p$ are independent and distributed according to $N(0, 1)$,

$$\phi_{\underline{Y}}(\underline{u}) = \left(\exp-\frac{1}{2}u_1^2\right)\cdots\left(\exp-\frac{1}{2}u_p^2\right), \text{ since } X \sim N(\mu, \sigma^2), \text{ and } \phi_X(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}.$$

$$= e^{-\frac{1}{2}(u_1^2 + \cdots + u_p^2)} = e^{-\frac{1}{2}\underline{u}'\underline{u}} \tag{2.10}$$

Thus

$$\phi_{\underline{X}}(\underline{t}) = E[e^{i\underline{t}'\underline{X}}] = E[e^{i\underline{t}'(C\underline{Y} + \underline{\mu})}] = e^{i\underline{t}'\underline{\mu}}\,E[e^{i\underline{t}'(C\underline{Y})}]$$

$$= e^{i\underline{t}'\underline{\mu}}\,E\,e^{i(C'\underline{t})'\underline{Y}} = e^{i\underline{t}'\underline{\mu}}\,e^{-\frac{1}{2}(C'\underline{t})'(C'\underline{t})} \text{ from equation (2.9) and (2.10)}$$

$$= e^{i\underline{t}'\underline{\mu}}\,e^{-\frac{1}{2}\underline{t}'CC'\underline{t}} = e^{i\underline{t}'\underline{\mu} - \frac{1}{2}\underline{t}'\Sigma\underline{t}}.$$

Corollary: Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, if $\underline{Z} = D\underline{X}$, then the characteristic function of vector $\underline{Z}$ is $e^{i\underline{t}'(D\underline{\mu}) - \frac{1}{2}\underline{t}'(D\Sigma D')\underline{t}}$.

Proof: The characteristic function of vector $\underline{Z}$ is defined as

$$\phi_{\underline{Z}}(\underline{t}) = E\,e^{i\underline{t}'\underline{Z}} = E\,e^{i\underline{t}'D\underline{X}} = E\,e^{i(D'\underline{t})'\underline{X}} = e^{i(D'\underline{t})'\underline{\mu} - \frac{1}{2}(D'\underline{t})'\Sigma D'\underline{t}}$$

$$= e^{i\underline{t}'(D\underline{\mu}) - \frac{1}{2}\underline{t}'(D\Sigma D')\underline{t}},$$

which is the characteristic function of $N_p(D\underline{\mu},\ D\Sigma D)$.

Theorem: If every linear combination of the components of a vector $\underline{X}$ is normally distributed, then $\underline{X}$ has normal distribution.

Proof: Consider a vector $\underline{X}$ of p-components with density function $f(\underline{x})$ and characteristic function $\phi_{\underline{X}}(\underline{u}) = E[e^{i\underline{u}'\underline{X}}]$ and suppose the mean of $\underline{X}$ is $\underline{\mu}$ and the covariance matrix is Σ.

Since $\underline{u}'\underline{X}$ is normally distributed for every $\underline{u}$. Then the characteristic function of $\underline{u}'\underline{X}$ is

$$E\,e^{it(\underline{u}'\underline{X})} = e^{it\underline{u}'\underline{\mu} - \frac{1}{2}t^2\underline{u}'\Sigma\underline{u}}, \text{ taking } t = 1, \text{ this reduces to}$$

$$E\,e^{i\underline{u}'\underline{X}} = e^{i\underline{u}'\underline{\mu} - \frac{1}{2}\underline{u}'\Sigma\underline{u}}.$$

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

Theorem: The moment generating function of a vector $\underline{X}$, which is distributed according to $N_p(\underline{\mu}, \Sigma)$ is $M_{\underline{X}}(\underline{t}) = e^{\underline{t}'\underline{\mu}+\frac{1}{2}\underline{t}'\Sigma\underline{t}}$.

Proof: Since Σ is a symmetric and positive definite, then there exits a non-singular matrix C such that

$$C'\Sigma^{-1}C = I \text{ and } \Sigma = CC'.$$

Make the nonsingular transformation

$$\underline{X} - \underline{\mu} = C\underline{Y}, \text{ then } \underline{Y} = C^{-1}(\underline{X} - \underline{\mu}) \text{ and } |J| = |C|.$$

Therefore, the density function of $\underline{Y}$ is

$$f(\underline{y}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(C\underline{y}+\underline{\mu}-\underline{\mu})'\Sigma^{-1}(C\underline{y}+\underline{\mu}-\underline{\mu})\right]|C|$$

$$= \frac{1}{(2\pi)^{p/2}}\exp\left[-\frac{1}{2}(\underline{y}'C'\Sigma^{-1}C\underline{y})\right], \text{ since } |C| = |\Sigma|^{1/2}$$

$$= \frac{1}{(2\pi)^{p/2}}\exp\left(-\frac{1}{2}\underline{y}'\underline{y}\right).$$

It shows that $Y_1, Y_2, \cdots, Y_p$ are independently normally distributed each with mean zero and variance one.

Now the moment generating function of $\underline{Y}$ is

$$M_{\underline{Y}}(\underline{u}) = E\,e^{\underline{u}'\underline{Y}} = E\,e^{(u_1Y_1+\cdots+u_pY_p)} = E\,e^{(u_1Y_1)}\cdots E\,e^{(u_pY_p)}$$

$$= \prod_{i=1}^{p} E\,e^{u_iY_i} = e^{\frac{1}{2}\underline{u}'\underline{u}}, \text{ since } Y_i \sim N(0, 1).$$

Thus we can say

$$\phi_{\underline{X}}(\underline{t}) = E\,e^{\underline{t}'\underline{X}} = E\,e^{\underline{t}'(C\underline{Y}+\underline{\mu})} = e^{\underline{t}'\underline{\mu}}E[e^{\underline{t}'(C\underline{Y})}] = e^{\underline{t}'\underline{\mu}}E[e^{(C'\underline{t})'\underline{Y}}]$$

$$= e^{\underline{t}'\underline{\mu}}e^{\frac{1}{2}(C'\underline{t})'(C'\underline{t})} = e^{\underline{t}'\underline{\mu}+\frac{1}{2}\underline{t}'CC'\underline{t}} = e^{\underline{t}'\underline{\mu}+\frac{1}{2}\underline{t}'\Sigma\underline{t}}.$$

Exercise: If $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, then obtain the distribution of quadratic form $Q = (\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})$. Also find its $r-$th mean.

Solution: Given $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, then the probability density function of $\underline{X}$ is

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right]$$

Since Σ is a symmetric and positive definite there exits a non-singular matrix C such that $C'\Sigma^{-1}C = I$ and $\Sigma = CC'$.

Let $\underline{X} - \underline{\mu} = C\underline{Y}$ a nonsingular transformation. Under this transformation the quadratic form expanded as

$$Q = (\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}) = (C\underline{y})'\Sigma^{-1}(C\underline{y}) = \underline{y}'C'\Sigma^{-1}C\underline{y} = \underline{y}'\underline{y}$$

and the Jacobian of the transformation is $|J| = |C|$.

Therefore, the density function of $\underline{Y}$ is

$$f(\underline{y}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}\underline{y}'\underline{y}\right)|C|, \text{ since } |\Sigma| = |CC'| = |C|^2, \text{ and } |C| = |\Sigma|^{1/2}.$$

$$= \frac{1}{(2\pi)^{p/2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{p}y_i^2\right) = \left(\frac{1}{(2\pi)^{1/2}}\exp-\frac{1}{2}y_1^2\right)\cdots\left(\frac{1}{(2\pi)^{1/2}}\exp-\frac{1}{2}y_p^2\right).$$

This shows that $Y_1, Y_2, \cdots, Y_p$ are independently standard normally variate i.e. $\underline{Y} \sim N_p(\underline{0}, I)$, then $Y_i \sim N(0, 1)$. Therefore, $Q = \sum_{i=1}^{p} y_i^2 \sim \chi_p^2$.

For $r-$th mean

$$E(Q^r) = \frac{1}{2^{p/2}\Gamma(p/2)}\int_0^\infty Q^r Q^{\frac{p}{2}-1}e^{-\frac{Q}{2}}dQ = \frac{1}{2^{p/2}\Gamma(p/2)}\int_0^\infty Q^{\frac{p}{2}+r-1}e^{-\frac{Q}{2}}dQ$$

$$= \frac{2^{\frac{p}{2}+r}}{2^{p/2}\Gamma(p/2)}\Gamma\left(\frac{p}{2}+r\right), \text{ as } \int_0^\infty x^{\alpha-1}e^{-x/\beta}dx = \frac{\Gamma\alpha}{(1/\beta)^\alpha}$$

$$= \frac{2^r\Gamma(\frac{p}{2}+r)}{\Gamma(p/2)} = \frac{2^r\left(\frac{p}{2}+r-1\right)\left(\frac{p}{2}+r-2\right)\cdots\left(\frac{p}{2}\right)\Gamma(p/2)}{\Gamma(p/2)} = 2^r\prod_{i=1}^{r}\left(\frac{p}{2}+r-i\right).$$

In particular $r = 1, 2$

$$\begin{pmatrix}\underline{Y}^{(1)}\\ \underline{Y}^{(2)}\end{pmatrix} = \begin{pmatrix}A\\ B\end{pmatrix}\underline{X}, \text{ and, } E(Q^2) = 2^2\left(\frac{p}{2}+1\right)\left(\frac{p}{2}\right) = p^2 + 2p.$$

Therefore,

$$V(Q) = E(Q^2) - [E(Q)]^2 = p^2 + 2p - p^2 = 2p.$$

Alternative method

Given $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, then the probability density function of $\underline{X}$ is

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right].$$

Let

$Y=(\underline{X}-\underline{\mu})'\Sigma^{-1}(\underline{X}-\underline{\mu})$, and its moment generating function

$$M_Y(t)=E(e^{tY})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\int e^{t(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})}e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})}\,d\underline{x}$$

$$=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\int e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'(1-2t)\Sigma^{-1}(\underline{x}-\underline{\mu})}\,d\underline{x}$$

$$=\frac{\left|(1-2t)\Sigma^{-1}\right|^{1/2}}{(2\pi)^{p/2}|\Sigma|^{1/2}\left|(1-2t)\Sigma^{-1}\right|^{1/2}}\int e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'(1-2t)\Sigma^{-1}(\underline{x}-\underline{\mu})}\,d\underline{x}$$

$$=\frac{1}{|\Sigma|^{1/2}\left|(1-2t)\Sigma^{-1}\right|^{1/2}}=\frac{1}{(1-2t)^{p/2}|\Sigma|^{1/2}\left|\Sigma^{-1}\right|^{1/2}}$$

$$=\frac{1}{(1-2t)^{p/2}}.$$

But this is the moment generating function of χ^2_p, therefore, $Y\sim\chi^2_p$.

Exercise: Suppose $\underline{X}\sim N_p(\underline{\mu},\Sigma)$ and let A be some $\underline{Y}\sim N_p(\underline{\upsilon},I)$ matrix of rank q. If $A\underline{X}=\underline{0}$, then the quadratic form $Q=(\underline{X}-\underline{\mu})'\Sigma^{-1}(\underline{X}-\underline{\mu})$ follows a chi-square distribution on $Y_i\sim N_1(\upsilon_i,1)$ degree of freedom.

Solution: Since Σ is positive definite and A is of full row rank, there exists a $(p-q)\times p$ matrix B of rank $p-q$ such that $A\Sigma B'=\underline{0}$ and that $C=\begin{pmatrix}A\\B\end{pmatrix}$ is nonsingular.

Consider the nonsingular linear transformation

$\underline{Y}=C\underline{X}$, therefore, $\underline{Y}\sim N_p(C\underline{\mu},\,C\Sigma C')$.

Now

$$\underline{Y}=\begin{pmatrix}A\\B\end{pmatrix}\underline{X}=\begin{pmatrix}A\underline{X}\\B\underline{X}\end{pmatrix}=\begin{pmatrix}\underline{Y}^{(1)}\\\underline{Y}^{(2)}\end{pmatrix},\text{ say}$$

$$C\underline{\mu}=\begin{pmatrix}A\\B\end{pmatrix}\underline{\mu},\text{ and }C\Sigma C'=\begin{pmatrix}A\\B\end{pmatrix}\Sigma(A'\quad B')=\begin{pmatrix}A\Sigma A' & A\Sigma B'\\B\Sigma A' & B\Sigma B'\end{pmatrix}=\begin{pmatrix}A\Sigma A' & \underline{0}\\\underline{0} & B\Sigma B'\end{pmatrix}.$$

This means that $\underline{Y}^{(1)}\sim N_q(A\underline{\mu},\,A\Sigma A')$ and $\underline{Y}^{(2)}\sim N_{p-q}(B\underline{\mu},\,B\Sigma B')$ independently.

Therefore,

$$Q=(\underline{X}-\underline{\mu})'\Sigma^{-1}(\underline{X}-\underline{\mu})=(C^{-1}\underline{Y}-\underline{\mu})'\Sigma^{-1}(C^{-1}\underline{Y}-\underline{\mu})$$

$$=(C^{-1}\underline{Y}-C^{-1}C\underline{\mu})'\Sigma^{-1}(C^{-1}\underline{Y}-C^{-1}C\underline{\mu})=(\underline{Y}-C\underline{\mu})'C^{-1'}\Sigma^{-1}C^{-1}(\underline{Y}-C\underline{\mu})$$

$$=\left[\begin{pmatrix}\underline{Y}^{(1)}\\\underline{Y}^{(2)}\end{pmatrix}-\begin{pmatrix}A\\B\end{pmatrix}\underline{\mu}\right]'(C\Sigma C')^{-1}\left[\begin{pmatrix}\underline{Y}^{(1)}\\\underline{Y}^{(2)}\end{pmatrix}-\begin{pmatrix}A\\B\end{pmatrix}\underline{\mu}\right]$$

$$=[(\underline{Y}^{(1)}-A\underline{\mu})'\quad(\underline{Y}^{(2)}-B\underline{\mu})']\begin{pmatrix}A\Sigma A' & 0\\0 & B\Sigma B'\end{pmatrix}^{-1}\begin{bmatrix}(\underline{Y}^{(1)}-A\underline{\mu})\\(\underline{Y}^{(2)}-B\underline{\mu})\end{bmatrix}.$$

Under the condition $A\underline{X}=\underline{0}$, we have $\underline{Y}^{(1)}=\underline{0}$, and hence $E\underline{Y}^{(1)}=\underline{0}$, we get

$$Q=[\underline{0}\quad(\underline{Y}^{(2)}-B\underline{\mu})']\begin{pmatrix}A\Sigma A' & 0\\0 & B\Sigma B'\end{pmatrix}^{-1}\begin{bmatrix}\underline{0}\\(\underline{Y}^{(2)}-B\underline{\mu})\end{bmatrix}$$

$=(\underline{Y}^{(2)}-B\underline{\mu})'(B\Sigma B')^{-1}(\underline{Y}^{(2)}-B\underline{\mu})$ and hence,

$Q=(\underline{X}-\underline{\mu})'\Sigma^{-1}(\underline{X}-\underline{\mu})\sim\chi^2_{p-q}$, since $\underline{Y}^{(2)}\sim N_{p-q}(B\underline{\mu},\,B\Sigma B')$.

Exercise: If $\underline{X}\sim N_p(\underline{\mu},\Sigma)$, find the distribution of quadratic form $Q=\underline{X}'\Sigma^{-1}\underline{X}$.

Solution: Since Σ is a symmetric and positive definite there exits a non-singular matrix C such that

$C'\Sigma^{-1}C=I$ and $\Sigma=CC'$.

Let $\underline{X}=C\underline{Y}$ a nonsingular transformation, then $\underline{Y}=C^{-1}\underline{X}$.

Since the transformation is nonsingular, the distribution of $\underline{Y}$ is also $p-$variate normal with

$E(\underline{Y})=C^{-1}E(\underline{X})=C^{-1}\underline{\mu}=\underline{\upsilon}$

and

$$\Sigma_{\underline{Y}}=E[\underline{Y}-E(\underline{Y})]\,[\underline{Y}-E(\underline{Y})]'=C^{-1}E(\underline{X}-\underline{\mu})\,(\underline{X}-\underline{\mu})'C^{-1'}=C^{-1}\Sigma\,C^{-1'}=I$$

This shows that $\underline{Y}\sim N_p(\underline{\upsilon},I)$, and $Y_i\sim N_1(\upsilon_i,1)$.

Therefore,

$$Q=\underline{X}'\Sigma^{-1}\underline{X}=(C\underline{Y})'\Sigma^{-1}(C\underline{Y})=\underline{Y}'(C'\Sigma^{-1}C)\underline{Y}=\underline{Y}'\underline{Y}$$

$$=\sum_{i=1}^{p}Y_i^2\sim\chi^2_{p,\sum_i\upsilon_i^2},\text{ where, }\sum_{i=1}^{p}\upsilon_i^2=\underline{\upsilon}'\underline{\upsilon}=\underline{\mu}'\Sigma^{-1}\underline{\mu}=\text{ non-centrality parameter.}$$

Alternative method

Given $\underline{X}\sim N_p(\underline{\mu},\Sigma)$, then the probability density function of $\underline{X}$ is

$$f(\underline{x})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right].$$

Let

$Y=\underline{X}'\Sigma^{-1}\underline{X}$, and its moment generating function

$$M_Y(t)=E(e^{tY})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\int e^{t\underline{x}'\Sigma^{-1}\underline{x}}\,e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})}\,d\underline{x}$$

$$=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\int e^{t\underline{x}'\Sigma^{-1}\underline{x}-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})}\,d\underline{x}.$$

Consider

$$Q=\underline{x}'\Sigma^{-1}\underline{x}-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})$$

$$=-\frac{1}{2}[\underline{x}'\Sigma^{-1}\underline{x}-\underline{\mu}'\Sigma^{-1}\underline{x}-\underline{x}'\Sigma^{-1}\underline{\mu}+\underline{\mu}'\Sigma^{-1}\underline{\mu}-2t\,\underline{x}'\Sigma^{-1}\underline{x}]$$

$$=-\frac{1}{2}[\underline{x}'(1-2t)\Sigma^{-1}\underline{x}-2\underline{x}'\Sigma^{-1}\underline{\mu}+\theta],\ \text{where}\ \theta=\underline{\mu}'\Sigma^{-1}\underline{\mu}.$$

We know that for A symmetric matrix

$$\underline{a}'A^{-1}\underline{a}-2\underline{a}'\underline{u}=(\underline{a}-A\underline{u})'A^{-1}(\underline{a}-A\underline{u})-\underline{u}'A\underline{u}$$

$$=\underline{a}'A^{-1}\underline{a}-\underline{u}'AA^{-1}\underline{a}-\underline{a}'A^{-1}A\underline{u}+\underline{u}'AA^{-1}A\underline{u}-\underline{u}'A\underline{u}$$

$$=\underline{a}'A^{-1}\underline{a}-2\underline{a}'\underline{u}.$$

Comparing $\underline{a}'A^{-1}\underline{a}-2\underline{a}'\underline{u}=(\underline{a}-A\underline{u})'A^{-1}(\underline{a}-A\underline{u})-\underline{u}'A\underline{u}$ with Q, we get

$$\underline{a}=\underline{x},\ A^{-1}=(1-2t)\Sigma^{-1},\ \Rightarrow\ A=\frac{1}{1-2t}\Sigma,\ \underline{u}=\Sigma^{-1}\underline{\mu}$$

Thus,

$$Q=-\frac{1}{2}\left[\left(\underline{x}-\frac{1}{(1-2t)}\Sigma\Sigma^{-1}\underline{\mu}\right)'(1-2t)\Sigma^{-1}\left(\underline{x}-\frac{1}{(1-2t)}\Sigma\Sigma^{-1}\underline{\mu}\right)-\underline{\mu}'\Sigma^{-1}\frac{\Sigma}{(1-2t)}\Sigma^{-1}\underline{\mu}+\theta\right]$$

$$=-\frac{1}{2}\left[\left(\underline{x}-\frac{\underline{\mu}}{(1-2t)}\right)'(1-2t)\Sigma^{-1}\left(\underline{x}-\frac{\underline{\mu}}{(1-2t)}\right)-\frac{\theta}{(1-2t)}+\theta\right].$$

Therefore,

$$M_Y(t)=\frac{e^{t\theta/(1-2t)}\left|(1-2t)\Sigma^{-1}\right|^{1/2}}{(2\pi)^{p/2}|\Sigma|^{1/2}\left|(1-2t)\Sigma^{-1}\right|^{1/2}}\int e^{-\frac{1}{2}\left(\underline{x}-\frac{\underline{\mu}}{1-2t}\right)'(1-2t)\Sigma^{-1}\left(\underline{x}-\frac{\underline{\mu}}{1-2t}\right)}\,d\underline{x}$$

$$=\frac{1}{(1-2t)^{p/2}}e^{(t\theta)/(1-2t)}.$$

But this is the *mgf* of $\chi^2_{(p,\,\theta)}$, $\Rightarrow\quad Y\sim\chi^2_{(p,\,\theta)}$.

Exercise: If $\underline{X}\sim N_p(\underline{\mu},\Sigma)$. Show that $M_{\underline{X}-\underline{\mu}}(\underline{t})=\exp\left(\frac{1}{2}\underline{t}'\Sigma\underline{t}\right)$, hence find the moment generating function of $\underline{X}$.

Solution: By definition

$$M_{(\underline{X}-\underline{\mu})}(t)=E\left[e^{\underline{t}'(\underline{X}-\underline{\mu})}\right]=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\int e^{\underline{t}'(\underline{x}-\underline{\mu})}e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})}\,d\underline{x}$$

$$=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\int e^{-\frac{1}{2}[(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})-2\underline{t}'(\underline{x}-\underline{\mu})]}\,d\underline{x}$$

We know that for A symmetric matrix

$$\underline{a}'A^{-1}\underline{a}-2\underline{a}'\underline{t}=(\underline{a}-A\underline{t})'A^{-1}(\underline{a}-A\underline{t})-\underline{t}'A\underline{t}$$

Comparing $\underline{a}'A^{-1}\underline{a}-2\underline{a}'\underline{u}$ with $(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})-2\underline{t}'(\underline{x}-\underline{\mu})$, we get

$$\underline{a}=(\underline{x}-\underline{\mu}),\quad A=\Sigma,\quad \underline{u}=\underline{t},\ \text{since}\ \underline{a}'\underline{u}=\underline{u}'\underline{a},\ \text{so that}$$

$$M_{(\underline{X}-\underline{\mu})}(\underline{t})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\int e^{-\frac{1}{2}[(\underline{x}-\underline{\mu}-\Sigma\underline{t})'\Sigma^{-1}(\underline{x}-\underline{\mu}-\Sigma\underline{t})-\underline{t}'\Sigma\underline{t}]}\,d\underline{x}$$

$$=\frac{e^{\frac{1}{2}\underline{t}'\Sigma\underline{t}}}{(2\pi)^{p/2}|\Sigma|^{1/2}}\int e^{-\frac{1}{2}(\underline{x}-\underline{\mu}^*)'\Sigma^{-1}(\underline{x}-\underline{\mu}^*)}\,d\underline{x}=e^{\frac{1}{2}\underline{t}'\Sigma\underline{t}}.$$

Also,

$$M_{(\underline{X}-\underline{\mu})}(\underline{t})=E\left[e^{\underline{t}'(\underline{X}-\underline{\mu})}\right]=e^{-\underline{t}'\underline{\mu}}E(e^{\underline{t}'\underline{X}})=e^{-\underline{t}'\underline{\mu}}M_{\underline{X}}(\underline{t})$$

$$\Rightarrow\ M_{\underline{X}}(\underline{t})=e^{\underline{t}'\underline{\mu}}M_{(\underline{X}-\underline{\mu})}(\underline{t})=e^{\underline{t}'\underline{\mu}+\frac{1}{2}\underline{t}'\Sigma\underline{t}}.$$

Exercise: For any vector $\underline{t}$ and any positive definite symmetric matrix W, show that $\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\underline{x}'W^{-1}\underline{x}+\underline{t}'\underline{x}\right)d\underline{x}=(2\pi)^{p/2}|W|^{1/2}e^{\frac{1}{2}\underline{t}'W\underline{t}}$. Hence find the moment generating function of $p-$variate normal distribution with positive definite covariance matrix Σ.

Solution: Consider the quadratic form

$-\frac{1}{2}(\underline{x}'W^{-1}\underline{x}-2\underline{t}'\underline{x})=Q$ (say). We know that for A symmetric matrix

$$\underline{a}'A^{-1}\underline{a}-2\underline{a}'\underline{u}=(\underline{a}-A\underline{u})'A^{-1}(\underline{a}-A\underline{u})-\underline{u}'A\underline{u}.$$

Comparing $\underline{a}'A^{-1}\underline{a}-2\underline{a}'\underline{u}$ with Q, we get $\underline{a}=\underline{x}$, $A=W$, $\underline{u}=\underline{t}$, since $\underline{u}'\underline{t}=\underline{t}'\underline{u}$.

Thus,

$$Q=-\frac{1}{2}[(\underset{\sim}{x}-W\underset{\sim}{t})'W^{-1}(\underset{\sim}{x}-W\underset{\sim}{t})-\underset{\sim}{t}'W\underset{\sim}{t}]=-\frac{1}{2}(\underset{\sim}{x}-W\underset{\sim}{t})'W^{-1}(\underset{\sim}{x}-W\underset{\sim}{t})+\frac{1}{2}\underset{\sim}{t}'W\underset{\sim}{t}$$

Therefore,

$$\frac{1}{(2\pi)^{p/2}|W|^{1/2}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}(\underset{\sim}{x}-W\underset{\sim}{t})'W^{-1}(\underset{\sim}{x}-W\underset{\sim}{t})\right]\exp\left(\frac{1}{2}\underset{\sim}{t}'W\underset{\sim}{t}\right)d\underset{\sim}{x}$$

$$=\frac{\exp\left(\frac{1}{2}\underset{\sim}{t}'W\underset{\sim}{t}\right)}{(2\pi)^{p/2}|W|^{1/2}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}(\underset{\sim}{x}-W\underset{\sim}{t})'W^{-1}(\underset{\sim}{x}-W\underset{\sim}{t})\right]d\underset{\sim}{x}$$

$$=e^{\frac{1}{2}\underset{\sim}{t}'W\underset{\sim}{t}}=e^{\frac{1}{2}\underset{\sim}{t}'\Sigma\underset{\sim}{t}}, \text{ since } \Sigma \text{ is positive definite symmetric matrix}$$

$$=M_{(\underset{\sim}{X}-\underset{\sim}{\mu})}(\underset{\sim}{t})=e^{-\underset{\sim}{t}'\underset{\sim}{\mu}}M_{\underset{\sim}{X}}(\underset{\sim}{t}) \text{ or } M_{\underset{\sim}{X}}(\underset{\sim}{t})=e^{\underset{\sim}{t}'\underset{\sim}{\mu}}M_{(\underset{\sim}{X}-\underset{\sim}{\mu})}(\underset{\sim}{t})=e^{\underset{\sim}{t}'\underset{\sim}{\mu}+\frac{1}{2}\underset{\sim}{t}'\Sigma\underset{\sim}{t}}.$$

Exercise: Let $\underset{\sim}{X}\sim N_p(\underset{\sim}{\mu},\Sigma)$ and $\underset{\sim}{Y}\sim N_p(\underset{\sim}{\delta},\Omega)$, if $\underset{\sim}{U}_1=\underset{\sim}{X}+\underset{\sim}{Y}$, and $\underset{\sim}{U}_2=\underset{\sim}{X}-\underset{\sim}{Y}$, find the joint distribution of $\underset{\sim}{U}_1$, and $\underset{\sim}{U}_2$.

Solution: Given

$$\underset{\sim}{Z}_{2p\times 1}=\begin{pmatrix}\underset{\sim}{X}\\ \underset{\sim}{Y}\end{pmatrix},\ E\underset{\sim}{Z}=\begin{pmatrix}\underset{\sim}{\mu}\\ \underset{\sim}{\delta}\end{pmatrix}, \text{ and } \Sigma_{\underset{\sim}{Z}_{2p\times 2p}}=\begin{pmatrix}\Sigma & \underset{\sim}{0}\\ \underset{\sim}{0} & \Omega\end{pmatrix}.$$

Consider the transformation

$$\underset{\sim}{U}=\begin{pmatrix}\underset{\sim}{U}_1\\ \underset{\sim}{U}_2\end{pmatrix}=\begin{pmatrix}I & I\\ I & -1I\end{pmatrix}_{2p\times 2p}\begin{pmatrix}\underset{\sim}{X}\\ \underset{\sim}{Y}\end{pmatrix}, \text{ i.e. } \underset{\sim}{U}=C_{2p\times 2p}\underset{\sim}{Z}, \text{ then}$$

$$E\underset{\sim}{U}=C\,E\underset{\sim}{Z}=C\begin{pmatrix}\underset{\sim}{\mu}\\ \underset{\sim}{\delta}\end{pmatrix}=\begin{pmatrix}I & I\\ I & -I\end{pmatrix}\begin{pmatrix}\underset{\sim}{\mu}\\ \underset{\sim}{\delta}\end{pmatrix}=\begin{pmatrix}\underset{\sim}{\mu}+\underset{\sim}{\delta}\\ \underset{\sim}{\mu}-\underset{\sim}{\delta}\end{pmatrix}$$

and

$$\Sigma_{\underset{\sim}{U}}=C\Sigma_{\underset{\sim}{Z}}C'=\begin{pmatrix}I & I\\ I & -I\end{pmatrix}\begin{pmatrix}\Sigma & \underset{\sim}{0}\\ \underset{\sim}{0} & \Omega\end{pmatrix}\begin{pmatrix}I & I\\ I & -I\end{pmatrix}=\begin{pmatrix}\Sigma+\Omega & \Sigma-\Omega\\ \Sigma-\Omega & \Sigma+\Omega\end{pmatrix}.$$

Exercise: Let $\underset{\sim}{X}_1,\underset{\sim}{X}_2,\underset{\sim}{X}_3$ and $\underset{\sim}{X}_4$ be independent $N_p(\underset{\sim}{\mu},\Sigma)$ random vectors, if $\underset{\sim}{U}_1=\frac{1}{4}\underset{\sim}{X}_1-\frac{1}{4}\underset{\sim}{X}_2+\frac{1}{4}\underset{\sim}{X}_3-\frac{1}{4}\underset{\sim}{X}_4$, and $\underset{\sim}{U}_2=\frac{1}{4}\underset{\sim}{X}_1+\frac{1}{4}\underset{\sim}{X}_2-\frac{1}{4}\underset{\sim}{X}_3-\frac{1}{4}\underset{\sim}{X}_4$, find the joint distribution of $\underset{\sim}{U}_1$, and $\underset{\sim}{U}_2$, and their marginals.

Solution: Consider the transformation

$$\underset{\sim}{U}=\begin{pmatrix}\underset{\sim}{U}_1\\ \underset{\sim}{U}_2\end{pmatrix}=\begin{pmatrix}\frac{1}{4}I & -\frac{1}{4}I & \frac{1}{4}I & -\frac{1}{4}I\\ \frac{1}{4}I & \frac{1}{4}I & -\frac{1}{4}I & -\frac{1}{4}I\end{pmatrix}_{2p\times 4p}\begin{pmatrix}\underset{\sim}{X}_1\\ \underset{\sim}{X}_2\\ \underset{\sim}{X}_3\\ \underset{\sim}{X}_4\end{pmatrix}, \text{ i.e. } \underset{\sim}{U}=C_{2p\times 4p}\underset{\sim}{X}_{4p\times 1}, \text{ then}$$

$$E\underset{\sim}{U}=D\,E\underset{\sim}{X}=\begin{pmatrix}\frac{1}{4}I & -\frac{1}{4}I & \frac{1}{4}I & -\frac{1}{4}I\\ \frac{1}{4}I & \frac{1}{4}I & -\frac{1}{4}I & -\frac{1}{4}I\end{pmatrix}_{4p\times 4p}\begin{pmatrix}\underset{\sim}{\mu}_1\\ \underset{\sim}{\mu}_2\\ \underset{\sim}{\mu}_3\\ \underset{\sim}{\mu}_4\end{pmatrix}=\begin{pmatrix}\underset{\sim}{0}\\ \underset{\sim}{0}\end{pmatrix}$$

and

$$\Sigma_{\underset{\sim}{U}}=D\Sigma_{\underset{\sim}{Z}}D'=\begin{pmatrix}\frac{1}{4}I & -\frac{1}{4}I & \frac{1}{4}I & -\frac{1}{4}I\\ \frac{1}{4}I & \frac{1}{4}I & -\frac{1}{4}I & -\frac{1}{4}I\end{pmatrix}\begin{pmatrix}\Sigma & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0}\\ \underset{\sim}{0} & \Sigma & \underset{\sim}{0} & \underset{\sim}{0}\\ \underset{\sim}{0} & \underset{\sim}{0} & \Sigma & \underset{\sim}{0}\\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} & \Sigma\end{pmatrix}\begin{pmatrix}\frac{1}{4}I & \frac{1}{4}I\\ -\frac{1}{4}I & \frac{1}{4}I\\ \frac{1}{4}I & -\frac{1}{4}I\\ -\frac{1}{4}I & -\frac{1}{4}I\end{pmatrix}$$

$$=\begin{pmatrix}\frac{1}{16}(4\Sigma) & \underset{\sim}{0}\\ \underset{\sim}{0} & \frac{1}{16}(4\Sigma)\end{pmatrix}=\begin{pmatrix}\frac{1}{4}\Sigma & \underset{\sim}{0}\\ \underset{\sim}{0} & \frac{1}{4}\Sigma\end{pmatrix}.$$

This shows that $\underset{\sim}{U}_1$ and $\underset{\sim}{U}_2$ are independent and $\underset{\sim}{U}_1\sim N_p(\underset{\sim}{0},\Sigma/4)$, $\underset{\sim}{U}_2\sim N_p(\underset{\sim}{0},\Sigma/4)$.

Note:

$$M_X(t)=E\,e^{tX}=\int e^{tx}f(x)\,dx=E\left[1+t\,X+\frac{t^2X^2}{2!}+\cdots+\frac{t^rX^r}{r!}+\cdots\right]$$

$$=1+t\,E(X)+\frac{t^2}{2!}E(X^2)+\cdots+\frac{t^r}{r!}E(X^r)+\cdots$$

$$=1+t\,\mu_1'+\frac{t^2}{2!}\mu_2'+\cdots+\frac{t^r}{r!}\mu_r'+\cdots \tag{2.11}$$

Differentiating r times equation (2.11) with respect to t and then putting $t=0$, we get

$$\left[\frac{\partial^r}{\partial t^r}M_X(t)\right]_{t=0}=\left(\mu_r'+\mu_{r+1}'t+\mu_{r+2}'\frac{t^2}{2!}+\cdots\right)_{t=0}\quad\Rightarrow\quad \mu_r'=\left.\frac{\partial^r}{\partial t^r}M_X(t)\right|_{t=0}.$$

Raw moments of multivariate normal distribution

$$E[X_n]=\frac{\partial M}{\partial t_n}\bigg|_{\underline{t}=0}=\frac{\partial}{\partial t_n}\left\{e^{\underline{t}'\underline{\mu}+\frac{1}{2}\underline{t}'\Sigma\underline{t}}\right\}\bigg|_{\underline{t}=\underline{0}}$$

$$=\frac{\partial}{\partial t_n}\left\{\exp\left(\sum_{k=1}^{p}t_k\,\mu_k+\frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{p}t_k\,t_j\,\sigma_{kj}\right)\right\}\bigg|_{\underline{t}=\underline{0}}$$

$$=M\left\{\mu_n+\frac{1}{2}\frac{\partial}{\partial t_n}\sum_{k=1}^{p}t_k\sum_{j=1}^{p}t_j\,\sigma_{kj}\right\}\bigg|_{\underline{t}=\underline{0}}$$

$$=M\left\{\mu_n+\frac{1}{2}\frac{\partial}{\partial t_n}(t_1(t_1\sigma_{11}+t_2\,\sigma_{12}+\cdots+t_n\sigma_{1n}+\cdots+t_p\sigma_{1p})\right.$$

$$+\cdots+t_n\,(t_1\sigma_{n1}+\cdots+t_n\sigma_{nn}+\cdots+t_p\sigma_{np})$$

$$\left.+\cdots+t_p\,(t_1\sigma_{p1}+\cdots+t_n\sigma_{pn}+\cdots+t_p\sigma_{pp}))\right\}_{\underline{t}=\underline{0}}$$

$$=M\left\{\mu_n+\frac{1}{2}(t_1\sigma_{1n}+\cdots+\sum_{j=1}^{p}t_j\sigma_{nj}+t_n\,\sigma_{nn}+\cdots+t_p\,\sigma_{pn})\right\}_{\underline{t}=\underline{0}}$$

$$=M\left\{\mu_n+\frac{1}{2}\left(\sum_{j=1}^{p}t_j\sigma_{jn}+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)\right\}_{\underline{t}=\underline{0}}=M\left\{\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right\}_{\underline{t}=\underline{0}}=\mu_n,\text{ as } e^0=1$$

Second moment

$$E[X_n\ \ X_l]=\frac{\partial^2 M}{\partial t_l\,\partial t_n}\bigg|_{\underline{t}=0}=\frac{\partial}{\partial t_l}\left\{\frac{\partial M}{\partial t_n}\right\}\bigg|_{\underline{t}=\underline{0}}=\frac{\partial}{\partial t_l}\left[M\left\{\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right\}\right]_{\underline{t}=\underline{0}}$$

$$=\left\{M\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)+M\,\sigma_{nl}\right\}_{\underline{t}=\underline{0}}=\mu_l\mu_n+\sigma_{nl}\,.$$

Therefore,

$$E(X_n^2)=\mu_n^2+\sigma_{nn}\quad\text{and}$$

$$V(X_n)=E(X_n^2)-[E(X_n)]^2=\mu_n^2+\sigma_{nn}-\mu_n^2=\sigma_{nn}\,.$$

$$Cov(X_n,X_l)=E(X_n\,X_l)-E(X_n)\,E(X_l)=\mu_n\,\mu_l+\sigma_{nl}-\mu_n\,\mu_l=\sigma_{nl}\,.$$

Third moment

$$E[X_n\ \ X_l\ \ X_r]=\frac{\partial^3 M}{\partial t_n\,\partial t_l\,\partial t_r}\bigg|_{\underline{t}=0}=\frac{\partial}{\partial t_r}\left\{\frac{\partial^2 M}{\partial t_n\,\partial t_l}\right\}\bigg|_{\underline{t}=\underline{0}}$$

$$=\frac{\partial}{\partial t_r}\left\{M\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)+M\,\sigma_{nl}\right\}_{\underline{t}=\underline{0}}$$

$$=\left[M\left(\mu_r+\sum_{j=1}^{p}t_j\,\sigma_{rj}\right)\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)\right.$$

$$\left.+M\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)\sigma_{lr}+M\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\sigma_{nr}+M\left(\mu_r+\sum_{j=1}^{p}t_j\,\sigma_{rj}\right)\sigma_{nl}\right]_{\underline{t}=\underline{0}}$$

$$=\mu_r\mu_l\mu_n+\mu_n\sigma_{lr}+\mu_l\sigma_{nr}+\mu_r\sigma_{nl}\,.$$

Fourth moment

$$E[X_n\ \ X_l\ \ X_r\ \ X_m]=\frac{\partial^4 M}{\partial t_n\,\partial t_l\,\partial t_r\,\partial t_m}\bigg|_{\underline{t}=0}=\frac{\partial}{\partial t_m}\left\{\frac{\partial^3 M}{\partial t_n\,\partial t_l\,\partial t_r}\right\}\bigg|_{\underline{t}=\underline{0}}$$

$$=\frac{\partial}{\partial t_m}\left[M\left(\mu_r+\sum_{j=1}^{p}t_j\,\sigma_{rj}\right)\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)\right.$$

$$\left.+M\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)\sigma_{lr}+M\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\sigma_{nr}+M\left(\mu_r+\sum_{j=1}^{p}t_j\,\sigma_{rj}\right)\sigma_{nl}\right]_{\underline{t}=\underline{0}}$$

$$=\left[M\left(\mu_m+\sum_{j=1}^{p}t_j\,\sigma_{mj}\right)\left(\mu_r+\sum_{j=1}^{p}t_j\,\sigma_{rj}\right)\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)\right.$$

$$+M\left\{\sigma_{rm}\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)+\sigma_{lm}\left(\mu_r+\sum_{j=1}^{p}t_j\,\sigma_{rj}\right)\right.$$

$$\left.\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)+\left(\mu_r+\sum_{j=1}^{p}t_j\,\sigma_{rj}\right)\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\sigma_{nm}\right\}$$

$$+M\left(\mu_m+\sum_{j=1}^{p}t_j\,\sigma_{mj}\right)\left(\mu_n+\sum_{j=1}^{p}t_j\,\sigma_{nj}\right)\sigma_{lr}+M\,\sigma_{nm}\,\sigma_{lr}$$

$$+M\left(\mu_m+\sum_{j=1}^{p}t_j\,\sigma_{mj}\right)\left(\mu_l+\sum_{j=1}^{p}t_j\,\sigma_{lj}\right)\sigma_{nr}+M\,\sigma_{lm}\,\sigma_{nr}$$

$$\left.+M\left(\mu_m+\sum_{j=1}^{p}t_j\,\sigma_{mj}\right)\left(\mu_r+\sum_{j=1}^{p}t_j\,\sigma_{rj}\right)\sigma_{nl}+M\,\sigma_{rm}\,\sigma_{nl}\right]_{\underline{t}=\underline{0}}$$

$$= \mu_m \mu_r \mu_l \mu_n + \mu_l \mu_n \sigma_{rm} + \mu_r \mu_n \sigma_{lm} + \mu_r \mu_l \sigma_{nm} + \mu_m \mu_n \sigma_{lr} + \sigma_{nm} \sigma_{lr}$$

$$+ \mu_m \mu_l \sigma_{nr} + \sigma_{lm} \sigma_{nr} + \mu_m \mu_r \sigma_{nl} + \sigma_{rm} \sigma_{nl} .$$

Moments about the mean

$$E[X_n - \mu_n] = \left.\frac{\partial M}{\partial t_n}\right|_{\underline{t}=0} = \frac{\partial}{\partial t_n}\left\{ e^{\frac{1}{2}\underline{t}'\Sigma \underline{t}} \right\}\Bigg|_{\underline{t}=\underline{0}} , \; E e^{\underline{t}'(\underline{X}-\underline{\mu})} = e^{\frac{1}{2}\underline{t}'\Sigma \underline{t}}$$

$$= \frac{\partial}{\partial t_n}\left(\exp \frac{1}{2} \sum_{k=1}^{p} \sum_{j=1}^{p} t_k\, t_j\, \sigma_{kj} \right)\Bigg|_{\underline{t}=\underline{0}} = M\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right)\Bigg|_{\underline{t}=\underline{0}} = 0 .$$

$$E[X_n - \mu_n][X_l - \mu_l] = \left.\frac{\partial^2 M}{\partial t_n\, \partial t_l}\right|_{\underline{t}=0} = \frac{\partial}{\partial t_n}\left\{ \frac{\partial M}{\partial t_l} \right\}\Bigg|_{\underline{t}=\underline{0}} = \frac{\partial}{\partial t_l}\left[M\left\{ \sum_{j=1}^{p} t_j\, \sigma_{nj} \right\} \right]_{\underline{t}=\underline{0}}$$

$$= \left\{ M\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right) + M\, \sigma_{nl} \right\}_{\underline{t}=\underline{0}} = \sigma_{nl} .$$

$$E[X_n - \mu_n]\,[X_l - \mu_l]\,[X_r - \mu_r] = \left.\frac{\partial^3 M}{\partial t_n\, \partial t_l\, \partial t_r}\right|_{\underline{t}=0} = \frac{\partial}{\partial t_r}\left\{ \frac{\partial^2 M}{\partial t_n\, \partial t_l} \right\}\Bigg|_{\underline{t}=\underline{0}}$$

$$= \frac{\partial}{\partial t_r}\left\{ M\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right) + M\, \sigma_{nl} \right\}_{\underline{t}=\underline{0}}$$

$$= \Bigg[M\left(\sum_{j=1}^{p} t_j\, \sigma_{rj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right)$$

$$+ M\left\{ \sigma_{lr}\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right) + \sigma_{nr}\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right) \right\} + M\left(\sum_{j=1}^{p} t_j\, \sigma_{rj} \right)\sigma_{nl} \Bigg]_{\underline{t}=\underline{0}} = 0 .$$

$$E[X_n - \mu_n]\,[X_l - \mu_l]\,[X_r - \mu_r]\,[X_m - \mu_m] = \left.\frac{\partial^4 M}{\partial t_n\, \partial t_l\, \partial t_r\, \partial t_m}\right|_{\underline{t}=0}$$

$$= \frac{\partial}{\partial t_m}\Bigg[M\left(\sum_{j=1}^{p} t_j\, \sigma_{rj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right) + M\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right)\sigma_{lr}$$

$$+ M\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right)\sigma_{nr} + M\left(\sum_{j=1}^{p} t_j\, \sigma_{rj} \right)\sigma_{nl} \Bigg]_{\underline{t}=\underline{0}}$$

$$= \Bigg[M\left(\sum_{j=1}^{p} t_j\, \sigma_{mj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{rj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right)$$

$$+ M\Bigg\{ \sigma_{rm}\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right) + \sigma_{lm}\left(\sum_{j=1}^{p} t_j\, \sigma_{rj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right)$$

$$+ \sigma_{nm}\left(\sum_{j=1}^{p} t_j\, \sigma_{rj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right) \Bigg\} + M\left(\sum_{j=1}^{p} t_j\, \sigma_{mj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{nj} \right)\sigma_{lr} + M\, \sigma_{nm}\, \sigma_{lr}$$

$$+ M\left(\sum_{j=1}^{p} t_j\, \sigma_{mj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{lj} \right)\sigma_{nr} + M\, \sigma_{lm}\, \sigma_{nr}$$

$$+ M\left(\sum_{j=1}^{p} t_j\, \sigma_{mj} \right)\left(\sum_{j=1}^{p} t_j\, \sigma_{rj} \right)\sigma_{nl} + M\, \sigma_{rm}\, \sigma_{nl} \Bigg]_{\underline{t}=\underline{0}}$$

$$= \sigma_{nm} \sigma_{lr} + \sigma_{lm} \sigma_{nr} + \sigma_{rm} \sigma_{nl} .$$

ESTIMATION OF PARAMETERS IN MULTIVARIATE NORMAL DISTRIBUTION

The normal distribution is completely specified if its mean vector $\underline{\mu}$ and covariance matrix Σ are known. In case of unknown parameters, the problems of their estimation arise. We estimate these parameters by the method of maximum likelihood estimation. For using this method we require a random sample of size n from the given $p-$variate normal population.

Let the random sample of size n from $N_p(\underline{\mu}, \Sigma)$ be $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_\alpha, \cdots, \underline{x}_n$, where $n > p$ and $\underline{x}_\alpha$ is $p \times 1$ vector, $\alpha = 1, 2, \cdots, n$. In extended vector notation the data are as follows:

Suppose we have n individuals $1, 2, \cdots, n$ and p characteristics $X_1, X_2, \cdots, X_p$ and each individuals studied.

Characteristic	Individual	Mean
	$1 \quad 2 \quad \cdots \quad \alpha \quad \cdots \quad n$	
X_1	$x_{11} \quad x_{12} \quad \cdots \quad x_{1\alpha} \quad \cdots \quad x_{1n}$	$\bar{x}_1$
X_2	$x_{21} \quad x_{22} \quad \cdots \quad x_{2\alpha} \quad \cdots \quad x_{2n}$	$\bar{x}_2$
$\vdots$	$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$	$\vdots$
X_i	$x_{i1} \quad x_{i2} \quad \cdots \quad x_{i\alpha} \quad \cdots \quad x_{in}$	$\bar{x}_i$
$\vdots$	$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$	$\vdots$
X_p	$x_{p1} \quad x_{p2} \quad \cdots \quad x_{p\alpha} \quad \cdots \quad x_{pn}$	$\bar{x}_p$
	$\underline{x}_1 \quad \underline{x}_2 \quad \cdots \quad \underline{x}_\alpha \quad \cdots \quad \underline{x}_n$	

Corresponding to $\alpha-$th individual there is a vector $\underline{x}_\alpha$ which is representing a point in $p-$dimensional space. So all these n point in E^n. Therefore,

$$\underline{\bar{x}} = \frac{1}{n}\sum_{\alpha=1}^{n} \underline{x}_\alpha = \begin{pmatrix} \frac{1}{n}\sum_{\alpha=1}^{n} x_{1\alpha} \\ \vdots \\ \frac{1}{n}\sum_{\alpha=1}^{n} x_{i\alpha} \\ \vdots \\ \frac{1}{n}\sum_{\alpha=1}^{n} x_{p\alpha} \end{pmatrix} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_i \\ \vdots \\ \bar{x}_p \end{pmatrix},$$ is the sample mean vector, and sample variance covariance matrix is

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix},$$

where $s_{ij} = \frac{1}{n-1}\sum_{\alpha=1}^{n}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)$, $\forall\, i$ and j.

Also

$$S = \frac{1}{n-1}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} = \frac{A}{n-1}.$$

The matrix A is called sum of square and cross products of deviations about the mean or $(SS\,\&\,CP)$.

where

$$a_{ij} = \sum_{\alpha=1}^{n}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j), \text{ for all } i \text{ and } j$$

$$= \sum_{\alpha=1}^{n}[x_{i\alpha}(x_{j\alpha} - \bar{x}_j) - \bar{x}_i(x_{j\alpha} - \bar{x}_j)] = \sum_{\alpha=1}^{n} x_{i\alpha}x_{j\alpha} - \bar{x}_j\sum_{\alpha=1}^{n} x_{i\alpha} - \bar{x}_i\sum_{\alpha=1}^{n} x_{j\alpha} + n\bar{x}_i\bar{x}_j$$

$$= \sum_{\alpha=1}^{n} x_{i\alpha}x_{j\alpha} - n\bar{x}_i\bar{x}_j.$$

Some results

1) Consider a quadratic form

$$Q = \underline{x}' A\, \underline{x} = \sum_{i,j}^{p} a_{ij}x_i x_j, \text{ where } A \text{ is symmetric, then}$$

$$\frac{\partial Q}{\partial \underline{x}} = \begin{pmatrix} \frac{\partial Q}{\partial x_1} \\ \frac{\partial Q}{\partial x_2} \\ \vdots \\ \frac{\partial Q}{\partial x_p} \end{pmatrix} = 2A\underline{x}.$$

In particular, for $p = 2$

$$Q = (x_1 \quad x_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (a_{11}x_1 + a_{21}x_2 \quad a_{12}x_1 + a_{22}x_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= (a_{11}x_1^2 + a_{21}x_1x_2 + a_{12}x_1x_2 + a_{22}x_2^2), \text{ then}$$

$$\frac{\partial Q}{\partial \underline{x}} = \begin{pmatrix} \frac{\partial Q}{\partial x_1} \\ \frac{\partial Q}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2a_{11}x_1 + a_{21}x_2 + a_{12}x_2 \\ a_{21}x_1 + a_{12}x_1 + 2a_{22}x_2 \end{pmatrix} = 2\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2A\underline{x}.$$

2) If $B = \underline{y}' A\,\underline{x} = \underline{x}' A' \underline{y}$, then $\frac{\partial B}{\partial \underline{y}} = A\,\underline{x}$, and $\frac{\partial B}{\partial \underline{x}} = A'\,\underline{y}$

3) If $Q=(\underline{x}-\underline{b})' A(\underline{x}-\underline{b})=(\underline{b}-\underline{x})' A(\underline{b}-\underline{x})$, then

$$\frac{\partial Q}{\partial \underline{x}}=2A(\underline{x}-\underline{b}), \text{ and } \frac{\partial Q}{\partial \underline{b}}=2A(\underline{b}-\underline{x}).$$

4) A **sub matrix** of A is a rectangular array obtained from A by deleting rows and columns. A **minor** is the determinant of square sub matrix of A.

$$|A|=\sum_{j=1}^{p} a_{ij} A_{ij}=\sum_{j=1}^{p} a_{jk} A_{jk},$$

where A_{ij}, is $(-1)^{i+j}$ times the minor of a_{ij}, and the minor of an element a_{ij} is the determinant of the sub matrix of a square matrix A obtained by deleting the $i-$th row and $j-$th column.

If $|A|\neq 0$, there exist a unique matrix B such that $AB=I$, B is called the inverse of A and is denoted by A^{-1}.

Let a^{ij} be the element of A^{-1} in the $i-$th row and $j-$th column, then

$$a^{ij}=\frac{A_{ji}}{|A|} \text{ and } a_{ij}=\frac{A^{ji}}{|A^{-1}|}.$$

Example: Let $A=\begin{pmatrix}2 & 4\\ 3 & 7\end{pmatrix}$, $|A|=14-12=2$, and $A^{-1}=\begin{pmatrix}7/2 & -2\\ -3/2 & 1\end{pmatrix}$, $|A^{-1}|=\frac{7}{2}-\frac{6}{2}=\frac{1}{2}$.

$$a^{11}=\frac{A_{11}}{|A|}=\frac{7}{2},\ a^{12}=\frac{A_{21}}{|A|}=\frac{-4}{2}=-2,\ a^{21}=\frac{A_{12}}{|A|}=\frac{-3}{2},\ a^{22}=\frac{A_{22}}{|A|}=\frac{2}{2}=1.$$

Also

$$a_{11}=\frac{A^{11}}{|A^{-1}|}=\frac{1}{1/2}=2, \qquad a_{12}=\frac{A^{21}}{|A^{-1}|}=\frac{2}{1/2}=4, \qquad a_{21}=\frac{A^{12}}{|A^{-1}|}=\frac{3/2}{1/2}=3,$$

$$a_{22}=\frac{A^{22}}{|A^{-1}|}=\frac{7/2}{1/2}=7.$$

5) A square matrix $A_{n\times n}$ is said to be orthogonal if $A'A=AA'=I_n$, and if the transformation $\underline{Y}=A\underline{X}$ transform $\underline{X}'\underline{X}$ to $\underline{Y}'\underline{Y}$. Also

$$\sum_{k=1}^{n} a_{ik}\times\frac{1}{\sqrt{n}}=0, \Rightarrow \sum_{k=1}^{n} a_{ik}=0,\ i=1,2,\cdots,n-1$$

and

$$\sum_{k=1}^{n} a_{ik}\times a_{jk}=\begin{cases}0, & i\neq j\\ 1, & i=j\end{cases} \text{ for } i,j=1,2,\cdots,n.$$

Example: Let orthogonal matrix of order 2

$$A=\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}, \text{ then } A'A=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}\times\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}=I$$

Here $n=2$

$$\sum_{k=1}^{2} a_{ik}=a_{11}+a_{12}=\frac{1}{\sqrt{2}}-\frac{1}{\sqrt{2}}=0,$$

and

$$\sum_{k=1}^{2} a_{ik}\times a_{jk}=a_{11}a_{21}+a_{12}a_{22}=\frac{1}{2}-\frac{1}{2}=0$$

$$\sum_{k=1}^{2} a_{ik}\times a_{jk}=a_{11}a_{11}+a_{12}a_{12}=\frac{1}{2}+\frac{1}{2}=1$$

$$\sum_{k=1}^{2} a_{ik}\times a_{jk}=a_{21}a_{21}+a_{22}a_{22}=\frac{1}{2}+\frac{1}{2}=1$$

Another example of orthogonal matrix of order 3

$$A=\begin{pmatrix}\frac{1}{\sqrt{2\times 1}} & -\frac{1}{\sqrt{2\times 1}} & \frac{1\times 0}{\sqrt{2\times 1}}\\ \frac{1}{\sqrt{3\times 2}} & \frac{1}{\sqrt{3\times 2}} & \frac{-2}{\sqrt{3\times 2}}\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\end{pmatrix}$$

$$A=\begin{pmatrix}\frac{1}{\sqrt{2\times 1}} & -\frac{1}{\sqrt{2\times 1}} & \frac{1\times 0}{\sqrt{2\times 1}} & 0 & \cdots & 0\\ \frac{1}{\sqrt{3\times 2}} & \frac{1}{\sqrt{3\times 2}} & \frac{-2}{\sqrt{3\times 2}} & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & & \frac{-(n-1)}{\sqrt{n(n-1)}}\\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & & \frac{1}{\sqrt{n}}\end{pmatrix}.$$

Maximum likelihood estimate of the mean vector

Let $\underline{x}_1,\underline{x}_2,\cdots,\underline{x}_\alpha,\cdots,\underline{x}_n$ be a random sample of size $n(>p)$ from $N_p(\underline{\mu},\Sigma)$.

The likelihood function is

$$\phi=f(\underline{x}_1)f(\underline{x}_2)\cdots f(\underline{x}_n)=\frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{n}(\underline{x}_\alpha-\underline{\mu})'\Sigma^{-1}(\underline{x}_\alpha-\underline{\mu})\right]$$

$$\log\phi = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{\alpha=1}^{n}(\underline{x}_\alpha - \underline{\mu})'\Sigma^{-1}(\underline{x}_\alpha - \underline{\mu}).$$

Differentiating with respect to $\underline{\mu}$ and equating to zero

$$\frac{\partial\log\phi}{\partial\underline{\mu}} = \underline{0} = -0 - 0 - \frac{1}{2}\sum_{\alpha=1}^{n}2\Sigma^{-1}(\underline{\mu} - \underline{x}_\alpha) \quad \text{or} \quad \Sigma^{-1}\sum_{\alpha=1}^{n}(\underline{\mu} - \underline{x}_\alpha) = 0$$

or $\hat{\underline{\mu}} = \frac{1}{n}\sum_{\alpha=1}^{n}\underline{x}_\alpha = \underline{\bar{x}}$.

Maximum likelihood estimate of variance covariance matrix

Let $\Sigma^{-1} = (\sigma^{ij})$ and $\Sigma = (\sigma_{ij})$ then the $|\Sigma^{-1}| = \sigma^{i1}\Sigma^{i1} + \cdots + \sigma^{ip}\Sigma^{ip}$, where Σ^{ij} is the cofactor of σ^{ij} in Σ^{-1}, therefore, $\frac{\Sigma^{ij}}{|\Sigma^{-1}|} = (i\,j)^{th}$ element of $(\Sigma^{-1})^{-1} = (i\,j)^{th}$ element of $\Sigma = \sigma_{ij}$.

Now, the logarithm of the likelihood function is

$$\log\phi = -\frac{np}{2}\log(2\pi) + \frac{n}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_{\alpha=1}^{n}(\underline{x}_\alpha - \underline{\mu})'\Sigma^{-1}(\underline{x}_\alpha - \underline{\mu})$$

$$= -\frac{np}{2}\log(2\pi) + \frac{n}{2}\log(\sigma^{i1}\Sigma^{i1} + \cdots + \sigma^{ij}\Sigma^{ij} + \cdots + \sigma^{ip}\Sigma^{ip})$$

$$-\frac{1}{2}\sum_{\alpha}\sum_{i,j}\sigma^{ij}(x_{i\alpha} - \mu_i)(x_{j\alpha} - \mu_j), \text{ since } \underline{x}'A\underline{x} = \sum_{i,j}a_{ij}x_i x_j$$

Differentiating with respect to σ^{ij} and equating to zero, we get

$$\frac{\partial\log\phi}{\partial\sigma^{ij}} = \underline{0} = \frac{n}{2}\frac{\Sigma^{ij}}{|\Sigma^{-1}|} - \frac{1}{2}\sum_{\alpha=1}^{n}(x_{i\alpha} - \mu_i)(x_{j\alpha} - \mu_j), \text{ because } \frac{\partial}{\partial x}\log f(x) = \frac{f'(x)}{f(x)}$$

or $\frac{n}{2}\sigma_{ij} = \frac{1}{2}\sum_{\alpha=1}^{n}(x_{i\alpha} - \mu_i)(x_{j\alpha} - \mu_j)$ or $\hat{\sigma}_{ij} = \frac{1}{n}\sum_{\alpha=1}^{n}(x_{i\alpha} - \hat{\mu}_i)(x_{j\alpha} - \hat{\mu}_j)$

Hence, $\hat{\Sigma} = \frac{A}{n}$.

Theorem: Given $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_\alpha, \cdots, \underline{x}_n$ be an independent random sample of size $n(>p)$ from $N_p(\underline{\mu}, \Sigma)$, then $\underline{\bar{x}} \sim N(\underline{\mu}, \Sigma/n)$.

Proof: $E(\underline{\bar{x}}) = E\left(\frac{1}{n}\sum_{\alpha}\underline{x}_\alpha\right) = \frac{1}{n}E(\underline{x}_1 + \cdots + \underline{x}_n) = \frac{1}{n}(n\underline{\mu}) = \underline{\mu}$ and

$$\Sigma_{\underline{\bar{x}}} = E(\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})' = E\left[\frac{1}{n}(\underline{x}_1 + \cdots + \underline{x}_n) - \underline{\mu}\right]\left[\frac{1}{n}(\underline{x}_1 + \cdots + \underline{x}_n) - \underline{\mu}\right]'$$

$$= \frac{1}{n^2}E(\underline{x}_1 + \cdots + \underline{x}_n - n\underline{\mu})(\underline{x}_1 + \cdots + \underline{x}_n - n\underline{\mu})'$$

$$= \frac{1}{n^2}[E(\underline{x}_1 - \underline{\mu})(\underline{x}_1 - \underline{\mu})' + \cdots + E(\underline{x}_n - \underline{\mu})(\underline{x}_n - \underline{\mu})' + 0], \text{ as } \underline{x}_\alpha\text{'s are independent.}$$

$$= \frac{1}{n^2}(n\Sigma) = \Sigma/n.$$

Thus,

$$\underline{\bar{x}} \sim N(\underline{\mu}, \Sigma/n).$$

Theorem: $\frac{A}{n-1}$ is an unbiased estimate of Σ.

Proof: We have

$$A = \sum_{\alpha}(\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = \sum_{\alpha}[(\underline{x}_\alpha - \underline{\mu}) - (\underline{\bar{x}} - \underline{\mu})][(\underline{x}_\alpha - \underline{\mu}) - (\underline{\bar{x}} - \underline{\mu})]'$$

$$= \sum_{\alpha}[(\underline{x}_\alpha - \underline{\mu})(\underline{x}_\alpha - \underline{\mu})' - (\underline{\bar{x}} - \underline{\mu})(\underline{x}_\alpha - \underline{\mu})' - (\underline{x}_\alpha - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})' + (\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})']$$

$$= \sum_{\alpha}(\underline{x}_\alpha - \underline{\mu})(\underline{x}_\alpha - \underline{\mu})' - n(\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})'.$$

Taking expectation on both the sides

$$E(A) = \sum_{\alpha}E(\underline{x}_\alpha - \underline{\mu})(\underline{x}_\alpha - \underline{\mu})' - nE(\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})' = n\Sigma - n\Sigma/n = (n-1)\Sigma$$

or $E\left(\frac{A}{n-1}\right) = \Sigma$, this shows that the maximum likelihood estimate of Σ is biased, i.e. $E\left(\frac{A}{n}\right) = \frac{n-1}{n}\Sigma$.

Theorem: Given a random sample $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_\alpha, \cdots, \underline{x}_n$ from $N_p(\underline{\mu}, \Sigma)$, $\underline{\bar{x}} = \frac{1}{n}\sum_{\alpha}\underline{x}_\alpha$ and $A = \sum_{\alpha}(\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})'$. Thus $\underline{\bar{x}}$ and A are independently distributed.

Proof: Make an orthogonal transformation

$$\begin{aligned}
\underline{y}_1 &= c_{11}\underline{x}_1 + \cdots + c_{1n}\underline{x}_n \\
&\vdots \\
\underline{y}_{n-1} &= c_{n-11}\underline{x}_1 + \cdots + c_{n-1n}\underline{x}_n \\
\underline{y}_n &= \frac{1}{\sqrt{n}}\underline{x}_1 + \cdots + \frac{1}{\sqrt{n}}\underline{x}_n
\end{aligned}$$

or $\underset{\rightarrow}{Y} = C \underset{\rightarrow}{X}$, where $\underset{\rightarrow}{Y} = \begin{pmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_n \end{pmatrix}$, $\underset{\rightarrow}{X} = \begin{pmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \end{pmatrix}$, and

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{n-11} & c_{n-12} & \cdots & c_{n-1n} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix} \text{ is orthogonal.}$$

Since C is orthogonal

$$\sum_{k=1}^{n} c_{ik} \frac{1}{\sqrt{n}} = 0, \quad i = 1, 2, \cdots, n-1 \quad \Rightarrow \sum_{k=1}^{n} c_{ik} = 0$$

Also

$$\sum_{k=1}^{n} c_{ik} \times c_{jk} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \text{ for } i, j = 1, 2, \cdots, n.$$

Now consider,

$$\underline{y}_n = \frac{1}{\sqrt{n}} (\underline{x}_1 + \cdots + \underline{x}_n) = \sqrt{n}\, \underline{\bar{x}} \tag{3.1}$$

So that

$$E\, \underline{y}_n = E \sqrt{n}\, \underline{\bar{x}} = \sqrt{n}\, \underline{\mu}, \text{ and}$$

$$E\, \underline{y}_i = E \sum_{k=1}^{n} c_{ik}\, \underline{x}_k = \underline{\mu} \sum_{k=1}^{n} c_{ik} = \underline{0}, \text{ for } i = 1, 2, \cdots, n-1$$

$$\begin{aligned} \operatorname{Cov}(\underline{y}_i, \underline{y}_j) &= E(\underline{y}_i - E\,\underline{y}_i)(\underline{y}_j - E\,\underline{y}_j)' \\ &= E[c_{i1}(\underline{x}_1 - \underline{\mu}) + \cdots + c_{in}(\underline{x}_n - \underline{\mu})][c_{j1}(\underline{x}_1 - \underline{\mu}) + \cdots + c_{jn}(\underline{x}_n - \underline{\mu})]' \\ &= c_{i1}\, c_{j1}\, E(\underline{x}_1 - \underline{\mu})(\underline{x}_1 - \underline{\mu})' + \cdots + c_{in}\, c_{jn}\, E(\underline{x}_n - \underline{\mu})(\underline{x}_n - \underline{\mu})' + 0 \end{aligned}$$

since $\underline{x}_1, \cdots, \underline{x}_n$ are independent

$$= (c_{i1}\, c_{j1} + \cdots + c_{in}\, c_{jn}) \Sigma = \Sigma \sum_{k=1}^{n} c_{ik}\, c_{jk} = \begin{cases} 0, & if \quad i \neq j \\ \Sigma, & if \quad i = j \end{cases}$$

This implies $\underline{y}_1, \cdots, \underline{y}_n$ are independent.

Now

$$\sum_{i=1}^{n} \underline{y}_i\, \underline{y}_i' = \sum_{i=1}^{n} (c_{i1}\, \underline{x}_1 + \cdots + c_{in}\, \underline{x}_n)(c_{i1}\, \underline{x}_1 + \cdots + c_{in}\, \underline{x}_n)'$$

$$= \underline{x}_1 \underline{x}_1' \sum_{i=1}^{n} c_{i1}^2 + \cdots + \underline{x}_1 \underline{x}_n' \sum_{i=1}^{n} c_{i1}\, c_{in} + \cdots + \underline{x}_n \underline{x}_1' \sum_{i=1}^{n} c_{in}\, c_{i1} + \cdots + \underline{x}_n \underline{x}_n' \sum_{i=1}^{n} c_{in}^2$$

$$= \underline{x}_1 \underline{x}_1' + \cdots + \underline{x}_n \underline{x}_n' = \sum_{i=1}^{n} \underline{x}_i \underline{x}_i'.$$

Therefore,

$$\sum_{i=1}^{n} \underline{y}_i \underline{y}_i' - \underline{y}_n \underline{y}_n' = \sum_{i=1}^{n} \underline{x}_i \underline{x}_i' - \underline{y}_n \underline{y}_n'$$

or
$$\sum_{i=1}^{n-1} \underline{y}_i \underline{y}_i' = \sum_{i=1}^{n} \underline{x}_i \underline{x}_i' - n \underline{\bar{x}}\, \underline{\bar{x}}' = A \tag{3.2}$$

In view of equations (3.1) and (3.2), $\underline{\bar{x}}$ and A depends on two mutually exclusive sets, which are independently distributed, therefore, $\underline{\bar{x}}$ and A are independently distributed.

Test for $\underline{\mu}$, when Σ is known

Given a random sample $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_\alpha, \cdots, \underline{x}_n$ from $N_p(\underline{\mu}, \Sigma)$. The hypothesis of interest is $H_0: \underline{\mu} = \underline{\mu}_0$, where $\underline{\mu}_0$ is a specified vector, then, under H_0, the test statistic is $n(\underline{\bar{x}} - \underline{\mu}_0)' \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu}_0) \sim \chi_p^2$.

Proof: Let C be a non-singular matrix such that

$C' \Sigma^{*-1} C = I$, and $CC' = \Sigma^* = \dfrac{\Sigma}{n}$. Make the transformation

$(\underline{\bar{x}} - \underline{\mu}_0) = C\, \underline{y} \quad \Rightarrow \underline{y} = C^{-1}(\underline{\bar{x}} - \underline{\mu}_0)$, and

$E\, \underline{y} = C^{-1} E(\underline{\bar{x}} - \underline{\mu}_0) = C^{-1}(\underline{\mu}_0 - \underline{\mu}_0) = \underline{0}$, under H_0.

$$\begin{aligned} \Sigma_{\underline{y}} &= E(\underline{y} - E\,\underline{y})(\underline{y} - E\,\underline{y})' = C^{-1} E(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)' C^{-1'} \\ &= C^{-1} \Sigma^* C^{-1'} = (C' \Sigma^{*-1} C)^{-1} = I. \end{aligned}$$

Therefore,

$\underline{y} \sim N_p(\underline{0}, I)$, i.e. $y_i \sim N(0,1)$, for all $i = 1, 2, \cdots, p$.

Now

$$\begin{aligned} n(\underline{\bar{x}} - \underline{\mu}_0)' \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu}_0) &= (\underline{\bar{x}} - \underline{\mu}_0)' (CC')^{-1} (\underline{\bar{x}} - \underline{\mu}_0) \\ &= [C^{-1}(\underline{\bar{x}} - \underline{\mu}_0)]' [C^{-1}(\underline{\bar{x}} - \underline{\mu}_0)] = \underline{y}' \underline{y} = \sum_{i=1}^{p} y_i^2 \sim \chi_p^2. \end{aligned}$$

Let $\chi_p^2(\alpha)$ be the number such that $\Pr[\chi_p^2 \geq \chi_p^2(\alpha)] = \alpha$, then for testing H_0, we use the critical region

$$n(\underline{\bar{x}} - \underline{\mu}_0)' \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu}_0) \geq \chi_p^2(\alpha).$$

For computational purpose use

$(\underline{\bar{x}} - \underline{\mu}_0) = \underline{d}$, then solve

$\Sigma \underline{\lambda} = \underline{d}$ (by Doolittle method)

$\Rightarrow \underline{\lambda} = \Sigma^{-1} \underline{d}$ and

$n \underline{d}' \underline{\lambda} = n(\underline{\bar{x}} - \underline{\mu}_0)' \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu}_0)$.

Note: If H_0 is not true, then

$E\,\underline{y} = C^{-1} E(\underline{\bar{x}} - \underline{\mu}_0) = C^{-1}(\underline{\mu} - \underline{\mu}_0) = \underline{\delta}$ (say), $\Sigma_{\underline{y}} = I$, and

$n(\underline{\bar{x}} - \underline{\mu}_0)' \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu}_0) = \sum_{i=1}^{p} y_i^2 \sim \chi^2_{p, \sum_i \delta_i^2}$, where

$\sum_{i=1}^{p} \delta_i^2 = \underline{\delta}\,\underline{\delta}' = n(\underline{\mu} - \underline{\mu}_0)' \Sigma^{-1} (\underline{\mu} - \underline{\mu}_0) =$ non-centrality parameter.

The confidence region for $\underline{\mu}$ is the set of possible values of $\underline{\mu}$ satisfying

$n(\underline{\bar{x}} - \underline{\mu})' \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu}) \le \chi_p^2(\alpha)$, this has confidence coefficient $1-\alpha$.

Two sample problem

Given $\underline{x}_1^{(1)}, \underline{x}_2^{(1)}, \cdots, \underline{x}_\alpha^{(1)}, \cdots, \underline{x}_{n_1}^{(1)}$ be a random sample from $N_p(\underline{\mu}^{(1)}, \Sigma)$ and $\underline{x}_1^{(2)}, \underline{x}_2^{(2)}, \cdots, \underline{x}_\alpha^{(2)}, \cdots, \underline{x}_{n_2}^{(2)}$ from $N_p(\underline{\mu}^{(2)}, \Sigma)$. $H_0: \underline{\mu}^{(1)} = \underline{\mu}^{(2)}$, then, in this case, the statistic is

$$\frac{n_1 n_2}{n_1 + n_2} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' \Sigma^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) \sim \chi_p^2 .$$

Proof: We know that

$\underline{\bar{x}}^{(1)} \sim N_p(\underline{\mu}^{(1)}, \Sigma / n_1)$, and $\underline{\bar{x}}^{(2)} \sim N_p(\underline{\mu}^{(2)}, \Sigma / n_2)$. Further, $\underline{\bar{x}}^{(1)}$ and $\underline{\bar{x}}^{(2)}$ are independent and

$$\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)} \sim N_p\left[\underline{\mu}^{(1)} - \underline{\mu}^{(2)}, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\Sigma\right].$$

Make the transformation (nonsingular0

$(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) = C\,\underline{y} \quad \Rightarrow \underline{y} = C^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})$

Since C is a non-singular matrix such that

$C' \Sigma^{*-1} C = I$, where $\Sigma^* = \dfrac{n_1 + n_2}{n_1 n_2} \Sigma = CC'$, and

$E\,\underline{y} = C^{-1} E(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) = \underline{0}$, under H_0.

$\Sigma_{\underline{y}} = E(\underline{y} - E\,\underline{y})(\underline{y} - E\,\underline{y})' = C^{-1} E(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' C^{-1'}$

$= C^{-1} \Sigma^* C^{-1'} = (C' \Sigma^{*-1} C)^{-1} = I$

Therefore,

$\underline{y} \sim N_p(\underline{0}, I)$, i.e. $y_i \sim N(0,1)$, for all $i = 1, 2, \cdots, p$.

Now

$$\frac{n_1 n_2}{n_1 + n_2} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' \Sigma^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) = (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' (CC')^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})$$

$$= [C^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})]' [C^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})] = \underline{y}'\,\underline{y} = \sum_{i=1}^{p} y_i^2 \sim \chi_p^2 .$$

Let $\chi_p^2(\alpha)$ be the number such that $\Pr[\chi_p^2 \ge \chi_p^2(\alpha)] = \alpha$

Then for testing H_0, we use the critical region

$$\frac{n_1 n_2}{n_1 + n_2} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' \Sigma^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) \ge \chi_p^2(\alpha) .$$

Note

i) If H_0 is not true, then

$E\,\underline{y} = C^{-1} E(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) = C^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) = \underline{\delta}$ (say), and $\Sigma_{\underline{y}} = I$

So each $y_i \sim N(\delta_i, 1)$, then

$$(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' \Sigma^{*-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) = \sum_{i=1}^{p} y_i^2 \sim \chi^2_{p, \sum_i \delta_i^2} ,$$

where

$$\sum_{i=1}^{p} \delta_i^2 = \underline{\delta}'\,\underline{\delta} = (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})' C^{-1'} C^{-1} (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$$

$$= \frac{n_1 n_2}{n_1 + n_2} (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})' \Sigma^{-1} (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$$

$$= \frac{n_1 n_2}{n_1 + n_2} \Delta = \text{non-centrality parameter.}$$

where Δ is the measure of distance between two populations defined by **Mahalanobis**.

The confidence region for $\underline{\upsilon} = \underline{\mu}^{(1)} - \underline{\mu}^{(2)}$ is the set of possible values of $\underline{\upsilon}$ satisfying the inequality

$$\frac{n_1 n_2}{n_1 + n_2} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)} - \underline{\upsilon})' \Sigma^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)} - \underline{\upsilon}) \le \chi_p^2(\alpha) .$$

ii) If the two populations have known covariance matrix Σ_1 and Σ_2, then

$$(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' \Sigma^{*-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) \sim \chi_p^2 \text{, where } \Sigma^* = \frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}.$$

Result:

If $\underline{t}$ is a sufficient statistic for $\underline{\theta}$ if $\prod_{\alpha=1}^{n} f(\underline{x}_\alpha; \underline{\theta}) = g(\underline{t}; \underline{\theta})\, h(\underline{x}_1, \cdots, \underline{x}_n)$, where $f(\underline{x}_\alpha; \underline{\theta})$ is the density of the $\alpha-$th observation, $g(\underline{t}; \underline{\theta})$ is the density of $\underline{t}$ and $h(\underline{x}_1, \cdots, \underline{x}_n)$ does not depend on $\underline{\theta}$.

Sufficient Statistics for $\underline{\mu}$ and Σ

The joint density function of $\underline{x}_1, \cdots, \underline{x}_n$ with $\underline{x}_\alpha \sim N_p(\underline{\mu}, \Sigma)$ is

$$\frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2} \sum_{\alpha=1}^{n} (\underline{x}_\alpha - \underline{\mu})' \Sigma^{-1} (\underline{x}_\alpha - \underline{\mu})\right].$$

Consider,

$$\sum_{\alpha=1}^{n} (\underline{x}_\alpha - \underline{\mu})' \Sigma^{-1} (\underline{x}_\alpha - \underline{\mu}) = tr \sum_{\alpha=1}^{n} (\underline{x}_\alpha - \underline{\mu})' \Sigma^{-1} (\underline{x}_\alpha - \underline{\mu}) = tr \sum_{\alpha=1}^{n} \Sigma^{-1} (\underline{x}_\alpha - \underline{\mu})' (\underline{x}_\alpha - \underline{\mu})$$

$$= tr \Sigma^{-1} \sum_{\alpha=1}^{n} (\underline{x}_\alpha - \underline{\mu})' (\underline{x}_\alpha - \underline{\mu}).$$

We can write

$$\sum_{\alpha=1}^{n} (\underline{x}_\alpha - \underline{\mu})' (\underline{x}_\alpha - \underline{\mu}) = \sum_{\alpha=1}^{n} [(\underline{x}_\alpha - \underline{\bar{x}}) + (\underline{\bar{x}} - \underline{\mu})] [(\underline{x}_\alpha - \underline{\bar{x}}) + (\underline{\bar{x}} - \underline{\mu})]'$$

$$= \sum_{\alpha=1}^{n} [(\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' + (\underline{x}_\alpha - \underline{\bar{x}})(\underline{\bar{x}} - \underline{\mu})' + (\underline{\bar{x}} - \underline{\mu})(\underline{x}_\alpha - \underline{\bar{x}})' + (\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})']$$

$$= \sum_{\alpha=1}^{n} (\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' + \left\{\sum_{\alpha} (\underline{x}_\alpha - \underline{\bar{x}})\right\} (\underline{\bar{x}} - \underline{\mu})' + (\underline{\bar{x}} - \underline{\mu}) \sum_{\alpha} (\underline{x}_\alpha - \underline{\bar{x}})'$$

$$+ n(\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})'$$

$$= A + n(\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})', \text{ because } \sum_{\alpha} (\underline{x}_\alpha - \underline{\bar{x}}) = \sum_{\alpha} \underline{x}_\alpha - n\,\underline{\bar{x}} = \underline{0}.$$

Thus the density of $\underline{x}_1, \cdots, \underline{x}_n$ can be written as

$$\frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2} tr \Sigma^{-1} [A + n(\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})']\right]$$

$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2} [n(\underline{\bar{x}} - \underline{\mu})' \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu}) + tr \Sigma^{-1} A]\right]$$

$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2} n(\underline{\bar{x}} - \underline{\mu})' \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu})\right]$$

$$\times \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2} tr \Sigma^{-1} A\right].$$

Thus, $\underline{\bar{x}}$ and $\frac{1}{n} A$ form a sufficient set of statistics for $\underline{\mu}$ and Σ. If Σ is known, $\underline{\bar{x}}$ is a sufficient statistic for $\underline{\mu}$. However, if $\underline{\mu}$ is known $\frac{1}{n} A$ is not a sufficient statistic for Σ, but $\frac{1}{n} \sum_{\alpha} (\underline{x}_\alpha - \underline{\mu})(\underline{x}_\alpha - \underline{\mu})'$ is a sufficient statistic for Σ.

WISHART DISTRIBUTION

If $x_1, x_2, \cdots, x_n$ are independent observations from $N(\mu, \sigma^2)$, it is well known $(n-1)\,s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \sigma^2\,\chi^2_{n-1}$. The multivariate analogue of $(n-1)\,s^2$ is the matrix A and is called Wishart matrix. In other words the Wishart matrix is defined as the $p \times p$ symmetric matrix of sums of squares and cross products (of deviations about the mean) of the sample observations, from a $p-$variate nonsingular normal distribution. The distribution of A when the multivariate distribution is assumed normal is called Wishart distribution and is a generalization of χ^2 distribution in the univariate case.

By definition of A, we mean the joint distribution of the $\frac{p\,(p+1)}{2}$ distinct elements a_{ij}, $(i, j = 1, 2, \cdots, p;\ i \le j)$ of the symmetric matrix A.

Results:

1) Given a positive definite symmetric matrix A, there exists a nonsingular triangular matrix B such that $A = BB'$.

2) The Jacobian of transformation for the distinct element of B is $\left|\frac{\partial A}{\partial B}\right| = 2^p \prod_{i=1}^{p}(b_{ii})^{p-(i-1)}$.

Proof: The equation $A = BB'$ can be written as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ b_{21} & b_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ 0 & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & b_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} b_{11}^2 & b_{11}b_{21} & \cdots & b_{11}b_{p1} \\ b_{21}b_{11} & b_{21}^2 + b_{22}^2 & \cdots & b_{21}b_{p1} + b_{22}b_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ b_{p1}b_{11} & b_{p1}b_{21} + b_{p2}b_{22} & \cdots & b_{p1}^2 + b_{p2}^2 + \cdots + b_{pp}^2 \end{pmatrix}$$

$$\frac{\partial A}{\partial B} = \frac{\partial\,(a_{11},\, a_{21}\, a_{22},\, a_{31}\, a_{32}\, a_{33}, \cdots, a_{p1}\, a_{p2}\, \cdots\, a_{pp})}{\partial\,(b_{11},\, b_{21}\, b_{22}, \cdots, b_{p1}\, b_{p2}\, \cdots\, b_{pp})}$$

∂	a_{11}	$a_{21}\ a_{22}$	$a_{31}\ a_{32}\ a_{33}$	$\cdots$	$a_{p1}\ a_{p2} \cdots a_{pp}$
b_{11}	$2b_{11}$			$\cdots$	
b_{21} b_{22}	0 0	b_{11} $0\quad 2b_{22}$		$\cdots$ $\cdots$	
b_{31} b_{32} b_{33}	0 0 0	$0\quad 0$ $0\quad 0$ $0\quad 0$	b_{11} $0\quad b_{22}$ $0\quad 0\quad 2b_{33}$	$\cdots$ $\cdots$ $\cdots$	
$\vdots$	$\vdots$	$\vdots\quad \vdots$	$\vdots\quad \vdots\quad \vdots$	$\cdots$	
b_{p1} b_{p2} $\vdots$ b_{pp}	0 0 $\vdots$ 0	$0\quad 0$ $0\quad 0$ $\vdots\quad \vdots$ $0\quad 0$	$0\quad 0\quad 0$ $0\quad 0\quad 0$ $\vdots\quad \vdots\quad \vdots$ $0\quad 0\quad 0$	$\cdots$ $\cdots$ $\cdots$ $\cdots$	b_{11} $0\quad b_{22}$ $\vdots\quad \vdots$ $0\quad 0\quad \cdots\quad 2b_{pp}$

$$\therefore\ \left|\frac{\partial A}{\partial B}\right| = 2^p\, b_{11}^p\, b_{22}^p \cdots b_{pp}^{p-(p-1)} = 2^p \prod_{i=1}^{p} b_{ii}^{p-(i-1)}.$$

Theorem: Let $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ be a random sample from $N_p(\underline{\mu}, I)$, $A = \sum_{\alpha=1}^{n}(\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = \sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}'_\alpha$, where $\underline{Z}_\alpha$ are independent, each distributed according to $N_p(\underline{0}, I)$. Then the density of $A = \sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}'_\alpha$ is

$$\frac{|\,A\,|^{(\upsilon-p-1)/2} \exp\left(-\frac{1}{2}\,tr\,A\right)}{2^{\upsilon p/2}\,\pi^{p(p-1)/4} \prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}, \text{ where } \upsilon = n-1.$$

Proof: Consider a nonsingular triangular matrix B such that

$$A = BB' \tag{4.1}$$

Let

$$B = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ b_{21} & b_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{pmatrix} = \begin{pmatrix} \underline{B}'_1 \\ \underline{B}'_2 \\ \vdots \\ \underline{B}'_p \end{pmatrix}, \text{ and } B_{rr} = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ b_{21} & b_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ b_{r1} & b_{r2} & \cdots & b_{rr} \end{pmatrix}, \quad \text{so that}$$

$B_{pp} = B$.

Let $f(B) =$ joint density function of nonzero elements of B, then we can write

$$f(B) = f(\underline{B}'_1)\, f(\underline{B}'_2 \,|\, B_{11}) \cdots f(\underline{B}'_p \,|\, B_{p-1\,p-1}) = \prod_{i=0}^{p-1} f(\underline{B}'_{i+1} \,|\, B_{ii}) \tag{4.2}$$

Let

$\underline{B}'_{.r} = (b_{r1}\ b_{r2}\ \cdots\ b_{r\,r-1})$, so that

$\underline{B}'_r = (\underline{B}'_{.r}\ b_{rr}\ 0 \cdots 0)$

The equation (4.1) can be explained as

$$\begin{pmatrix} a_{11} & \cdots & a_{1i+1} & \cdots & a_{1p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{ii+1} & \cdots & a_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{p1} & \cdots & a_{pi+1} & \cdots & a_{pp} \end{pmatrix} = \begin{pmatrix} b_{11} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{i1} & \cdots & b_{ii} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{p1} & \cdots & b_{pi} & \cdots & b_{pp} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{i+11} & \cdots & b_{p1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & b_{i+1i} & \cdots & b_{pi} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & b_{pp} \end{pmatrix}$$

Let

$$\underline{a}_{.i+1} = \begin{pmatrix} a_{1i+1} \\ a_{2i+1} \\ \vdots \\ a_{ii+1} \end{pmatrix}, \quad \text{then} \quad \underline{a}_{.i+1} = B_{ii}\, \underline{B}_{.i+1}$$

$$= B_{ii}\, B'_{ii}\, (B'_{ii})^{-1}\, \underline{B}_{.i+1} \tag{4.3}$$

$$= A_{ii}\, \underline{C} \tag{4.4}$$

where

$$\underline{C} = (B'_{ii})^{-1}\, \underline{B}_{.i+1}, \text{ and } \quad A_{ii} = B_{ii}\, B'_{ii} \tag{4.5}$$

Equation (4.4) can be identified to be the normal equation for estimating the regression coefficient of Z_{i+1} on $Z_1, Z_2, \cdots, Z_i$, thus $\underline{C}$ is the vector of regression coefficient.

$\therefore \quad \underline{C} \sim N_i(E\,\underline{C}, \Sigma_{\underline{C}})$

Since $Z_1, Z_2, \cdots, Z_{i+1}$ are independent, it follows that

$E\,\underline{C} = 0$, $\quad \Sigma_{\underline{C}} = A_{ii}^{-1}$, because $\quad \underline{Q} = S\,\underline{\beta}$, then $E\underline{\hat{\beta}} = \underline{\beta}$ and $\Sigma_{\underline{\hat{\beta}}} = \sigma^2 S^{-1}$

Now

$\underline{B}_{.i+1} = B'_{ii}\, \underline{C}$, from equation (4.5), since B'_{ii} is fixed with respect to Z_{i+1}, then

$E\,\underline{B}_{.i+1} = B'_{ii}\; E\,\underline{C} = \underline{0}$, $\Sigma_{\underline{B}.i+1} = B'_{ii}\; A_{ii}^{-1}\; B_{ii} = B'_{ii}\,(B_{ii}\, B'_{ii})^{-1}\, B_{ii} = I$.

Further, $\underline{B}_{.i+1}$ being a linear combination of regression coefficients which are themselves normally distributed, we conclude that

$$\underline{B}_{.i+1} \sim N_i\,(\underline{0}, I) \tag{4.6}$$

$\mathrm{Re}\,g\;\; SS = \underline{C}'\;\underline{a}_{.i+1}$, because if the normal equation $\underline{Q} = S\,\underline{\beta}$, then $\mathrm{Re}\,g\;\; SS$ is $\underline{\beta}'\,\underline{Q}$

$$= [(B'_{ii})^{-1}\, \underline{B}_{.i+1}]'\; B_{ii}\, \underline{B}_{.i+1} = \underline{B}'_{.i+1}\,(B_{ii})^{-1} B_{ii}\; \underline{B}_{.i+1} = \underline{B}'_{.i+1}\; \underline{B}_{.i+1}$$

$$= b^2_{i+11} + b^2_{i+12} + \cdots + b^2_{i+1i} \sim \chi^2_i \tag{4.7}$$

$Error\;\; SS = Total\;\; SS - \mathrm{Re}\,g\;\; SS$

$$= a_{i+1i+1} - (b^2_{i+11} + b^2_{i+12} + \cdots + b^2_{i+1i})$$

$$= (b^2_{i+11} + b^2_{i+12} + \cdots + b^2_{i+1i+1}) - (b^2_{i+11} + b^2_{i+12} + \cdots + b^2_{i+1i}), \text{since } A = BB'$$

$$= b^2_{i+1i+1} \sim \chi^2_{\upsilon - i} \tag{4.8}$$

As *reg SS* and *error SS* are independent, so that b_{i+1i+1} is independent of $\underline{B}_{.i+1}$.

Therefore,

$$f(\underline{B}_{i+1} \mid B_{ii}) = f(\underline{B}_{.i+1} \mid B_{ii})\, f(b_{i+1\,i+1} \mid B_{ii}), \text{ because } \underline{B}'_{r} = (\underline{B}'_{.r}\;\; b_{rr}\;\; 0 \cdots 0)$$

$$= \frac{1}{(2\pi)^{i/2}} \exp\left\{-\frac{1}{2}(b^2_{1i+1} + b^2_{2i+1} + \cdots + b^2_{ii+1})\right\}$$

$$\frac{1}{2^{(\upsilon-i-2)/2}\,\Gamma(\upsilon - i)/2} \exp\left\{-\frac{1}{2}(b^2_{i+1i+1})\right\}(b_{i+1i+1})^{\upsilon-i-1}$$

$$\text{because, } f(\chi) = \frac{1}{2^{(\upsilon-2)/2}\,\Gamma\upsilon/2} \exp\left\{-\frac{1}{2}\chi^2\right\}(\chi)^{\upsilon-1}$$

$$= \frac{1}{\pi^{i/2}\, 2^{(\upsilon-2)/2}\,\Gamma(\upsilon - i)/2} \exp\left\{-\frac{1}{2}\left(b^2_{i+1i+1} + \sum_{j=1}^{i} b^2_{ji+1}\right)\right\}(b_{i+1i+1})^{\upsilon-i-1}$$

and

$$f(B) = \prod_{i=0}^{p-1} f(\underline{B}_{i+1} \mid B_{ii}) = \frac{\prod_{i=0}^{p-1}(b_{i+1i+1})^{\upsilon-i-1}}{2^{p(\upsilon-2)/2}\,\pi^{p(p-1)/4} \prod_{i=0}^{p-1}\Gamma(\upsilon - i)/2} \exp\left\{-\frac{1}{2}\sum_{i=0}^{p-1}\sum_{j=1}^{i+1} b^2_{ji+1}\right\}$$

$$\text{because } \prod_{i=0}^{p-1}\pi^{i/2} = \pi^{[1+2+\cdots+(p-1)]/2} = \pi^{p(p-1)/4}.$$

$$f(B) = \frac{\prod_{i=1}^{p}(b_{ii})^{\upsilon-i}}{2^{p(\upsilon i-2)/2}\,\pi^{p(p-1)/4} \prod_{i=1}^{p}\Gamma(\upsilon - i + 1)/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{i} b^2_{ji}\right\}$$

Now

$$\left|\frac{\partial B}{\partial A}\right| = \frac{1}{\left|\frac{\partial A}{\partial B}\right|} = \frac{1}{2^p \prod_{i=1}^{p}(b_{ii})^{p-i+1}}.$$

Also, $tr\,A =$ the sum of the diagonal elements of $A = \sum_{i=1}^{p} A_{ii} = \sum_{i=1}^{p}\left(\sum_{j=1}^{i} b^2_{ji}\right)$.

Therefore,

$$f(A)=\frac{f(B)}{2^p\prod_{i=1}^{p}(b_{ii})^{p-i+1}}=\frac{\prod_{i=1}^{p}(b_{ii})^{\upsilon-p-1}}{2^{\upsilon p/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}\exp\left\{-\frac{1}{2}\,tr\,A\right\}$$

Consider,

$$\prod_{i=1}^{p}(b_{ii})^{\upsilon-p-1}=(b_{11}b_{22}\cdots b_{pp})^{\upsilon-p-1}=|B|^{\upsilon-p-1}$$

$$|A|=|B||B'|=|B|^2 \Rightarrow |B|=|A|^{1/2} \Rightarrow \prod_{i=1}^{p}(b_{ii})^{\upsilon-p-1}=|A|^{(\upsilon-p-1)/2}$$

$$f(A)=\frac{|A|^{(\upsilon-p-1)/2}}{2^{\upsilon p/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}\exp\left\{-\frac{1}{2}\,tr\,A\right\}.$$

This is the form of wishart distribution when $\Sigma=I$ and is denoted by $W_p(\upsilon,\ I)$.

Theorem: Suppose the $p-$component vectors $\underline{x}_1,\underline{x}_2,\cdots,\underline{x}_n$ $(n>p)$ are independent, each distributed according to $N_p(\underline{\mu},\Sigma)$, then the density of $A=\sum_{\alpha=1}^{n}(\underline{x}_\alpha-\underline{\overline{x}})(\underline{x}_\alpha-\underline{\overline{x}})'=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha\underline{Z}'_\alpha$, where, $\underline{Z}_\alpha\sim N_p(\underline{0},\Sigma)$ is

$$\frac{|A|^{(\upsilon-p-1)/2}\exp(-\frac{1}{2}tr\,A\Sigma^{-1})}{2^{\upsilon p/2}\pi^{p(p-1)/4}|\Sigma|^{\upsilon/2}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2},\ \text{where } \upsilon=n-1.$$

Proof: Since Σ is symmetric and positive definite there exist a nonsingular triangular matrix B such that

$$B\Sigma B'=I \ \Rightarrow\ B'B=\Sigma^{-1}.$$

Make the transformation

$$\underline{Z}^*_\alpha=B\underline{Z}_\alpha,\quad \alpha=1,2,\cdots,n-1.$$

$$E(\underline{Z}^*_\alpha)=B\,E(\underline{Z}_\alpha)=\underline{0},\quad \Sigma_{\underline{Z}^*_\alpha}=B\,E(\underline{Z}_\alpha\underline{Z}'_\alpha)B'=B\Sigma B'=I.$$

Thus,

$$\underline{Z}^*_\alpha\sim N_p(\underline{0},\ I),\ \text{and}\ \sum_{\alpha=1}^{n-1}\underline{Z}^*_\alpha\underline{Z}^{*\prime}_\alpha=B\left(\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha\underline{Z}'_\alpha\right)B'=BAB'=A^*$$

$$\Rightarrow A^*\sim W_p(\upsilon,\ I)$$

and

$$f(A^*)=\frac{|A^*|^{(\upsilon-p-1)/2}}{2^{\upsilon p/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}\exp\left\{-\frac{1}{2}\,tr\,A^*\right\}$$

$$f(A)=f[A^*(A)]\frac{\partial A^*}{\partial A}.$$

Now

$$\frac{\partial A^*}{\partial A}=\frac{\partial\,BAB}{\partial A}=|B|^{p+1},\ \text{and}\ \ tr\,A^*=tr\,BAB'=tr\,AB'B=tr\,A\Sigma^{-1}$$

From $B\Sigma B'=I$, we have $1=|B\Sigma B'|=|B||\Sigma||B'|=|\Sigma||BB'|=|\Sigma||B|^2$

and

$$|BB'|=\frac{1}{|\Sigma|},\quad |B|=\frac{1}{|\Sigma|^{1/2}}\ \text{and}\ |A^*|=|BAB'|=\frac{|A|}{|\Sigma|}.$$

Therefore,

$$f(A)=\frac{|A|^{(\upsilon-p-1)/2}\exp(-\frac{1}{2}tr\,A\Sigma^{-1})}{2^{\upsilon p/2}\pi^{p(p-1)/4}|\Sigma|^{(\upsilon-p-1)/2}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}\frac{1}{|\Sigma|^{(p+1)/2}}$$

$$=\frac{|A|^{(\upsilon-p-1)/2}\exp(-\frac{1}{2}tr\,A\Sigma^{-1})}{2^{\upsilon p/2}\pi^{p(p-1)/4}|\Sigma|^{\upsilon/2}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2},\ \Rightarrow\ A\sim W_p(\upsilon,\ \Sigma).$$

Theorem: Let $\underline{x}_1,\underline{x}_2,\cdots,\underline{x}_n$ $(n\ge p+1)$ are distributed independently, each according to $N_p(\underline{\mu},\Sigma)$. Then the distribution of $S=\frac{1}{n-1}\sum_{\alpha=1}^{n}(\underline{x}_\alpha-\underline{\overline{x}})(\underline{x}_\alpha-\underline{\overline{x}})'$ is $W_p(\upsilon,\ \Sigma/\upsilon)$, where $\upsilon=n-1$.

Proof: Clearly,

$$S=\frac{A}{n-1}=\sum_{\alpha=1}^{n}\left(\frac{1}{\sqrt{n-1}}\underline{Z}_\alpha\right)\left(\frac{1}{\sqrt{n-1}}\underline{Z}_\alpha\right)',\ \text{where}\ \frac{1}{\sqrt{n-1}}\underline{Z}_\alpha\sim N_p(\underline{0},\ \Sigma/n-1).$$

Since Σ is symmetric and positive definite there exist a nonsingular triangular matrix B such that

$$B\Sigma^*B'=I,\ \text{where}\ \Sigma^*=\Sigma/n-1\ \text{and}\ B'B=\Sigma^{*-1}.$$

Make the transformation

$$\underline{Z}^*_\alpha=B\left(\frac{1}{\sqrt{n-1}}\underline{Z}_\alpha\right),\qquad \alpha=1,2,\cdots,n-1.$$

$$E(\underline{Z}_\alpha^*)=\underline{0}, \quad \Sigma_{\underline{Z}_\alpha^*}=B\,E\left(\frac{1}{\sqrt{n-1}}\underline{Z}_\alpha\right)\left(\frac{1}{\sqrt{n-1}}\underline{Z}_\alpha\right)' B'=B\Sigma^* B'=I.$$

Thus,

$$\underline{Z}_\alpha^* \sim N_p(\underline{0}, I), \text{ and } \sum_{\alpha=1}^{n-1}\underline{Z}_\alpha^*\underline{Z}_\alpha^{*\,\prime}=B\left[\sum_{\alpha=1}^{n-1}\left(\frac{1}{\sqrt{n-1}}\underline{Z}_\alpha\right)\left(\frac{1}{\sqrt{n-1}}\underline{Z}_\alpha\right)'\right]B'=B\,S\,B'=S^*$$

Hence,

$$S^*\sim W_p(\upsilon, I), \text{ and } \quad f(S^*)=\frac{\left|S^*\right|^{(\upsilon-p-1)/2}}{2^{\upsilon p/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}\exp\left\{-\frac{1}{2}\,tr\,S^*\right\}$$

$$f(S)=f[S^*(S)]\frac{\partial S^*}{\partial S}.$$

Now

$$\frac{\partial S^*}{\partial S}=\frac{\partial\, BSB'}{\partial S}=\left|B\right|^{p+1}, \quad tr\,S^*=tr\,BSB'=tr\,S\,B'B=tr\,S\left(\frac{\Sigma}{n-1}\right)^{-1}$$

From $B\Sigma^*B'=I$, we have $1=\left|B\Sigma^*B'\right|=\left|B\right|\left|\Sigma^*\right|\left|B'\right|=\left|\Sigma^*\right|\left|BB'\right|=\left|\Sigma^*\right|\left|B\right|^2$, thus,

$$\left|BB'\right|=\frac{1}{\left|\Sigma^*\right|}, \; \left|B\right|=\frac{1}{\left|\Sigma^*\right|^{1/2}}, \text{ and } \left|S^*\right|=\left|BSB'\right|=\frac{\left|S\right|}{\left|\Sigma^*\right|}=\frac{\left|S\right|}{\left|\frac{\Sigma}{n-1}\right|}.$$

Therefore,

$$f(S)=\frac{\left|S\right|^{(\upsilon-p-1)/2}\exp\left[-\frac{1}{2}tr\,S\left(\frac{\Sigma}{n-1}\right)^{-1}\right]}{2^{\upsilon p/2}\pi^{p(p-1)/4}\left|\frac{\Sigma}{n-1}\right|^{\upsilon/2}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}.$$

This is the density of $W_p(\upsilon, \Sigma/\upsilon)$.

Characteristic function

Consider

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}, \text{ the elements of the matrix are } A_{11}, A_{22}, 2A_{12}, \text{ because } A_{12}=A_{21}.$$

Introduce a real matrix $\Theta=(\theta_{ij})$, with $\theta_{ij}=\theta_{ji}$

$$\Theta=\begin{pmatrix}\theta_{11} & \theta_{12}\\ \theta_{21} & \theta_{22}\end{pmatrix} \text{ then, } A\Theta=\begin{pmatrix}A_{11}\theta_{11}+A_{12}\theta_{21} & A_{11}\theta_{12}+A_{12}\theta_{22}\\ A_{21}\theta_{11}+A_{22}\theta_{21} & A_{21}\theta_{12}+A_{22}\theta_{22}\end{pmatrix}$$

$$tr\,A\Theta=A_{11}\theta_{11}+A_{12}\theta_{21}+A_{21}\theta_{12}+A_{22}\theta_{22}=A_{11}\theta_{11}+A_{22}\theta_{22}+2A_{12}\theta_{12} \tag{4.9}$$

We know that the characteristic function of a vector $\underline{x}'=(x_1, x_2, \cdots, x_p)$ is defined as

$$Ee^{i\underline{t}'\underline{x}}, \text{ where } \underline{t}'\underline{x}=t_1x_1+t_2x_2+\cdots+t_px_p \tag{4.10}$$

In view of (4.10), we can write (4.9) as $E(e^{itr\,A\Theta})$.

Theorem: If $\underline{Z}_1, \underline{Z}_2, \cdots, \underline{Z}_{n-1}$ are independent, each with distribution $N_p(\underline{0}, \Sigma)$, the characteristic function of $A_{11}, A_{22}, \cdots, A_{pp}, 2A_{12}, \cdots, 2A_{p-1p}$, where $A=(A_{ij})=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha\underline{Z}_\alpha'$ is given by

$$\phi_A(\Theta)=E\,e^{itr\,A\Theta}=\left|I-2i\Theta\Sigma\right|^{-\upsilon/2}, \text{ where } n-1=\upsilon.$$

Proof: We have

$$\phi_A(\Theta)=E\,e^{itr\sum_{\alpha=1}^{\upsilon}\underline{Z}_\alpha\underline{Z}_\alpha'\Theta}=E\,e^{itr\sum_{\alpha=1}^{\upsilon}\underline{Z}_\alpha'\Theta\underline{Z}_\alpha}, \text{ by virtue of the fact that}$$

$$tr\,EFG=\sum e_{ij}f_{jk}g_{ki}=tr\,FGE$$

$$=E\,e^{i\sum_{\alpha=1}^{\upsilon}\underline{Z}_\alpha'\Theta\underline{Z}_\alpha}=\prod_{\alpha=1}^{\upsilon}E\,e^{i\,\underline{Z}_\alpha'\Theta\underline{Z}_\alpha}, \text{ as } \underline{Z}_\alpha\text{'s are identical distributed random vector.}$$

$$=E(e^{i\,\underline{Z}'\Theta\underline{Z}})^\upsilon, \text{ where } \underline{Z}\sim N_p(\underline{0}, \Sigma) \tag{4.11}$$

Since Θ is real and Σ is positive definite, there exist a real nonsingular matrix C such that $C'\Sigma^{-1}C=I$, and $C'\Theta C=D$, where D is a real diagonal matrix, diagonal elements of D is d_{jj}.

Consider the transformation

$$\underline{Z}=C\underline{Y}, \text{ then } E\underline{Y}=\underline{0}, \text{ and } \Sigma_{\underline{Y}}=C^{-1}E\underline{Z}\,\underline{Z}'(C^{-1})'=(C'\Sigma^{-1}C)^{-1}=I$$

$\Rightarrow \underline{Y}\sim N_p(\underline{0}, I)$ i.e. Y_j, $j=1,2,\cdots,p$ are independently distributed as $N(0, 1)$, therefore,

$$E\,e^{i\,\underline{Z}'\Theta\underline{Z}}=E\,e^{i\,\underline{Y}'C'\Theta C\underline{Y}}=E\,e^{i\,\underline{Y}'D\underline{Y}}=E\,e^{i\sum_{j=1}^{p}d_{jj}\,y_j^2}$$

$$=\prod_{j=1}^{p}E\,e^{i\,d_{jj}\,y_j^2}=\prod_{j=1}^{p}(\text{ch. function of } \chi_1^2)=\prod_{j=1}^{p}(1-2i\,d_{jj})^{-1/2}$$

$$=\left|I-2i\,D\right|^{-1/2}, \text{ as } (I-2i\,D) \text{ is a diagonal matrix.} \tag{4.12}$$

Moreover,

$$\left|I-2i\,D\right|=\left|C'\Sigma^{-1}C-2i\,C'\Theta C\right|=\left|C'(\Sigma^{-1}-2i\Theta)C\right|=\left|C\right|^2\left|\Sigma^{-1}-2i\Theta\right|.$$

But

$$C'\Sigma^{-1}C = I \quad \text{or} \quad \left| C'\Sigma^{-1}C \right| = 1 \quad \text{or} \quad |C'| \left|\Sigma^{-1}\right| |C| = 1 \quad \text{or} \quad |C'| |C| \left|\Sigma^{-1}\right| = 1$$

$$|C|^2 \left|\Sigma^{-1}\right| = 1 \quad \text{or} \quad |C|^2 = \frac{1}{\left|\Sigma^{-1}\right|} = |\Sigma|$$

Hence,

$$|I - 2iD| = \left|\Sigma^{-1} - 2i\Theta\right| |\Sigma| = |I - 2i\Theta\Sigma| \tag{2.13}$$

Using equations (4.12) and (4.13) in (4.11) gives, the characteristic function of $A_{11}, A_{22}, \cdots, A_{pp}, 2A_{12}, \cdots, 2A_{p-1p}$, as $\phi_A(\Theta) = |I - 2i\Theta\Sigma|^{-\upsilon/2}$.

Alternative proof

By definition,

$$\phi_A(\Theta) = Ee^{i\,tr\,\Theta A} = \int \frac{e^{i\,tr\,\Theta A} |A|^{(\upsilon-p-1)/2} \exp(-\frac{1}{2} tr\, A\Sigma^{-1})}{2^{\upsilon p/2} \pi^{p(p-1)/4} |\Sigma|^{\upsilon/2} \prod_{i=1}^{p} \Gamma(\upsilon - i + 1)/2} dA$$

$$= \frac{K}{|\Sigma|^{\upsilon/2}} \int |A|^{(\upsilon-p-1)/2} e^{-\frac{1}{2} tr(\Sigma^{-1} - 2i\Theta)A} dA$$

$$= \frac{K \left|\Sigma^{-1} - 2i\Theta\right|^{\upsilon/2}}{|\Sigma|^{\upsilon/2} \left|\Sigma^{-1} - 2i\Theta\right|^{\upsilon/2}} \int |A|^{(\upsilon-p-1)/2} e^{-\frac{1}{2} tr(\Sigma^{-1} - 2i\Theta)A} dA$$

$$= \frac{1}{|\Sigma|^{\upsilon/2} \left|\Sigma^{-1} - 2i\Theta\right|^{\upsilon/2}}, \text{ since } K\left|\Sigma^{-1}\right|^{\upsilon/2} \int |A|^{(\upsilon-p-1)/2} e^{-\frac{1}{2} tr\, A\Sigma^{-1}} dA = 1$$

$$= \frac{1}{|I - 2i\Theta\Sigma|^{\upsilon/2}} = |I - 2i\Theta\Sigma|^{-\upsilon/2}.$$

This shows that, if $A \sim W_p(\upsilon, \Sigma)$, and then the characteristic function of A is $|I - 2i\Theta\Sigma|^{-\upsilon/2}$.

Properties of Wishart distribution

Theorem: Suppose A_i $(i = 1, 2)$ are distributed independently according to $W_p(\upsilon_i, \Sigma)$ respectively, then $A_1 + A_2 \sim W_p(\upsilon_1 + \upsilon_2, \Sigma)$.

Proof: We know that the characteristic function of A_1, if $A_1 \sim W_p(\upsilon_1, \Sigma)$, is

$$\phi_{A_1}(\Theta) = |I - 2i\Theta\Sigma|^{-\upsilon_1/2}.$$

Similarly, the characteristic function of A_2 will be

$$\phi_{A_2}(\Theta) = |I - 2i\Theta\Sigma|^{-\upsilon_2/2}$$

Since A_1 and A_2 are independently distributed, so

$$\phi_{A_1+A_2}(\Theta) = \phi_{A_1}(\Theta)\phi_{A_2}(\Theta) = |I - 2i\Theta\Sigma|^{-(\upsilon_1+\upsilon_2)/2}$$

But this is the characteristic function of $W_p(\upsilon_1 + \upsilon_2, \Sigma)$, therefore,

$$A_1 + A_2 \sim W_p(\upsilon_1 + \upsilon_2, \Sigma).$$

Alternative Solution

$$A_1 = \sum_{\alpha=1}^{n_1-1} \underline{Z}_\alpha \underline{Z}_\alpha' = \underline{Z}_1 \underline{Z}_1' + \cdots + \underline{Z}_{\upsilon_1} \underline{Z}_{\upsilon_1}', \text{ where } \underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma),\ \upsilon_1 = n_1 - 1, \text{ and}$$

$$A_2 = \sum_{\alpha=1}^{n_2-1} \underline{Y}_\alpha \underline{Y}_\alpha' = \underline{Y}_1 \underline{Y}_1' + \cdots + \underline{Y}_{\upsilon_2} \underline{Y}_{\upsilon_2}', \text{ where } \underline{Y}_\alpha \sim N_p(\underline{0}, \Sigma),\ \upsilon_2 = n_2 - 1$$

$\underline{Y}_\alpha$ and $\underline{Z}_\alpha$ are independent, since A_1 and A_2 are independent, therefore,

$$A_1 + A_2 = \underline{Z}_1 \underline{Z}_1' + \cdots + \underline{Z}_{\upsilon_1} \underline{Z}_{\upsilon_1}' + \underline{Y}_1 \underline{Y}_1' + \cdots + \underline{Y}_{\upsilon_2} \underline{Y}_{\upsilon_2}'$$

$$= \sum_{\alpha=1}^{\upsilon_1+\upsilon_2} \underline{Z}_\alpha^* \underline{Z}_\alpha^{*\prime}, \text{ each } \underline{Z}_\alpha^* \sim N_p(\underline{0}, \Sigma).$$

Hence,

$$A_1 + A_2 \sim W_p(\upsilon_1 + \upsilon_2, \Sigma).$$

Theorem: If A_j, $j = 1, 2, \cdots, q$ are independently distributed according to $W_p(\upsilon_j, \Sigma)$, then $A = \sum_{j=1}^{q} A_j \sim W_p\left(\sum_{j=1}^{q} \upsilon_j, \Sigma\right)$.

Proof: We know that the characteristic function of A_j is

$\phi_{A_j}(\Theta) = |I - 2i\Theta\Sigma|^{-\upsilon_j/2}$. Now consider the characteristic function of $A = \sum_{j=1}^{q} A_j$

$$\phi_A(\Theta) = \phi_{\sum_j A_j}(\Theta) = \prod_j \phi_{A_j}(\Theta), \text{ since } A_j\text{'s are independently distributed}$$

$$= \prod_{j=1}^{q} |I - 2i\Theta\Sigma|^{-\upsilon_j/2} = |I - 2i\Theta\Sigma|^{-\sum_j \upsilon_j/2}.$$

Which is the characteristic function of $W_p\left(\sum_j \upsilon_j, \Sigma\right)$.

Therefore,

$$A=\sum_{j=1}^{q}A_j \sim W_p\left(\sum_{j=1}^{q}\upsilon_j,\ \Sigma\right).$$

Theorem: If $A \sim W_p(n-1,\ \Sigma)$, then the distribution of $\underline{l}'A\underline{l} \sim (\underline{l}'\Sigma\,\underline{l})\chi^2_{n-1}$, where $\underline{l}$ is a known vector.

Proof: Given $A \sim W_p(n-1,\ \Sigma)$, then $A=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}_\alpha'$, where $\underline{Z}_\alpha \sim N_p(\underline{0},\ \Sigma)$, and

$$\underline{l}'A\underline{l}=\sum_{\alpha=1}^{n-1}\underline{l}'\underline{Z}_\alpha \underline{Z}_\alpha'\underline{l}=\sum_{\alpha=1}^{n-1}(\underline{l}'\underline{Z}_\alpha)(\underline{l}'\underline{Z}_\alpha)'=\sum_{\alpha=1}^{n-1}V_\alpha^2\text{, where } V_\alpha=\underline{l}'\underline{Z}_\alpha \text{ is } N(0,\ \underline{l}'\Sigma\,\underline{l}).$$

Therefore,

$$\underline{l}'A\underline{l} \sim (\underline{l}'\Sigma\,\underline{l})\chi^2_{n-1}.$$

Theorem: If $A \sim W_p(n-1,\Sigma)$, and if a is a positive constant, then $aA \sim W_p(n-1, a\Sigma)$.

Proof: Since $A \sim W_p(n-1,\ \Sigma)$, there are $n-1$ independent $p-$component random vectors $\underline{Z}_1,\cdots,\underline{Z}_{n-1}$ each distributed as $N_p(\underline{0},\Sigma)$, such that

$A=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}_\alpha'$. Evidently, we have

$aA=\sum_{\alpha=1}^{n-1}(\sqrt{a}\,\underline{Z}_\alpha)(\sqrt{a}\,\underline{Z}_\alpha)'$ and $\sqrt{a}\,\underline{Z}_1,\cdots,\sqrt{a}\,\underline{Z}_{n-1}$ are independently identically distributed as normal $N_p(\underline{0}, a\Sigma)$. Thus

$$aA \sim W_p(n-1, a\Sigma).$$

Theorem: Let A and Σ be partitioned into q and $p-q$ rows and columns as

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

If A is distributed according to $W_p(\upsilon,\ \Sigma)$, then A_{11} is distributed according to $W_q(\upsilon,\ \Sigma_{11})$ (the marginal distribution of some sets of elements of A).

Proof: Given $A \sim W_p(\upsilon,\ \Sigma)$, where

$A=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}_\alpha'$, and $\underline{Z}_\alpha$ are independent, each with distribution $N_p(\underline{0},\ \Sigma)$.

Partition $\underline{Z}_\alpha$ into sub vectors of q and $p-q$ components as

$$\underline{Z}_\alpha=\begin{pmatrix}\underline{Z}_\alpha^{(1)}\\ \underline{Z}_\alpha^{(2)}\end{pmatrix}.$$

Then $\underline{Z}_\alpha^{(1)}$ are independent, each with distribution $N_q(\underline{0},\ \Sigma_{11})$, because $\underline{Z}_\alpha^{(1)}$ and $\underline{Z}_\alpha^{(2)}$ are independent, so that $\Sigma_{12}=\Sigma_{21}=0$, and

$$A_{11}=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha^{(1)}\underline{Z}_\alpha^{(1)'},\quad \Rightarrow\ A_{11} \sim W_q(\upsilon,\ \Sigma_{11}).$$

Similarly,

$$A_{22} \sim W_{p-q}(\upsilon,\ \Sigma_{22}).$$

Theorem: Let A and Σ be partitioned into $p_1, p_2,\cdots p_q$ rows and columns $(p_1+p_2+\cdots+p_q=p)$

$$A=\begin{pmatrix}A_{11} & \cdots & A_{1q}\\ \vdots & \vdots & \vdots\\ A_{q1} & \cdots & A_{qq}\end{pmatrix},\qquad \Sigma=\begin{pmatrix}\Sigma_{11} & \cdots & \Sigma_{1q}\\ \vdots & \vdots & \vdots\\ \Sigma_{q1} & \cdots & \Sigma_{qq}\end{pmatrix}$$

If $\Sigma_{ij}=\underline{0}$ for $i \neq j$ and if A is distributed according to $W_p(\upsilon,\ \Sigma)$, then $A_{11}, A_{22},\cdots, A_{qq}$ are independently distributed and A_{jj} is distributed according to $W(\upsilon,\ \Sigma_{jj})$.

Proof: Let $A=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}_\alpha'$, where $\underline{Z}_\alpha$ are independent, each with distribution $N_p(\underline{0},\ \Sigma)$. Let us partition $\underline{Z}_\alpha$ as

$\underline{Z}_\alpha=\begin{pmatrix}\underline{Z}_\alpha^{(1)}\\ \vdots\\ \underline{Z}_\alpha^{(q)}\end{pmatrix}$. Since $\Sigma_{ij}=\underline{0}$ for $i \neq j$, so $\underline{Z}_\alpha^{(i)}$ and $\underline{Z}_\alpha^{(j)}$ are independent, then

$A_{ii}=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha^{(i)}\underline{Z}_\alpha^{(i)'}$ is independent of $A_{jj}=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha^{(j)}\underline{Z}_\alpha^{(j)'}$, where

$\underline{Z}_\alpha^{(i)} \sim N_p(\underline{0},\ \Sigma_{ii})$, and $\underline{Z}_\alpha^{(j)} \sim N_p(\underline{0},\ \Sigma_{jj})$, thus, $\underline{Z}_\alpha^{(1)}, \underline{Z}_\alpha^{(2)},\cdots,\underline{Z}_\alpha^{(q)}$ are independent.

Hence,

$$A_{ii}=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha^{(i)}\underline{Z}_\alpha^{(i)'},\ i=1,2,\cdots,q \text{ are independently distributed.}$$

Theorem: Let A and Σ be partitioned into q and $p-q$ rows and columns

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

If A is distributed according to $W_p(\upsilon,\ \Sigma)$, then $A_{11}-A_{12}A_{22}^{-1}A_{21}$ is distributed according to $W_q(\upsilon-(p-q),\ \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ (the conditional distribution of some sets of elements of A).

Theorem: Suppose A is distributed according to $W_p(\upsilon, \Sigma)$ and let $A^* = BAB'$, where B is a matrix of order $q \times p$, then $A^* \sim W_q(\upsilon, \Phi)$, where $\Phi = B\Sigma B'$.

Proof: We have

$$A = \sum_{\alpha=1}^{n-1} \underline{Z}_\alpha \underline{Z}_\alpha', \text{ where } \underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma)$$

$$BAB' = \sum_{\alpha=1}^{n-1} B\,\underline{Z}_\alpha \underline{Z}_\alpha'\, B' = \sum_{\alpha=1}^{n-1} (B\,\underline{Z}_\alpha)(B\underline{Z}_\alpha)'$$

$$= \sum_{\alpha=1}^{n-1} \underline{V}_\alpha\, \underline{V}_\alpha', \text{ where } \underline{V}_\alpha = B\,\underline{Z}_\alpha \sim N_q(\underline{0}, B\Sigma B').$$

Therefore,

$$A^* = BAB' = \sum_{\alpha=1}^{n-1} \underline{V}_\alpha\, \underline{V}_\alpha' \sim W_q(\upsilon, B\Sigma B').$$

Theorem: Suppose $A \sim W_p(\upsilon, \Sigma)$ and let $A = CBC'$, where C is a nonsingular matrix of order p, then $B \sim W_p(\upsilon, \Phi)$, where $\Phi = (C'\Sigma^{-1}C)^{-1}$.

Proof: We have $A = \sum_{\alpha=1}^{n-1} \underline{Z}_\alpha \underline{Z}_\alpha'$, where $\underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma)$.

Let

$$\underline{Y}_\alpha = C^{-1}\underline{Z}_\alpha, \text{ then } E\,\underline{Y}_\alpha = \underline{0},\ \Sigma_{\underline{Y}_\alpha} = C^{-1}\, E\underline{Z}_\alpha \underline{Z}_\alpha' C^{-1'} = C^{-1}\,\Sigma\, C^{-1'} = \Phi, \text{ and}$$

$$\sum_{\alpha=1}^{n-1} \underline{Y}_\alpha \underline{Y}_\alpha' = C^{-1}\left(\sum_{\alpha=1}^{n-1} \underline{Z}_\alpha \underline{Z}_\alpha'\right) C^{-1'} = C^{-1}\, A\, C^{-1'} = B$$

Let D be nonsingular triangular matrix such that $D\Phi D' = I$, and $D'D = \Phi^{-1}$.
Make the transformation

$$\underline{Y}_\alpha^* = D\underline{Y}_\alpha, \text{ then } E\,\underline{Y}_\alpha^* = \underline{0},\ \Sigma_{\underline{Y}_\alpha^*} = D\, E\underline{Y}_\alpha \underline{Y}_\alpha' D' = D\Phi D' = I \quad \Rightarrow \underline{Y}_\alpha^* \sim N_p(\underline{0}, I),$$

and

$$\sum_\alpha \underline{Y}_\alpha^* \underline{Y}_\alpha^{*'} = D\left(\sum_\alpha \underline{Y}_\alpha \underline{Y}_\alpha'\right) D' = DBD' = B^* \quad \Rightarrow B^* \sim W_p(\upsilon, I).$$

Thus,

$$f(B^*) = \frac{\left|B^*\right|^{(\upsilon-p-1)/2} \exp\left(-\frac{1}{2} trB^*\right)}{2^{\upsilon p/2}\, \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma(\upsilon - i + 1)/2}.$$

$$f(B) = f(B^*(B)) \frac{\partial B^*}{\partial B}, \text{ where } \frac{\partial B^*}{\partial B} = \left|D\right|^{p+1},\ tr\,B^* = tr\,DBD' = tr\,BD'D = tr\,B\Phi^{-1}.$$

From $D\Phi D' = I \quad \Rightarrow \left|DD'\right| = \frac{1}{\left|\Phi\right|} \quad \Rightarrow \left|D\right| = \frac{1}{\left|\Phi\right|^{1/2}}$, and $\left|B^*\right| = \left|DBD'\right| = \frac{\left|B\right|}{\left|\Phi\right|}$.

Therefore,

$$f(B) = \frac{\left|B\right|^{(\upsilon-p-1)/2} \exp\left(-\frac{1}{2} trB\Phi^{-1}\right)}{2^{\upsilon p/2}\, \pi^{p(p-1)/4} \left|\Phi\right|^{(\upsilon-p-1)/2} \prod_{i=1}^{p} \Gamma(\upsilon - i + 1)/2} \frac{1}{\left|\Phi\right|^{(p+1)/2}}$$

$$= \frac{\left|B\right|^{(\upsilon-p-1)/2} \exp\left(-\frac{1}{2} tr\,B\,\Phi^{-1}\right)}{2^{\upsilon p/2}\, \pi^{p(p-1)/4} \left|\Phi\right|^{\upsilon/2} \prod_{i=1}^{p} \Gamma(\upsilon - i + 1)/2}.$$

This is the density of $W_p(\upsilon, \Phi)$, i.e. $B \sim W_p(\upsilon, \Phi)$.

Exercise: Let $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ be a random sample from $N_p(\underline{\mu}, \Sigma)$, then $n(\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})' \sim W_p(1, \Sigma)$.

Solution: Let $\underline{Y} = \sqrt{n}\,(\underline{\bar{x}} - \underline{\mu})$, then $E\,\underline{Y} = \sqrt{n}\, E\,(\underline{\bar{x}} - \underline{\mu}) = \underline{0}$, and $\Sigma_{\underline{Y}} = E\,\underline{Y}\,\underline{Y}' = n\,\Sigma / n = \Sigma$.

We know that, if $\underline{Y}_\alpha \sim N_p(\underline{0}, \Sigma)$, and $A = \sum_{\alpha=1}^{n-1} \underline{Y}_\alpha \underline{Y}_\alpha'$, then $A \sim W_p(n-1, \Sigma)$, therefore,

$$\underline{Y}\,\underline{Y}' = n(\underline{\bar{x}} - \underline{\mu})(\underline{\bar{x}} - \underline{\mu})' \sim W_p(1, \Sigma).$$

Exercise: Show that wishart distribution is a generalization of the chi-square distribution.

Solution: We have $A \sim W_p(\upsilon, \Sigma)$, the characteristic function of A is

$$\phi_A(\Theta) = \left|I - 2i\,\Theta\,\Sigma\right|^{-\upsilon/2}.$$

For $p = 1$, $A = a_{11}$, and $\Sigma = \sigma_{11} = \sigma^2$, $\Theta = \theta_{11} = \theta$.

$\phi_a(\theta) = \left|1 - 2i\,\theta\,\sigma^2\right|^{-\upsilon/2}$, which is the characteristic function of $\sigma^2\chi^2_\upsilon$, therefore,

$$a_{11} \sim \sigma^2\chi^2_\upsilon \quad \text{or} \quad \frac{a_{11}}{\sigma^2} \sim \chi^2_\upsilon.$$

Moments of the elements of the matrix A

Let $\underline{Z}_1, \underline{Z}_2, \cdots, \underline{Z}_\upsilon$ are independent, each with distribution $N_p(\underline{0}, \Sigma)$, then $A_{ij} = \sum_{\alpha=1}^{n-1} Z_{i\alpha}\, Z_{j\alpha}$.

$$E(A_{ij})=\sum_{\alpha=1}^{\upsilon}E(Z_{i\alpha}\,Z_{j\alpha})\,,\ \text{where}\ n-1=\upsilon$$

$$=\sum_{\alpha=1}^{\upsilon}\sigma_{ij}=\upsilon\,\sigma_{ij}\ \text{(we have infact prove that } E(A/n-1)=\Sigma\text{)}$$

$$E(A_{ij}\,A_{kl})=E\left(\sum_{\alpha=1}^{\upsilon}Z_{i\alpha}\,Z_{j\alpha}\right)\left(\sum_{\beta=1}^{\upsilon}Z_{k\beta}\,Z_{l\beta}\right)$$

$$=E\sum_{\alpha=1}^{\upsilon}(Z_{i\alpha}\,Z_{j\alpha}\,Z_{k\alpha}\,Z_{l\alpha})+E\sum_{\alpha,\beta=1}^{\upsilon}(Z_{i\alpha}\,Z_{j\alpha}\,Z_{k\beta}\,Z_{l\beta})\,,\ \alpha\neq\beta$$

$$=\sum_{\alpha=1}^{\upsilon}\ \text{(Fourth moment of a vector)}+\sum_{\alpha,\beta=1}^{\upsilon}E(Z_{i\alpha}\,Z_{j\alpha})\,E(Z_{k\beta}\,Z_{l\beta})\,,\ \alpha\neq\beta$$

$$=\upsilon(\sigma_{ij}\,\sigma_{kl}+\sigma_{ik}\,\sigma_{jl}+\sigma_{il}\,\sigma_{jk})+\upsilon(\upsilon-1)\sigma_{ij}\,\sigma_{kl}\,.$$

Therefore,

$$Cov(A_{ij},A_{kl})=E[A_{ij}-E(A_{ij})][A_{kl}-E(A_{kl})]=E(A_{ij}-\upsilon\sigma_{ij})(A_{kl}-\upsilon\sigma_{kl})$$

$$=E(A_{ij}\,A_{kl})-\upsilon\sigma_{kl}\,E(A_{ij})-\upsilon\sigma_{ij}\,E(A_{kl})+\upsilon^2\,\sigma_{ij}\,\sigma_{kl}$$

$$=\upsilon^2\sigma_{ij}\,\sigma_{kl}+\upsilon\sigma_{ik}\,\sigma_{jl}+\upsilon\sigma_{il}\,\sigma_{jk}-\upsilon^2\sigma_{ij}\,\sigma_{kl}-\upsilon^2\sigma_{ij}\,\sigma_{kl}+\upsilon^2\,\sigma_{ij}\,\sigma_{kl}$$

$$=\upsilon(\sigma_{ik}\,\sigma_{jl}+\sigma_{il}\,\sigma_{jk})$$

If $i=k$ and $j=l$, we obtain the variance of A_{ij}

$$Var(A_{ij})=E[A_{ij}-E(A_{ij})]^2=\upsilon(\sigma_{ii}\,\sigma_{jj}+\sigma_{ij}^2)\,.$$

Generalized variance

The multivariate analogue of the variance σ^2 of a univariate distribution is the covariance matrix Σ, and the determinant of covariance matrix is termed as generalized variance of the multivariate distribution. Similarly, the generalized variance of a sample $\underline{x}_1,\underline{x}_2,\cdots,\underline{x}_n$ is defined as

$$|\,S\,|=\left|\frac{1}{n-1}\sum_{\alpha=1}^{n}(\underline{x}_\alpha-\underline{\bar{x}})(\underline{x}_\alpha-\underline{\bar{x}})'\right|.$$

Distribution of the generalized variance

By definition, the distribution of $|\,S\,|$ is the same as the distribution of $\dfrac{1}{(n-1)^p}|\,A\,|$, where $A=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha\underline{Z}_\alpha'$, with $\underline{Z}_\alpha\sim N_p(\underline{0},\,\Sigma)$. Since Σ is symmetric and positive definite there exist a nonsingular matrix C such that $C\Sigma C'=I$.

Make the transformation

$\underline{Z}_\alpha^*=C\,\underline{Z}_\alpha$, then, $\underline{Z}_\alpha^*\sim N_p(\underline{0},\,I)$, are independent, and

$$\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha^*\underline{Z}_\alpha^{*\prime}=C\left(\sum_{\alpha}\underline{Z}_\alpha\underline{Z}_\alpha'\right)C'=C\,A\,C'=B\ \text{(say)}$$

We have

$$|\,C\,A\,C'\,|=|\,B\,|\quad\text{or}\quad|\,A\,|=\frac{1}{|\,C\,|^2}|\,B\,|.\ \text{But}\ C\Sigma C'=I\text{, then}$$

$$|\,C\,\Sigma\,C'\,|=1,\quad\text{or}\quad|\,C\,|^2=\frac{1}{|\Sigma|}\qquad\Rightarrow\ |\,A\,|=|\,B\,||\,\Sigma\,|.$$

Let

$$B=\begin{pmatrix}b_{11}&b_{12}&\cdots&b_{1p}\\b_{21}&b_{22}&\cdots&b_{2p}\\\vdots&\vdots&\vdots&\vdots\\b_{p1}&b_{p2}&\cdots&b_{pp}\end{pmatrix}$$

and $B_{ii}=\begin{pmatrix}b_{ii}&\underline{b}'_{(i)}\\\underline{b}_{(i)}&B_{i+1i+1}\end{pmatrix}$, where $\underline{b}'_{(i)}=(b_{ii+1},b_{ii+2},\cdots,b_{ip})$.

So that

$B_{11}=B$ and $B_{pp}=b_{pp}$, and

$$b_{ii.i+1,\cdots,p}=b_{ii}-\underline{b}'_{(i)}\,B_{i+1i+1}^{-1}\,\underline{b}_{(i)}$$

$=\dfrac{|\,B_{ii}\,|}{|\,B_{i+1i+1}\,|}$, as $A=\begin{pmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{pmatrix}$ and assume that A_{22} is square and nonsingular, then

$$|\,A\,|=\left|\,A_{11}-A_{12}\,A_{22}^{-1}\,A_{21}\right||\,A_{22}\,|=|\,A_{11.2}\,||A_{22}\,|.$$

Also

$$|\,B\,|=|\,B_{11}\,|=\frac{|\,B_{11}\,|}{|\,B_{22}\,|}\frac{|\,B_{22}\,|}{|\,B_{33}\,|}\cdots\frac{|\,B_{p-1p-1}\,|}{|\,B_{pp}\,|}|\,B_{pp}\,|$$

$$=(b_{11.2,\cdots,p})(b_{22.3,\cdots,p})\cdots(b_{p-1p-1.p})\,b_{pp}\,.$$

Since

$$B=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha^*\underline{Z}_\alpha^{*\prime}\,,\ \text{with}\ \underline{Z}_\alpha^*\sim N_p(\underline{0},\,I)$$

$$\Rightarrow\ b_{11.2,\cdots,p}=\sum_{\alpha=1}^{n-1-(p-1)}V_\alpha^2\,,\ \text{where}\ V_\alpha\sim N(0,\,1)\,.$$

Thus,

$b_{11.2,\cdots,p}$ has χ^2 – distribution with $(n-p)$ degree of freedom.

Since $\underline{Z}_\alpha^*$'s are independently distributed

$\Rightarrow$ $(b_{11.2,\cdots,p})(b_{22.3,\cdots,p})\cdots(b_{p-1\,p-1.\,p})b_{pp}$ are independently distributed

$\Rightarrow$ $b_{ii.i+1,\cdots,p}$ has χ^2 – distribution with $n-1-(p-i)$ degrees of freedom.

Therefore,

$|B|$ is distributed as $\chi^2_{n-1-(p-1)}\cdot\chi^2_{n-1-(p-2)}\cdot\cdots\cdot\chi^2_{n-2}\cdot\chi^2_{n-1}$.

Since $|S|=\dfrac{1}{(n-1)^p}|A|=\dfrac{1}{(n-1)^p}|\Sigma||B|$, then,

$|S|$ is distributed as $\dfrac{1}{(n-1)^p}|\Sigma|\ \chi^2_{n-1}\cdot\chi^2_{n-2}\cdots\chi^2_{n-1-(p-2)}\cdot\chi^2_{n-p}$.

For $p=1$

$$|S|=\frac{|\Sigma|}{n-1}\chi^2_{n-1}\ \Rightarrow\ \frac{(n-1)s_{11}}{\sigma_{11}}\sim\chi^2_{n-1}\quad\text{or}\quad\frac{a_{11}}{\sigma^2}\sim\chi^2_{n-1}.$$

Moments of sample generalized variance

Since $|S|$ is distributed as $\dfrac{|\Sigma|}{(n-1)^p}(y_1\,y_2\cdots y_p)$, where $y_i\sim\chi^2_{n-i}$, $i=1,2,\cdots,p$, are independent, thus

$$E|S|^h=\frac{|\Sigma|^h}{(n-1)^{ph}}E(y_1)^h\,E(y_2)^h\cdots E(y_p)^h.$$

Consider,

$y\sim\chi^2_\upsilon$, where, $\upsilon=n-1$, then

$$E(y^h)=\int y^h\,f(y)\,dy=\frac{1}{2^{\upsilon/2}\,\Gamma(\upsilon/2)}\int_0^\infty y^{(\upsilon/2)+h-1}e^{-y/2}\,dy$$

$$=\frac{1}{2^{\upsilon/2}\,\Gamma(\upsilon/2)}\left(\frac{\Gamma[(\upsilon/2)+h]}{(1/2)^{(\upsilon/2)+h}}\right),\ \text{because}\ \int_0^\infty x^{m-1}\,e^{-\lambda x}\,dx=\frac{\Gamma m}{\lambda^m}$$

$$=\frac{2^h\,\Gamma[(\upsilon/2)+h]}{\Gamma(\upsilon/2)}.$$

Therefore,

$$E|S|^h=\frac{|\Sigma|^h}{(n-1)^{ph}}\frac{2^h\,\Gamma\left(\frac{n-1}{2}+h\right)}{\Gamma(n-1)/2}\cdots\frac{2^h\,\Gamma\left(\frac{n-p}{2}+h\right)}{\Gamma(n-p)/2}$$

$$=\frac{|\Sigma|^h}{(n-1)^{ph}}\prod_{i=1}^{p}\frac{2^h\,\Gamma\left(\frac{n-i}{2}+h\right)}{\Gamma(n-i)/2}.$$

For $h=1$

$$E|S|=\frac{|\Sigma|}{(n-1)^p}\prod_{i=1}^{p}\frac{2\Gamma\left(\frac{n-i}{2}+1\right)}{\Gamma(n-i)/2}=\frac{|\Sigma|}{(n-1)^p}\prod_{i=1}^{p}\frac{2\left(\frac{n-i}{2}\right)\Gamma\left(\frac{n-i}{2}\right)}{\Gamma(n-i)/2}$$

$$=\frac{|\Sigma|}{(n-1)^p}\prod_{i=1}^{p}(n-i).$$

For $h=2$

$$E|S|^2=\frac{|\Sigma|^2}{(n-1)^{2p}}\prod_{i=1}^{p}\frac{4\Gamma\left(\frac{n-i}{2}+2\right)}{\Gamma(n-i)/2}=\frac{|\Sigma|^2}{(n-1)^{2p}}\prod_{i=1}^{p}\frac{4\left(\frac{n-i}{2}+1\right)\left(\frac{n-i}{2}\right)\Gamma\left(\frac{n-i}{2}\right)}{\Gamma(n-i)/2}$$

$$=\frac{|\Sigma|^2}{(n-1)^{2p}}\prod_{i=1}^{p}(n-i+2)(n-i).$$

$$V(|S|)=E|S|^2-(E|S|)^2=\frac{|\Sigma|^2}{(n-1)^{2p}}\prod_{i=1}^{p}(n-i+2)(n-i)-\frac{|\Sigma|^2}{(n-1)^{2p}}\prod_{i=1}^{p}(n-i)^2$$

$$=\frac{|\Sigma|^2}{(n-1)^{2p}}\prod_{i=1}^{p}2(n-i).$$

Note: For $p=2$

$$E|S|^h=\frac{|\Sigma|^h\,2^h}{(n-1)^{2h}}\prod_{i=1}^{2}\frac{\Gamma\left(\frac{n-i}{2}+h\right)}{\Gamma(n-i)/2}=\frac{|\Sigma|^h\,2^{2h}}{(n-1)^{2h}}\frac{\Gamma\left(\frac{n+2h-1}{2}\right)}{\Gamma(n-1)/2}\frac{\Gamma\left(\frac{n+2h-2}{2}\right)}{\Gamma(n-2)/2}$$

$$=\frac{|\Sigma|^h\,2^{2h}}{(n-1)^{2h}}\frac{\Gamma\left(\frac{n+2h-2}{2}+\frac{1}{2}\right)\Gamma\left(\frac{n+2h-2}{2}\right)}{\Gamma\left(\frac{n-2}{2}+\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}.$$

Using legendre's duplication formula

$$\Gamma\left(\alpha+\frac{1}{2}\right)\Gamma(\alpha)=\frac{\sqrt{\pi}\ \Gamma(2\alpha)}{2^{2\alpha}}$$

$$E|S|^h=\frac{|\Sigma|^h\,2^{2h}}{(n-1)^{2h}}\left(\frac{\sqrt{\pi}\ \Gamma(n+2h-2)/2^{(n+2h-2)}}{\sqrt{\pi}\ \Gamma(n-2)/2^{2(n-2)}}\right)$$

$$=\frac{|\Sigma|^h}{(n-1)^{2h}}\left(\frac{\Gamma(n+2h-2)}{\Gamma(n-2)}\right)\tag{4.14}$$

Let

$$X \sim \chi^2_{2m}, \text{ where } m \text{ is positive integer}$$

$$E(X^{2h}) = \frac{1}{2^m\,\Gamma(m)} \int_0^\infty X^{m+2h-1}\, e^{-X/2}\, dX = \frac{2^{2h}}{\Gamma(m)} \Gamma(m+2h)$$

$$\Rightarrow \quad E\left(\frac{X}{2}\right)^{2h} = \frac{\Gamma(m+2h)}{\Gamma(m)}$$

In view of equation (4.14) and $m = n-2$, we get

$$E|S|^h = \frac{|\Sigma|^h}{(n-1)^{2h}} E\left(\frac{X}{2}\right)^{2h} \Rightarrow E\left(\frac{2(n-1)|S|^{1/2}}{|\Sigma|^{1/2}}\right)^{2h} = E(X^{2h})$$

$$\Rightarrow \frac{2(n-1)|S|^{1/2}}{|\Sigma|^{1/2}} \sim \chi^2_{(2n-4)} \quad \text{or} \quad 2\left(\frac{|A|^{1/2}}{|\Sigma|^{1/2}}\right) \sim \chi^2_{(2n-4)}, \text{ because } X \sim \chi^2_{2(n-2)}.$$

Exercise: Let $\underline{X} \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ and let s_1^2 and s_2^2 denote the sample variances and r the sample correlation coefficient. Show that the statistic

$$u = \frac{2(n-1)\, s_1\, s_2}{\sigma_1\, \sigma_2} \left(\frac{1-r^2}{1-\rho^2}\right)^{1/2} \sim \chi^2_{2n-4}.$$

Theorem: Let $\underline{x}_1, \cdots, \underline{x}_n \quad (n \ge p+1)$ are independent, each with distribution $N_p(\underline{\mu}, \delta_{ij}\,\sigma_{ii})$, then the density of the sample correlation coefficients is

$$\frac{[\Gamma(\upsilon/2)]^p\, |r_{ij}|^{(\upsilon-p-1)/2}}{\pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(\upsilon-i+1)/2}, \text{ where } \upsilon = n-1, \quad i, j = 1, 2, \cdots, p \text{ and } \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \ne j \end{cases}$$

Proof: Let $A = \sum_{\alpha=1}^n (\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = \sum_{\alpha=1}^{n-1} \underline{Z}_\alpha \underline{Z}_\alpha'$, where $\underline{Z}_\alpha$ are independent, each with distribution $N_p(\underline{0}, \delta_{ij}\,\sigma_{ii})$, so $A \sim W_p(\upsilon, \delta_{ij}\,\sigma_{ii})$, and

$$f(A) = \frac{|a_{ij}|^{(\upsilon-p-1)/2}}{2^{\upsilon p/2}\, \pi^{p(p-1)/4} \left(\prod_{i=1}^p \sigma_{ii}\right)^{\upsilon/2} \prod_{i=1}^p \Gamma(\upsilon-i+1)/2} \exp\left\{-\frac{1}{2}\sum_i \frac{a_{ii}}{\sigma_{ii}}\right\}, \text{ because}$$

$$|\Sigma| = \begin{vmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{vmatrix} = \prod_{i=1}^p \sigma_{ii}.$$

Make the transformation

$$a_{ij} = \sqrt{a_{ii}}\,\sqrt{a_{jj}}\; r_{ij}, \quad i \ne j, \quad i < j \tag{4.15}$$

$$a_{ii} = a_{ii} \tag{4.16}$$

The Jacobian is the product of the Jacobian of (4.15) and (4.16) for a_{ii} fixed, the Jacobian of (4.15) is the determinant of a $\frac{p(p-1)}{2}$ – order diagonal matrix with diagonal elements $\sqrt{a_{ii}}\,\sqrt{a_{jj}}$

$$J = \begin{vmatrix} \sqrt{a_{11}}\sqrt{a_{22}} &&&&&&& \\ & \sqrt{a_{11}}\sqrt{a_{33}} &&&&&& \\ && \ddots &&&&& \\ &&& \sqrt{a_{11}}\sqrt{a_{pp}} &&&& \\ &&&& \sqrt{a_{22}}\sqrt{a_{33}} &&& \\ &&&&& \ddots && \\ &&&&&& \sqrt{a_{22}}\sqrt{a_{pp}} & \\ &&&&&&& \ddots \end{vmatrix}$$

$$= \prod_{i=1}^p a_{ii}^{(p-1)/2}, \text{ because each particulars subscript appears } i < j,\ (p-1) \text{ times.}$$

Thus,

$$f(A) = \frac{\left|\sqrt{a_{ii}}\,\sqrt{a_{jj}}\; r_{ij}\right|^{(\upsilon-p-1)/2}}{2^{\upsilon p/2}\, \pi^{p(p-1)/4} \left(\prod_{i=1}^p \sigma_{ii}\right)^{\upsilon/2} \prod_{i=1}^p \Gamma(\upsilon-i+1)/2} \exp\left\{-\frac{1}{2}\sum_i \frac{a_{ii}}{\sigma_{ii}}\right\} \prod_{i=1}^p a_{ii}^{(p-1)/2}.$$

But

$$\left|\sqrt{a_{ii}}\,\sqrt{a_{jj}}\; r_{ij}\right|$$

$$= \left| \begin{pmatrix} \sqrt{a_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{a_{22}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{a_{pp}} \end{pmatrix} \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \sqrt{a_{11}} & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \sqrt{a_{pp}} \end{pmatrix} \right|$$

$$= |r_{ij}| \prod_{i=1}^p a_{ii}, \text{ then}$$

$$f(A) = \frac{|r_{ij}|^{(\upsilon-p-1)/2}}{\pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(\upsilon-i+1)/2} \prod_{i=1}^p \left(\frac{a_{ii}^{(\upsilon-1)/2} \exp\left\{-\frac{1}{2}\sum_i \frac{a_{ii}}{\sigma_{ii}}\right\}}{2^{\upsilon/2}\, (\sigma_{ii})^{\upsilon/2}} \right)$$

$$= \frac{|r_{ij}|^{(\upsilon-p-1)/2}}{\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}\prod_{i=1}^{p}\int_0^\infty \frac{a_{ii}^{(\upsilon-1)/2}\exp\left\{-\frac{1}{2}\frac{a_{ii}}{\sigma_{ii}}\right\}}{2^{\upsilon/2}(\sigma_{ii})^{\upsilon/2}}da_{ii}.$$

Consider

$$B = \int_0^\infty \frac{a_{ii}^{(\upsilon-1)/2}\exp\left\{-\frac{1}{2}\frac{a_{ii}}{\sigma_{ii}}\right\}}{2^{\upsilon/2}(\sigma_{ii})^{\upsilon/2}}da_{ii}.$ Put $\frac{a_{ii}}{2\sigma_{ii}} = u_i$, or $d\,a_{ii} = 2\sigma_{ii}\,du_i$, then

$$B = \int_0^\infty (u_i)^{(\upsilon/2)-1}\exp(-u_i)\,du_i = \Gamma(\upsilon/2), \quad \forall\ i.$$

Hence, the density function of r_{ij} is

$$f(r_{ij}) = \frac{|r_{ij}|^{(\upsilon-p-1)/2}\prod_{i=1}^{p}\Gamma(\upsilon/2)}{\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2} = \frac{|r_{ij}|^{(\upsilon-p-1)/2}[\Gamma(\upsilon/2)]^p}{\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}.$$

Exercise: Find the $E|A|^h$ directly from $W(\upsilon, \Sigma)$.

Solution: We know that, if $A \sim W(\upsilon, \Sigma)$, then

$$f(A) = \frac{|A|^{(\upsilon-p-1)/2}}{2^{\upsilon p/2}\pi^{p(p-1)/4}|\Sigma|^{\upsilon/2}\prod_{i=1}^{p}\Gamma\{(\upsilon-i+1)/2\}}\exp\left\{-\frac{1}{2}tr\,A\Sigma^{-1}\right\}.$$

Now

$$\int_A f(A)\,dA = 1$$

$$\Rightarrow \int_A |A|^{(\upsilon-p-1)/2}\exp\left\{-\frac{1}{2}tr\,A\Sigma^{-1}\right\}dA = 2^{\upsilon p/2}\pi^{p(p-1)/4}|\Sigma|^{\upsilon/2}\prod_{i=1}^{p}\Gamma\{(\upsilon-i+1)/2\} \tag{4.17}$$

$$E|A|^h = \int_A |A|^h f(A)\,dA$$

$$= \frac{|A|^{(\upsilon-p-1+2h)/2}}{2^{\upsilon p/2}\pi^{p(p-1)/4}|\Sigma|^{\upsilon/2}\prod_{i=1}^{p}\Gamma\{(\upsilon-i+1)/2\}}\exp\left\{-\frac{1}{2}tr\,A\Sigma^{-1}\right\}dA$$

Now using equation (4.17), we have

$$\int_A |A|^{[(\upsilon+2h)-p-1]/2}\exp\left\{-\frac{1}{2}tr\,A\Sigma^{-1}\right\}dA$$

$$= 2^{(\upsilon+2h)p/2}\pi^{p(p-1)/4}|\Sigma|^{(\upsilon+2h)/2}\prod_{i=1}^{p}\Gamma[(\upsilon+2h)-i+1]/2.$$

Therefore,

$$E|A|^h = \frac{2^{(\upsilon+2h)p/2}\pi^{p(p-1)/4}|\Sigma|^{(\upsilon+2h)/2}\prod_{i=1}^{p}\Gamma[(\upsilon+2h)-i+1]/2}{2^{\upsilon p/2}\pi^{p(p-1)/4}|\Sigma|^{\upsilon/2}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}$$

$$= \frac{2^{ph/2}|\Sigma|^h\prod_{i=1}^{p}\Gamma\ (\upsilon+2h-i+1)/2}{\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}$$

$$= \frac{2^{ph/2}|\Sigma|^h\prod_{i=1}^{p}\Gamma\left(\frac{n-i}{2}+h\right)}{\prod_{i=1}^{p}\Gamma(n-i)/2}.$$

Result:

Let us suppose that $x_1,\cdots,x_n$ constitute a random sample of size n from a population whose density at x is $f(x;\theta)$ and Ω is the set of values, which can be taken on by the parameter θ (Ω is the parameter space for θ) and the parameter space for θ is partitioned into the disjoint sets ω and ω', according to the null hypothesis

$H_0: \theta \in \omega$, against $H_A: \theta \in \omega'$, if

$\lambda = \frac{\max L_0}{\max L}$ is referred to as a value of the **likelihood ratio statistic** λ,

where $\max L_0$ and $\max L$ are the maximum values of the likelihood function for all values of θ in ω and Ω respectively. Since $\max L_0$ and $\max L$ are both values of a likelihood function, and therefore, never negative, it follows that $\lambda \ge 0$, also, since ω is a subset of the Ω, it follows that $\lambda \le 1$, then the critical region of the form $\lambda \le k$, $(0<k<1)$, defines a **likelihood ratio test** of the null hypothesis $H_0: \theta \in \omega$ against the alternative hypothesis $H_A: \theta \in \omega'$. Such that

$\Pr[\lambda \le k \mid H_0] = \alpha$.

Example: Let $H_0: \mu = \mu_0$, against $H_A: \mu \ne \mu_0$

On the basis of a random sample of size n from normal population with the known variance σ^2, the critical region of the likelihood ratio test is obtained as follows:

Since ω contains only μ_0, it follows that $\mu = \mu_0$, and Ω is the set of all real numbers, it follows that $\hat{\mu} = \bar{x}$. Thus

$$\max L_0 = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2\right]$$

and $\max L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \bar{x})^2\right]$, then the value of the likelihood ratio statistic becomes,

$$\lambda = \frac{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \bar{x})^2\right]} = \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2\right]$$

Hence, the critical region of the likelihood ratio test is

$$\exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2\right] \le k \text{, since } \lambda \le k$$

and after taking logarithms and dividing by $\frac{n}{2\sigma^2}$, it becomes

$$(\bar{x} - \mu_0)^2 \ge -\frac{2\sigma^2}{n}\ln k \text{, } \ln k \text{ is negative in view of } 0 < k < 1$$

or $|\bar{x} - \mu_0| \ge K$,

where K will have to be determined so that the size of the critical region is α i.e.

$$\Pr[|\bar{x} - \mu_0| \ge K] = \alpha. \tag{a}$$

Since $\bar{x}$ has a normal distribution with mean μ_0 and variance σ^2/n, i.e.

$$\bar{x} \sim N(\mu_0, \sigma^2/n)\text{, then } Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

For the given value of α, we find a value Z_α (say) of standard normal variate by the following equation

$$\Pr[|Z| \ge Z_{\alpha/2}] = \alpha \quad \text{or} \quad \Pr\left[|\bar{x} - \mu_0| \ge \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right] = \alpha \tag{b}$$

From equation (a) and (b), we get

$$K = \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}.$$

Therefore, the critical region of the likelihood ratio test is

$$\Pr\left[|\bar{x} - \mu_0| \ge \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right] = \alpha.$$

Note: It was easy to find the constant that made the size of the critical region equal to α, because we were able to refer to the known distribution of $\bar{x}$, and did not have to derive the distribution of the likelihood ratio statistic λ, itself. Since the distribution of λ is generally quite complicated, which makes it difficult to evaluate k, it is often preferable to use the following approximation. For large n, the distribution of $-2\ln\lambda$ approaches, under very general conditions the chi-square distribution with 1 degree of freedom, i.e. $-2\ln\lambda \sim \chi_1^2$. In above example, we find that

$$-2\ln\lambda = \frac{n}{\sigma^2}(\bar{x} - \mu_0)^2 = \left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)^2.$$

Testing independence of sets of variates

Let the $p-$component vector $\underline{X}$ be distributed according to $N_p(\underline{\mu}, \Sigma)$. We partition $\underline{X}$ into q sub vectors with $p_1, p_2, \cdots, p_q$ components respectively, that is

$\underline{X} = \begin{pmatrix} \underline{X}^{(1)} \\ \vdots \\ \underline{X}^{(q)} \end{pmatrix}$, $p_1 + p_2 + \cdots + p_q = p$. The vector of mean $\underline{\mu}$ and the covariance matrix Σ are partitioned similarly,

$$\underline{\mu} = \begin{pmatrix} \underline{\mu}^{(1)} \\ \vdots \\ \underline{\mu}^{(q)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1q} \\ \vdots & \vdots & \vdots \\ \Sigma_{q1} & \cdots & \Sigma_{qq} \end{pmatrix}.$$

$H_0:$ $\underline{X}^{(1)}, \underline{X}^{(2)}, \cdots, \underline{X}^{(q)}$ are mutually independently distributed or equivalently $\Sigma_{ij} = \underline{0}$, for $i \ne j$, where $\Sigma_{ij} = E(\underline{X}^{(i)} - \underline{\mu}^{(i)})(\underline{X}^{(j)} - \underline{\mu}^{(j)})'$.

Under H_0, Σ is of the form

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \Sigma_{qq} \end{pmatrix} = \Sigma_0 \text{ (say)}$$

Let $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ be n independent observations on $\underline{X}$, the likelihood ratio criterion is

$$\lambda = \frac{\max L(\underline{\mu}, \Sigma_0)}{\max L(\underline{\mu}, \Sigma)} = \frac{\max L_0}{\max L}, \tag{4.18}$$

where

$$L(\underline{\mu}, \Sigma) = \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underline{x}_\alpha - \underline{\mu})'\Sigma^{-1}(\underline{x}_\alpha - \underline{\mu})\right],$$

and the numerator of (4.18) is the maximum of L for $\underline{\mu}, \Sigma \in \omega$ restricted by H_0 and the denominator is the maximum of L over the entire parametric space Ω. Now consider the likelihood function over the entire parametric space

$$\max L_\Omega = \frac{1}{(2\pi)^{np/2}\left|\frac{A}{n}\right|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})'\left(\frac{A}{n}\right)^{-1}(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})\right]$$

$$= \frac{n^{np/2}}{(2\pi)^{np/2}|A|^{n/2}} \exp\left[-\frac{1}{2} tr\left(\frac{A}{n}\right)^{-1}\sum_\alpha (\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})'\right]$$

$$= \frac{n^{np/2}}{(2\pi)^{np/2}|A|^{n/2}} \exp\left(-\frac{1}{2} tr\, n\, I\right)$$

$$= \frac{n^{np/2}}{(2\pi)^{np/2}|A|^{n/2}} \exp\left(-\frac{1}{2} np\right), \text{ where trace } nI_{p\times p} = np$$

Similarly, under H_0, the likelihood function is

$$\max L_\omega = \frac{1}{(2\pi)^{np/2}\left|\hat{\Sigma}_0\right|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underset{\sim}{x}_\alpha - \underset{\sim}{\hat{\mu}})'\hat{\Sigma}_0^{-1}(\underset{\sim}{x}_\alpha - \underset{\sim}{\hat{\mu}})\right]$$

$$= \prod_{i=1}^q \frac{1}{(2\pi)^{np_i/2}\left|\hat{\Sigma}_{ii}\right|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underset{\sim}{x}_\alpha^{(i)} - \underset{\sim}{\hat{\mu}}^{(i)})'\hat{\Sigma}_{ii}^{-1}(\underset{\sim}{x}_\alpha^{(i)} - \underset{\sim}{\hat{\mu}}^{(i)})\right]$$

$$= \prod_{i=1}^q \frac{1}{(2\pi)^{np_i/2}\left|\frac{A_{ii}}{n}\right|^{n/2}} \exp\left(-\frac{1}{2} np_i\right)$$

$$= \frac{1}{(2\pi)^{np/2}\prod_{i=1}^q \frac{1}{n^{np_i/2}}\left|A_{ii}\right|^{n/2}} \exp\left(-\frac{1}{2} np\right).$$

Thus, the likelihood criterion becomes

$$\lambda = \frac{\frac{1}{n^{np/2}}|A|^{n/2}}{\frac{1}{n^{np/2}}\prod_{i=1}^q |A_{ii}|^{n/2}} = \frac{|A|^{n/2}}{\prod_{i=1}^q |A_{ii}|^{n/2}} = V^{n/2}$$

and $H_0 : \Sigma_{ij} = \underset{\sim}{0}$ is rejected if $\lambda < \lambda_0$, where λ_0 is so chosen so as to have level α.

Limiting results

By the large sample general result about likelihood ratio criterion is

$-2\ln\lambda \sim \chi_f^2$ approximately, where,

$$f = \frac{p(p+1)}{2} - \sum_{i=1}^q \frac{p_i(p_i+1)}{2} = \frac{1}{2}\left(p^2 - \sum_{i=1}^q p_i^2\right).$$

A better approximation to the distribution of λ is obtained by looking at the moments of λ (or of V), one infact has

$-m\ln V \sim \chi_f^2$ approximately, where

$$f = \frac{p(p+1)}{2} - \sum_{i=1}^q \frac{p_i(p_i+1)}{2} = \frac{1}{2}\left(p^2 - \sum_{i=1}^q p_i^2\right), \text{ and } m = n - \frac{3}{2} - \frac{p^3 - \sum_{i=1}^q p_i^3}{3\left(p^2 - \sum_{i=1}^q p_i^2\right)}.$$

Exercise: Show that the likelihood ratio criterion λ for testing independence of sets of vectors can be written as

$$\lambda = \frac{|R|^{n/2}}{\prod_{i=1}^q |R_{ii}|^{n/2}}, \text{ where } R = (r_{jk}) = \begin{pmatrix} R_{11} & \cdots & R_{1q} \\ \vdots & \vdots & \vdots \\ R_{q1} & \cdots & R_{qq} \end{pmatrix}.$$

Solution: We have, $\lambda = \dfrac{|A|^{n/2}}{\prod_{i=1}^q |A_{ii}|^{n/2}} = V^{n/2}$.

Define,

$$r_{jk} = \frac{a_{jk}}{\sqrt{a_{jj}}\sqrt{a_{kk}}}, \Rightarrow a_{jk} = r_{jk}\sqrt{a_{jj}\, a_{kk}}$$

$$\Rightarrow |A| = |R|\prod_{j=1}^p a_{jj}, \text{ where } p = p_1 + p_2 + \cdots + p_q, \text{ and } |A_{ii}| = |R_{ii}| \prod_{k=p_1+\cdots+p_{i-1}+1}^{p_1+\cdots+p_{i-1}+p_i} a_{kk}$$

$$\Rightarrow V = \frac{|A|}{\prod_{i=1}^q |A_{ii}|} = \frac{|R|\prod_{j=1}^p a_{jj}}{\prod_{i=1}^q\left(|R_{ii}|\prod_{k=p_1+\cdots+p_{i-1}+1}^{p_1+\cdots+p_{i-1}+p_i} a_{kk}\right)} = \frac{|R|\prod_{j=1}^p a_{jj}}{\prod_{i=1}^q |R_{ii}|\prod_{k=1}^p a_{kk}} = \frac{|R|}{\prod_{i=1}^q |R_{ii}|}.$$

Thus, $\lambda = \dfrac{|R|^{n/2}}{\prod_{i=1}^q |R_{ii}|^{n/2}}$.

Exercise: Let C_i be an arbitrary nonsingular matrix of order p_i and let

$$C = \begin{pmatrix} C_1 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & C_q \end{pmatrix} \text{ and } \underset{\sim}{x}_\alpha^* = C\,\underset{\sim}{x}_\alpha + \underset{\sim}{d}.$$

Show that the likelihood ratio criterion for independence interms of $\underline{x}_\alpha^*$ is identical to the criterion interms of $\underline{x}_\alpha$.

Solution: The likelihood ratio criterion for independence interms of $\underline{x}_\alpha$ is $V = \dfrac{|A|}{\prod_{i=1}^{q} |A_{ii}|}$,

define,

$$A^* = \sum_{\alpha=1}^{n} (\underline{x}_\alpha^* - \underline{\bar{x}}^*)(\underline{x}_\alpha^* - \underline{\bar{x}}^*)', \text{ where, } \underline{\bar{x}}_\alpha^* = \frac{1}{n}\sum_{\alpha=1}^{n}(C\,\underline{x}_\alpha + \underline{d}) = C\,\underline{\bar{x}} + \underline{d}$$

$$A^* = \sum_{\alpha=1}^{n} C(\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})'C' = C\,A\,C'.$$

Similarly,

$$A_{ij}^* = \sum_{\alpha=1}^{n} (\underline{x}_\alpha^{*(i)} - \underline{\bar{x}}^{*(i)})(\underline{x}_\alpha^{*(j)} - \underline{\bar{x}}^{*(j)})'$$

$$= C_i \left(\sum_{\alpha=1}^{n} (\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(j)} - \underline{\bar{x}}^{(j)})' \right) C_j' = C_i A_{ij} C_j'.$$

Thus,

$$V^* = \frac{|A^*|}{\prod_{i=1}^{q} |A_{ii}^*|} = \frac{|C\,A\,C'|}{\prod_{i=1}^{q} |C_i\, A_{ii}\, C_i'|}$$

$$= \frac{|C||A||C'|}{\prod_{i=1}^{q} |C_i||A_{ii}||C_i'|} = \frac{|A|}{\prod_{i=1}^{q} |A_{ii}|} = V, \text{ because } \prod_{i=1}^{q} |C_i| = |C|.$$

Therefore, the test is invariant with respect to linear transformation within each set.

Testing equality of covariance matrices

Let $\underline{x}_\alpha^{(i)}$ $\alpha = 1, 2, \cdots, n_i;$ $i = 1, 2, \cdots, q$ be an observation from $N_p(\underline{\mu}^{(i)}, \Sigma_i)$

$H_0:\ \Sigma_1 = \Sigma_2 = \cdots = \Sigma_q = \Sigma$ (say), versus $H_A:\ \Sigma_i \neq \Sigma_j$, for $i \neq j$.

Let

$$A_i = \sum_{\alpha=1}^{n_i} (\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})', \text{ and } A = \sum_{i=1}^{q} A_i.$$

The likelihood function is

$$L(\underline{\mu}, \Sigma) = \prod_{i=1}^{q} \left[\frac{1}{(2\pi)^{n_i p/2} |\Sigma_i|^{n_i/2}} \exp\left\{ -\frac{1}{2} \sum_{\alpha=1}^{n_i} (\underline{x}_\alpha^{(i)} - \underline{\mu}^{(i)})' \Sigma_i^{-1} (\underline{x}_\alpha^{(i)} - \underline{\mu}^{(i)}) \right\} \right]$$

and the likelihood ratio criterion is

$$\lambda = \frac{\max L(\underline{\mu}, \Sigma_0)}{\max L(\underline{\mu}, \Sigma)} = \frac{\max L_0}{\max L}$$

the numerator is the maximum of L for $\underline{\mu}, \Sigma \in \omega$ restricted by H_0 and the denominator is the maximum of L over the entire parametric space Ω. Now

$$\max L_\Omega = \prod_{i=1}^{q} \left[\frac{1}{(2\pi)^{n_i p/2} \left| \frac{A_i}{n_i} \right|^{n_i/2}} \exp\left\{ -\frac{1}{2} \sum_{\alpha=1}^{n_i} (\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})' \left(\frac{A_i}{n_i} \right)^{-1} (\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)}) \right\} \right]$$

$$= \prod_{i=1}^{q} \left(\frac{1}{(2\pi)^{n_i p/2} \left| \frac{A_i}{n_i} \right|^{n_i/2}} \exp\left\{ -\frac{1}{2} n_i p \right\} \right) = \frac{1}{(2\pi)^{np/2} \prod_{i=1}^{q} \left| \frac{A_i}{n_i} \right|^{n_i/2}} \exp\left(-\frac{1}{2} np \right).$$

Similarly, under H_0, the likelihood function is

$$\max L_\omega = \prod_{i=1}^{q} \left[\frac{1}{(2\pi)^{n_i p/2} |\hat{\Sigma}|^{n_i/2}} \exp\left\{ -\frac{1}{2} \sum_{\alpha=1}^{n_i} (\underline{x}_\alpha^{(i)} - \underline{\hat{\mu}}^{(i)})' \hat{\Sigma}^{-1} (\underline{x}_\alpha^{(i)} - \underline{\hat{\mu}}^{(i)}) \right\} \right]$$

$$= \frac{1}{(2\pi)^{np/2} \left| \frac{A}{n} \right|^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{q} \sum_{\alpha=1}^{n_i} (\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})' \left(\frac{A}{n} \right)^{-1} (\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)}) \right)$$

$$= \frac{1}{(2\pi)^{np/2} \left| \frac{A}{n} \right|^{n/2}} \exp\left[-\frac{1}{2} np \right].$$

Thus, the likelihood criterion becomes

$$\lambda = \frac{\prod_{i=1}^{q} \left| \frac{A_i}{n_i} \right|^{n_i/2}}{\left| \frac{A}{n} \right|^{n/2}} = \frac{\prod_{i=1}^{q} \frac{1}{n_i^{n_i p/2}} |A_i|^{n_i/2}}{\frac{1}{n^{np/2}} |A|^{n/2}} = \frac{n^{np/2}}{\prod_{i=1}^{q} n_i^{n_i p/2}} \left(\frac{\prod_{i=1}^{q} |A_i|^{n_i/2}}{|A|^{n/2}} \right).$$

Bartlett has suggested that the likelihood ratio criterion be modified by replacing the sample size n_i by the degree of freedom $\upsilon_i = n_i - 1$, the modified criterion is

$$\lambda^* = \frac{\upsilon^{\upsilon p/2}}{\prod_{i=1}^{q} \upsilon_i^{\upsilon_i p/2}} \left(\frac{\prod_{i=1}^{q} |A_i|^{\upsilon_i/2}}{|A|^{\upsilon/2}} \right) = \frac{\prod_{i=1}^{q} |S_i|^{\upsilon_i/2}}{|S|^{\upsilon/2}}, \text{ where, } \upsilon = \sum_{i=1}^{q} \upsilon_i = n - q \text{ and}$$

$S_i = \dfrac{A_i}{n_i - 1}$, an unbiased estimate of Σ_i, and $S = \dfrac{A}{\upsilon} = \dfrac{1}{\upsilon}\sum_i A_i = \dfrac{1}{\upsilon}\sum_i \upsilon_i S_i$, be the pooled estimate of the common covariance matrix.

The test for $H_0: \Sigma_1 = \Sigma_2 = \cdots = \Sigma_q$ is to reject H_0 if $\lambda^* < \lambda_0^*$, where λ_0^* is so chosen so as to have $\Pr[\lambda^* < \lambda_0^* \mid H_0] = \alpha$.

Box has shown that

$$-2\rho \ln \lambda^* \sim \chi_f^2 \text{ approx.},$$

where, $\rho = 1 - \left(\sum_{i=1}^{q} \dfrac{1}{\upsilon_i} - \dfrac{1}{\upsilon}\right)\left(\dfrac{2p^2 + 3p - 1}{6(p+1)(q-1)}\right)$, and $f = p(p+1)(q-1)/2$.

Testing the hypothesis that a covariance matrix is equal to a given matrix

Let $\underline{x}_\alpha$ be a random sample of size n from $N_p(\underline{\mu}, \Sigma)$ and $H_0: \Sigma = I$.

The likelihood ratio criterion is

$$\lambda = \frac{\max L(\underline{\mu}, \Sigma_0)}{\max L(\underline{\mu}, \Sigma)} = \frac{\max L_0}{\max L},$$

the numerator is the maximum of L for $\underline{\mu}, \Sigma \in \omega$ restricted by H_0 and the denominator is the maximum of L over the entire parametric space Ω,

where,

$$L(\underline{\mu}, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underline{x}_\alpha - \underline{\mu})' \Sigma^{-1} (\underline{x}_\alpha - \underline{\mu})\right]$$

The likelihood function over the entire parametric space is

$$\max L_\Omega = \frac{1}{(2\pi)^{np/2} \left|\dfrac{A}{n}\right|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underline{x}_\alpha - \underline{\bar{x}})' \left(\frac{A}{n}\right)^{-1} (\underline{x}_\alpha - \underline{\bar{x}})\right]$$

$$= \frac{n^{np/2}}{(2\pi)^{np/2} |A|^{n/2}} \exp\left(-\frac{1}{2} np\right).$$

The likelihood function over ω, the parametric space as restricted by $H_0: \Sigma = I$ is

$$\max L_\omega = \frac{1}{(2\pi)^{np/2} |I|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underline{x}_\alpha - \underline{\bar{x}})' I (\underline{x}_\alpha - \underline{\bar{x}})\right]$$

$$= \frac{1}{(2\pi)^{np/2}} \exp\left[-\frac{1}{2} tr\, I \sum_\alpha (\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})'\right] = \frac{1}{(2\pi)^{np/2}} \exp\left(-\frac{1}{2} tr\, A\right).$$

Therefore, the likelihood criterion becomes

$$\lambda = \frac{\exp\left(-\dfrac{1}{2} tr\, A\right)}{\dfrac{n^{np/2}}{|A|^{n/2}} \exp\left(-\dfrac{1}{2} np\right)} = \left(\frac{e}{n}\right)^{np/2} |A|^{n/2} \exp\left(-\frac{1}{2} tr\, A\right)$$

the null hypothesis $H_0: \Sigma = I$ is rejected if $\lambda < \lambda_0$, where λ_0 is so chosen so as to have level α.

An approximation to the distribution of λ is obtained by looking at the moment of λ, one infact has

$$-2 \ln \lambda \sim \chi^2_{p(p+1)/2}.$$

Theorem: Given $p-$component observation vector $\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_n$ from $N_p(\underline{\upsilon}, \Phi)$, the likelihood ratio criterion for testing the hypothesis $H_0: \underline{\upsilon} = \underline{\upsilon}_0$, $\Phi = \Phi_0$ is

$$\lambda = \left(\frac{e}{n}\right)^{np/2} \left|B\Phi_0^{-1}\right|^{n/2} \exp\left\{-\frac{1}{2}[tr\, B\Phi_0^{-1} + n(\underline{\bar{y}} - \underline{\upsilon}_0)'\Phi_0^{-1}(\underline{\bar{y}} - \underline{\upsilon}_0)]\right\},$$ where,

$B = \sum_{\alpha=1}^{n} (\underline{y}_\alpha - \underline{\bar{y}})(\underline{y}_\alpha - \underline{\bar{y}})'$, when the null hypothesis is true, then, $-2\ln\lambda \sim \chi^2_{p(p+1)+p}$.

Proof: Given $\underline{y}_\alpha \sim N_p(\underline{\upsilon}, \Phi)$, and $H_0: \underline{\upsilon} = \underline{\upsilon}_0$, $\Phi = \Phi_0$.

Let

$\underline{X} = C(\underline{Y} - \underline{\upsilon}_0)$, with C is a nonsingular matrix

$E\,\underline{X} = \underline{0}$, and $\Sigma_{\underline{X}} = C\Phi_0^{-1}C' = I$, under H_0.

$\Rightarrow C'C = \Phi_0^{-1}$,

then $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ constitutes a sample from $N_p(\underline{\mu}, \Sigma)$ and the hypothesis is

$H_0: \underline{\mu} = \underline{0}$, $\Sigma = I$

$$\max L_\Omega = \frac{n^{np/2}}{(2\pi)^{np/2}|A|^{n/2}} \exp\left(-\frac{1}{2}np\right), \text{ and}$$

$$\max L_\omega = \frac{1}{(2\pi)^{np/2}} \exp\left(-\frac{1}{2}\sum_\alpha \underline{x}_\alpha' \underline{x}_\alpha\right).$$

Thus, the likelihood ratio criterion is

$$\lambda = \frac{\max L(\underline{\mu}, \Sigma_0)}{\max L(\underline{\mu}, \Sigma)} = \frac{\exp\left(-\dfrac{1}{2}\sum_\alpha \underline{x}_\alpha' \underline{x}_\alpha\right)}{\dfrac{n^{np/2}}{|A|^{n/2}} \exp\left(-\dfrac{1}{2}np\right)} = \left(\frac{e}{n}\right)^{np/2} |A|^{n/2} \exp\left(-\frac{1}{2}\sum_\alpha \underline{x}_\alpha' \underline{x}_\alpha\right).$$

Now let us return to the observations $\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_n$, then

$$\sum_\alpha \underline{x}_\alpha' \underline{x}_\alpha = \sum_\alpha (\underline{y}_\alpha - \underline{\upsilon}_0)' C' C (\underline{y}_\alpha - \underline{\upsilon}_0) = \sum_\alpha (\underline{y}_\alpha - \underline{\upsilon}_0)' \Phi_0^{-1} (\underline{y}_\alpha - \underline{\upsilon}_0)$$

$$= tr \sum_\alpha (\underline{y}_\alpha - \underline{\upsilon}_0)' \Phi_0^{-1} (\underline{y}_\alpha - \underline{\upsilon}_0) = tr\, \Phi_0^{-1} \sum_\alpha (\underline{y}_\alpha - \underline{\upsilon}_0)(\underline{y}_\alpha - \underline{\upsilon}_0)'$$

$$= tr\, \Phi_0^{-1} \left[\sum_\alpha (\underline{y}_\alpha - \underline{\bar{y}})(\underline{y}_\alpha - \underline{\bar{y}})' + n(\underline{\bar{y}} - \underline{\upsilon}_0)(\underline{\bar{y}} - \underline{\upsilon}_0)' \right]$$

$$= tr\, \Phi_0^{-1} [B + n(\underline{\bar{y}} - \underline{\upsilon}_0)(\underline{\bar{y}} - \underline{\upsilon}_0)'] = tr\,(\Phi_0^{-1} B) + tr\, \Phi_0^{-1}\, n(\underline{\bar{y}} - \underline{\upsilon}_0)(\underline{\bar{y}} - \underline{\upsilon}_0)'$$

$$= tr\,(\Phi_0^{-1} B) + n(\underline{\bar{y}} - \underline{\upsilon}_0) \Phi_0^{-1} (\underline{\bar{y}} - \underline{\upsilon}_0)'$$

and

$$A = \sum_\alpha (\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = \sum_\alpha [\{C(\underline{y}_\alpha - \underline{\upsilon}_0) - C(\underline{\bar{y}} - \underline{\upsilon}_0)\}\{C(\underline{y}_\alpha - \underline{\upsilon}_0) - C(\underline{\bar{y}} - \underline{\upsilon}_0)\}']$$

$$= \sum_\alpha [C(\underline{y}_\alpha - \underline{\bar{y}})]\,[C(\underline{y}_\alpha - \underline{\bar{y}})]' = C \sum_\alpha (\underline{y}_\alpha - \underline{\bar{y}})(\underline{y}_\alpha - \underline{\bar{y}})' C' = C B C'.$$

Thus,

$\left| A \right| = \left| C B C' \right| = \dfrac{|B|}{|\Phi_0|} = \left| B \Phi_0^{-1} \right|$, because $C \Phi_0 C' = I$, then $\left| C \Phi_0 C' \right| = 1$, and

$$\left| C C' \right| = \left| \Phi_0 \right| = 1 \;\Rightarrow\; \left| C C' \right| = \frac{1}{|\Phi_0|}.$$

Therefore, the likelihood ratio criterion becomes

$$\lambda = \left(\frac{e}{n}\right)^{np/2} \left| B \Phi_0^{-1} \right|^{n/2} \exp\left\{ -\frac{1}{2}[tr\, B \Phi_0^{-1} + n(\underline{\bar{y}} - \underline{\upsilon}_0)' \Phi_0^{-1} (\underline{\bar{y}} - \underline{\upsilon}_0)] \right\}.$$

MULTIPLE AND PARTIAL CORRELATIONS

Result:

$X_1^* = E(X_1 \mid X_2 = x_2) = a + bx_2$ is the regression line of X_1 on X_2, b is the regression coefficient. It follows that

$$\int x_1\, f(x_1 \mid x_2)\, dx_1 = a + bx_2 \tag{5.1}$$

On multiplying both the sides of equation (5.1) by $f(x_2)$, and integrating with respect to x_2, we get

$$\int\int x_1\, f(x_1 | x_2)\, f(x_2)\, dx_1\, dx_2 = a\int f(x_2)\, dx_2 + b\int x_2\, f(x_2)\, dx_2$$

or $\int\int x_1 [f(x_1, x_2)\, dx_2]\, dx_1 = a + b\,\mu_2$ or $\int x_1\, f(x_1)\, dx_1 = a + b\,\mu_2$

or $\mu_1 = a + b\,\mu_2$ (5.2)

On multiplying both the sides of equation (5.1) by $x_2\, f(x_2)$, and integrating with respect to x_2, we get

$$\int\int x_1\, x_2\, f(x_1, x_2)\, dx_2\, dx_1 = a\int x_2\, f(x_2)\, dx_2 + b\int x_2^2\, f(x_2)\, dx_2$$

or $E(X_1, X_2) = a\,\mu_2 + b\,E(X_2^2)$

or $\sigma_{12} + \mu_1\mu_2 = a\,\mu_2 + b(\sigma_2^2 + \mu_2^2)$ (5.3)

By solving equations (5.2) and (5.3), we get

$$\mu_1\mu_2 = a\,\mu_2 + b\,\mu_2^2$$

$$\sigma_{12} + \mu_1\mu_2 = a\,\mu_2 + b(\sigma_2^2 + \mu_2^2)$$

$$\overline{\qquad\sigma_{12} = b[(\sigma_2^2 + \mu_2^2) - \mu_2^2]\qquad}$$

$b = \dfrac{\sigma_{12}}{\sigma_2^2} = \rho\dfrac{\sigma_1}{\sigma_2}$. After substituting the value of b in equation (5.1), we get

$a = \mu_1 - \rho\dfrac{\sigma_1}{\sigma_2}\mu_2$, therefore,

$$X_1^* = E(X_1 \mid X_2 = x_2) = \mu_1 - \rho\frac{\sigma_1}{\sigma_2}\mu_2 + \rho\frac{\sigma_1}{\sigma_2}x_2 = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2).$$

Multiple correlation

If X_1 is the first component of $\underline{X}$ and $\underline{X}^{(2)}$ the vector of remaining $(p-1)$ components. We first express X_1 as a linear combination of $\underline{X}^{(2)}$ defined by the relation

$X_1^* = \mu_1 + \underline{\beta}'(\underline{X}^{(2)} - \underline{\mu}^{(2)})$, the coefficient vector $\underline{\beta}$ is determined by minimizing

$$U = E[X_1 - X_1^*]^2 = E[X_1 - \mu_1 - \underline{\beta}'(\underline{X}^{(2)} - \underline{\mu}^{(2)})]^2.$$

Differentiating with respect to $\underline{\beta}$ and equating to zero

$$-2E(\underline{X}^{(2)} - \underline{\mu}^{(2)})[(X_1 - \mu_1) - \underline{\beta}'(\underline{X}^{(2)} - \underline{\mu}^{(2)})] = 0$$

or $E(\underline{X}^{(2)} - \underline{\mu}^{(2)})(X_1 - \mu_1) - E(\underline{X}^{(2)} - \underline{\mu}^{(2)})(\underline{X}^{(2)} - \underline{\mu}^{(2)})'\underline{\beta} = 0$

or $\underline{\sigma}_{12} = \Sigma_{22}\underline{\beta}$ and

$$\hat{\underline{\beta}}' = \underline{\sigma}_{12}'\Sigma_{22}^{-1}.$$

Therefore, the best linear function of X_1 in terms of $\underline{X}^{(2)}$ is

$$X_1^* = \mu_1 + \underline{\sigma}_{12}'\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) = \hat{X}_1.$$

The correlation coefficient between X_1 and its best linear estimate in terms of $\underline{X}^{(2)}$ is called **Multiple correlation** between X_1 and $X_2, \cdots, X_p$. This is denoted by

$$\rho_{1(2,3,\cdots,p)} = \frac{Cov(X_1, \hat{X}_1)}{\sqrt{V(X_1)\,V(\hat{X}_1)}}, \text{ where } \quad V(X_1) = E[X_1 - E(X_1)]^2 = \sigma_{11}.$$

$$\begin{aligned}
V(\hat{X}_1) &= E[\hat{X}_1 - E(\hat{X}_1)][\hat{X}_1 - E(\hat{X}_1)]' \\
&= E[\mu_1 + \underline{\sigma}_{12}'\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) - \mu_1][\mu_1 + \underline{\sigma}_{12}'\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) - \mu_1]' \\
&\qquad \text{since } E(\hat{X}_1) = \mu_1 + \underline{\sigma}_{12}'\Sigma_{22}^{-1}E(\underline{X}^{(2)} - \underline{\mu}^{(2)}) = \mu_1 \\
&= \underline{\sigma}_{12}'\Sigma_{22}^{-1}E(\underline{X}^{(2)} - \underline{\mu}^{(2)})(\underline{X}^{(2)} - \underline{\mu}^{(2)})'\Sigma_{22}^{-1}\underline{\sigma}_{12} \\
&= \underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{12}.
\end{aligned}$$

$$\begin{aligned}
Cov(X_1, \hat{X}_1) &= E[X_1 - E(X_1)][\hat{X}_1 - E(\hat{X}_1)]' \\
&= E(X_1 - \mu_1)[\mu_1 + \underline{\sigma}_{12}'\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) - \mu_1]' \\
&= E(X_1 - \mu_1)(\underline{X}^{(2)} - \underline{\mu}^{(2)})'\Sigma_{22}^{-1}\underline{\sigma}_{12} = \underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{12}.
\end{aligned}$$

Hence,

$$\rho_{1(2,3,\cdots,p)} = \frac{\underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{12}}{\sqrt{\sigma_{11}(\underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{12})}} = \sqrt{\frac{\underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{12}}{\sigma_{11}}} = \sqrt{\frac{\underline{\beta}'\Sigma_{22}\underline{\beta}}{\sigma_{11}}},$$

where $\underline{\sigma}_{12}$ and Σ_{22} are defined as $\begin{pmatrix} \sigma_{11} & \underline{\sigma}_{12}' \\ \underline{\sigma}_{21} & \Sigma_{22} \end{pmatrix}$.

Note: Since the numerator is $\sqrt{V(\hat{X}_1)}$, therefore, $\rho_{1(2,3,\cdots,p)} \ge 0$ i.e. $0 \le \rho_{1(2,3,\cdots,p)} \le 1$. This is so because $\hat{X}_1$ is an estimate of X_1.

Estimation of multiple correlation coefficient

The multiple correlation in the population is

$$\rho_{1(2,\cdots,p)} = \sqrt{\frac{\underline{\sigma}_{12}' \Sigma_{22}^{-1} \underline{\sigma}_{12}}{\sigma_{11}}} = \sqrt{\frac{\underline{\beta}' \Sigma_{22} \underline{\beta}}{\sigma_{11}}}.$$

Given $\underline{x}_\alpha$ $(\alpha = 1, \cdots, n)$, $n > p$. We estimate Σ by $\hat{\Sigma} = \frac{A}{n} = \frac{n-1}{n} S$,

where, $A = \sum_\alpha (\underline{x}_\alpha - \overline{\underline{x}})(\underline{x}_\alpha - \overline{\underline{x}})'$.

Now A is partitioned as follows

$$\frac{A}{n} = \begin{pmatrix} \frac{a_{11}}{n} & \frac{\underline{a}_{12}'}{n} \\ \frac{\underline{a}_{12}}{n} & \frac{A_{22}}{n} \end{pmatrix}, \text{ and the estimate of } \underline{\beta} \text{ is } \hat{\underline{\beta}}' = \underline{\sigma}_{12}' \Sigma_{22}^{-1} = \frac{\underline{a}_{12}'}{n} \left(\frac{A_{22}}{n}\right)^{-1} = \underline{a}_{12}' A_{22}^{-1}.$$

Using the above estimate, the sample multiple correlation coefficient of X_1 on $X_2, \cdots, X_p$ is

$$R_{1(2,\cdots,p)} = \sqrt{\frac{\hat{\underline{\sigma}}_{12}' \hat{\Sigma}_{22}^{-1} \underline{\sigma}_{12}}{\hat{\sigma}_{11}}} = \sqrt{\frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}}}$$

and

$$1 - R_{1(2,\cdots,p)}^2 = \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}} = \frac{\left| a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{12} \right| \left| A_{22} \right|}{a_{11} \left| A_{22} \right|} = \frac{|A|}{a_{11} |A_{22}|}.$$

Distribution of sample multiple correlation coefficient in null case

The sample multiple correlation coefficient between X_1 and $\underline{X}^{(2)}$ is defined by relation

$$R^2 = \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}} \quad \text{and} \quad 1 - R^2 = \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}},$$

where $R^2 = R_{1(2,3,\cdots,p)}^2$ and $A = \begin{pmatrix} a_{11} & \underline{a}_{12}' \\ \underline{a}_{12} & A_{22} \end{pmatrix}$.

Therefore,

$$\frac{R^2}{1 - R^2} = \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11.2}}.$$

We know that, if A is partitioned as

$$A = \begin{array}{c} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \end{array} \begin{array}{l} q \\ p-q \end{array} \quad \text{and} \quad A \sim W_p(n-1, \Sigma), \text{ then, } A_{11} \sim W_q(n-1, \Sigma_{11}) \text{ and}$$

$$A_{11} - A_{12} A_{22}^{-1} A_{21} \sim W_q(n-1-(p-q), \Sigma_{11.2}).$$

Thus, in our case

$$a_{11} \sim W_1(n-1, \sigma_{11}), \Rightarrow \frac{a_{11}}{\sigma_{11}} \sim \chi_{n-1}^2.$$

In null case $\rho_{1(2,3,\cdots,p)} = 0$

$$\Sigma_{11.2} = \sigma_{11} - \underline{\sigma}_{12}' \Sigma_{22}^{-1} \underline{\sigma}_{21} = \sigma_{11}, \text{ since } \underline{\sigma}_{12}' = 0, \text{ so that}$$

$$a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{21} \sim W_1(n-1-(p-1), \sigma_{11})$$

$$\Rightarrow \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{21}}{\sigma_{11}} \sim \chi_{n-p}^2.$$

Consider

$$\frac{a_{11}}{\sigma_{11}} = \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{21}}{\sigma_{11}} + \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{21}}{\sigma_{11}}$$

or $Q = Q_1 + Q_2$, (say),

where $Q \sim \chi_{n-1}^2$, and $Q_1 \sim \chi_{n-p}^2$.

From Fisher Cochran theorem Q_2 is independently distributed as $\chi_{n-1-(n-p)}^2$, i.e. $Q_2 \sim \chi_{p-1}^2$ and is independent of Q_1, hence,

$$F = \frac{R^2}{1 - R^2} \times \frac{n-p}{p-1} = \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12} / \sigma_{11}}{a_{11.2} / \sigma_{11}} \times \frac{n-p}{p-1} = \frac{\chi_{p-1}^2 / p-1}{\chi_{n-p}^2 / n-p} \sim F_{p-1, n-p}.$$

The distribution of the statistic F is

$$df(F) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} F^{\frac{\nu_1}{2} - 1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) \left(1 + \frac{\nu_1}{\nu_2} F\right)^{(\nu_1 + \nu_2)/2}} dF,$$

where $\nu_1 = p - 1$, $\nu_2 = n - p$.

In this put $F = \frac{R^2}{1 - R^2} \frac{\nu_2}{\nu_1}$, then $dF = \frac{dR^2}{(1 - R^2)^2} \frac{\nu_2}{\nu_1}$

$$df(R^2) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \left(\frac{R^2}{1 - R^2} \frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{2} - 1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) \left(1 + \frac{R^2}{1 - R^2}\right)^{(\nu_1 + \nu_2)/2}} \frac{\nu_2}{\nu_1} \frac{dR^2}{(1 - R^2)^2}$$

$$= \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}-\frac{\nu_1}{2}+1-1}\left(\frac{R^2}{1-R^2}\right)^{\frac{\nu_1}{2}-1}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)\left(\frac{1}{1-R^2}\right)^{(\nu_1+\nu_2)/2}}\frac{dR^2}{(1-R^2)^2}$$

$$= \frac{1}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}(R^2)^{\frac{\nu_1}{2}-1}(1-R^2)^{\frac{\nu_1+\nu_2}{2}-\frac{\nu_1}{2}+1-2}\,dR^2$$

$$= \frac{1}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}(R^2)^{\frac{\nu_1}{2}-1}(1-R^2)^{\frac{\nu_2}{2}-1}\,dR^2 .$$ Put $dR^2 = 2R\,dR$, thus the distribution of R.

$$df(R) = \frac{2R^{(\nu_1-1)}(1-R^2)^{\frac{\nu_2}{2}-1}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}dR = \frac{2R^{p-2}(1-R^2)^{\frac{n-p}{2}-1}}{B\left(\frac{p-1}{2},\frac{n-p}{2}\right)}dR, \qquad 0<R<1.$$

Likelihood ratio criteria for testing $H_0:\ \rho_{1(2,3,\ldots,p)}=0$.

The likelihood function of the sample $\underline{x}_\alpha$ $(\alpha=1,2,\ldots,n>p)$ from $N(\underline{\mu},\Sigma)$ is

$$L(\underline{\mu},\Sigma) = \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}}\exp\left[-\frac{1}{2}\sum_\alpha(\underline{x}_\alpha-\underline{\mu})'\Sigma^{-1}(\underline{x}_\alpha-\underline{\mu})\right]$$

and the likelihood ratio criterion

$\lambda = \dfrac{\max L_0}{\max L}$, where the numerator is the maximum of L for $\underline{\mu},\Sigma\in\omega$ restricted by $H_0:\ \rho_{1(2,3,\ldots,p)}=0$ (*i.e.* $\underline{\beta}=\underline{0}$, $\Rightarrow$ $\underline{\sigma}_{12}=\underline{0}$) and the denominator is the maximum of L over the entire parametric space Ω.

Now

$$\max L_\Omega = \frac{1}{(2\pi)^{np/2}\left|\frac{A}{n}\right|^{n/2}}\exp\left[-\frac{1}{2}\sum_\alpha(\underline{x}_\alpha-\underline{\bar{x}})'\left(\frac{A}{n}\right)^{-1}(\underline{x}_\alpha-\underline{\bar{x}})\right]$$

$$= \frac{n^{np/2}}{(2\pi)^{np/2}|A|^{n/2}}\exp\left[-\frac{1}{2}tr\left(\frac{A}{n}\right)^{-1}\sum_\alpha(\underline{x}_\alpha-\underline{\bar{x}})(\underline{x}_\alpha-\underline{\bar{x}})'\right]$$

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

where trace $nI_{p\times p}=np$.

Similarly,

$$\max L_\omega = \frac{n^{np/2}}{(2\pi)^{np/2}(a_{11}|A_{22}|)^{n/2}}\exp\left(-\frac{1}{2}Q\right),\ \text{where}$$

$$Q = trace\begin{pmatrix} n\,a_{11}^{-1} & \underline{0} \\ \underline{0} & n\,A_{22}^{-1}\end{pmatrix}\begin{pmatrix} a_{11} & \underline{0} \\ \underline{0} & A_{22}\end{pmatrix} = tr\,nI = np,\ \text{thus,}$$

$$\max L_\omega = \frac{n^{np/2}}{(2\pi)^{np/2}(a_{11}|A_{22}|)^{n/2}}\exp\left(-\frac{1}{2}np\right).$$

Hence, the ratio

$$\lambda = \frac{|A|^{n/2}}{(a_{11}|A_{22}|)^{n/2}} \quad\Rightarrow\quad \lambda^{2/n} = \frac{|A|}{a_{11}|A_{22}|} = 1-R^2$$

The likelihood ratio test is defined by the critical region $\lambda\le\lambda_0$, where

$\Pr[\lambda\le\lambda_0\,|\,H_0]=\alpha$

$\Rightarrow\ \lambda^{2/n}\le\lambda_o^{2/n}\ \Rightarrow\ 1-\lambda^{2/n}=R^2>1-\lambda_o^{2/n}=R_0^2$, (say)

$$\Rightarrow\ \frac{R^2}{1-R^2}\frac{n-p}{p-1} > \frac{R_0^2}{1-R_0^2}\frac{n-p}{p-1}$$

$\Rightarrow\ F > F_{p-1,\,n-p}(\alpha)$.

Theorem: Multiple correlation is invariant under the non-singular linear transformation.

Proof: We know that

$$\rho^2_{1(2,\cdots,p)} = \frac{\underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{12}}{\sigma_{11}},\ \text{where}\quad \underline{X}=\begin{pmatrix} X_1 \\ \underline{X}^{(2)}\end{pmatrix},\ \text{and}\ \Sigma=\begin{pmatrix}\sigma_{11} & \underline{\sigma}_{12}' \\ \underline{\sigma}_{12} & \Sigma_{22}\end{pmatrix}$$

Let

$Y_1=a_{11}X_1$, and $\underline{Y}^{(2)}=A_{22}\underline{X}^{(2)}$

or $\underline{Y}=\begin{pmatrix} a_{11} & \underline{0} \\ \underline{0} & A_{22}\end{pmatrix}\underline{X}$, where $a_{11}\ne 0$, $|A_{22}|\ne 0$.

Assume that $E\underline{X}=\underline{0}$, and

$$\Sigma_{\underline{Y}} = \begin{pmatrix} a_{11} & \underline{0} \\ \underline{0} & A_{22}\end{pmatrix}\begin{pmatrix}\sigma_{11} & \underline{\sigma}_{12}' \\ \underline{\sigma}_{12} & \Sigma_{22}\end{pmatrix}\begin{pmatrix} a_{11} & \underline{0} \\ \underline{0} & A_{22}\end{pmatrix} = \begin{pmatrix} a_{11}^2\sigma_{11} & a_{11}A_{22}\underline{\sigma}_{12}' \\ a_{11}A_{22}\underline{\sigma}_{12} & A_{22}\Sigma_{22}A_{22}\end{pmatrix}$$

Therefore,

$$\rho^2_{1(2,3,\ldots,p)}(\underline{Y}) = \frac{a_{11}A_{22}\underline{\sigma}_{12}'(A_{22}\Sigma_{22}A_{22})^{-1}a_{11}A_{22}\underline{\sigma}_{12}}{a_{11}^2\sigma_{11}}$$

$$=\frac{\underline{\sigma}_{12}{}' \Sigma_{22}^{-1} \underline{\sigma}_{12}}{\sigma_{11}} = \rho^2_{1(2,3,\ldots,p)}(\underline{X}).$$

Theorem: Let $\underline{X} \sim N(\underline{0}, \Sigma)$, $\underline{X} = \begin{pmatrix} X_1 \\ \underline{X}^{(2)} \end{pmatrix}$, and $\Sigma = \begin{pmatrix} \sigma_{11} & \underline{\sigma}_{12}{}' \\ \underline{\sigma}_{12} & \Sigma_{22} \end{pmatrix}$, then of all linear combinations of $X_2, X_3, \ldots, X_p$ has the maximum correlation with X_1.

Proof: We know that, for any c nonzero scalar and $\underline{\gamma}$

$$E(X_1 - \underline{\beta}' \underline{X}^{(2)})^2 \le E(X_1 - c\,\underline{\gamma}' \underline{X}^{(2)})^2,$$

because $\underline{\beta}' = \underline{\sigma}_{12}{}' \Sigma_{22}^{-1}$ is determined by minimizing $E(X_1 - \underline{\beta}' \underline{X}^{(2)})^2$.

Hence,

$$E[X_1^2 + (\underline{\beta}' \underline{X}^{(2)})^2 - 2X_1(\underline{\beta}' \underline{X}^{(2)})] \le E[X_1^2 + (c\,\underline{\gamma}' \underline{X}^{(2)})^2 - 2c\,X_1(\underline{\gamma}' \underline{X}^{(2)})]$$

or $$E X_1^2 + E[\underline{\beta}' \underline{X}^{(2)}]^2 - 2E X_1(\underline{\beta}' \underline{X}^{(2)}) \le E X_1^2 + c^2 E(\underline{\gamma}' \underline{X}^{(2)})^2 - 2c\,E X_1(\underline{\gamma}' \underline{X}^{(2)}).$$

Let

$$c^2 = \frac{E(\underline{\beta}' \underline{X}^{(2)})^2}{E(\underline{\gamma}' \underline{X}^{(2)})^2}, \text{ then}$$

$$E[\underline{\beta}' \underline{X}^{(2)}]^2 - 2EX_1(\underline{\beta}' \underline{X}^{(2)}) \le \frac{E(\underline{\beta}' \underline{X}^{(2)})^2}{E(\underline{\gamma}' \underline{X}^{(2)})^2} E(\underline{\gamma}' \underline{X}^{(2)})^2 - 2\sqrt{\frac{E(\underline{\beta}' \underline{X}^{(2)})^2}{E(\underline{\gamma}' \underline{X}^{(2)})^2}} EX_1(\underline{\gamma}' \underline{X}^{(2)})$$

or $$-2EX_1(\underline{\beta}' \underline{X}^{(2)}) \le -2\sqrt{\frac{E(\underline{\beta}' \underline{X}^{(2)})^2}{E(\underline{\gamma}' \underline{X}^{(2)})^2}}\, EX_1(\underline{\gamma}' \underline{X}^{(2)})$$

$$EX_1(\underline{\beta}' \underline{X}^{(2)}) \ge \sqrt{\frac{E(\underline{\beta}' \underline{X}^{(2)})^2}{E(\underline{\gamma}' \underline{X}^{(2)})^2}}\, EX_1(\underline{\gamma}' \underline{X}^{(2)})$$

Dividing both the sides by $\sqrt{EX_1^2\, E(\underline{\beta}' \underline{X}^{(2)})^2}$, we get

$$\frac{EX_1(\underline{\beta}' \underline{X}^{(2)})}{\sqrt{EX_1^2\, E(\underline{\beta}' \underline{X}^{(2)})^2}} \ge \frac{EX_1(\underline{\gamma}' \underline{X}^{(2)})}{\sqrt{EX_1^2\, E(\underline{\gamma}' \underline{X}^{(2)})^2}}$$

Corr. $(X_1, \underline{\beta}' \underline{X}^{(2)}) \ge$ Corr. $(X_1, \underline{\gamma}' \underline{X}^{(2)})$.

Exercise: Let $\underline{X} \sim N(\underline{0}, \Sigma)$, $\underline{X} = \begin{pmatrix} X_1 \\ \underline{X}^{(2)} \end{pmatrix}$, and $\Sigma = \begin{pmatrix} \sigma_{11} & \underline{\sigma}_{12}{}' \\ \underline{\sigma}_{12} & \Sigma_{22} \end{pmatrix}$. The difference $X_1 - \underline{\sigma}_{12}{}' \Sigma_{22}^{-1} \underline{X}^{(2)}$ (called the residual of X_1 from its mean square regression line on $\underline{X}^{(2)}$) is uncorrelated with any of the independent variable $X_2, X_3, \ldots, X_p$.

Solution:

$$\begin{aligned}\text{Cov}[(X_1 - \underline{\sigma}_{12}{}' \Sigma_{22}^{-1} \underline{X}^{(2)}), \underline{X}^{(2)}] &= E[(X_1 - \underline{\sigma}_{12}{}' \Sigma_{22}^{-1} \underline{X}^{(2)})(\underline{X}^{(2)})'] \\ &= E[X_1(\underline{X}^{(2)})'] - \underline{\sigma}_{12}{}' \Sigma_{22}^{-1} E[\underline{X}^{(2)}(\underline{X}^{(2)})'] \\ &= \underline{\sigma}_{12}{}' - \underline{\sigma}_{12}{}' \Sigma_{22}^{-1} \Sigma_{22} = \underline{0}.\end{aligned}$$

Exercise: Derive the sampling distribution of sample correlation coefficient r on the basis of a random sample of size n from a bivariate normal distribution $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, if $\rho = 0$. Also find $E(r^k)$.

Solution: The sample correlation coefficient between X_1 and X_2 is defined by the relation

$$r^2 = \frac{a_{12}\, a_{22}^{-1}\, a_{12}}{a_{11}} \text{ and } 1 - r^2 = \frac{a_{11} - a_{12}\, a_{22}^{-1}\, a_{12}}{a_{11}}, \text{ where } r^2 = r_{12}^2, \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Therefore,

$$\frac{r^2}{1 - r^2} = \frac{a_{12}\, a_{22}^{-1}\, a_{12}}{a_{11} - a_{12}\, a_{22}^{-1}\, a_{12}}.$$

We know that, if

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}, \text{ then, } A_{11} \sim W_q(n-1, \Sigma_{11}), \text{ and}$$

$$A_{11} - A_{12} A_{22}^{-1} A_{21} \sim W_q(n - 1 - (p - q), \Sigma_{11.2}).$$

At $p = 2$, and $q = 1$

$$a_{11} \sim W_1(n-1, \sigma_{11}), \Rightarrow \frac{a_{11}}{\sigma_{11}} \sim \chi^2_{n-1}.$$

In null case $\rho = 0$

$$\sigma_{11.2} = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21} = \sigma_{11}, \text{ since } \sigma_{12} = 0.$$

So that

$$a_{11} - a_{12}\, a_{22}^{-1}\, a_{21} \sim W_1(n - 1 - (2 - 1), \sigma_{11})$$

$$\Rightarrow \frac{a_{11} - a_{12}\, a_{22}^{-1}\, a_{21}}{\sigma_{11}} \sim \chi^2_{n-2}.$$

Consider

$$\frac{a_{11}}{\sigma_{11}} = \frac{a_{11} - a_{12}\, a_{22}^{-1}\, a_{21}}{\sigma_{11}} + \frac{a_{12}\, a_{22}^{-1}\, a_{21}}{\sigma_{11}}$$

or $Q = Q_1 + Q_2$, (say),

where $Q \sim \chi^2_{n-1}$ and $Q_1 \sim \chi^2_{n-2}$, therefore, $Q_2 \sim \chi^2_1$, by the additive property of χ^2. Thus,

$$\frac{r^2}{1-r^2} \times \frac{n-2}{1} = \frac{a_{12}\, a_{22}^{-1}\, a_{21} / \sigma_{11}}{a_{11.2} / \sigma_{11}} \times \frac{n-2}{1} = \frac{\chi^2_1}{\chi^2_{n-2} / n-2} \sim F_{1,n-2}\,.$$

Hence,

$$\sqrt{(n-2)}\, \frac{r}{\sqrt{1-r^2}} \sim t_{n-2}\,.$$

Partial correlation

The correlation between two variables X_1 and X_2 is measured by the correlation coefficient which sometimes called the total correlation coefficient between X_1 and X_2. If X_1 and X_2 are considered in the conjunction with $p-2$ other variables $X_3, X_4, \ldots, X_p$, we may regard the variation of X_1 and X_2 as to certain extents due to the variation of the other variables. Let $X_{1.3\ldots p}$ and $X_{2.3\ldots p}$ represent these parts of variation of X_1 and X_2 respectively, which remains after subs traction of the best linear estimate in terms of $X_3, X_4, \ldots, X_p$.

Thus we may regards the correlation coefficient between $X_{1.3\ldots p}$ and $X_{2.3\ldots p}$ as a measure of correlation between X_1 and X_2 after removal of any part of the variation due to the influence of $X_3, X_4, \ldots, X_p$, this correlation is called partial correlation of X_1 and X_2 with respect to $X_3, X_4, \ldots, X_p$. Let

$$\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \underline{X}^{(3)} \end{pmatrix}, \quad \underline{\mu} = \underline{0}, \text{ and } \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \vdots & \underline{\sigma}_{13}' \\ \sigma_{12} & \sigma_{22} & \vdots & \underline{\sigma}_{23}' \\ \cdots & \cdots & \cdots & \cdots \\ \underline{\sigma}_{13} & \underline{\sigma}_{23} & \vdots & \Sigma_{33} \end{pmatrix}$$

and the best linear estimates of X_1 and X_2 interms of $\underline{X}^{(3)}$ are $\hat{X}_1 = \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{X}^{(3)}$, and $\hat{X}_2 = \underline{\sigma}_{23}'\, \Sigma_{33}^{-1}\, \underline{X}^{(3)}$ respectively.

Define,

$$X_{1.3\ldots p} = X_1 - \hat{X}_1, \text{ and } \quad X_{2.3\ldots p} = X_2 - \hat{X}_2, \text{ then}$$

$$V(X_{1.3\ldots p}) = E[X_1 - \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{X}^{(3)}][X_1 - \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{X}^{(3)}]'$$

$$= E[X_1^2 - 2\underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, X_1\, \underline{X}^{(3)} + \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{X}^{(3)}\, \underline{X}^{(3)\prime}\, \Sigma_{33}^{-1}\, \underline{\sigma}_{13}]$$

$$= \sigma_{11} - 2\underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{13} + \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \Sigma_{33}\, \Sigma_{33}^{-1}\, \underline{\sigma}_{13} = \sigma_{11} - \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{13}\,.$$

Similarly, we have

$$V(X_{2.3\ldots p}) = \sigma_{22} - \underline{\sigma}_{23}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{23}, \text{ and}$$

$$Cov(X_{1.3\ldots p}, X_{2.3\ldots p}) = E[X_1 - \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{X}^{(3)}][X_2 - \underline{\sigma}_{23}'\, \Sigma_{33}^{-1}\, \underline{X}^{(3)}]'$$

$$= \sigma_{12} - \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{23}\,.$$

Therefore,

$$\rho_{12.3\ldots p} = \frac{\sigma_{12} - \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{23}}{\sqrt{(\sigma_{11} - \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{13})(\sigma_{22} - \underline{\sigma}_{23}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{23})}}\,.$$

Alternative proof

Consider the var. cov. matrix of conditional distribution of $\underline{X}^{(1)}$ given $\underline{X}^{(2)}$, which is

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \text{ in our case}$$

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} - \begin{pmatrix} \underline{\sigma}_{13}' \\ \underline{\sigma}_{23}' \end{pmatrix} \Sigma_{33}^{-1} (\underline{\sigma}_{13} \quad \underline{\sigma}_{23})$$

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} - \begin{pmatrix} \underline{\sigma}_{13}'\Sigma_{33}^{-1}\underline{\sigma}_{13} & \underline{\sigma}_{13}'\Sigma_{33}^{-1}\underline{\sigma}_{23} \\ \underline{\sigma}_{23}'\Sigma_{33}^{-1}\underline{\sigma}_{13} & \underline{\sigma}_{23}'\Sigma_{33}^{-1}\underline{\sigma}_{23} \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{11} - \underline{\sigma}_{13}'\Sigma_{33}^{-1}\underline{\sigma}_{13} & \sigma_{12} - \underline{\sigma}_{13}'\Sigma_{33}^{-1}\underline{\sigma}_{23} \\ \sigma_{12} - \underline{\sigma}_{13}'\Sigma_{33}^{-1}\underline{\sigma}_{23} & \sigma_{22} - \underline{\sigma}_{23}'\Sigma_{33}^{-1}\underline{\sigma}_{23} \end{pmatrix} = \begin{pmatrix} \sigma_{11.3\ldots p} & \sigma_{12.3\ldots p} \\ \sigma_{12.3\ldots p} & \sigma_{22.3\ldots p} \end{pmatrix}$$

We know that simple correlation coefficient is

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\,\sqrt{\sigma_{22}}}, \text{ where } \sigma_{11},\ \sigma_{12}, \text{ and } \sigma_{22} \text{ are defined as } \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

Therefore, the partial correlation coefficient is obtained like simple correlation coefficient from $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{21}\Sigma_{22}^{-1}\Sigma_{12}$

$$\rho_{12.3\ldots p} = \frac{\sigma_{12.3\ldots p}}{\sqrt{\sigma_{11.3\ldots p}}\,\sqrt{\sigma_{22.3\ldots p}}} = \frac{\sigma_{12} - \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{23}}{\sqrt{(\sigma_{11} - \underline{\sigma}_{13}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{13})(\sigma_{22} - \underline{\sigma}_{23}'\, \Sigma_{33}^{-1}\, \underline{\sigma}_{23})}}\,.$$

In general,

Let the partition of $\underline{X}$ and Σ as follows

$$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_q \\ X_{q+1} \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{pmatrix}, \text{ and } \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

The partial correlation coefficient between the variable X_i and X_j $(X_i, X_j \in \underline{X}^{(1)})$ holding the components of $\underline{X}^{(2)}$ fixed (there will be total of ${}^qC_2$ partial correlation coefficients), is often denoted by

$$\rho_{ij.q+1\ldots p} = \frac{\sigma_{ij.q+1\ldots p}}{\sqrt{(\sigma_{ii.q+1\ldots p})(\sigma_{jj.q+1\ldots p})}}, \; i, j = 1, 2, \ldots, q .$$

Estimation of partial correlation coefficient

We know that the population partial correlation coefficient between X_i and X_j holding the components of $\underline{X}^{(2)}$ fixed, is given by

$$\rho_{ij.q+1\ldots p} = \frac{\sigma_{ij.q+1\ldots p}}{\sqrt{(\sigma_{ii.q+1\ldots p})(\sigma_{jj.q+1\ldots p})}} .$$

Given a sample $\underline{x}_\alpha$ $(\alpha = 1, 2, \ldots, n > p)$ from $N(\underline{\mu}, \Sigma)$, the maximum likelihood estimate of $\rho_{ij.q+1\ldots p}$ is

$$\hat{\rho}_{ij.q+1\ldots p} = \frac{\hat{\sigma}_{ij.q+1\ldots p}}{\sqrt{(\hat{\sigma}_{ii.q+1\ldots p})(\hat{\sigma}_{jj.q+1\ldots p})}},$$

i.e. $r_{ij.q+1\ldots p} = \dfrac{a_{ij.q+1\ldots p}}{\sqrt{(a_{ii.q+1\ldots p})(a_{jj.q+1\ldots p})}}$ is called the sample partial correlation coefficient between X_i and X_j holding $X_{q+1}, \cdots, X_p$ fixed,

where, $(a_{ij.q+1\ldots p}) = A_{11} - A_{12}A_{22}^{-1}A_{21} = A_{11.2}$, i.e. $a_{ij.q+1\ldots p}$ is the $i-$th and $j-$th element of $A_{11.2}$, and $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$.

Distribution of sample partial correlation coefficient

The sample partial correlation coefficient between X_i and X_j holding $X_{q+1}, \cdots, X_p$ fixed is defined as

$$r_{ij.q+1\ldots p} = \frac{a_{ij.q+1\ldots p}}{\sqrt{(a_{ii.q+1\ldots p})(a_{jj.q+1\ldots p})}},$$

where, $(a_{ij.q+1\ldots p}) = A_{11} - A_{12}A_{22}^{-1}A_{21} = A_{11.2}$, i.e. $a_{ij.q+1\ldots p}$ is the $i-$th and $j-$th element of $A_{11.2}$, and $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$.

Let

$$A = \sum_\alpha (\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' \sim W_p(n-1, \Sigma), \text{ then,}$$

$$A_{11} - A_{12}A_{22}^{-1}A_{21} \sim W_q(n-1-(p-q), \Sigma_{11.2}) .$$

The distribution of the sample partial correlation $r_{ij.q+1\ldots p}$ based on a sample of size n from a distribution with population correlation $\rho_{ij.q+1\ldots p}$ is same as the distribution of ordinary correlation coefficient r_{ij} based on a sample of size $n-(p-q)$ from a distribution with the corresponding population partial correlation $\rho_{ij.q+1\ldots p} = \rho$. Thus,

$$\sqrt{n-(p-q)-2}\,\frac{r}{\sqrt{1-r^2}} \sim t_{n-(p-q)-2} .$$

Test of hypothesis and confidence region for partial correlation coefficient

Case I: $H_0 : \rho_{ij.q+1\cdots p} = \rho_0$, (a specified value), against $H_A : \rho_{ij.q+1\cdots p} \neq \rho_0$. For testing H_0 on the basis of sample of size $n-(p-q)$, we use the following test statistic

$$U = \frac{Z - \xi_0}{1/\sqrt{n-3}} \sim N(0, 1), \text{ where}$$

$$Z = \frac{1}{2}\ln\frac{1 + r_{ij.q+1\cdots p}}{1 - r_{ij.q+1\cdots p}} = \tan h^{-1} r_{ij.q+1\cdots p}, \text{ and } \quad \xi_0 = \frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0} = \tan h^{-1} \rho_0 .$$

If the absolute value of U is greater than 1.96, we reject H_0 at $\alpha = 0.05$ level of significance otherwise accept H_0. For confidence region when $\rho_{ij.q+1\cdots p}$ is unknown, then we can write

$$\Pr[-U_{\alpha/2} \leq \sqrt{(n-3)}\,(Z - \xi) \leq U_{\alpha/2}] = 1-\alpha, \text{ where}$$

$$\xi = \frac{1}{2}\ln\frac{1+\rho_{ij.q+1\cdots p}}{1-\rho_{ij.q+1\cdots p}} = \tan h^{-1} \rho_{ij.q+1\cdots p}, \text{ then}$$

$$\Pr\left[Z - \frac{U_{\alpha/2}}{\sqrt{(n-3)}} \leq \tan h^{-1}\rho_{ij.q+1\cdots p} \leq Z + \frac{U_{\alpha/2}}{\sqrt{(n-3)}}\right] = 1-\alpha$$

or $$\Pr\left[\tan h\left(Z - \frac{U_{\alpha/2}}{\sqrt{(n-3)}}\right) \leq \rho_{ij.q+1\cdots p} \leq \tan h\left(Z + \frac{U_{\alpha/2}}{\sqrt{(n-3)}}\right)\right] = 1-\alpha .$$

Case II: $H_0 : \rho_{ij.q+1\cdots p} = 0$, against $H_A : \rho_{ij.q+1\cdots p} \neq 0$. For testing H_0 on the basis of sample of size $n-(p-q)$, we use the following test statistic

$$t = \frac{r_{ij.q+1\cdots p}}{\sqrt{1 - r^2_{ij.q+1\cdots p}}}\sqrt{n-(p-q)-2} \sim t_{n-(p-q)-2} .$$

If $|t| > t_{n-(p-q)-2}(\alpha)$, we reject H_0 at α level of significance otherwise accept H_0.

Theorem: Partial correlation coefficient is invariant under the nonsingular linear transformation.

Proof: We know that

$$\rho_{12.3\ldots p}(\underline{X}) = \frac{\sigma_{12} - \underline{\sigma}_{13}'\,\Sigma_{33}^{-1}\,\underline{\sigma}_{23}}{\sqrt{(\sigma_{11} - \underline{\sigma}_{13}'\,\Sigma_{33}^{-1}\,\underline{\sigma}_{13})(\sigma_{22} - \underline{\sigma}_{23}'\,\Sigma_{33}^{-1}\,\underline{\sigma}_{23})}},$$

where $\Sigma_{\underline{X}} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \underline{\sigma}_{13}' \\ \sigma_{12} & \sigma_{22} & \underline{\sigma}_{23}' \\ \underline{\sigma}_{13} & \underline{\sigma}_{23} & \Sigma_{33} \end{pmatrix}$.

Let

$$Y_1 = a_{11} X_1, \quad a_{11} \neq 0,\ Y_2 = a_{22} X_2, a_{22} \neq 0, \text{ and } \ \underline{Y}^{(3)} = A_{33} \underline{X}^{(3)}, \quad |A_{33}| \neq 0$$

or $\underline{Y} = \begin{pmatrix} a_{11} & 0 & \underline{0} \\ 0 & a_{22} & \underline{0} \\ \underline{0} & \underline{0} & A_{33} \end{pmatrix} \underline{X}$.

Assume that $E\underline{X} = \underline{0}$, and

$$\Sigma_{\underline{Y}} = \begin{pmatrix} a_{11} & 0 & \underline{0} \\ 0 & a_{22} & \underline{0} \\ \underline{0} & \underline{0} & A_{33} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \underline{\sigma}_{13}' \\ \sigma_{12} & \sigma_{22} & \underline{\sigma}_{23}' \\ \underline{\sigma}_{13} & \underline{\sigma}_{23} & \Sigma_{33} \end{pmatrix} \begin{pmatrix} a_{11} & 0 & \underline{0} \\ 0 & a_{22} & \underline{0} \\ \underline{0} & \underline{0} & A_{33} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}^2 \sigma_{11} & a_{11} a_{22} \sigma_{12} & a_{11} \underline{\sigma}_{13}' A_{33} \\ a_{11} a_{22} \sigma_{12} & a_{22}^2 \sigma_{22} & a_{22} \underline{\sigma}_{23}' A_{33} \\ a_{11} \underline{\sigma}_{13} A_{33} & A_{33} \underline{\sigma}_{23} a_{22} & A_{33} \Sigma_{33} A_{33} \end{pmatrix}$$

by the analogy $\rho_{12.3\ldots p}(\underline{X})$

$$\rho_{12.3\ldots p}(\underline{Y}) = \frac{a_{11} a_{22} \sigma_{12} - a_{11} \underline{\sigma}_{13}' A_{33} (A_{33} \Sigma_{33} A_{33})^{-1} A_{33} \underline{\sigma}_{23} a_{22}}{\sqrt{a_{11}^2 \sigma_{11} - a_{11} \underline{\sigma}_{13}' A_{33} (A_{33} \Sigma_{33} A_{33})^{-1} A_{33} \underline{\sigma}_{13} a_{11}}}$$

$$\times \frac{1}{\sqrt{a_{22}^2 \sigma_{22} - a_{22} \underline{\sigma}_{23}' A_{33} (A_{33} \Sigma_{33} A_{33})^{-1} A_{33} \underline{\sigma}_{23} a_{22}}}.$$

Taking common a_{11} and a_{22}, then we get

$$\rho_{12.3\ldots p}(\underline{Y}) = \frac{\sigma_{12} - \underline{\sigma}_{13}' \Sigma_{33}^{-1} \underline{\sigma}_{23}}{\sqrt{(\sigma_{11} - \underline{\sigma}_{13}' \Sigma_{33}^{-1} \underline{\sigma}_{13})(\sigma_{22} - \underline{\sigma}_{23}' \Sigma_{33}^{-1} \underline{\sigma}_{23})}} = \rho_{12.3\ldots p}(\underline{X}).$$

Note:

$$1 - \rho_{1(2,3,\cdots,p)}^2 = (1 - \rho_{12}^2)(1 - \rho_{13.2}^2)(1 - \rho_{14.23}^2) \cdots (1 - \rho_{1p.23\cdots p-1}^2).$$

Exercise: If all the total correlation coefficient in a $p-$variate normal distribution are equal to $\rho \neq 0$, show that

i) $\rho \geq -\dfrac{1}{p-1}$, and ii) $\rho_{1(2,3,\cdots,p)}^2 = \dfrac{(p-1)\rho^2}{1 + (p-2)\rho}$.

Solution:

i) We are given that

$$\rho_{ij} = \rho, \quad ij = 1, 2, \cdots, p;\ i \neq j, \text{ we have}$$

$$\rho_{ij.k} = \frac{\sigma_{ij} - \sigma_{ik} \sigma_{kk}^{-1} \sigma_{jk}}{\sqrt{\sigma_{ii} - \sigma_{ik} \sigma_{kk}^{-1} \sigma_{ik}} \sqrt{\sigma_{jj} - \sigma_{jk} \sigma_{kk}^{-1} \sigma_{jk}}}$$

$$= \frac{\sigma_i \sigma_j \rho_{ij} - \sigma_i \sigma_k \rho_{ik} \dfrac{1}{\sigma_k \sigma_k} \sigma_j \sigma_k \rho_{jk}}{\sqrt{\sigma_{ii} - \sigma_i \sigma_k \rho_{ik} \dfrac{1}{\sigma_k \sigma_k} \sigma_i \sigma_k \rho_{jk}} \sqrt{\sigma_{jj} - \sigma_j \sigma_k \rho_{jk} \dfrac{1}{\sigma_k \sigma_k} \sigma_j \sigma_k \rho_{jk}}}$$

$$= \frac{\sigma_i \sigma_j (\rho_{ij} - \rho_{ik} \rho_{jk})}{\sigma_i \sigma_j \sqrt{1 - \rho_{ik}^2} \sqrt{1 - \rho_{jk}^2}} = \frac{\rho - \rho^2}{\sqrt{1 - \rho^2} \sqrt{1 - \rho^2}} = \frac{\rho}{1 + \rho}.$$

Thus every partial correlation coefficient of order 1 is $\dfrac{\rho}{1+\rho}$. Similarly,

$$\rho_{ij.kl} = \frac{\rho_{ij.l} - \rho_{ik.l} \rho_{jk.l}}{\sqrt{1 - \rho_{ik.l}^2} \sqrt{1 - \rho_{jk.l}^2}} = \frac{\left(\dfrac{\rho}{1+\rho}\right) - \left(\dfrac{\rho}{1+\rho}\right)^2}{1 - \left(\dfrac{\rho}{1+\rho}\right)^2} = \frac{\dfrac{\rho}{1+\rho}\left(1 - \dfrac{\rho}{1+\rho}\right)}{\left(1 + \dfrac{\rho}{1+\rho}\right)\left(1 - \dfrac{\rho}{1+\rho}\right)} = \frac{\rho}{1 + 2\rho}$$

Thus every partial correlation coefficient of order 1 is $\dfrac{\rho}{1+2\rho}$.

The partial correlation coefficient of the highest order in $p-$variate distribution is $p-2$, by the method of induction, the every partial correlation coefficient of order $p-2$ is $\dfrac{\rho}{1+(p-2)\rho}$. Since $\left| \rho_{ij.(p-1) components} \right| \leq 1$, so that

$$-1 \leq \rho_{ij.(p-1) components} \leq 1$$

We have on considering the lower limit

$$-1 \leq \frac{\rho}{1 + (p-2)\rho}, \quad \text{or} \quad -(1 + p\rho - 2\rho) \leq \rho \quad \text{or} \quad -1 \leq \rho + p\rho - 2$$

$$\text{or} \quad -1 \leq \rho(p-1) \quad \text{or} \quad \rho \geq -\frac{1}{p-1}.$$

ii) We know that

$$1 - \rho_{1(2,3,\cdots,p)}^2 = (1 - \rho_{12}^2)(1 - \rho_{13.2}^2)(1 - \rho_{14.23}^2) \cdots (1 - \rho_{1p.23\cdots p-1}^2),$$

since $\rho_{12} = \rho$, and $\rho_{13.2} = \dfrac{\rho}{1+\rho}$.

$$\rho^2_{1(2,3)} = (1-\rho^2)\left[1-\left(\frac{\rho}{1+\rho}\right)^2\right] = (1-\rho^2)\left[\frac{(1+\rho)^2-\rho^2}{(1+\rho)^2}\right] = \frac{(1-\rho)(1+2\rho)}{(1+\rho)}.$$

Similarly

$$1-\rho^2_{1(2,3,4)} = (1-\rho^2_{12})(1-\rho^2_{13.2})(1-\rho^2_{14.23}) = \frac{(1-\rho)(1+2\rho)}{(1+\rho)}\left[1-\left(\frac{\rho}{1+2\rho}\right)^2\right]$$

$$= \frac{(1-\rho)(1+2\rho)}{(1+\rho)}\left(\frac{1+3\rho^2+4\rho}{(1+2\rho)^2}\right) = \frac{(1-\rho)(1+3\rho)}{(1+2\rho)}.$$

By the method of induction

$$1-\rho^2_{1(2,3,\cdots,p)} = (1-\rho)\left(\frac{1+(p-1)\rho}{1+(p-2)\rho}\right)$$

$$\rho^2_{1(2,3,\cdots,p)} = \frac{1+p\rho-2\rho-[1+p\rho-2\rho-(p-1)\rho^2]}{1+(p-2)\rho} = \frac{(p-1)\rho^2}{1+(p-2)\rho}.$$

HOTELLING'S $-T^2$

If X is univariate normal with mean μ and standard deviation σ, then $U=\frac{\sqrt{n}\,(\bar{x}-\mu)}{\sigma}\sim N(0,1)$, and $V=\frac{1}{\sigma^2}\sum_i (x_i-\bar{x})^2=\frac{(n-1)\,s^2}{\sigma^2}\sim \chi^2_{n-1}$, where s^2 is the sample variance from a sample of size n. If U and V are independently distributed, then **Student's** $-t$ is defined as

$$t=\frac{U}{\sqrt{V/n-1}}=\frac{\sqrt{n}\,(\bar{x}-\mu)/\sigma}{\sqrt{(n-1)\,s^2/(n-1)\,\sigma^2}}=\frac{\sqrt{n}\,(\bar{x}-\mu)}{s}\sim t_{n-1}.$$

The multivariate analogue of Student's $-t$ is Hotelling's T^2.

If $\underline{x}_\alpha$ $(\alpha=1,2,\cdots,n)$ is an independent sample of size n from $N_p(\underline{\mu},\Sigma)$ and, if $\underline{\bar{x}}$ is the sample mean vector, S the matrix of variance covariance, then the Hotelling's $-T^2$ is defined by the relation

$$T^2=n\,(\underline{\bar{x}}-\underline{\mu})'\,S^{-1}\,(\underline{\bar{x}}-\underline{\mu}).$$

Result:

Let the square matrix A be partitioned as

$A=\begin{pmatrix} A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}$, so that A_{22} is square. If A_{22} is nonsingular, let

$C=\begin{pmatrix} I & -A_{12}A_{22}^{-1}\\ \underline{0} & I\end{pmatrix}$, then $CAC'=\begin{pmatrix} A_{11}-A_{12}A_{22}^{-1}A_{21} & \underline{0}\\ \underline{0} & A_{22}\end{pmatrix}=\begin{pmatrix} A_{11.2} & \underline{0}\\ \underline{0} & A_{22}\end{pmatrix}$

$\Rightarrow\ A=C^{-1}\begin{pmatrix} A_{11.2} & \underline{0}\\ \underline{0} & A_{22}\end{pmatrix}C'^{-1}$, then

$$A^{-1}=C'\begin{pmatrix} A_{11.2} & \underline{0}\\ \underline{0} & A_{22}\end{pmatrix}^{-1}C=\begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}A_{12}A_{22}^{-1}\\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22}^{-1}A_{21}A_{11.2}^{-1}A_{12}A_{22}^{-1}+A_{22}^{-1}\end{pmatrix}.$$

Distribution of Hotelling's $-T^2$

Let $\underline{x}_1,\cdots,\underline{x}_n$ be an independent sample drawn from $N_p(\underline{\mu},\Sigma)$, then $A=(n-1)\,S=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha\underline{Z}_\alpha'$ with $\underline{Z}_\alpha$ independent, each with distribution $N_p(\underline{0},\Sigma)$.

By definition,

$$T^2=n\,(\underline{\bar{x}}-\underline{\mu})'\,S^{-1}\,(\underline{\bar{x}}-\underline{\mu}).$$

Let

$\underline{Y}=\sqrt{n}\,(\underline{\bar{x}}-\underline{\mu}_0)$, then $E(\underline{Y})=\sqrt{n}\,E(\underline{\bar{x}}-\underline{\mu}_0)=\sqrt{n}\,(\underline{\mu}-\underline{\mu}_0)=\underline{\upsilon}$ (say)

and

$$\Sigma_{\underline{Y}}=E[\underline{Y}-E(\underline{Y})][\underline{Y}-E(\underline{Y})]'=\Sigma.$$

Therefore,

$\underline{Y}\sim N_p(\underline{\upsilon},\Sigma)$, then

$$T^2=\underline{Y}'S^{-1}\underline{Y}.$$

Since Σ is positive definite matrix, there exits a nonsingular matrix C, such that

$$C\,\Sigma\,C'=I\quad\Rightarrow\quad C'C=\Sigma^{-1}.$$

Define,

$\underline{Y}^*=C\underline{Y}$, $S^*=C\,S\,C'$, and $\underline{\upsilon}^*=C\underline{\upsilon}$, then

$$E(\underline{Y}^*)=C\sqrt{n}\,E(\underline{\bar{x}}-\underline{\mu}_0)=C\sqrt{n}\,(\underline{\mu}-\underline{\mu}_0)=C\underline{\upsilon}=\underline{\upsilon}^*$$

and

$$\Sigma_{\underline{Y}^*}=E[\underline{Y}^*-E(\underline{Y}^*)][\underline{Y}^*-E(\underline{Y}^*)]'=C\,\Sigma\,C'=I.$$

Therefore,

$\underline{Y}^*\sim N_p(\underline{\upsilon}^*,I)$, then

$$T^2=\underline{Y}^{*\prime}S^{*-1}\underline{Y}^*$$

and

$(n-1)\,S^*=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha^*\,\underline{Z}_\alpha^{*\prime}$, where $\underline{Z}_\alpha^*=C\,\underline{Z}_\alpha\sim N(\underline{0},\,I)$.

Now consider a random orthogonal matrix Q of order $p\times p$

$$Q=\begin{pmatrix} \frac{Y_1^*}{\sqrt{\underline{Y}^{*\prime}\underline{Y}^*}} & \frac{Y_2^*}{\sqrt{\underline{Y}^{*\prime}\underline{Y}^*}} & \cdots & \frac{Y_p^*}{\sqrt{\underline{Y}^{*\prime}\underline{Y}^*}}\\ q_{21} & q_{22} & \cdots & q_{2p}\\ \vdots & \vdots & \cdots & \vdots\\ q_{p1} & q_{p2} & \cdots & q_{pp}\end{pmatrix}.$$

Let

$\underline{U}=Q\underline{Y}^*$, be an orthogonal transformation, also $B=(b_{ij})=Q\,(n-1)\,S^*Q'$.

So that

$$U_1=\frac{Y_1^{*2}+Y_2^{*2}+\ldots+Y_p^{*2}}{\sqrt{\underline{Y}^{*\prime}\underline{Y}^*}}=\frac{\underline{Y}^{*\prime}\underline{Y}^*}{\sqrt{\underline{Y}^{*\prime}\underline{Y}^*}}=\sqrt{\underline{Y}^{*\prime}\underline{Y}^*}$$

and

$$U_j=\sum_{i=1}^{p} q_{ji}\, Y_i^* = \sqrt{\underline{Y}^{*'}\,\underline{Y}^*}\sum_{i=1}^{p} q_{ji}\,\frac{Y_i^*}{\sqrt{\underline{Y}^{*'}\,\underline{Y}^*}}\,,\quad j=2,3,\ldots,p$$

$$=\underline{0}\qquad \forall\ \ j=2,3,\ldots,p\text{, by using the property of an orthogonal matrix.}$$

Thus,

$$T^2=\underline{Y}^{*'}S^{*-1}\underline{Y}^*=(Q^{-1}\underline{U})'S^{*-1}(Q^{-1}\underline{U})=\underline{U}'(QS^*Q')^{-1}\underline{U}$$

$$=(n-1)\underline{U}'[Q(n-1)S^*Q']^{-1}\underline{U}=(n-1)\underline{U}'B^{-1}\underline{U}\,.$$

$$\Rightarrow\quad \frac{T^2}{n-1}=\underline{U}'B^{-1}\underline{U}=(U_1\quad 0\quad \cdots\quad 0)\begin{pmatrix} b^{11} & b^{12} & \cdots & b^{1p}\\ b^{21} & b^{22} & \cdots & b^{2p}\\ \vdots & \vdots & \vdots & \vdots\\ b^{p1} & b^{p2} & \cdots & b^{pp}\end{pmatrix}\begin{pmatrix}U_1\\ 0\\ \vdots\\ 0\end{pmatrix}$$

$$=U_1^2\, b^{11},\ \text{where}\ (b^{ij})=B^{-1}.$$

We know that if the square matrix say B is partition as

$$B=\begin{pmatrix} b_{11} & \underline{b}_{12}'\\ \underline{b}_{12} & B_{22}\end{pmatrix},\ \text{and if } B_{22} \text{ is nonsingular, then}$$

$$B^{-1}=\begin{pmatrix} b_{11.2}^{-1} & -b_{11.2}^{-1}\,\underline{b}_{12}\,B_{22}^{-1}\\ -B_{22}^{-1}\,\underline{b}_{21}\,b_{11.2}^{-1} & B_{22}^{-1}\,\underline{b}_{21}\,b_{11.2}^{-1}\,\underline{b}_{12}'\,B_{22}^{-1}+B_{22}^{-1}\end{pmatrix}.$$

Thus,

$$U_1^2\, b^{11}=\frac{\underline{Y}^{*'}\,\underline{Y}^*}{b_{11.2}}\,,$$

where $b^{11}=\dfrac{1}{b_{11.2}}=\dfrac{1}{b_{11}-\underline{b}_{12}'\,B_{22}^{-1}\,\underline{b}_{12}}$.

Therefore,

$$\frac{T^2}{n-1}=\frac{\underline{Y}^{*'}\,\underline{Y}^*}{b_{11.2}}\,.$$

But

$$B=Q\,(n-1)\,S^*Q'=\sum_{\alpha=1}^{n-1}(Q\,\underline{Z}_\alpha^*)(Q\,\underline{Z}_\alpha^*)'=\sum_{\alpha=1}^{n-1}\underline{V}_\alpha\,\underline{V}_\alpha'\,,$$

where, $\underline{V}_\alpha=Q\,\underline{Z}_\alpha^*\sim N_p\,(\underline{0},I)$, for given Q.

$\Rightarrow\ b_{11.2}$ is distributed as $\displaystyle\sum_{\alpha=1}^{n-1-(p-1)} w_\alpha^2$, where w_α are independent $N(0,\,1)$.

$\Rightarrow\ b_{11.2}$ is a χ^2 with $(n-p)$ degree of freedom, but the conditional distribution of $b_{11.2}$ does not depend upon Q. Therefore $b_{11.2}$ is unconditionally distributed as a χ^2 with $(n-p)$ degree of freedom.

Since $\underline{Y}^*$ is distributed according to $N_p\,(\underline{\upsilon}^*,I)$

$$\Rightarrow\quad \underline{Y}^{*'}\,\underline{Y}^*\sim\chi_p^2\,(\underline{\upsilon}^{*'}\,\underline{\upsilon}^*),\ \text{where}\ \underline{\upsilon}^{*'}\,\underline{\upsilon}^*=\underline{\upsilon}'\,\Sigma^{-1}\,\underline{\upsilon}=\lambda^2\ \text{(say).}$$

Thus the $\dfrac{T^2}{n-1}$ is the ratio of a noncentral χ^2 with p degree of freedom to an independent χ^2 with $(n-p)$ degree of freedom, i.e.

$$\frac{T^2}{n-1}\,\frac{n-p}{p}=\frac{\chi_p^2\,(\lambda^2)/p}{\chi_{n-p}^2\,/(n-p)}\sim F_{p,n-p}\,(\lambda^2)\,.$$

If $\underline{\mu}=\underline{\mu}_0$, then the $F-$distribution is central.

Alternative proof

By definition,

$$T^2=n\,(\underline{\bar{x}}-\underline{\mu})'\,S^{-1}\,(\underline{\bar{x}}-\underline{\mu})\quad\text{or}\quad \frac{T^2}{n-1}=n\,(\underline{\bar{x}}-\underline{\mu}_0)'A^{-1}(\underline{\bar{x}}-\underline{\mu}_0)\,.$$

Let

$$\underline{d}=\sqrt{n}\,(\underline{\bar{x}}-\underline{\mu}_0)\,,\ \Rightarrow\ \underline{d}\sim N_p\,(\underline{0},\,\Sigma)$$

$$\Rightarrow\ \underline{d}\,\underline{d}'=Q\sim W_p\,(1,\,\Sigma)\tag{6.1}$$

and $\quad A\sim W_p\,(\upsilon,\,\Sigma)$ (6.2)

Now

$$\begin{vmatrix} 1 & \underline{d}'\\ -\underline{d} & A\end{vmatrix}=|\,A\,|\,\left|1+\underline{d}'A^{-1}\underline{d}\right|\qquad\text{or}\qquad \left|A+\underline{d}\,\underline{d}'\right|$$

$$\Rightarrow\quad \frac{|\,A\,|}{\left|A+\underline{d}\,\underline{d}'\right|}=\frac{1}{\left|1+\underline{d}'A^{-1}\underline{d}\right|}=\frac{1}{1+n\,(\underline{\bar{x}}-\underline{\mu}_0)'A^{-1}(\underline{\bar{x}}-\underline{\mu}_0)}=\frac{1}{1+\dfrac{T^2}{n-1}}\tag{6.3}$$

We determine the distribution $\Phi=\dfrac{1}{1+\dfrac{T^2}{n-1}}=\dfrac{|\,A\,|}{|\,A+Q\,|}$.

A and Q are independent as Q is based on $\underline{\bar{x}}$. Thus from (6.1), (6.2) and the additive property of wishart distribution, we have

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

We find the $r-$th moment of Φ

$$E(\Phi^r)=\int \Phi^r\, f(A,Q)\,dA\,dQ=\int \Phi^r\, f(A)\,f(Q)\;dA\;dQ$$

$$=\int \frac{|A|^r}{|A+Q|^r}\left(\frac{C(p,\upsilon)}{|\Sigma|^{\upsilon/2}}|A|^{(\upsilon-p-1)/2}\exp\left(-\frac{1}{2}tr\,A\Sigma^{-1}\right)\right)f(Q)\,dA\;dQ,$$

where,

$$C(p,\upsilon)=\frac{1}{2^{\upsilon\, p/2}(\pi)^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{\upsilon-i+1}{2}\right)}.$$

$$E(\Phi^r)=C(p,\upsilon)|\Sigma|^{-\upsilon/2}\int \frac{|A|^{(\upsilon+2r-p-1)/2}}{|A+Q|^r}\exp\left(-\frac{1}{2}tr\,A\Sigma^{-1}\right)f(Q)\,dA\;dQ$$

$$=\frac{C(p,\upsilon)|\Sigma|^{-\upsilon/2}}{C(p,\upsilon+2r)|\Sigma|^{-(\upsilon+2r)/2}}\int\frac{1}{|A+Q|^r}\left(\frac{C(p,\upsilon+2r)}{|\Sigma|^{(\upsilon+2r)/2}}\right.$$

$$\left.|A|^{(\upsilon+2r-p-1)/2}e^{-\frac{1}{2}tr\,A\Sigma^{-1}}\right)f(Q)\,dA\;dQ \qquad (6.4)$$

Note that expression inside the bracket is $W_p(\upsilon+2r,\,\Sigma)$.

Let $A+Q=U$, we will now integrate over the constant surface of $A+Q=U$, from the additive property of wishart distribution

$$E(\Phi^r)=\frac{C(p,\upsilon)}{C(p,\upsilon+2r)|\Sigma|^{-r}}\int_u\frac{C(p,\upsilon+2r+1)}{|U|^r|\Sigma|^{(\upsilon+2r+1)/2}}|U|^{(\upsilon+2r+1-p-1)/2}e^{-\frac{1}{2}trU\Sigma^{-1}}\;du$$

$$=\frac{C(p,\upsilon)\,C(p,\upsilon+2r+1)}{C(p,\upsilon+2r)\,C(p,\upsilon+1)}\int\frac{C(p,\upsilon+1)}{|U|^r|\Sigma|^{(\upsilon+1)/2}}|U|^{(\upsilon+1-p-1)/2}e^{-\frac{1}{2}trU\Sigma^{-1}}\;du$$

$$=\frac{2^{(\upsilon+2r)p/2}(\pi)^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{\upsilon+2r-i+1}{2}\right)2^{(\upsilon+1)p/2}}{2^{\upsilon p/2}(\pi)^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{\upsilon-i+1}{2}\right)2^{(\upsilon+2r+1)p/2}}$$

$$\times\frac{(\pi)^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{\upsilon+1-i+1}{2}\right)}{(\pi)^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{\upsilon+2r+1-i+1}{2}\right)}$$

$$=\frac{\Gamma\left(\frac{\upsilon+2r}{2}\right)\Gamma\left(\frac{\upsilon+2r-1}{2}\right)\cdots\Gamma\left(\frac{\upsilon+2r-p+2}{2}\right)\Gamma\left(\frac{\upsilon+2r-p+1}{2}\right)}{\Gamma\left(\frac{\upsilon}{2}\right)\Gamma\left(\frac{\upsilon-1}{2}\right)\cdots\Gamma\left(\frac{\upsilon-p+2}{2}\right)\Gamma\left(\frac{\upsilon-p+1}{2}\right)}$$

$$\times\frac{\Gamma\left(\frac{\upsilon+1}{2}\right)\Gamma\left(\frac{\upsilon}{2}\right)\cdots\Gamma\left(\frac{\upsilon-p+3}{2}\right)\Gamma\left(\frac{\upsilon-p+2}{2}\right)}{\Gamma\left(\frac{\upsilon+2r+1}{2}\right)\Gamma\left(\frac{\upsilon+2r}{2}\right)\cdots\Gamma\left(\frac{\upsilon+2r-p+2}{2}\right)\Gamma\left(\frac{\upsilon+2r-p+2}{2}\right)}$$

$$=\frac{\Gamma\left(\frac{\upsilon+2r-p+1}{2}\right)\Gamma\left(\frac{\upsilon+1}{2}\right)}{\Gamma\left(\frac{\upsilon-p+1}{2}\right)\Gamma\left(\frac{\upsilon+2r+1}{2}\right)} \qquad (6.5)$$

We compare this $r-$th moment with the $r-$th moment of Beta one, the density function of $\beta_1(l,\,m)$ is given by

$$f(x\colon l,m)=\frac{1}{\beta(l,m)}x^{l-1}(1-x)^{m-1},\quad 0<x<1.$$

$$E(x^r)=\frac{1}{\beta(l,m)}\int_0^1 x^r x^{l-1}(1-x)^{m-1}\,dx=\frac{\beta(r+l,m)}{\beta(l,m)}=\frac{\Gamma(r+l)\Gamma(m)\Gamma(l+m)}{\Gamma(r+l+m)\Gamma(l)\Gamma(m)}$$

$$=\frac{\Gamma(r+l)\Gamma(l+m)}{\Gamma(r+l+m)\Gamma(l)}$$

So if we write

$$l=\frac{\upsilon-p+1}{2},\text{ and } m=\frac{p}{2},\qquad\Rightarrow\; l+m=\frac{\upsilon+1}{2}$$

$$\Rightarrow\;\Phi\sim\beta_1\left(\frac{\upsilon-p+1}{2},\,\frac{p}{2}\right).$$

We know that if $X\sim\beta_1\left(\frac{\upsilon_1^*}{2},\frac{\upsilon_2^*}{2}\right)$, then $\frac{1-X}{X}\times\frac{\upsilon_1^*}{\upsilon_2^*}\sim F_{\upsilon_2^*,\;\upsilon_1^*}$.

So that

$$\frac{1-\Phi}{\Phi}\times\frac{\upsilon-p+1}{p}\sim F_{p,\upsilon-p+1}$$

$$\Rightarrow\;\frac{1-\frac{1}{1+T^2/(n-1)}}{\frac{1}{1+T^2/(n-1)}}\times\frac{n-p}{p}\sim F_{p,n-p}$$

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

Properties of Hotelling's $-T^2$

T^2 – Statistic as a function of likelihood ratio criterion

Let $\underline{x}_\alpha$ $(\alpha = 1, 2, \cdots, n > p)$ be a random sample of size n from $N_p(\underline{\mu}, \Sigma)$. The likelihood function is

$$L(\underline{\mu}, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underline{x}_\alpha - \underline{\mu})' \Sigma^{-1} (\underline{x}_\alpha - \underline{\mu})\right]$$

and the likelihood ratio criterion

$$\lambda = \frac{\max L_\omega(\underline{\mu}, \Sigma)}{\max L_\Omega(\underline{\mu}, \Sigma)} = \frac{\max L_0}{\max L}.$$

In the parameter space Ω, the maximum of L occurs when the parameters $\underline{\mu}$ and Σ are estimated by their maximum likelihood estimators i.e. $\hat{\underline{\mu}} = \underline{\bar{x}}$, and $\hat{\Sigma} = A/n$. In the space ω, we have $\underline{\mu} = \underline{\mu}_0$, and $\hat{\Sigma} = \frac{1}{n}\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'$, therefore,

$$\max L_\Omega = \frac{1}{(2\pi)^{np/2} \left|\frac{A}{n}\right|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underline{x}_\alpha - \underline{\bar{x}})' \left(\frac{A}{n}\right)^{-1} (\underline{x}_\alpha - \underline{\bar{x}})\right]$$

$$= \frac{n^{np/2}}{(2\pi)^{np/2} |A|^{n/2}} \exp\left(-\frac{1}{2}np\right).$$

Similarly,

$$\max L_\omega = \frac{1}{(2\pi)^{np/2} \left|\frac{1}{n}\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'\right|^{n/2}}$$

$$\exp\left[-\frac{1}{2} tr\, n\left\{\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'\right\}^{-1}\left\{\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'\right\}\right]$$

$$= \frac{n^{np/2}}{(2\pi)^{np/2} \left|\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'\right|^{n/2}} \exp\left(-\frac{1}{2}np\right).$$

Consider

$$\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)' = \sum_\alpha [(\underline{x}_\alpha - \underline{\bar{x}}) + (\underline{\bar{x}} - \underline{\mu}_0)][(\underline{x}_\alpha - \underline{\bar{x}}) + (\underline{\bar{x}} - \underline{\mu}_0)]'$$

$$= A + n(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)'.$$

Hence,

$$\max L_\omega = \frac{n^{np/2}}{(2\pi)^{np/2} \left|A + n(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)'\right|^{n/2}} \exp\left(-\frac{1}{2}np\right)$$

Thus, the likelihood ratio criterion is

$$\lambda = \frac{|A|^{n/2}}{\left|A + n(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)'\right|^{n/2}}$$

or $$\lambda^{2/n} = \frac{|A|}{\begin{vmatrix} 1 & -\sqrt{n}\,(\underline{\bar{x}} - \underline{\mu}_0)' \\ \sqrt{n}\,(\underline{\bar{x}} - \underline{\mu}_0) & A \end{vmatrix}} = \frac{|A|}{|A|\left|1 + n(\underline{\bar{x}} - \underline{\mu}_0)' A^{-1}(\underline{\bar{x}} - \underline{\mu}_0)\right|},$$

since $|\Sigma| = |\Sigma_{22}|\,\left|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right|$.

$$= \frac{1}{1 + n(\underline{\bar{x}} - \underline{\mu}_0)' A^{-1}(\underline{\bar{x}} - \underline{\mu}_0)}$$

$$= \frac{1}{1 + \frac{n}{n-1}(\underline{\bar{x}} - \underline{\mu}_0)' S^{-1}(\underline{\bar{x}} - \underline{\mu}_0)} = \frac{1}{1 + \frac{T^2}{n-1}},$$

where, $T^2 = n\,(\underline{\bar{x}} - \underline{\mu}_0)' S^{-1}(\underline{\bar{x}} - \underline{\mu}_0) = n(n-1)(\underline{\bar{x}} - \underline{\mu}_0)' A^{-1}(\underline{\bar{x}} - \underline{\mu}_0)$.

The likelihood ratio test is defined by the critical region $\lambda \le \lambda_0$, where, λ_0 is so chosen so as to have level α, i.e. $\Pr[\lambda \le \lambda_0 \mid H_0] = \alpha$.

Thus

$$\lambda^{2/n} \le \lambda_0^{2/n}, \text{ or } \frac{1}{1 + T^2/(n-1)} \le \lambda_0^{2/n}, \quad \text{or } 1 + \frac{T^2}{n-1} \ge \lambda_0^{-n/n},$$

or $T^2 \ge (n-1)(\lambda_0^{-2/n} - 1) = T_0^2$, (say)

$\Rightarrow T^2 \ge T_0^2$.

Therefore, $\Pr[T^2 \ge T_0^2 \mid H_0] = \alpha$.

Invariance property of T^2

Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, then $T_x^2 = n\,(\underline{\bar{x}} - \underline{\mu}_{0x})' S_x^{-1}(\underline{\bar{x}} - \underline{\mu}_{0x})$,

where

$$S_x = \frac{1}{n-1}\sum_\alpha (\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = \frac{1}{n-1}(\underline{x}_1\underline{x}_1' + \cdots + \underline{x}_n\underline{x}_n' - n\,\underline{\bar{x}}\,\underline{\bar{x}}')$$

Make a non-singular transformation

$$\underline{Y}=C\,\underline{X}\,,\quad\Rightarrow\quad \underline{y}_\alpha=C\,\underline{x}_\alpha$$

Now

$$S_y=\frac{1}{n-1}\sum_\alpha(\underline{y}_\alpha-\underline{\bar{y}})(\underline{y}_\alpha-\underline{\bar{y}})'=\frac{1}{n-1}(\underline{y}_1\underline{y}_1'+\cdots+\underline{y}_n\underline{y}_n'-n\,\underline{\bar{y}}\,\underline{\bar{y}}')$$

$$=\frac{1}{n-1}(C\,\underline{x}_1\underline{x}_1'C'+\cdots+C\,\underline{x}_n\underline{x}_n'C'-n\,C\,\underline{\bar{x}}\,\underline{\bar{x}}'C')$$

$$=C\left[\frac{1}{n-1}(\underline{x}_1\underline{x}_1'+\cdots+\underline{x}_n\underline{x}_n'-n\,\underline{\bar{x}}\,\underline{\bar{x}}')\right]C'$$

$$=C\,S_x\,C'.$$

By definition

$$T_y^2=n\,(\underline{\bar{y}}-\underline{\mu}_{0y})'S_y^{-1}(\underline{\bar{y}}-\underline{\mu}_{0y})=n\,(C\,\underline{\bar{x}}-C\,\underline{\mu}_{0x})'(C\,S_x\,C')^{-1}(C\,\underline{\bar{x}}-C\,\underline{\mu}_{0x})$$

$$=n\,(\underline{\bar{x}}-\underline{\mu}_{0x})'C'C'^{-1}S_x^{-1}C^{-1}C\,(\underline{\bar{x}}-\underline{\mu}_{0x})$$

$$=n\,(\underline{\bar{x}}-\underline{\mu}_{0x})'S_x^{-1}(\underline{\bar{x}}-\underline{\mu}_{0x})=T_x^2.$$

Uses of T^2 – statistic

One sample problem

Let $\underline{x}_1,\cdots,\underline{x}_n$ be a random sample from $N_p(\underline{\mu},\,\Sigma)$, when the variance covariance matrix Σ is unknown. Suppose we are required to test

$H_0:\underline{\mu}=\underline{\mu}_0$ (specific mean vector).

Let

$\underline{Y}=\sqrt{n}\,(\underline{\bar{x}}-\underline{\mu}_0)$, then, $E\underline{Y}=\underline{0}$, under H_0, and $\Sigma_{\underline{Y}}=\Sigma$. Thus, $\underline{Y}\sim N_p(\underline{0},\,\Sigma)$ and

$$(n-1)\,S=\sum_{\alpha=1}^{n}(\underline{x}_\alpha-\underline{\bar{x}})(\underline{x}_\alpha-\underline{\bar{x}})'=A=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha\underline{Z}_\alpha',\ \text{with}\ \ \underline{Z}_\alpha\sim N_p(\underline{0},\,\Sigma).$$

Therefore, by definition

$$T^2=\underline{Y}'S^{-1}\underline{Y},\ \text{and}\ \frac{T^2}{n-1}\,\frac{n-p}{p}\sim F_{p,n-p}.$$

Thus adopting a significance level of size α, the null hypothesis is rejected if

$$T^2\ge T_0^2,\ \text{where,}\ T_0^2=\frac{(n-1)\,p}{n-p}F_{p,n-p}(\alpha).$$

From the sample T^2 is computed easily as follows:

Let $(\underline{\bar{x}}-\underline{\mu}_0)=\underline{d}$, then solve

$S\,\underline{\lambda}=\underline{d}$ (by Doolittle method) $\Rightarrow$ $\underline{\lambda}=S^{-1}\underline{d}$, and $T^2=n\,\underline{d}'\,\underline{\lambda}$.

Let A be a matrix of order 3, then augmented matrix, say $B=\begin{pmatrix}A & \underline{d}\\ \underline{d}' & 0\end{pmatrix}\begin{matrix}3\\1\end{matrix}$

$$|\,B\,|=|\,A\,|\,\left|\,0-\underline{d}'A^{-1}\underline{d}\,\right|,\ \text{since}\ \begin{vmatrix}A & B\\ C & D\end{vmatrix}=\begin{cases}|\,A\,|\,|\,D-CA^{-1}B\,|\\ |\,D\,|\,|\,A-BD^{-1}C\,|\end{cases}$$

$$\Rightarrow\ \underline{d}'A^{-1}\underline{d}=-\frac{|\,B\,|}{|\,A\,|},$$

where

$|\,B\,|=a_{11}\,b_{11}\,c_{11}\,d_{11}$ (Doolittle method), $|\,A\,|=a_{11}\,b_{11}\,c_{11}$, then $\underline{d}'A^{-1}\underline{d}=-d_{11}$.

Note:

Sometimes we are given the $S\,S$ and $C\,P$ matrix A, then we solve

$A\underline{\lambda}^*=\underline{d}$ (by Doolittle method) $\Rightarrow$ $\underline{\lambda}^*=A^{-1}\underline{d}$ and $T^2=n\,(n-1)\,\underline{d}'\,\underline{\lambda}^*$.

Two sample problem

Let $\underline{x}_1^{(i)},\cdots,\underline{x}_n^{(i)}$ be a random sample from $N_p(\underline{\mu}^{(i)},\,\Sigma)$, $i=1,2$, where the variance covariance matrices are assumed equal but unknown. The hypothesis of interest is $H_0:\underline{\mu}^{(1)}=\underline{\mu}^{(2)}$.

Let

$$\underline{\bar{x}}^{(i)}=\frac{1}{n_i}\sum_{\alpha=1}^{n_i}\underline{x}_\alpha^{(i)}\ \text{ be the sample mean vector, and}$$

$$\underline{\bar{x}}^{(i)}\sim N_p(\underline{\mu}^{(i)},\Sigma/n_i),\ \text{then,}\ \underline{\bar{x}}^{(1)}-\underline{\bar{x}}^{(2)}\sim N_p\left[\underline{0},\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma\right],\ \text{under}\ H_0.$$

$$\Rightarrow\sqrt{\frac{n_1n_2}{n_1+n_2}}(\underline{\bar{x}}^{(1)}-\underline{\bar{x}}^{(2)})\sim N_p(\underline{0},\,\Sigma).$$

Let

$$\underline{Y}=\sqrt{\frac{n_1n_2}{n_1+n_2}}(\underline{\bar{x}}^{(1)}-\underline{\bar{x}}^{(2)}),\ \text{then,}$$

$$E\underline{Y}=\underline{0},\ \text{under}\ H_0,\ \text{and}\quad \Sigma_{\underline{Y}}=\frac{n_1n_2}{n_1+n_2}E\,(\underline{\bar{x}}^{(1)}-\underline{\bar{x}}^{(2)})(\underline{\bar{x}}^{(1)}-\underline{\bar{x}}^{(2)})'=\Sigma.$$

Therefore, $\underline{Y}\sim N_p(\underline{0},\,\Sigma)$.

Let

$$S^{(i)}=\frac{1}{n_i-1}\sum_{\alpha=1}^{n_i}(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})' \quad \text{and}$$

$$S=\frac{1}{n_1+n_2-2}\sum_{i=1}^{2}\sum_{\alpha=1}^{n_i}(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})'$$

$$=\frac{(n_1-1)S^{(1)}+(n_2-1)S^{(2)}}{n_1+n_2-2} \text{ be the pooled sample variance covariance matrix.}$$

or $(n_1+n_2-2)S=(n_1-1)S^{(1)}+(n_2-1)S^{(2)}$

$$=A^{(1)}+A^{(2)}=\sum_{\alpha=1}^{n_1+n_2-2}\underline{Z}_\alpha\underline{Z}_\alpha', \text{ with } \underline{Z}_\alpha\sim N_p(\underline{0},\Sigma).$$

Hence, $(n_1+n_2-2)S$ is distributed as $\sum_{\alpha=1}^{n_1+n_2-2}\underline{Z}_\alpha\underline{Z}_\alpha'$.

Therefore, by definition

$$T^2=\underline{Y}'S^{-1}\underline{Y}=\frac{n_1n_2}{n_1+n_2}(\underline{\bar{x}}^{(1)}-\underline{\bar{x}}^{(2)})'S^{-1}(\underline{\bar{x}}^{(1)}-\underline{\bar{x}}^{(2)})$$

and

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

Thus adopting a significance level of size α, the null hypothesis is rejected if $T^2\ge T_0^2$, where, $T_0^2=\frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}(\alpha)$.

$q-$sample problem

Let $\underline{x}_\alpha^{(i)}$ $(\alpha=1,2,\cdots,n_i;\ i=1,2,\cdots,q)$ be a random sample from $N_p(\underline{\mu}^{(i)},\Sigma)$ respectively. Suppose we are required to test

$$H_0:\sum_{i=1}^{q}\beta_i\,\underline{\mu}^{(i)}=\underline{\mu}, \text{ where } \beta_1,\cdots,\beta_q \text{ are given scalars and } \underline{\mu} \text{ is given vector.}$$

Let

$$\underline{\bar{x}}^{(i)}=\frac{1}{n_i}\sum_{\alpha=1}^{n_i}\underline{x}_\alpha^{(i)} \text{ be the sample mean vector, and } \underline{\bar{x}}^{(i)}\sim N_p(\underline{\mu}^{(i)},\Sigma/n_i)\text{, then,}$$

$$\sum_{i=1}^{q}\beta_i\,\underline{\bar{x}}^{(i)}\sim N_p\left(\underline{\mu},\frac{1}{C}\Sigma\right), \text{ under } H_0, \text{ where } E\left(\sum_i\beta_i\,\underline{\bar{x}}^{(i)}\right)=\underline{\mu}$$

and the variance covariance matrix is

$$\text{Var Cov}\left(\sum_i\beta_i\,\underline{\bar{x}}^{(i)}\right)=\sum_i\beta_i^2\text{ Var Cov}(\underline{\bar{x}}^{(i)})=\left(\sum_i\frac{1}{n_i}\beta_i^2\right)\Sigma=\frac{1}{C}\Sigma.$$

Let

$$\underline{Y}=\sqrt{C}\left(\sum_i\beta_i\,\underline{\bar{x}}^{(i)}-\underline{\mu}\right)\text{, then, } \underline{Y}\sim N_p(\underline{0},\Sigma).$$

Consider

$$S^{(i)}=\frac{1}{n_i-1}\sum_{\alpha=1}^{n_i}(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})'$$

and

$$S=\frac{1}{\sum_{i=1}^{q}n_i-q}\sum_{i=1}^{q}\sum_{\alpha=1}^{n_i}(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})'$$

$$\left(\sum_{i=1}^{q}n_i-q\right)S=\sum_{i=1}^{q}\sum_{\alpha=1}^{n_i}(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)}-\underline{\bar{x}}^{(i)})'$$

$$=\sum_{\alpha=1}^{\sum_i n_i-q}\underline{Z}_\alpha\underline{Z}_\alpha'\text{, where, } \underline{Z}_\alpha\sim N_p(\underline{0},\Sigma)$$

Therefore, by definition

$$T^2=\underline{Y}'S^{-1}\underline{Y}=C\left(\sum_i\beta_i\,\underline{\bar{x}}^{(i)}-\underline{\mu}\right)'S^{-1}\left(\sum_i\beta_i\,\underline{\bar{x}}^{(i)}-\underline{\mu}\right) \text{ is distributed as } T^2 \text{ with}$$

$\sum_i n_i-q$ degree of freedom, i.e.

$$\frac{T^2}{\sum_i n_i-q}\,\frac{\sum_i n_i-q-(p-1)}{p}\sim F_{p,\sum_i n_i-q-p+1}.$$

Thus adopting a significance level of size α, the null hypothesis is rejected if $T^2\ge T_0^2$,

where

$$T_0^2=\frac{\left(\sum_i n_i-q\right)p}{\sum_i n_i-q-p+1}F_{p,\sum_i n_i-q-p+1}(\alpha).$$

A problem of symmetry

Given a random sample $\underline{x}_1,\cdots,\underline{x}_n$ from $N_p(\underline{\mu},\Sigma)$, where $\underline{\mu}'=(\mu_1,\mu_2,\cdots,\mu_p)$. The hypothesis of interest is $H_0:\mu_1=\mu_2=\cdots=\mu_p$.

Let C be a matrix of order $(p-1)\times p$ such that $C\underline{\eta}=\underline{0}$, where $\underline{\eta}'=(1,1,\cdots,1)$.

A matrix satisfying this condition is

$$C=\begin{pmatrix}1&0&\cdots&0&-1\\0&1&\cdots&0&-1\\ \vdots&\vdots&\vdots&\vdots&\vdots\\0&0&\cdots&1&-1\end{pmatrix}_{(p-1)\times p},$$

where, C is called a contrast matrix.

Let

$$\underline{y}_\alpha=C\,\underline{x}_\alpha\text{, then,}$$

$$E\,\underline{y}_\alpha=C\,\underline{\mu}\text{, and }\Sigma_{\underline{y}_\alpha}=E[\underline{y}_\alpha-E(\underline{y}_\alpha)][\underline{y}_\alpha-E(\underline{y}_\alpha)]'=C\Sigma C',$$

with this transformation we can write

$$H_0:C\,\underline{\mu}=0\text{, i.e. }\begin{pmatrix}1&0&\cdots&0&-1\\0&1&\cdots&0&-1\\ \vdots&\vdots&\vdots&\vdots&\vdots\\0&0&\cdots&1&-1\end{pmatrix}_{(p-1)\times p}\begin{pmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_p\end{pmatrix}=0,\quad\text{or}\quad\begin{matrix}\mu_1-\mu_p=0\\ \mu_2-\mu_p=0\\ \vdots\\ \mu_{p-1}-\mu_p=0\end{matrix}$$

Therefore,

$$\underline{y}_\alpha\sim N_{p-1}(\underline{0},\,C\Sigma C')\text{ under }H_0\text{, then}$$

$$\underline{\bar{y}}\sim N_{p-1}\left(\underline{0},\frac{C\Sigma C'}{n}\right)\Rightarrow\quad\sqrt{n}\,\underline{\bar{y}}\sim N_{p-1}(\underline{0},\,C\Sigma C').$$

Let

$$S_{\underline{y}}=\frac{1}{n-1}\sum_{\alpha=1}^{n}(\underline{y}_\alpha-\underline{\bar{y}})(\underline{y}_\alpha-\underline{\bar{y}})'=\frac{1}{n-1}\sum_{\alpha=1}^{n}C(\underline{x}_\alpha-\underline{\bar{x}})(\underline{x}_\alpha-\underline{\bar{x}})'C'=CSC'$$

$$(n-1)S_{\underline{y}}=\sum_{\alpha=1}^{n-1}\underline{Z}^*_\alpha\underline{Z}^{*\prime}_\alpha\text{, with }\underline{Z}^*_\alpha\sim N_{p-1}(\underline{0},\,C\Sigma C').$$

Thus, by definition,

$T^2=n\,\underline{\bar{y}}'S_{\underline{y}}^{-1}\underline{\bar{y}}$ is distributed as T^2 with $(n-1)$ degree of freedom, and the critical region of size α for testing $H_0:C\,\underline{\mu}=0$ is

$$T^2\ge\frac{(n-1)(p-1)}{n-1-(p-2)}F_{p-1,n-p+1}(\alpha).$$

Behrens-Fisher problem

Let $\underline{x}_\alpha^{(i)}$ $(\alpha=1,2,\cdots,n_i;\ i=1,2)$ be random sample from $N_p(\underline{\mu}^{(i)},\Sigma_i)$. Hypothesis of interest is $H_0:\underline{\mu}^{(1)}=\underline{\mu}^{(2)}$. The mean $\underline{\bar{x}}^{(1)}$ of the first sample is normally distributed with expected value

$$E(\underline{\bar{x}}^{(1)})=\underline{\mu}^{(1)}\text{, and covariance matrix}$$

$$E(\underline{\bar{x}}^{(1)}-\underline{\mu}^{(1)})(\underline{\bar{x}}^{(1)}-\underline{\mu}^{(1)})'=\frac{1}{n_1}\Sigma_1\text{, i.e. }\underline{\bar{x}}^{(1)}\sim N_p(\underline{\mu}^{(1)},\Sigma_1/n_1).$$

Similarly,

$$\underline{\bar{x}}^{(2)}\sim N_p(\underline{\mu}^{(2)},\Sigma_2/n_2).$$

Thus,

$$(\underline{\bar{x}}^{(1)}-\underline{\bar{x}}^{(2)})\sim N_p\left[\underline{\mu}^{(1)}-\underline{\mu}^{(2)},\left(\frac{1}{n_1}\Sigma_1+\frac{1}{n_2}\Sigma_2\right)\right].$$

If $n_1=n_2=n$

Let $\underline{y}_\alpha=\underline{x}_\alpha^{(1)}-\underline{x}_\alpha^{(2)}$, (assuming the numbering of the observations in the two samples is independent of the observations themselves), then $\underline{y}_\alpha\sim N_p(\underline{0},\,\Sigma_1+\Sigma_2)$ under H_0.

$$\Rightarrow\ \underline{\bar{y}}=\frac{1}{n}\sum_{\alpha=1}^{n}\underline{y}_\alpha=(\underline{\bar{x}}^{(1)}-\underline{\bar{x}}^{(2)})\sim N_p\left(\underline{0},\frac{\Sigma_1+\Sigma_2}{n}\right)\quad\Rightarrow\sqrt{n}\,\underline{\bar{y}}\sim N_p(\underline{0},\,\Sigma_1+\Sigma_2).$$

Let

$$S_{\underline{y}}=\frac{1}{n-1}\sum_{\alpha=1}^{n}(\underline{y}_\alpha-\underline{\bar{y}})(\underline{y}_\alpha-\underline{\bar{y}})'$$

or $(n-1)S_{\underline{y}}=\sum_{\alpha=1}^{n-1}\underline{Z}_\alpha\underline{Z}_\alpha'$, where, $\underline{Z}_\alpha\sim N_{p-1}(\underline{0},\,\Sigma_1+\Sigma_2)$.

Thus, by definition, $T^2=n\,\underline{\bar{y}}'S_{\underline{y}}^{-1}\underline{\bar{y}}$ has T^2-distribution with $(n-1)$ degrees of freedom. The critical region is $T^2\ge\frac{(n-1)\,p}{n-p}F_{p,n-p}(\alpha)$.

If $n_1\ne n_2$, and $n_1<n_2$.

Define,

$$\underline{y}_\alpha=\underline{x}_\alpha^{(1)}-\sqrt{\frac{n_1}{n_2}}\,\underline{x}_\alpha^{(2)}+\frac{1}{\sqrt{n_1n_2}}\sum_{\beta=1}^{n_1}\underline{x}_\beta^{(2)}-\frac{1}{n_2}\sum_{\gamma=1}^{n_2}\underline{x}_\gamma^{(2)},\ \alpha=1,2,\cdots,n_1\text{, then}$$

$$E\,\underline{y}_\alpha=\underline{\mu}^{(1)}-\sqrt{\frac{n_1}{n_2}}\,\underline{\mu}^{(2)}+\frac{1}{\sqrt{n_1n_2}}\sum_{\beta=1}^{n_1}\underline{\mu}^{(2)}-\frac{1}{n_2}\sum_{\gamma=1}^{n_2}\underline{\mu}^{(2)}$$

$$=\underline{\mu}^{(1)}-\sqrt{\frac{n_1}{n_2}}\,\underline{\mu}^{(2)}+\sqrt{\frac{n_1}{n_2}}\,\underline{\mu}^{(2)}-\underline{\mu}^{(2)}=\underline{\mu}^{(1)}-\underline{\mu}^{(2)}.$$

The covariance matrix of $\underline{y}_\alpha$ and $\underline{y}_\beta$ is

$$E[\underline{y}_\alpha-E(\underline{y}_\alpha)][\underline{y}_\beta-E(\underline{y}_\beta)]'$$

$$=E\left[(\underline{x}_\alpha^{(1)}-\underline{\mu}^{(1)})-\sqrt{\frac{n_1}{n_2}}(\underline{x}_\alpha^{(2)}-\underline{\mu}^{(2)})+\frac{1}{\sqrt{n_1n_2}}\sum_{\gamma=1}^{n_1}(\underline{x}_\gamma^{(2)}-\underline{\mu}^{(2)})-\frac{1}{n_2}\sum_{\gamma=1}^{n_2}(\underline{x}_\gamma^{(2)}-\underline{\mu}^{(2)})\right]$$

$$\left[(\underline{x}_\beta^{(1)}-\underline{\mu}^{(1)})'-\sqrt{\frac{n_1}{n_2}}(\underline{x}_\beta^{(2)}-\underline{\mu}^{(2)})'+\frac{1}{\sqrt{n_1n_2}}\sum_{\gamma=1}^{n_1}(\underline{x}_\gamma^{(2)}-\underline{\mu}^{(2)})'-\frac{1}{n_2}\sum_{\gamma=1}^{n_2}(\underline{x}_\gamma^{(2)}-\underline{\mu}^{(2)})'\right]$$

$$=\Sigma_1+\frac{n_1}{n_2}\Sigma_2+\frac{1}{n_1n_2}n_1\Sigma_2+\frac{1}{n_2^2}n_2\Sigma_2-2\sqrt{\frac{n_1}{n_2}}\frac{1}{\sqrt{n_1n_2}}\Sigma_2$$

$$+2\sqrt{\frac{n_1}{n_2}}\frac{1}{n_2}\Sigma_2-2\frac{1}{\sqrt{n_1n_2}}\frac{1}{n_2}n_1\Sigma_2.$$

$$=\Sigma_1+\Sigma_2\left(\frac{n_1}{n_2}+\frac{1}{n_2}+\frac{1}{n_2}-\frac{2}{n_2}+\frac{2}{n_2}\sqrt{\frac{n_1}{n_2}}-\frac{2}{n_2}\sqrt{\frac{n_1}{n_2}}\right)$$

$$=\Sigma_1+\frac{n_1}{n_2}\Sigma_2.$$

Hence,

$$\underline{y}_\alpha\sim N_p\left(\underline{0},\,\Sigma_1+\frac{n_1}{n_2}\Sigma_2\right)\text{ under }H_0\text{, then}$$

$$\underline{\bar{y}}=\frac{1}{n_1}\sum_{\alpha=1}^{n_1}\underline{y}_\alpha\sim N_p\left[\underline{0},\frac{1}{n_1}\left(\Sigma_1+\frac{n_1}{n_2}\Sigma_2\right)\right]$$

$$\Rightarrow\ \sqrt{n_1}\,\underline{\bar{y}}\sim N_p\left(\underline{0},\Sigma_1+\frac{n_1}{n_2}\Sigma_2\right).$$

Consider

$$(n_1-1)S=\sum_{\alpha=1}^{n_1}(\underline{y}_\alpha-\underline{\bar{y}})(\underline{y}_\alpha-\underline{\bar{y}})'=\sum_{\alpha=1}^{n_1-1}\underline{Z}_\alpha\underline{Z}_\alpha',\text{ with }\underline{Z}_\alpha\sim N_p\left(\underline{0},\,\Sigma_1+\frac{n_1}{n_2}\Sigma_2\right).$$

Therefore, by definition,

$T^2=n_1\,\underline{\bar{y}}'\,S^{-1}\,\underline{\bar{y}}$. This statistic has T^2-distribution with (n_1-1) degree of freedom.

The critical region of size α is

$$T^2\geq\frac{(n_1-1)\,p}{n_1-p}\,F_{p,n_1-p}(\alpha).$$

Behrens-Fisher problem for $q-$sample

Suppose $\underline{x}_\alpha^{(i)}$ $(\alpha=1,2,\ldots,n_i\,;\ i=1,2,\ldots,q)$ be random sample from $N_p(\underline{\mu}^{(i)},\Sigma_i)$ respectively.

Consider testing the hypothesis

$$H_0:\sum_{i=1}^{q}\beta_i\underline{\mu}^{(i)}=\underline{\mu},$$

where $\beta_1,\beta_2,\ldots,\beta_q$ are given scalars and $\underline{\mu}$ is a given vector. If n_i are unequal, take n_1 to be the smallest.

Define,

$$\underline{y}_\alpha=\beta_1\,\underline{x}_\alpha^{(1)}+\sum_{i=2}^{q}\beta_i\sqrt{\frac{n_1}{n_i}}\left(\underline{x}_\alpha^{(i)}-\frac{1}{n_1}\sum_{\beta=1}^{n_1}\underline{x}_\beta^{(i)}+\frac{1}{\sqrt{n_1n_i}}\sum_{\gamma=1}^{n_i}\underline{x}_\gamma^{(i)}\right),\ \alpha=1,2,\ldots,n_1.$$

The expected value of $\underline{y}_\alpha$ is

$$E(\underline{y}_\alpha)=\beta_1\,\underline{\mu}^{(1)}+\sum_{i=2}^{q}\beta_i\sqrt{\frac{n_1}{n_i}}\left(\underline{\mu}^{(i)}-\frac{1}{n_1}n_1\,\underline{\mu}^{(i)}+\frac{n_i}{\sqrt{n_1n_i}}\underline{\mu}^{(i)}\right)$$

$$=\beta_1\,\underline{\mu}^{(1)}+\sum_{i=2}^{q}\beta_i\,\underline{\mu}^{(i)}=\sum_{i=1}^{q}\beta_i\,\underline{\mu}^{(i)}$$

and the variance covariance matrix of $\underline{y}_\alpha$ and $\underline{y}_\beta$ is

$$E[\underline{y}_\alpha-E(\underline{y}_\alpha)][\underline{y}_\beta-E(\underline{y}_\beta)]'=\sum_{i=1}^{q}\frac{n_1\beta_i^2}{n_i}\Sigma_i.$$

So that

$$\underline{y}_\alpha\sim N_p\left(\underline{\mu},\ \sum_{i=1}^{q}\frac{n_1\beta_i^2}{n_i}\Sigma_i\right)\text{ under }H_0$$

and

$$\underline{\bar{y}}=\frac{1}{n_1}\sum_{\alpha=1}^{n_1}\underline{y}_\alpha\sim N_p\left(\underline{\mu},\ \frac{1}{n_1}\sum_{i=1}^{q}\frac{n_1\beta_i^2}{n_i}\Sigma_i\right)\quad\Rightarrow\quad\sqrt{n_1}\,(\underline{\bar{y}}-\underline{\mu})\sim N_p\left(\underline{0},\sum_{i=1}^{q}\frac{n_1\beta_i^2}{n_i}\Sigma_i\right).$$

Consider

$$(n_1-1)S=\sum_{\alpha=1}^{n_1}(\underline{y}_\alpha-\underline{\bar{y}})(\underline{y}_\alpha-\underline{\bar{y}})'.$$

Therefore, $T^2=n_1(\underline{\bar{y}}-\underline{\mu})'S^{-1}(\underline{\bar{y}}-\underline{\mu})$ has T^2-distribution with (n_1-1) degree of freedom. The critical region is $T^2\geq\dfrac{(n_1-1)\,p}{n_1-p}F_{p,n_1-p}(\alpha)$, with α level of significance.

Test for equality of population mean of two sub vectors

Let $\underline{x}=\begin{pmatrix}\underline{x}^{(1)}\\ \underline{x}^{(2)}\end{pmatrix}\begin{matrix}q\\ q\end{matrix}$ be distributed normally with mean $\underline{\mu}=\begin{pmatrix}\underline{\mu}^{(1)}\\ \underline{\mu}^{(2)}\end{pmatrix}$ and covariance matrix $\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$, where both $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$ are $q-$dimensional vector (each of $q-$components). The hypothesis of interest is $H_0:\underline{\mu}^{(1)}=\underline{\mu}^{(2)}$. Let

$\underline{y}=\underline{x}^{(1)}-\underline{x}^{(2)}$, then the expectation of $\underline{y}$ is $E(\underline{y})=E(\underline{x}^{(1)}-\underline{x}^{(2)})=\underline{\mu}^{(1)}-\underline{\mu}^{(2)}$

and the variance covariance matrix of $\underline{y}$ is

$$E[\underline{y}-E(\underline{y})][\underline{y}-E(\underline{y})]'=E[(\underline{x}^{(1)}-\underline{x}^{(2)})-(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})]\,[(\underline{x}^{(1)}-\underline{x}^{(2)})-(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})]'$$

$$=E[(\underline{x}^{(1)}-\underline{\mu}^{(1)})(\underline{x}^{(1)}-\underline{\mu}^{(1)})'-(\underline{x}^{(1)}-\underline{\mu}^{(1)})(\underline{x}^{(2)}-\underline{\mu}^{(2)})'$$

$$-(\underline{x}^{(2)}-\underline{\mu}^{(2)})(\underline{x}^{(1)}-\underline{\mu}^{(1)})'+(\underline{x}^{(2)}-\underline{\mu}^{(2)})(\underline{x}^{(2)}-\underline{\mu}^{(2)})']$$

$$=\Sigma_{11}-\Sigma_{12}-\Sigma_{21}+\Sigma_{22}.$$

Thus,

$$\underline{y}\sim N(\underline{\mu}^{(1)}-\underline{\mu}^{(2)},\ \Sigma_{11}-\Sigma_{21}-\Sigma_{12}+\Sigma_{22})$$

and

$$\underline{\bar{y}}\sim N\left(\underline{\mu}^{(1)}-\underline{\mu}^{(2)},\ \frac{\Sigma_{11}-\Sigma_{21}-\Sigma_{12}+\Sigma_{22}}{n}\right)$$

$\Rightarrow\ \sqrt{n}\,\underline{\bar{y}}\sim N(\underline{0},\ \Sigma_{11}-\Sigma_{12}-\Sigma_{21}+\Sigma_{22})$, under H_0.

Consider,

$$(n-1)\,S=\sum_{\alpha=1}^{n}(\underline{y}_\alpha-\underline{\bar{y}})\,(\underline{y}_\alpha-\underline{\bar{y}})'$$

Therefore, by definition, $T^2=n\,\underline{\bar{y}}'\,(S_{11}-S_{12}-S_{21}+S_{22})^{-1}\,\underline{\bar{y}}$, which has T^2-distribution with $(n-1)$ degree of freedom. The critical region is

$$T^2\geq\frac{(n-1)\,q}{n-1-(q-1)}\,F_{q,n-q}(\alpha)\ \text{ with } \alpha \text{ level of significance.}$$

Result:

If $\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$, then $|\Sigma|=|\Sigma_{22}|\,\left|\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right|$

or $\dfrac{|\Sigma|}{|\Sigma_{22}|}=\left|\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right|=\dfrac{1}{\sigma^{11}}$, where σ^{11} is the leading term of Σ^{-1}.

Example: Let $\Sigma=\begin{pmatrix}2 & 1\\ 1 & 5\end{pmatrix}$, then $|\Sigma|=9$, and $\Sigma^{-1}=\begin{pmatrix}5/9 & -1/9\\ -1/9 & 2/9\end{pmatrix}$, so that

$$\frac{|\Sigma|}{|\Sigma_{22}|}=9\times\frac{1}{5}=9/5\text{, also } \frac{1}{\sigma^{11}}=\frac{1}{5/9}=9/5.$$

Exercise: Let $A\sim W_p(n-1,\Sigma)$ with $n>p$ and $|\Sigma|>0$. Let $\underline{u}$ be distributed independently of A. Prove that $\underline{u}'\Sigma^{-1}\underline{u}/\underline{u}'A^{-1}\underline{u}$ has χ^2_{n-p} distribution, independent of $\underline{u}$. Hence deduce the distribution of Hotelling's T^2.

Solution: Consider the variance covariance of conditional distribution of X_1 given $\underline{X}^{(2)}$, which is $\sigma_{11}-\underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{21}$, where

$$\Sigma=\begin{pmatrix}\sigma_{11} & \underline{\sigma}_{12}'\\ \underline{\sigma}_{21} & \Sigma_{22}\end{pmatrix}\begin{matrix}1\\ p-1\end{matrix}\text{, and } |\Sigma|=|\Sigma_{22}|\,\left|\sigma_{11}-\underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{21}\right|.$$

Thus,

$$\frac{|\Sigma|}{|\Sigma_{22}|}=\left|\sigma_{11}-\underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{21}\right|=\frac{1}{\sigma^{11}}\text{, where } \sigma^{11} \text{ is the leading term of } \Sigma^{-1}.$$

Let $A=\sum_\alpha(\underline{x}_\alpha-\underline{\bar{x}})(\underline{x}_\alpha-\underline{\bar{x}})'$ and partitioned according as Σ

$$A=\begin{pmatrix}a_{11} & \underline{a}_{12}'\\ \underline{a}_{21} & A_{22}\end{pmatrix}\begin{matrix}1\\ p-1\end{matrix}\text{, and } |A|=|A_{22}|\,\left|a_{11}-\underline{a}_{12}'A_{22}^{-1}\underline{a}_{21}\right|$$

or
$$\frac{|A|}{|A_{22}|}=\left|a_{11}-\underline{a}_{12}'A_{22}^{-1}\underline{a}_{21}\right|=\frac{1}{a^{11}} \tag{6.6}$$

where a^{11} is the leading term of A^{-1}.

Since $A\sim W_p(n-1,\Sigma)$, and, is partitioned as

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}\begin{matrix}q\\ p-q\end{matrix}\text{, then, } A_{11}-A_{12}A_{22}^{-1}A_{21}\sim W_q(n-1-(p-q),\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$$\Rightarrow\ a_{11}-\underline{a}_{12}'A_{22}^{-1}\underline{a}_{21}\sim W_1(n-1-(p-1),\ \sigma_{11}-\underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{21}) \tag{6.7}$$

In view of equation (6.6), equation (6.7) reduces as

$$\frac{1}{a^{11}}\sim W_1\left(n-p,\ \frac{1}{\sigma^{11}}\right)\ \Rightarrow\ \frac{\sigma^{11}}{a^{11}}\sim\chi^2_{n-p}.$$

Of course we could obtain the var. cov. of conditional distribution of any component X_j $(j=1,2,\cdots,p)$ of $\underline{X}$ given the other components. Hence, $\dfrac{\sigma^{jj}}{a^{jj}}\sim\chi^2_{n-p}$, where, σ^{jj} and a^{jj} are the leading term of Σ^{-1} and A^{-1} respectively.

Let C be a nonsingular matrix and let

$$B = C^{-1}, \Rightarrow B^{-1'} = C', \text{ as } A \sim W_p(n-1, \Sigma), \text{ and } BAB' \sim W_p(n-1, B\Sigma B').$$

Now

$$(BAB')^{-1} = B'^{-1}A^{-1}B^{-1} = C'A^{-1}C, \text{ and } (B\Sigma B')^{-1} = B'^{-1}\Sigma^{-1}B^{-1} = C'\Sigma^{-1}C$$

Let $C = \underline{C}_1' \cdots \underline{C}_p'$, where $\underline{C}_i$ is a $p-$component vector, then the leading term in $(BAB')^{-1}$ is $\underline{C}_1' A^{-1}\underline{C}_1$ and leading term in $(B\Sigma B')^{-1}$ is $\underline{C}_1'\Sigma^{-1}\underline{C}_1$, thus, $\dfrac{\underline{C}_1'\Sigma^{-1}\underline{C}_1}{\underline{C}_1' A^{-1}\underline{C}_1} \sim \chi^2_{n-p}$, as both the matrices are of order 1×1.

The result is true for any $\underline{C}_1$ even when $\underline{C}_1$ is obtained by observing a random vector, which is distributed independently of A.

Hence, if $\underline{u}$ and A are independently distributed then, $\dfrac{\underline{u}'\Sigma^{-1}\underline{u}}{\underline{u}' A^{-1}\underline{u}} \sim \chi^2_{n-p}$.

We know that

$$V_1 = n(\underline{\bar{x}} - \underline{\mu}_0)'\Sigma^{-1}(\underline{\bar{x}} - \underline{\mu}_0) \sim \chi^2_p$$

Let $\underline{u} = (\underline{\bar{x}} - \underline{\mu}_0)$

$$V_2 = \frac{(\underline{\bar{x}} - \underline{\mu}_0)'\Sigma^{-1}(\underline{\bar{x}} - \underline{\mu}_0)}{(\underline{\bar{x}} - \underline{\mu}_0)' A^{-1}(\underline{\bar{x}} - \underline{\mu}_0)} \sim \chi^2_{n-p}$$

$$= \frac{(n-1)(\underline{\bar{x}} - \underline{\mu}_0)'\Sigma^{-1}(\underline{\bar{x}} - \underline{\mu}_0)}{(\underline{\bar{x}} - \underline{\mu}_0)' S^{-1}(\underline{\bar{x}} - \underline{\mu}_0)} \times \frac{n}{n} \sim \chi^2_{n-p}, \text{ where } A = (n-1)S$$

V_1 and V_2 are independently distributed as χ^2

$$T^2 = n(\underline{\bar{x}} - \underline{\mu}_0)' S^{-1}(\underline{\bar{x}} - \underline{\mu}_0) = \frac{V_1/n}{V_2/n(n-1)} = (n-1)\frac{V_1}{V_2}$$

$$\Rightarrow \frac{T^2}{n-1} = \frac{V_1}{V_2} \quad \text{or} \quad \frac{T^2}{n-1} \times \frac{n-p}{n} = \frac{V_1/p}{V_2/n-p} \sim F_{p,\, n-p}.$$

Distribution of Mahalanobis's D^2

The quantity $(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$ is denoted by Δ^2 and was proposed by Mahalanobis as a measure of the distance between the two populations, $N_p(\underline{\mu}^{(1)}, \Sigma)$, and $N_p(\underline{\mu}^{(2)}, \Sigma)$. If the parameters are replaced by their unbiased estimates, is denoted by D^2, which is given by

$$D^2 = (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' S^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) \text{ and is known as Mahalanobis's } D^2,$$

where

$$S = \frac{(n_1-1)S^{(1)} + (n_2-1)S^{(2)}}{n_1+n_2-2} = \frac{1}{n_1+n_2-2}\sum_{i=1}^{2}\sum_{\alpha=1}^{n_i}(\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})',$$

where $\underline{\bar{x}}^{(i)} = \dfrac{1}{n_i}\displaystyle\sum_{\alpha=1}^{n_i}\underline{x}_\alpha^{(i)}$, $i = 1, 2$.

It is obvious that

$$T^2 = \frac{n_1 n_2}{n_1+n_2} D^2,$$

i.e. two-sample T^2 and D^2 are almost the same, except for the constant $k^2 = \dfrac{n_1 n_2}{n_1+n_2}$.

Let

$$\underline{Y} = k(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}), \text{ then expected value of } \underline{Y} \text{ is } E(\underline{Y}) = k(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) = \underline{\delta}$$

and the variance covariance matrix of $\underline{Y}$ is

$$\Sigma_{\underline{Y}} = k^2 E[(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) - (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})][(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) - (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})]'$$

$$= k^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\Sigma = \Sigma, \text{ because } \frac{n_1+n_2}{n_1 n_2} = \frac{1}{k^2}.$$

Therefore,

$$\underline{Y} \sim N_p(\underline{\delta}, \Sigma), \text{ then, } \quad k^2 D^2 = \underline{Y}' S^{-1}\underline{Y}.$$

Since Σ is positive definite matrix there exist a nonsingular matrix C such that

$$C\Sigma C' = I \Rightarrow CC' = \Sigma^{-1}.$$

Define,

$$\underline{Y}^* = C\underline{Y}, \; S^* = C\,S\,C', \text{ and } \underline{\delta}^* = C\,\underline{\delta}, \text{ then,}$$

$k^2 D^2 = \underline{Y}^{*'} S^{*-1}\underline{Y}^*$, and the expected value of $\underline{Y}^*$ is

$E(\underline{Y}^*) = C\,E(\underline{Y}) = C\underline{\delta} = \underline{\delta}^*$, and the variance covariance matrix of $\underline{Y}^*$ is

$$\Sigma_{\underline{Y}^*} = C\,E[\underline{Y} - E(\underline{Y})][(\underline{Y} - E(\underline{Y})]'C' = C\,\Sigma\,C' = I.$$

Thus,

$$\underline{Y}^* \sim N_p(\underline{\delta}^*, I), \Rightarrow \underline{Y}^{*'}\underline{Y}^* \sim \chi^2_p(\underline{\delta}^{*'}\underline{\delta}^*),$$

where

$$\underline{\delta}^{*'}\underline{\delta}^* = \underline{\delta}' C' C\,\underline{\delta} = \underline{\delta}'\Sigma^{-1}\underline{\delta} = \lambda^2.$$

Let

$$(n_1+n_2-2)\,S=\sum_{\alpha=1}^{n_1+n_2-2}\underline{Z}_\alpha\underline{Z}_\alpha{}',\ \text{where } \underline{Z}_\alpha\sim N_p(\underline{0},\,\Sigma)$$

$$\Rightarrow\ (n_1+n_2-2)\,S^*=\sum_{\alpha=1}^{n_1+n_2-2}(C\,\underline{Z}_\alpha)(C\,\underline{Z}_\alpha)',\ \text{where } C\,\underline{Z}_\alpha\sim N_p(\underline{0},\,I)\,.$$

Therefore,

$$k^2D^2=\underline{Y}^*S^{*-1}\underline{Y}^*=(n_1+n_2-2)\frac{\chi_p^2(\lambda^2)}{\chi^2_{n_1+n_2-2-(p-1)}}$$

$$\Rightarrow\ \frac{n_1n_2}{n_1+n_2}\frac{n_1+n_2-p-1}{(n_1+n_2-2)\,p}D^2=\frac{\chi_p^2(\lambda^2)\,/\,p}{\chi^2_{n_1+n_2-2-(p-1)}\,/\,n_1+n_2-p-1}\sim F_{p,n_1+n_2-p-1}(\lambda^2)\,.$$

If $\underline{\mu}^{(1)}=\underline{\mu}^{(2)}$, then the $F-$ distribution is central.

DISCRIMINANT ANALYSIS

The problem of discriminant analysis deals with assigning an individual to one of several categories on the basis of measurements on a $p-$component vector of variable $\underline{x}$ on that individual. For example, we take certain measurements on the skull of an animal and want to know whether it was male or female, a patient is to be classified as diabetic or non diabetic on the basis of certain tests such as blood, urine, blood pressure etc., a salesman is to be classified as successful or unsuccessful on different psychological tests.

Procedure of classification into one of the two populations with known probability distribution

Let R denote the entire $p-$dimensional space in which the point of observation $\underline{x}$ falls. We then have to divided R into two, say, R_1 and R_2 so that

If $\underline{x}$ falls in R_1, then assign the individual to population π_1

If $\underline{x}$ falls in R_2, assign the individual to population π_2

Obviously, with any such procedure an error of misclassification is inevitable (unavoidably) i.e. the rule may assign an individual to π_2, when he really belongs to π_1 and vice versa. A rule should control this error of discrimination.

Let $f_1(\underline{x})$ and $f_2(\underline{x})$ are the probability density function of $\underline{x}$ in the two populations π_1 and π_2. Let

$q_1 =$ a priori probability that an individual comes from π_1

$q_2 =$ a priori probability that an individual comes from π_2

$\Pr(1|2) = \Pr($ an individual belongs to π_2 is misclassified to $\pi_1)$

$\Pr(2|1) = \Pr($ an individual belongs to π_1 is misclassified to $\pi_2)$.

Obviously,

$$\Pr(1|2) = \int_{R_1} f_2(\underline{x})\, d\underline{x}, \text{ and } \Pr(2|1) = \int_{R_2} f_1(\underline{x})\, d\underline{x}.$$

Since the probability of drawing an observation from π_1 is q_1 and from π_2 is q_2, we have

$\Pr($ drawing an observation from π_1 and is misclassified as from $\pi_2) = q_1 \Pr(2|1)$

$\Pr($ drawing an observation from π_2 and is misclassified as from $\pi_1) = q_2 \Pr(1|2)$

Then the total chance of misclassification, say ϕ, is

$$\phi = q_1 \Pr(2|1) + q_2 \Pr(1|2) \tag{7.1}$$

We choose regions R_1 and R_2 such that equation (7.1) is minimized. The procedure that minimize (7.1) for a given q_1 and q_2 is called a Bayes procedure. Consider,

$$\begin{aligned}\phi &= q_1 \Pr(2|1) + q_2 \Pr(1|2) = q_1 \int_{R_2} f_1(\underline{x})\, d\underline{x} + q_2 \int_{R_1} f_2(\underline{x})\, d\underline{x} \\ &= q_1 \int_{R_2} f_1(\underline{x})\, d\underline{x} + q_1 \int_{R_1} f_1(\underline{x})\, d\underline{x} - q_1 \int_{R_1} f_1(\underline{x})\, d\underline{x} + q_2 \int_{R_1} f_2(\underline{x})\, d\underline{x} \\ &= q_1 \int_R f_1(\underline{x})\, d\underline{x} + \int_{R_1} [q_2 f_2(\underline{x}) - q_1 f_1(\underline{x})]\, d\underline{x}\end{aligned}$$

$\Rightarrow \phi$ is minimize, when $q_2 f_2(\underline{x}) \le q_1 f_1(\underline{x})$, so we divide R such as

$$R_1 = \{\underline{x} | q_2 f_2(\underline{x}) \le q_1 f_1(\underline{x})\} = \left\{\underline{x} \Big| \frac{q_1 f_1(\underline{x})}{q_2 f_2(\underline{x})} \ge 1\right\} = \left\{\underline{x} \Big| \frac{f_1(\underline{x})}{f_2(\underline{x})} \ge \frac{q_2}{q_1}\right\}.$$

Similarly,

$$R_2 = \left\{\underline{x} \Big| \frac{q_1 f_1(\underline{x})}{q_2 f_2(\underline{x})} < 1\right\} = \left\{\underline{x} \Big| \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{q_2}{q_1}\right\}.$$

Further, if the cost of misclassification is given, $C(2|1)$ cost of misclassification to π_2 when it actually comes from π_1, $C(1|2)$ the cost of misclassification to π_1 when it actually comes from π_2. Example, if a potentially good candidate for admission to a medical school is rejected, the nation will suffer a shortage in medical persons, but, on the contrary, if a bad candidate is admitted, he may not be able to complete the course successfully and money, resources equipment used by him will be a waste. Total expected cost from misclassification

$E(C) = C(2|1)\Pr(2|1) q_1 + C(1|2)\Pr(1|2) q_2$ and

classification rule will be

$$R_1 = \left\{\underline{x} \Big| \frac{f_1(\underline{x})}{f_2(\underline{x})} \ge \frac{q_2\, C(1|2)}{q_1\, C(2|1)}\right\}, \text{ and } R_2 = \left\{\underline{x} \Big| \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{q_2\, C(1|2)}{q_1\, C(2|1)}\right\}.$$

Classification into one of two known multivariate normal populations

Let $f_1(\underline{x}) =$ density function of $N_p(\underline{\mu}^{(1)}, \Sigma)$ and $f_2(\underline{x}) =$ density function of $N_p(\underline{\mu}^{(2)}, \Sigma)$ and the region R_1 of the classification into the population first is given by

$$R_1 = \left\{\underline{x} \,\Big|\, \frac{f_1(\underline{x})}{f_2(\underline{x})} \ge k\right\}, \text{ where } k = \frac{q_2\, C(1|2)}{q_1\, C(2|1)}. \text{ Consider}$$

$$\ln \frac{f_1(\underline{x})}{f_2(\underline{x})} \ge \ln k \tag{7.2}$$

The left hand side of (7.2) can be expanded as

$$\begin{aligned}\ln f_1(\underline{x}) - \ln f_2(\underline{x}) &= -\frac{1}{2}\{(\underline{x} - \underline{\mu}^{(1)})'\Sigma^{-1}(\underline{x} - \underline{\mu}^{(1)}) - (\underline{x} - \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{x} - \underline{\mu}^{(2)})\} \\ &= -\frac{1}{2}\{(\underline{x}'\Sigma^{-1}\underline{x} - \underline{x}'\Sigma^{-1}\underline{\mu}^{(1)} - \underline{\mu}^{(1)'}\Sigma^{-1}\underline{x} + \underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)}) \\ &\quad - (\underline{x}'\Sigma^{-1}\underline{x} - \underline{x}'\Sigma^{-1}\underline{\mu}^{(2)} - \underline{\mu}^{(2)'}\Sigma^{-1}\underline{x} + \underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)})\} \\ &= -\frac{1}{2}\{-2\,\underline{x}'\Sigma^{-1}\underline{\mu}^{(1)} + 2\,\underline{x}'\Sigma^{-1}\underline{\mu}^{(2)} + \underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)} - \underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)}\} \\ &= \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)} - \underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)}).\end{aligned}$$

Thus,

$$R_1 = \{\underline{x} | \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)} - \underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)}) \ge \ln k\}.$$

In particular, when

$C(1|2)=C(2|1)$, and $q_1=q_2=1/2$, then $k=1$, $\ln k=0$.

Then the region of classification into population first is

$$R_1=\{\underline{x}\mid \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})-\frac{1}{2}(\underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)}-\underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)})\geq 0\}$$

$$=\{\underline{x}\mid \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})\geq\frac{1}{2}(\underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)}-\underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)})\}$$

$$=\{\underline{x}\mid \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})\geq\frac{1}{2}(\underline{\mu}^{(1)}+\underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})\}.$$

The term $\underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})$ is known as the Fisher's discriminant function.

Similarly, the region of classification into population second is

$$R_2=\{\underline{x}\mid \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})<\frac{1}{2}(\underline{\mu}^{(1)}+\underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})\}$$

The regions are computed easily as follows:

Consider

$\Sigma^{-1}\underline{d}=\underline{\delta}$, then solve, $\Sigma\underline{\delta}=\underline{d}$ (by Doolittle method), where $\underline{d}=\underline{\mu}^{(1)}-\underline{\mu}^{(2)}$.

Probability of misclassification (Two known $p-$variate normal population)

If the observation is from π_1, then

$$\Pr(2\mid 1)=\int_{R_2}f_1(\underline{x})\,d\underline{x}=\int_{R_2}f(\underline{x}:\underline{\mu}^{(1)},\Sigma)\,d\underline{x},\text{ where}$$

$$R_2=\{\underline{x}\mid \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})<\frac{1}{2}(\underline{\mu}^{(1)}+\underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})\}.\text{ Put}$$

$U=\underline{x}'\underline{\delta}$, where $\underline{\delta}=\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})$, and $h=\frac{1}{2}(\underline{\mu}^{(1)}+\underline{\mu}^{(2)})\underline{\delta}$

$\Rightarrow R_2=\{\underline{x}\mid U<h\}$.

Since $\underline{x}\sim N_p(\underline{\mu}^{(1)},\Sigma)$, then $U=\underline{x}'\underline{\delta}$ is the univariate normal with following parameters

$$E(U)=\underline{\mu}^{(1)'}\underline{\delta},\text{ and }\operatorname{Var}(U)=\underline{\delta}'\Sigma\underline{\delta}=(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})=\Delta^2.$$

Therefore,

$$\Pr(2|1)=\int_{-\infty}^{h}\frac{1}{\Delta\sqrt{2\pi}}\exp\left\{-\frac{1}{2\Delta^2}(U-\underline{\mu}^{(1)'}\underline{\delta})^2\right\}du$$

Make a transformation

$$\frac{U-\underline{\mu}^{(1)'}\underline{\delta}}{\Delta}=y,\quad\Rightarrow du=\Delta\,dy.$$

Thus,

$$\Pr(2|1)=\int_{-\infty}^{(h-\underline{\mu}^{(1)'}\underline{\delta})/\Delta}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(\underline{\mu}^{(2)}-\underline{\mu}^{(1)})'\underline{\delta}/2\Delta}e^{-y^2/2}\,dy.$$

Similarly,

$\Pr(1|2)=\int_{R_1}f_2(\underline{x})\,d\underline{x}=\int_{R_1}f(\underline{x}:\underline{\mu}^{(2)},\Sigma)\,d\underline{x}$, where $R_1=\{\underline{x}\mid U\geq h\}$, $E(U)=\underline{\mu}^{(2)'}\underline{\delta}$ and $\operatorname{Var}(U)=\Delta^2$. Put

$$\frac{U-\underline{\mu}^{(2)'}\underline{\delta}}{\Delta}=y,\quad\Rightarrow du=\Delta\,dy.\text{ Therefore,}$$

$$\Pr(1|2)=\frac{1}{\sqrt{2\pi}}\int_{(h-\underline{\mu}^{(2)'}\underline{\delta})/\Delta}^{\infty}e^{-y^2/2}\,dy=\frac{1}{\sqrt{2\pi}}\int_{(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})'\underline{\delta}/2\Delta}^{\infty}e^{-y^2/2}\,dy.$$

Sample discriminant function

Suppose that we have a sample $\underline{x}_1^{(1)},\cdots,\underline{x}_{n_1}^{(1)}$ from $N_p(\underline{\mu}^{(1)},\Sigma)$ and a sample $\underline{x}_1^{(2)},\cdots,\underline{x}_{n_2}^{(2)}$ from $N_p(\underline{\mu}^{(2)},\Sigma)$, and the unbiased estimate of $\underline{\mu}^{(1)}$ is

$$\underline{\overline{x}}^{(1)}=\frac{1}{n_1}\sum_{\alpha=1}^{n_1}\underline{x}_\alpha^{(1)},\text{ and }\underline{\mu}^{(2)}\text{ is }\underline{\overline{x}}^{(2)}=\frac{1}{n_2}\sum_{\alpha=1}^{n_2}\underline{x}_\alpha^{(2)},\text{ and of }\Sigma\text{ is }S\text{ defined by}$$

$$S=\frac{1}{n_1+n_2-2}\left[\sum_{\alpha=1}^{n_1}(\underline{x}_\alpha^{(1)}-\underline{\overline{x}}^{(1)})(\underline{x}_\alpha^{(1)}-\underline{\overline{x}}^{(1)})'+\sum_{\alpha=1}^{n_1}(\underline{x}_\alpha^{(2)}-\underline{\overline{x}}^{(2)})(\underline{x}_\alpha^{(2)}-\underline{\overline{x}}^{(2)})'\right].$$

Substitute these estimates for the parameters in the function $\underline{x}'\underline{\delta}$, Fisher's discriminant function becomes

$$\underline{x}'S^{-1}(\underline{\overline{x}}^{(1)}-\underline{\overline{x}}^{(2)}).$$

This is known as sample discriminant function. The classification procedure now becomes

i) Compute $\underline{x}'S^{-1}(\underline{\overline{x}}^{(1)}-\underline{\overline{x}}^{(2)})=\underline{x}'\underline{\delta}$

ii) Compute $\frac{1}{2}(\underline{\overline{x}}^{(1)}+\underline{\overline{x}}^{(2)})'S^{-1}(\underline{\overline{x}}^{(1)}-\underline{\overline{x}}^{(2)})=\frac{1}{2}(\underline{\overline{x}}^{(1)}+\underline{\overline{x}}^{(2)})'\underline{\delta}$

iii) Assign the individual with measurements $\underline{x}$ to population first or population second, according as $\underline{x}'\underline{\delta}-\frac{1}{2}(\underline{\overline{x}}^{(1)}+\underline{\overline{x}}^{(2)})'\underline{\delta}$ is ≥ 0 or <0.

Classification into one of several populations

Consider the m populations say $\pi_1,\cdots,\pi_m$ with the priori probabilities $q_1,\cdots,q_m$ and the density functions $f_1(\underline{x}),\cdots,f_m(\underline{x})$. We wish to divide the $p-$dimensional space R, in which the point of observation $\underline{x}$ falls, into m mutually exclusive and exhaustive regions $R_1,\cdots,R_m$.

If $\underline{x}$ falls in R_i, then assign the individual to population π_i.

Let

$$\Pr(j|i) = \Pr(\text{an individual belongs to } \pi_i \text{ is misclassified to } \pi_j) = \int_{R_j} f_i(\underline{x})\, d\underline{x}$$

and

$$C(j|i) = \text{Cost of misclassifying an observation from } \pi_i \text{ as coming from } \pi_j.$$

Since the probability that an observation comes from π_i is q_i.

$$\Pr(\text{an observation belongs to } \pi_i \text{ and is classified as coming from } \pi_j)$$

$$= \Pr(\text{an observation comes from } \pi_i) \times \Pr(\text{misclassified it as coming from } \pi_j)$$

$$= q_i \Pr(j|i).$$

The total expected cost from misclassification

$$\sum_{i=1}^{m}\left[\sum_{j=1}^{m} C(j|i)\, q_i \Pr(j|i)\right], \quad i \neq j.$$

We would like to choose regions $R_1, \cdots, R_m$ to make this expected cost minimum.

It can be seen that the classification rule comes out to be

Assign $\underline{x}$ to π_k if

$$\sum_{\substack{i=1\\ i\neq k}}^{m} q_i f_i(\underline{x})\, C(k|i) < \sum_{\substack{i=1\\ i\neq j}}^{m} q_i f_i(\underline{x})\, C(j|i), \; j = 1, 2, \cdots, m; \quad j \neq k$$

Let $\underline{X} \sim N_p(\underline{\mu}^{(i)}, \Sigma)$, $i = 1, 2, \cdots, m$, let the cost of misclassification be equal. Then the rule is assign $\underline{x}$ to π_k if

$$\sum_{\substack{i=1\\ i\neq k}}^{m} q_i f_i(\underline{x}) < \sum_{\substack{i=1\\ i\neq j}}^{m} q_i f_i(\underline{x}), \; j = 1, 2, \cdots, m; \quad j \neq k$$

$$\sum_{\substack{i=1\\ i\neq k}}^{m} q_i f_i(\underline{x}) + q_k f_k(\underline{x}) - q_k f_k(\underline{x}) < \sum_{\substack{i=1\\ i\neq j}}^{m} q_i f_i(\underline{x}) + q_j f_j(\underline{x}) - q_j f_j(\underline{x})$$

$$\Rightarrow \sum_{\substack{i=1\\ i\neq k}}^{m} q_i f_i(\underline{x}) + q_k f_k(\underline{x}) = \sum_{i=1}^{m} q_i f_i(\underline{x}) = \sum_{\substack{i=1\\ i\neq j}}^{m} q_i f_i(\underline{x}) + q_j f_j(\underline{x})$$

$$\Rightarrow q_k f_k(\underline{x}) > q_j f_j(\underline{x})$$

$$\Rightarrow \frac{f_k(\underline{x})}{f_j(\underline{x})} > \frac{q_j}{q_k}.$$

or assign $\underline{x}$ to π_j if $\dfrac{f_j(\underline{x})}{f_k(\underline{x})} > \dfrac{q_k}{q_j}$

or assign $\underline{x}$ to π_j if $U_{jk}(\underline{x}) = \ln \dfrac{f_j(\underline{x})}{f_k(\underline{x})} > \ln \dfrac{q_k}{q_j}$

$$\ln \frac{f_j(\underline{x})}{f_k(\underline{x})} = \ln \frac{\exp[-1/2(\underline{x} - \underline{\mu}^{(j)})' \Sigma^{-1} (\underline{x} - \underline{\mu}^{(j)})]}{\exp[-1/2(\underline{x} - \underline{\mu}^{(k)})' \Sigma^{-1} (\underline{x} - \underline{\mu}^{(k)})]}$$

$$= \left[\underline{x} - \frac{1}{2}(\underline{\mu}^{(j)} + \underline{\mu}^{(k)})\right]' \Sigma^{-1} (\underline{\mu}^{(j)} - \underline{\mu}^{(k)}).$$

Note that

$$U_{kj}(\underline{x}) = \left[\underline{x} - \frac{1}{2}(\underline{\mu}^{(k)} + \underline{\mu}^{(j)})\right]' \Sigma^{-1} (\underline{\mu}^{(k)} - \underline{\mu}^{(j)})$$

Therefore,

$$U_{jk}(\underline{x}) = -U_{kj}(\underline{x}).$$

There are ${}^{m}C_2$ combinations of $U_{jk}(\underline{x})$ for m categories

For $m = 3$, combinations are $U_{12}\; U_{13}\; U_{23}$.

$$\left.\begin{aligned} \ln \frac{f_1(\underline{x})}{f_2(\underline{x})} &= U_{12} > \ln \frac{q_2}{q_1} \\ \ln \frac{f_1(\underline{x})}{f_3(\underline{x})} &= U_{13} > \ln \frac{q_3}{q_1} \end{aligned}\right\} \text{assign } \underline{x} \text{ to } \pi_1$$

$$\left.\begin{aligned} \ln \frac{f_2(\underline{x})}{f_1(\underline{x})} &= U_{21} > \ln \frac{q_1}{q_2} \\ \ln \frac{f_2(\underline{x})}{f_3(\underline{x})} &= U_{23} > \ln \frac{q_3}{q_2} \end{aligned}\right\} \text{assign } \underline{x} \text{ to } \pi_2$$

$$\left.\begin{aligned} \ln \frac{f_3(\underline{x})}{f_1(\underline{x})} &= U_{31} > \ln \frac{q_1}{q_3} \\ \ln \frac{f_3(\underline{x})}{f_2(\underline{x})} &= U_{32} > \ln \frac{q_2}{q_3} \end{aligned}\right\} \text{assign } \underline{x} \text{ to } \pi_3.$$

Sample discriminant function

$\underline{X} \sim N_p(\underline{\mu}^{(i)}, \Sigma)$, $\underline{\mu}^{(i)}$, Σ are unknown, then replace $\underline{\mu}^{(i)}$ by $\underline{\bar{x}}^{(i)}$ and Σ by S, where

$$S = \frac{1}{\sum_{i=1}^{m} n_i - m} \sum_{i=1}^{m} \sum_{\alpha=1}^{n_i} (\underline{x}_{i\alpha} \underline{x}'_{i\alpha} - n_i \underline{\bar{x}}^{(i)} \underline{\bar{x}}^{(i)'}) \quad \text{or} \quad \left(\sum_{i=1}^{m} n_i - m\right) S = A_1 + A_2 + \cdots + A_m.$$

Substitute these estimates for the parameters in the function, the function becomes

$$V_{jk}(\underline{x}) = \left[\underline{x} - \frac{1}{2}(\underline{\bar{x}}^{(j)} + \underline{\bar{x}}^{(k)})\right]' S^{-1}(\underline{\bar{x}}^{(j)} - \underline{\bar{x}}^{(k)}).$$

Tests associated with discriminant functions

1) Goodness of fit of a hypothetical discriminant function

H_0: A given function $\underline{x}'\underline{\delta}$ is good enough for discriminating between two populations.

We use the test statistic

$$\frac{n_1+n_2-p-1}{p-1}\left[\frac{k^2(D_p^2 - D_1^2)}{(n_1+n_2-2)+k^2D_1^2}\right] \sim F_{p-1,\, n_1+n_2-p-1},$$

where $k^2 = \frac{n_1 n_2}{n_1+n_2}$, $D_p^2 = (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' S^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})$ is the studentized D^2 of Mahalanobis, based on the p characters $\underline{x}$, and

$$D_1^2 = \frac{(\underline{\delta}'\underline{d})^2}{\underline{\delta}' S \underline{\delta}}, \quad \text{where } \underline{d} = (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}),\ \underline{\delta} = S^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}).$$

2) Test for additional information

$H_0: \Delta_p^2 = \Delta_q^2$ (i.e. $p-q$ components do not provide additional information).

We use the test statistic

$$\frac{n_1+n_2-p-1}{p-q}\left[\frac{k^2(D_p^2 - D_q^2)}{(n_1+n_2-2)+k^2D_q^2}\right] \sim F_{p-q,\, n_1+n_2-p-1},$$

where, $D_q^2 = \underline{d}_1' S_{11}^{-1} \underline{d}_1$, $\underline{d} = \begin{pmatrix}\underline{d}_1 \\ \underline{d}_2\end{pmatrix}$, and $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22}\end{pmatrix}$.

Exercise: If $f_i(\underline{x})$ is a $p-$variate normal density with parameters $\underline{\mu}^{(i)}$ and Σ for $i=1,2$ with $|\Sigma|>0$, and the region R_1 of classification is $U = \left\{\underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})\right\} \ge 0$, where $q_1 = q_2$, and $C(1|2) = C(2|1)$. Find the distribution of U, if $\underline{x} \sim N_p(\underline{\mu}^{(1)}, \Sigma)$.

Solution: Given

$$\underline{x} \sim N_p(\underline{\mu}^{(1)}, \Sigma), \text{ and } U = \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$$

$$E(U) = \underline{\mu}^{(1)'}\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$$

$$= \frac{1}{2}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}), \text{ and}$$

$$V(U) = E(U - EU)(U - EU)'$$

$$= E\Big[\underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})\Big]$$

$$\Big[\underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})\Big]'$$

$$= E\left[\left\{\underline{x}' - \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})' - \frac{1}{2}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\right\}\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})\right]$$

$$\left[\left\{\underline{x}' - \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})' - \frac{1}{2}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\right\}\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})\right]'$$

$$= E[(\underline{x} - \underline{\mu}^{(1)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})][(\underline{x} - \underline{\mu}^{(1)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})]'$$

$$= (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\Sigma^{-1} E(\underline{x} - \underline{\mu}^{(1)})(\underline{x} - \underline{\mu}^{(1)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$$

$$= (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) = \alpha \text{ (say).}$$

Therefore,

$$U \sim N_p(\alpha/2, \alpha)$$

Similarly, if $\underline{x} \sim N_p(\underline{\mu}^{(2)}, \Sigma)$, then

$$E(U) = \underline{\mu}^{(2)'}\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$$

$$= \frac{1}{2}(\underline{\mu}^{(2)} - \underline{\mu}^{(1)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) = -\frac{1}{2}\alpha, \text{ and}$$

$$V(U) = (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) = \alpha.$$

Thus,

$$U \sim N_p(-\alpha/2, \alpha).$$

PRINCIPAL COMPONENTS

In multivariate analysis the dimension of $\underline{X}$ causes problems in obtaining suitable statistical analysis to analyze a set of observations (data) on $\underline{X}$. It is natural to look for method for rearranging the data, so that with as little loss of information as possible, the dimension of problem is considerably reduced. This reduction is possible by transforming the original variable to a new set of variables which are uncorrelated. These new variables are known as principal components.

Principal components are normalized linear combination's of original variables which has specified properties interms of variance. They are uncorrelated and are ordered, so that the first component displays the largest amount of variation, the second components displays the second largest amount of variation, and so on.

If there are $p-$variables, then p components are required to reproduce (rearrange) the total variability present in the data, this variability can be accounted for by a small number, k $(k<p)$ of the components. If this is so, there is almost as much information in the k components as there is in original p variables and then k components can be replace the original p variables. . That is why; this is considered as linear reduction technique. This technique produces best results when the original variables are highly correlated, positively or negatively.

For example, suppose we are interested in finding the level of performance in mathematics of the tenth grade students of a certain school. We may then record their scores in mathematics: i.e., we consider just one characteristic of each student. Now suppose we are instead interested in overall performance and select some p characteristics, such as mathematics, English, history, science, etc. These characteristics, although related to each other, may not all contain the same amount of information, and in fact some characteristics may be completely redundant. Obviously, this will result in a loss of information and waste of resources in analyzing the data. Thus, we should select only those characteristics that will truly discriminate one student from another, while those least discriminatory should be discarded.

Determination

Let $\underline{X}$ be a $p-$component vector with covariance matrix Σ be positive definite (all the characteristic roots are positive and distinct). Since we are interested only in the variances and covariances, we shall assume that $E(\underline{X})=\underline{0}$. Let $\underline{\beta}$ is a $p-$component column vector such that $\underline{\beta}'\underline{\beta}=1$ (in order to obtain a unique solution).

Define,

$$U=\underline{\beta}'\underline{X}\text{, then } EU=E(\underline{\beta}'\underline{X})=0\text{, and } V(\underline{\beta}'\underline{X})=\underline{\beta}'E(\underline{X}\,\underline{X}')\underline{\beta}=\underline{\beta}'\Sigma\underline{\beta} \qquad (8.1)$$

The normalized linear combination $\underline{\beta}'\underline{X}$ with maximum variance is therefore obtained by maximizing equation (8.1) subject to condition $\underline{\beta}'\underline{\beta}=1$. We shall use technique of Lagrange multiplier and maximize

$$\phi=\underline{\beta}'\Sigma\underline{\beta}-\lambda(\underline{\beta}'\underline{\beta}-1)=\sum_{i,\,j}\beta_i\sigma_{ij}\beta_j-\lambda\left(\sum_i\beta_i^2-1\right),$$

where λ is a Lagrange multiplier, the vector of partial derivatives $\left(\dfrac{\partial\phi}{\partial\beta_i}\right)$ is

$$\frac{\partial\phi}{\partial\underline{\beta}}=2\Sigma\underline{\beta}-2\lambda\underline{\beta}\text{, equating this to zero, gives}$$

$$\Sigma\underline{\beta}-\lambda\underline{\beta}=\underline{0},$$

$$\Rightarrow\ (\Sigma-\lambda I)\underline{\beta}=\underline{0}. \qquad (8.2)$$

The equation (8.2) admits a non-zero solution in $\underline{\beta}$ if only,

$$|\Sigma-\lambda I|=\underline{0}. \qquad (8.3)$$

i.e. $(\Sigma-\lambda I)$ is a singular matrix.

The function $|\Sigma-\lambda I|$ is a polynomial in λ of degree p (i.e. $\lambda^p+a_1\lambda^{p-1}+\cdots+a_{p-1}\lambda+a_p$). Therefore, the equation (8.3) will give p solutions in λ, let these be $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_p$.

Pre-multiplying (8.2) with $\underline{\beta}'$, gives

$$\underline{\beta}'\Sigma\underline{\beta}-\lambda\underline{\beta}'\underline{\beta}=0$$

$$\Rightarrow\ \underline{\beta}'\Sigma\underline{\beta}=\lambda. \qquad (8.4)$$

This shows if $\underline{\beta}$ satisfies equation (8.2) and $\underline{\beta}'\underline{\beta}=1$, then the variance of $\underline{\beta}'\underline{X}$ is λ.

Thus for the maximum variance we should use in (8.2) the largest root λ_1. Let $\underline{\beta}^{(1)}$ be a normalized solution of

$$(\Sigma-\lambda_1 I)\underline{\beta}=\underline{0}\quad\text{ i.e. }\Sigma\underline{\beta}^{(1)}=\lambda_1\underline{\beta}^{(1)}\text{, then}$$

$$U_1=\underline{\beta}^{(1)'}\underline{X}$$

is a normalized linear combination (first principal component) with maximum variance i.e.

$$V(U_1)=\underline{\beta}^{(1)'}\Sigma\underline{\beta}^{(1)}=\lambda_1 \text{ [from equation (9.4)].}$$

We have next to find a normalized linear combination $\underline{\beta}'\underline{X}$ that has maximum variance of all linear combinations un-correlated with U_1. Non-correlation with U_1 is same as

$$0=\text{Cov}(\underline{\beta}'\underline{X},U_1)=E\,(\underline{\beta}'\underline{X}-E\underline{\beta}'\underline{X})(\underline{\beta}^{(1)'}\underline{X}-E\underline{\beta}^{(1)'}\underline{X})'$$

$$=E\,(\underline{\beta}'\underline{X}\,\underline{X}'\underline{\beta}^{(1)})=\underline{\beta}'\Sigma\underline{\beta}^{(1)}.$$

Since $\Sigma\underline{\beta}^{(1)}=\lambda_1\underline{\beta}^{(1)}$, then

$$\underline{\beta}'\Sigma\underline{\beta}^{(1)}=\lambda_1\underline{\beta}'\underline{\beta}^{(1)}=0. \qquad (8.5)$$

Thus we have to determine $\underline{\beta}$ as to maximize the function

$$\phi_2 = \underline{\beta}' \Sigma \underline{\beta} - \lambda (\underline{\beta}' \underline{\beta} - 1) - 2\upsilon_1 (\underline{\beta}' \Sigma \underline{\beta}^{(1)}),$$

where λ and υ_1 are Lagrange multipliers. The vector of partial derivatives is

$$\frac{\partial \phi_2}{\partial \underline{\beta}} = 2\Sigma \underline{\beta} - 2\lambda \underline{\beta} - 2\upsilon_1 \Sigma \underline{\beta}^{(1)}, \text{ and we set this to } \underline{0}.$$

$$\Rightarrow \Sigma \underline{\beta} - \lambda \underline{\beta} - \upsilon_1 \Sigma \underline{\beta}^{(1)} = \underline{0}.$$

Premultiplying by $\underline{\beta}^{(1)'}$

$$\underline{\beta}^{(1)'} \Sigma \underline{\beta} - \lambda \underline{\beta}^{(1)'} \underline{\beta} - \upsilon_1 \underline{\beta}^{(1)'} \Sigma \underline{\beta}^{(1)} = 0.$$

Since $\underline{\beta}' \Sigma \underline{\beta}^{(1)} = \lambda_1 \underline{\beta}' \underline{\beta}^{(1)} = 0$, so that

$$-\upsilon_1 \underline{\beta}^{(1)'} \Sigma \underline{\beta}^{(1)} = 0 \quad \text{or} \quad \upsilon_1 \lambda_1 = 0, \text{ because } \underline{\beta}' \Sigma \underline{\beta} = \lambda \quad \Rightarrow \quad \underline{\beta}^{(1)'} \Sigma \underline{\beta}^{(1)} = \lambda_1.$$

$\Rightarrow \upsilon_1 = 0$, because $\lambda_1 \neq 0$.

This means that the condition of uncorrelated ness is itself satisfied when $\underline{\beta}$ satisfied (8.2) and λ satisfied (8.3).

So the procedure to find the second principal component is to solve (8.2) for $\lambda = \lambda_2$, and call the vector as $\underline{\beta}^{(2)}$ and the corresponding linear combination

$U_2 = \underline{\beta}^{(2)'} \underline{X}$, clearly,

$$\mathrm{V}(U_2) = \underline{\beta}^{(2)'} \Sigma \underline{\beta}^{(2)} = \lambda_2$$

Similarly, $\underline{\beta}^{(3)}, \underline{\beta}^{(4)}, \cdots$, and their linear combinations

$$\underline{\beta}^{(3)'} \underline{X}, \underline{\beta}^{(4)'} \underline{X}, \cdots$$

and their variances with $\lambda_3, \lambda_4, \cdots$

As a general case, we want to find a vector $\underline{\beta}$ such that $\underline{\beta}' \underline{X}$ has maximum variance of all the normalized linear combinations which are uncorrelated with $U_1, U_2, \cdots, U_r$, that is

$$0 = \mathrm{Cov}(\underline{\beta}' \underline{X}, U_i) = E(\underline{\beta}' \underline{X}\, U_i) = E(\underline{\beta}' \underline{X}\, \underline{X}' \underline{\beta}^{(i)}) = \underline{\beta}' \Sigma \underline{\beta}^{(i)}, \quad i = 1, 2, \cdots, r$$

$$= \lambda_i \underline{\beta}' \underline{\beta}^{(i)}. \tag{8.6}$$

We want to maximize

$$\phi_{r+1} = \underline{\beta}' \Sigma \underline{\beta} - \lambda (\underline{\beta}' \underline{\beta} - 1) - 2 \sum_{i=1}^{r} \upsilon_i \underline{\beta}' \Sigma \underline{\beta}^{(i)},$$

where λ and υ_i are Lagrange multipliers, the vector of partial derivatives is

$$\frac{\partial \phi_{r+1}}{\partial \underline{\beta}} = \underline{0} = 2\Sigma \underline{\beta} - 2\lambda \underline{\beta} - 2 \sum_{i=1}^{r} \upsilon_i \Sigma \underline{\beta}^{(i)}. \tag{8.7}$$

Premultiplying by $\underline{\beta}^{(j)'}$, we obtain

$$\underline{\beta}^{(j)'} \Sigma \underline{\beta} - \lambda \underline{\beta}^{(j)'} \underline{\beta} - \upsilon_j \underline{\beta}^{(j)'} \Sigma \underline{\beta}^{(j)} = 0, \text{ using equation (8.5)}$$

$$\Rightarrow -\upsilon_j \underline{\beta}^{(j)'} \Sigma \underline{\beta}^{(j)} = 0$$

$\Rightarrow \upsilon_j \lambda_j = 0$, because $\underline{\beta}' \Sigma \underline{\beta} = \lambda$

$\Rightarrow \upsilon_j = 0$, because $\lambda_j \neq 0$.

Thus equation (8.7) becomes

$$\Sigma \underline{\beta} - \lambda \underline{\beta} = \underline{0}$$

$$\Rightarrow (\Sigma - \lambda I) \underline{\beta} = \underline{0}$$

and therefore, $\underline{\beta}$ satisfies (8.2) and λ must satisfies (8.3).

Let λ_{r+1} be the maximum of $\lambda_1, \lambda_2, \cdots, \lambda_p$, such that there exists a vector $\underline{\beta}$ satisfying $(\Sigma - \lambda_{r+1} I) \underline{\beta} = \underline{0}$, $\underline{\beta}' \underline{\beta} = 1$ and equation (8.6), call this vector $\underline{\beta}^{(r+1)}$ and corresponding linear combination $U_{r+1} = \underline{\beta}^{(r+1)'} \underline{X}$.

We shall write,

$$\mathrm{B}_{p \times p} = (\underline{\beta}^{(1)} \quad \underline{\beta}^{(2)} \quad \cdots \quad \underline{\beta}^{(p)})$$

and

$$\Lambda = diag(\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_p) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

$$\underline{U} = \mathrm{B}' \underline{X}.$$

Since $\Sigma \underline{\beta}^{(r)} = \lambda_r \underline{\beta}^{(r)}$ for $r = 1, 2, \cdots, p$ can be written in matrix form as

$$\Sigma \mathrm{B} = \mathrm{B} \Lambda \tag{8.8}$$

and the equations $\underline{\beta}^{(r)'} \underline{\beta}^{(r)} = 1$ and $\underline{\beta}^{(r)'} \underline{\beta}^{(s)} = 0$, $r \neq s$ can be written as

$$\mathrm{B}' \mathrm{B} = \mathrm{I} \tag{8.9}$$

i.e. B is an orthogonal matrix

From equations (8.8) and (8.9) we obtain

$$\mathrm{B}' \Sigma \mathrm{B} = \mathrm{B}' \mathrm{B} \Lambda = \Lambda.$$

Thus given a positive definite matrix Σ, there exist an orthogonal matrix B such that $B'\Sigma B$ is in diagonal form and the diagonal elements of $B'\Sigma B$ are the characteristic roots of Σ.

Therefore, we proceed as follows:

Solve $|\Sigma - \lambda I| = 0$. Let the roots be $\lambda_1 > \lambda_2 > \cdots > \lambda_p$, then solve

$(\Sigma - \lambda_1 I)\underline{\beta} = \underline{0}$, get a solution $\underline{\beta}^{(1)}$ i.e. $\Sigma\underline{\beta}^{(1)} = \lambda_1 \underline{\beta}^{(1)}$.

First principal component

$$U_1 = \underline{\beta}^{(1)'} \underline{X}.$$

Again solve

$(\Sigma - \lambda_2 I)\underline{\beta} = \underline{0}$, get a solution $\underline{\beta}^{(2)}$ i.e. $\Sigma\underline{\beta}^{(2)} = \lambda_2 \underline{\beta}^{(2)}$.

Second principal component

$$U_2 = \underline{\beta}^{(2)'} \underline{X}$$

$$\vdots$$

r – th principal component

$$U_r = \underline{\beta}^{(r)'} \underline{X}$$

Stop at m $(\le p)$ for which λ_m is negligible as compared to $\lambda_1, \cdots, \lambda_{m-1}$. Thus we get a fewer (countable less) linear combinations or at most as many linear combinations as the original variable (p).

Contribution of first principal component $= \dfrac{\lambda_1}{\sum_{i=1}^{p} \lambda_i}$

Contribution of $(I + II)$ principal component $= \dfrac{\lambda_1 + \lambda_2}{\sum_{i=1}^{p} \lambda_i}$

Contribution of $(I + II + III)$ principal component $= \dfrac{\lambda_1 + \lambda_2 + \lambda_3}{\sum_{i=1}^{p} \lambda_i}$.

Theorem: An orthogonal transformation $\underline{V} = C\underline{X}$ of a random vector $\underline{X}$ leaves invariant the generalized variance and the sum of the variances of the components.

Proof: Let $E\,\underline{X} = \underline{0}$, and $E\,\underline{X}\,\underline{X}' = \Sigma$, then $E(\underline{V}) = \underline{0}$ and $\Sigma_V = E(\underline{V}\,\underline{V}') = E(C\underline{X}\,\underline{X}'C') = C\Sigma C'$.

The generalized variance of $\underline{V}$ is

$|C\Sigma C'| = |C||\Sigma||C'| = |\Sigma||CC'| = |\Sigma|$, because C is orthogonal matrix, $\Rightarrow CC' = I$

$=$ generalized variance of $\underline{X}$.

The sum of the variances of the components of $\underline{V}$ is

$$\sum_{i=1}^{p} E(V_i)^2 = tr(C\Sigma C') = tr\,\Sigma C'C = tr\,\Sigma I = tr\,\Sigma = \sum_{i=1}^{p} E(X_i)^2.$$

Corollary: The generalized variance of the vector of principal components is the generalized variance of the original vector.

Proof: We have,

$\underline{U} = B'\underline{X}$, and $\Sigma_U = B'\Sigma B$, then

$|\Sigma_U| = |B'\Sigma B| = |\Sigma||B'B| = |\Sigma||I| = |\Sigma|$.

Corollary: The sum of the variances of the principal components is equal to the sum of the variances of the original variates.

Proof: We have, $\underline{U} = B'\underline{X}$, and $\Sigma_U = B'\Sigma B$, then

$tr\,\Sigma_U = tr\,B'\Sigma B = tr\,\Sigma B'B = tr\,\Sigma I = tr\,\Sigma$.

But

$tr\,\Sigma_U = V(u_1) + \cdots V(u_p)$, and $tr\,\Sigma_X = V(x_1) + \cdots V(x_p)$.

Exercise: Let $R = \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix}$. Solve

$|R - \lambda I| = 0$, or $\begin{vmatrix} 1-\lambda & r_{12} \\ r_{12} & 1-\lambda \end{vmatrix}$, or $(1-\lambda)^2 - r_{12}^2 = 0$

or $\lambda^2 - 2\lambda + 1 - r^2 = 0$, $r_{12} = r$

or $\lambda = \dfrac{2 \pm \sqrt{4 - 4(1 - r^2)}}{2} = 1 \pm r$

If $r > 0$ $\quad \lambda_1 = 1 + r$, $\lambda_2 = 1 - r$

If $r < 0$ $\quad \lambda_1 = 1 - r$, $\lambda_2 = 1 + r$

If $r = 0$ $\quad \lambda_1 = 1 = \lambda_2$,

i.e. in case of perfect correlation, we need one principal component which explains fully but in case of zero correlation, no principal component.

Exercise: Find the variance of the first principle component of the covariance matrix Σ defined by

$\Sigma = (1-\rho)I + \rho\,\underline{e}\,\underline{e}'$, where $\underline{e}' = (1 \;\; 1 \;\; \cdots \;\; 1)$.

Exercise: Find the characteristic vector of $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ corresponding to the characteristic roots $1 + \rho$ and $1 - \rho$.