BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Multivariate Analysis

Course Code: 23ST12CC

Unit-I

Multivariate Normal Distribution and Properties

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. I. Angel Agnes Mary
Guest Faculty
Department of Statistics

UNIT – I

MULTIVARIATE NORMAL DISTRIBUTION AND PROPERTIES

Introduction to Multivariate Analysis

- Multivariate analysis is a set of techniques used for analysis of data that contain more than one variable.
- Multivariate analysis is a branch of statistics concerned with the analysis of multiple measurements, made on one or several samples of individuals. For example, we may wish to measure length, width, and weight of a product.
- Multivariate analysis provides a more accurate view of the behavior between variables that are highly correlated, and can detect potential problems in a product or process.
- Many decisions are based on univariate analysis, but only multivariate analysis reveals relationships that help you detect problems that are not obvious by looking at the variables individually.
- Multivariate analysis is a statistical technique used to examine relationships among multiple variables (three or more) simultaneously. It helps to identify patterns, correlations and dependencies between variables, and to understand how they interact and influence each other.

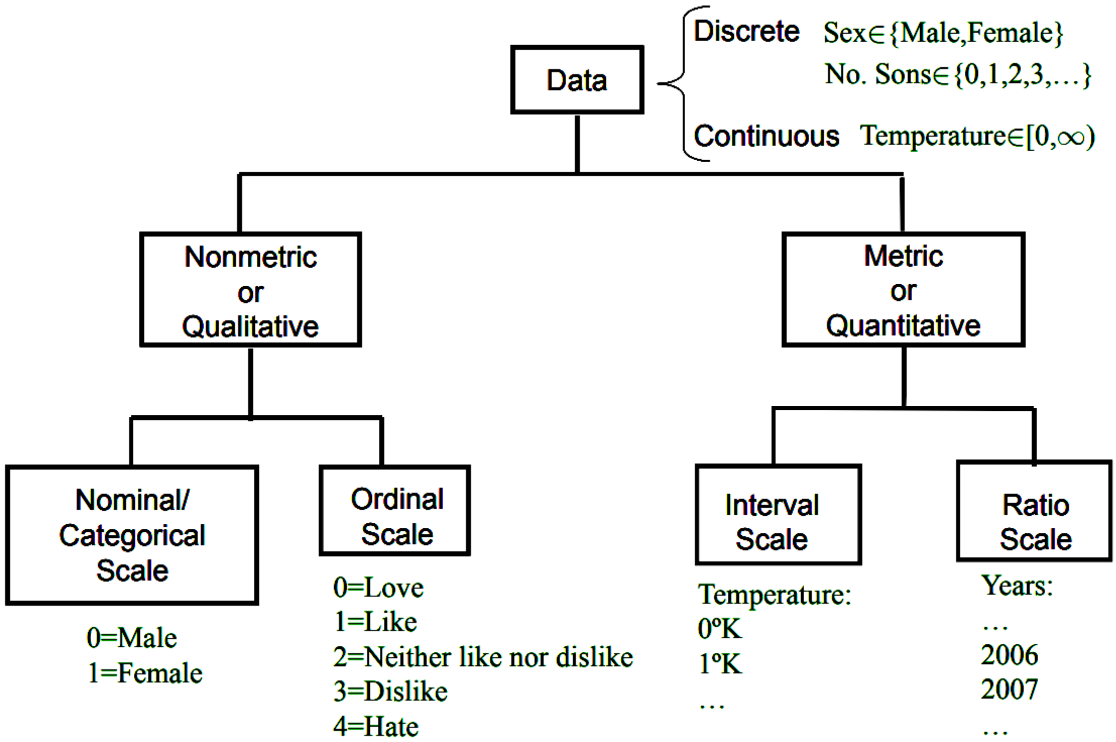
Types of Data

Data

Data can be defined as a systematic record of a particular quantity. It is the different values of that quantity represented together in a set. It is a collection of facts and figures to be used for a specific purpose such as a survey or analysis. When arranged in an organized form, can be called information. The source of data (primary data, secondary data) is also an important factor. Data can be classified into two types. these are

- **Qualitative Data:** Qualitative Data represent some characteristics or attributes. They depict descriptions that may be observed but cannot be computed or calculated. For example, data on attributes such as intelligence, honesty, wisdom and cleanliness. They are more exploratory than conclusive in nature.

- **Quantitative Data:** These can be measured and not simply observed. They can be numerically represented and calculations can be performed on them. For example, data on the number of students playing different sports from your class gives an estimate of how many of the total students play which sport. This information is numerical and can be classified as quantitative.



Types of measurements

Nominal Data

Nominal Data is used to label variables without any order or quantitative value. for example, the color of hair can be considered nominal data, as one color can't be compared with another color.

Examples of Nominal Data

- Colour of hair (Blonde, red, Brown, Black, etc.,)
- Marital status (Single, Widowed, Married)
- Nationality (Indian, German, American, etc.,)
- Gender (Male, Female, Others)
- Eye Color (Black, Brown, etc.,)

Ordinal Data

Ordinal data have natural ordering where a number is present in some kind of order by their position on the scale. These data are used for observation like customer satisfaction, happiness, etc.,

Examples of Ordinal Data

- When companies ask for feedback, experience, or satisfaction on a scale of 1 to 10
- Letter grades in the exam (A, B, C, D, etc.)
- Ranking of people in a competition (First, Second, Third, etc.)
- Economic Status (High, Medium, and Low)
- Education Level (Higher, Secondary, Primary)

Interval data

Interval data refers to information measured along a scale with equal distances. The distances or spaces in between the adjacent values are called intervals. So, the interval scale represents information about the order and it gives meaning to the difference between two values.

Examples of Interval Data

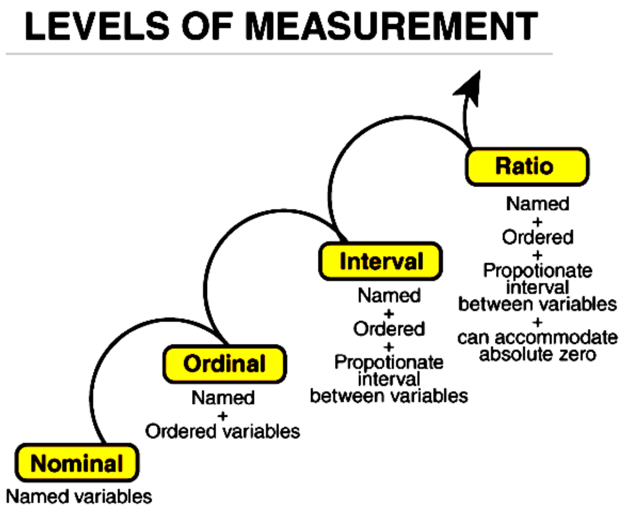
- Celsius and Fahrenheit are examples of interval scales. Each value on these scales differs from the adjacent values by intervals of exactly 1 degree.
- The difference between 20 and 21 degrees is identical to the difference between 225 and 226 degrees.

Ratio data

Ratio data is quantitative data that has an equal and definitive ratio between each value. Unlike interval data, ratio data has an absolute zero. It means ratio variables can’t have negative values, and zero means none of that variable is present.

Examples of Ratio Data

- The measurement of height is considered ratio data, and it’s not applicable to have a negative number for height.
- Age is a ratio variable, and a 40-year-old person is twice the age of someone who’s 20.



Multiple measurement or observation as row or column vector

A multiple measurement or observation may be expressed as

$$x = [4 \quad 2 \quad 0.6]$$

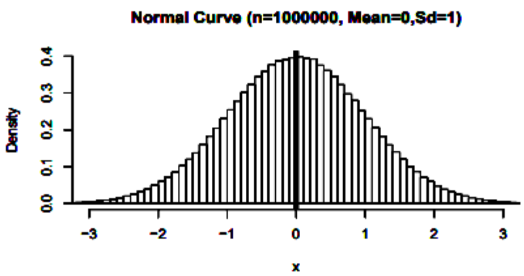
referring to the physical properties of length, width, and weight, respectively.

The collection of measurements on x is called a vector. In this case it is a row vector. We could have written x as a column vector.

$$x = \begin{bmatrix} 4 \\ 2 \\ 0.6 \end{bmatrix}$$

Multivariate Distributions

- A multivariate distribution describes the underlying random structure of a vector of random variables.
- From it we can derive marginal properties of the individual variables.
- It also describes relationships between variables or groups of variables.
- As in much of statistics, we are generally interested in making inferences about this distribution based on a sample



Normal curve with $n = 1000000, \mu = 0, \sigma = 1$

Structure of Multivariate Data

- Suppose that we have measurements on *p* variables for each of *n* experimental units.
- We will use x_{ij} to denote the observed value of the j^{th} variable $(j = 1,\ldots, p)$ on the i^{th} unit $(i = 1,\ldots,n)$.
- We will typically gather the information into a $n \times p$ matrix.

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n1} & \cdots & x_{nj} & \cdots & x_{p} \end{pmatrix}$$

Singular and non-singular

If A is square matrix and of full rank, then A is said to be nonsingular, and A has a unique inverse denoted by A^{-1} with the property that $AA^{-1} = A^{-1}A = I$. If A is square and of less than full rank, then an inverse does not exist and A is said to be singular.

Example (Singular)

$$\text{Let } A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} = (1\times2) - (1\times2) = 2 - 2 = 0$$

Now, Matrix A said to be a singular, because its determinant is equal to zero.

Example (Non-Singular)

$$\text{Let } A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} = (3\times2) - (1\times2) = 6 - 2 = 4$$

Now, Matrix A said to be a non-singular, because its determinant is 4 (Which is not equal to zero).

Types of Multivariate Techniques

There are many different techniques for multivariate analysis and they can be divided into two categories:

- Dependence techniques
- Interdependence techniques

Dependence methods

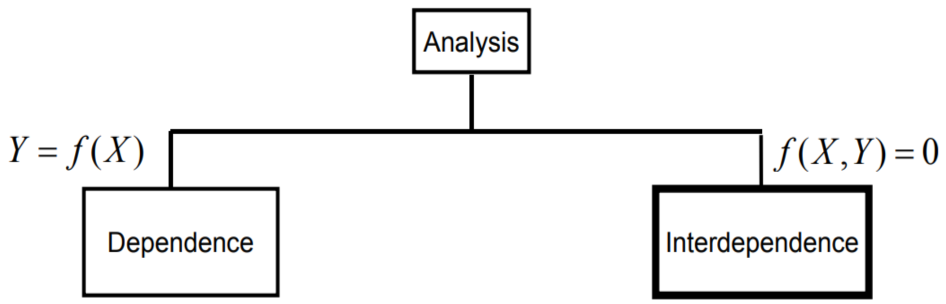
Dependence methods are used when one or some of the variables are dependent on others. Dependence looks at cause and effect; **in other words, can the values of two or more independent variables be used to explain, describe, or predict the value of another, dependent variable?** To give a simple example, the dependent variable of "weight" might be predicted by independent variables such as "height" and "age."

Interdependence methods

Interdependence methods are used to understand the structural makeup and underlying patterns within a dataset. In this case, no variables are dependent on others, so you are not looking for causal relationships. Rather, interdependence methods seek to give meaning to a set of variables or to group them together in meaningful ways.

The classifications of Multivariate Techniques are,

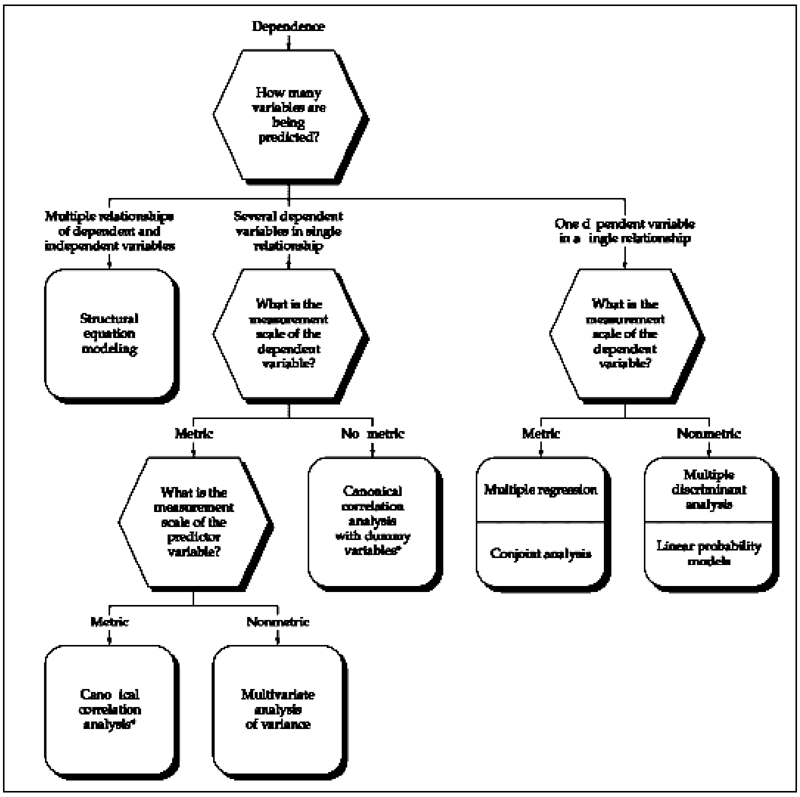
- Principal Components and Common Factor Analysis
- Cluster Analysis
- Multidimensional Scaling (perceptual mapping)
- Correspondence Analysis
- Canonical Correlation
- Multiple Discriminant Analysis
- Logit/Logistic Regression
- Multivariate Analysis of Variance (MANOVA) and Covariance
- Conjoint Analysis
- Canonical Correlation
- Multiple Regression
- Structural Equations Modeling (SEM)

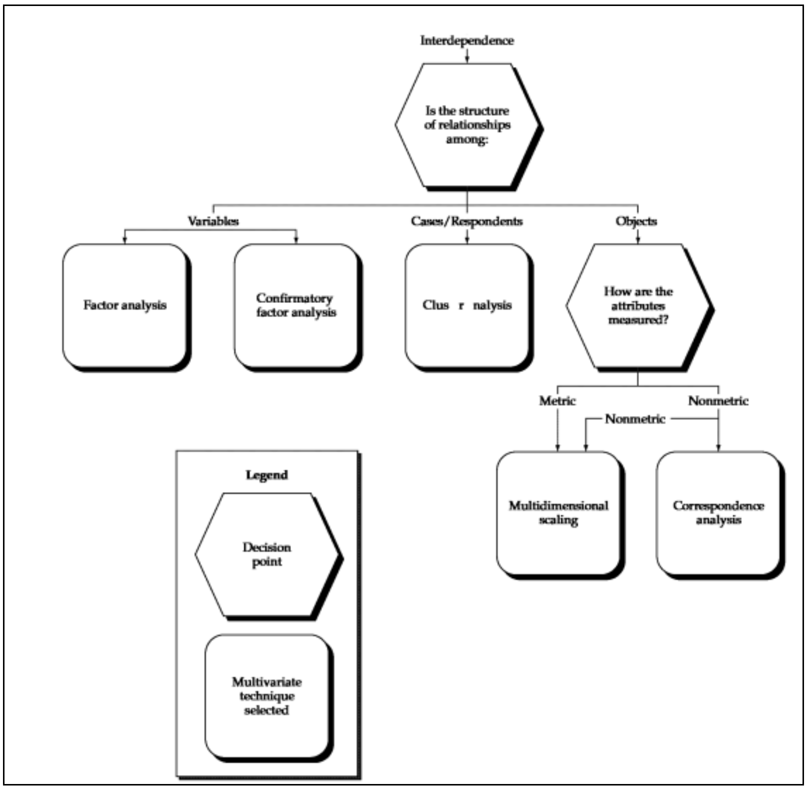


A variable or set of variables is identified as the dependent variable to be predicted or explained by other variables known as independent variables.

Example of Dependence: (No. Sons, House Type) = f(Income, Social Status, Studies)

Example of Interdependence: Who is similar to whom? (No. Sons, House Type, Income, Social Status, Studies, …)





Multivariate Techniques help to:

- Reduce dimensionality
- Identify patterns and relationships
- Predict outcomes
- Classify observations
- Identify correlations and dependencies

Applications of Multivariate Analysis

Multivariate analysis is used in various fields including:

- ***Social sciences:*** to study relationships between demographic, economic and social variables.
- ***Marketing:*** to analyze customer behavior, preferences and demographics.
- ***Healthcare:*** to investigate relationships between symptoms, treatments and outcomes.
- ***Finance:*** to analyze portfolio performance, risk management and asset pricing.
- ***Biology:*** to study genetic associations, protein interactions and ecological relationships.

Multivariate normal distributions

The multivariate normal distribution (also known as the multivariate Gaussian distribution) is a generalization of the univariate normal distribution to multiple variables. A multivariate normal distribution is a vector in multiple normally distributed variables, such that any linear combination of the variables is also normally distributed.

It is mostly useful in extending the central limit theorem to multiple variables, but also has applications to Bayesian inference and thus machine learning, where the multivariate normal distribution is used to approximate the features of some characteristics.

Applications

Multivariate normal distributions are widely used in various fields, including:

- Statistics: to model correlated data and perform inference
- Machine learning: as a prior distribution for Bayesian models
- Finance: to model asset returns and portfolio risk
- Engineering: to model complex systems with correlated variables

- Data analysis and visualization
- Regression analysis and prediction
- Cluster analysis and classification
- Dimensionality reduction and feature extraction

Multivariate normal density and its Properties

The multivariate normal density is a generalization of the univariate normal density to p ≥ 2 dimensions. The univariate normal distribution, with mean μ and variance σ^2 has the probability density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad -\infty < x < \infty$$

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

This can be generalized for $p\times 1$ vector x of observations on several variables as

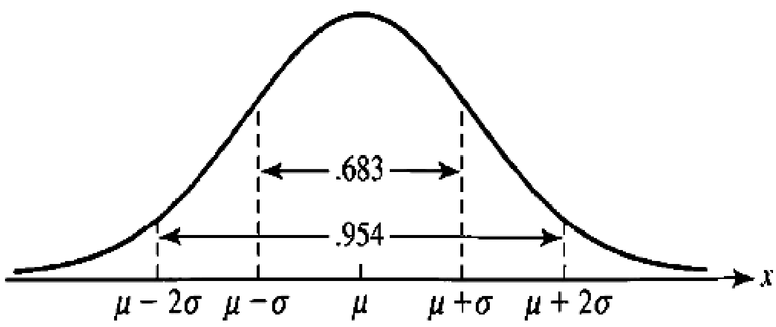
$$(x-\mu)'\Sigma^{-1}(x-\mu)$$

The $p\times 1$ vector μ represents the expected value of the random vector X and the $p\times q$ matrix Σ is the variance-covariance matrix of X.

A p-dimensional vector of random variables, $X = X_1, X_2, \dots, X_p$, $-\infty < X_i < \infty$, $i = 1, \dots, p$ is said to have a multivariate normal distribution if its density function $f(X)$ is of the form

$$f(x) = f(x_1, x_2, \cdots, x_p) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

where $\mu = (\mu_1, \dots, \mu_p)$ is the vector of means and Σ is the variance-covariance matrix of the multivariate normal distribution. The shortcut notation for this density is $X = N_p(\mu, \Sigma)$.



A plot of this function yields the familiar bell-shaped curve. Also shown in the figure are appropriate areas under curve with in ±1 standard deviations and ±2 standard deviations of the mean.

The areas represent the probabilities, and thus, for the normal random variable X.

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$$

The multivariate normal distribution has several important properties:

- **Marginal distributions:** Each individual variable follows a univariate normal distribution.
- **Linear combinations:** Any linear combination of the variables also follows a univariate normal distribution.
- **Independence:** If the covariance matrix is diagonal, the variables are independent.
- **Correlation:** The correlation matrix can be derived from the covariance matrix.

Additional Properties of Multivariate normal distributions

The following are true for a random vector X having a multivariate normal distribution:

- Linear combinations of the components of X are normally distributed.
- All Subsets of the components of X have a Multivariate Normal distribution.
- Zero covariance implies that the corresponding components are independently distributed.
- The conditional distribution of the components are Multivariate Normal.

Result – 1

If X is distributed as $N_p(\mu, \Sigma)$, then any linear combination of variables $a'X = a_1X_1+a_2X_2+\cdots+a_pX_p$ is distributed as $N(a'\mu, a'\Sigma a)$. Also if $a'X$ is distributed as $N(a'\mu, a'\Sigma a)$ for every a, then X must be $N_p(\mu, \Sigma)$.

<u>Example-1</u>: The distribution of a linear combination of the component of a normal random vector.

Consider the linear combination $a'\,X$ of a multivariate normal random vector determined by the choice $a'=[1, 0, \ldots, 0]$. Since

$$a'X = [1, 0, \cdots, 0]\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = X_1 \text{ and } a'\mu = [1, 0, \cdots, 0]\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mu_1$$

$$a'\Sigma a = [1, 0, \cdots, 0]\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} = \sigma_{11}$$

Note that X_1 is distributed as $N(\mu_1, \sigma_{11})$. Generally, the marginal distribution of any component X_i (i = 1, 2, . . ., p) of X_p is $N(\mu_p, \sigma_{pp})$.

<u>Example-2</u>: Considers several linear combinations of a multivariate normal vector *X*.

If *X* is distributed as $N_p(\mu, \Sigma)$, the *q* linear combinations

$$A_{(p\times q)}X_{(p\times 1)} = \begin{bmatrix} a_{11}X_1 + \cdots\cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots\cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots\cdots + a_{pq}X_p \end{bmatrix}$$

are distributed as $N_q(A\mu, A\Sigma A')$. Also $X_{p\times 1} + d_{p\times 1}$, where d is a vector of constants, is distributed as $N_p(\mu + d, \Sigma)$.

<u>Example-3</u>: The distribution of two linear combinations of the components of a normal random vector.

For *X* distributed as $N_3(\mu, \Sigma)$, find the distribution of

$$\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = AX$$

The distribution of *AX* is multivariate normal with mean

$$A\mu = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{bmatrix}$$

And covariance matrix

$$A\Sigma A' = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$A\Sigma A' = \begin{bmatrix} \sigma_{11} - \sigma_{12} & \sigma_{12} - \sigma_{22} & \sigma_{13} - \sigma_{23} \\ \sigma_{12} - \sigma_{13} & \sigma_{22} - \sigma_{23} & \sigma_{23} - \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} \\ \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} & \sigma_{22} - 2\sigma_{23} + \sigma_{33} \end{bmatrix}$$

Alternatively, the mean vector $A\mu$ and covariance matrix $A\sum A'$ may be verified by direct calculation of the means and covariances of the two random variables $Y_1 = X_1 - X_2$ and $Y_2 = X_2 - X_3$.

Result – 2: The distribution of a subset of a normal random vector

All subsets of X are normally distributed. If we respectively partition X, its mean vector μ, and its covariance matrix Σ as

$$X_{(p\times1)} = \begin{bmatrix} X_1 \\ (q\times1) \\ \cdots\cdots \\ X_2 \\ (p-q)\times1 \end{bmatrix}, \quad \mu_{(p\times1)} = \begin{bmatrix} \mu_1 \\ (q\times1) \\ \cdots\cdots \\ \mu_2 \\ (p-q)\times1 \end{bmatrix} \text{ and } X_{(p\times1)} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ (q\times1) & (q\times(p-q)) \\ \cdots\cdots & \cdots\cdots \\ \Sigma_{21} & \Sigma_{22} \\ ((p-q)\times q) & ((p-q)\times(p-q)) \end{bmatrix}$$

Then X_1 is distributed as $N_q(\mu_1, \Sigma_{11})$.

<u>Example-3</u>**: The distribution of a subset of a normal random vector**

If X is distributed as $N_5(\mu, \sum)$, find the distribution of $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$. We set $X_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$, $\mu_1 = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \Sigma_{22} & \Sigma_{24} \\ \Sigma_{24} & \Sigma_{44} \end{bmatrix}$ and note that with this assignment, X, μ and $\sum$ can respectively be rearranged and partitioned as

$$X = \begin{bmatrix} X_2 \\ X_4 \\ \cdots \\ X_1 \\ X_3 \\ X_5 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_2 \\ \mu_4 \\ \cdots \\ \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix}, \quad \Sigma = \left[\begin{array}{cc:ccc} \sigma_{22} & \sigma_{24} & \sigma_{12} & \sigma_{23} & \sigma_{25} \\ \sigma_{24} & \sigma_{44} & \sigma_{14} & \sigma_{34} & \sigma_{45} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{12} & \sigma_{14} & \sigma_{11} & \sigma_{13} & \sigma_{15} \\ \sigma_{23} & \sigma_{34} & \sigma_{13} & \sigma_{33} & \sigma_{35} \\ \sigma_{25} & \sigma_{45} & \sigma_{15} & \sigma_{35} & \sigma_{55} \end{array}\right]$$

or

$$X = \begin{bmatrix} X_1 \\ (2\times1) \\ \cdots\cdots \\ X_2 \\ (3\times1) \end{bmatrix}, \; \mu = \begin{bmatrix} \mu_1 \\ (2\times1) \\ \cdots\cdots \\ \mu_2 \\ (3\times1) \end{bmatrix} \text{ and } X_{(p\times1)} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ (2\times2) & (2\times3) \\ \cdots\cdots & \cdots\cdots \\ \Sigma_{21} & \Sigma_{22} \\ (3\times2) & (3\times3) \end{bmatrix}$$

Thus, $X_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$, Then X_1 is distributed as $N_2(\mu_1, \Sigma_{11}) = N_2\left(\begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \begin{bmatrix} \Sigma_{22} & \Sigma_{24} \\ \Sigma_{24} & \Sigma_{44} \end{bmatrix}\right)$.
Therefore, the normal distribution for any subset can be expressed by simply selecting the appropriate means and covariances from the original μ and $\sum$.

Theorem

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

Proof

The probability density function of p-variates normal random vector is

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

Given

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

$$\left[(x_1-\mu_1), \cdots, (x_p-\mu_p)\right]_{(1\times p)} \begin{bmatrix} 1/\sigma_1^2 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & 1/\sigma_p^2 \end{bmatrix}_{(p\times p)} \begin{bmatrix} (x_1-\mu_1) \\ \vdots \\ (x_p-\mu_p) \end{bmatrix}_{(p\times1)} = \sum_{i=1}^{p}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$

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,

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

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

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

Theorem

If X (with p components) be distributed according to $N(\mu, \Sigma)$. Then $Y = CX$ (nonsingular transformation) is distributed according to $N(C\mu, C\Sigma C^T)$ for C nonsingular.

Proof

We have

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

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

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

The quadratic form in the exponent of g(y) is

$$(C^{-1}y-\mu)^T \Sigma^{-1} (C^{-1}y-\mu) = (C^{-1}y - C^{-1}C\mu)^T \Sigma^{-1} (C^{-1}y - C^{-1}C\mu)$$

$$= \left[C^{-1}(y-C\mu)\right]^T \Sigma^{-1} \left[C^{-1}(y-C\mu)\right]$$

$$= (y-C\mu)^T\, C^{-1^T}\, \Sigma^{-1}\, (y-C\mu) C^{-1}, \text{ Since } (AB)^T = B^T A^T$$

$$= (y-C\mu)^T\, C^{-1^T}\, \Sigma^{-1}\, C^{-1} (y-C\mu), \text{ Since } (C^{-1})^T = (C^T)^{-1}$$

$$= (y-C\mu)^T\, (C \Sigma C^T)^{-1} (y-C\mu), \text{ Since } (AB)^{-1} = B^{-1} A^{-1}.$$

And the Jacobian of the transformation, which is

$$\left|C^{-1}\right| = \frac{1}{|C|} = \sqrt{\frac{1}{|C|^2}} = \sqrt{\frac{1}{|C|\left|C^T\right|}} = \sqrt{\frac{|\Sigma|}{|C||\Sigma|\left|C^T\right|}} = \frac{|\Sigma|^{1/2}}{\left|C\Sigma C^T\right|^{1/2}}, \text{ Since } |AB| = |B|\,|A|$$

Thus, the density function of Y is

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

Therefore, $Y \sim N(C\mu,\ C\Sigma C^T)$

Transformation of variables

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

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

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

$$g(y_1, y_2, \ldots, y_p) = f[x_1(y_1, y_2, \ldots, y_p), \ldots, x_p(y_1, y_2, \ldots, y_p)]\ |J|$$

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

Joint and Marginal cumulative distribution function

Let $X = (X_1, \ldots X_p)$ be an p-dimensional vector of random variables. We have the following definitions and statements.

Joint CDF

Let X and Y be two random variables. The joint cumulative distribution function 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 *p* random variables $X_1, \ldots X_p$. The cumulative distribution function is

$$F_X(X) = F_X(x_1, \cdots, x_p) = P(X_1 \le x_1, \cdots\cdots, X_p \le x_p)$$

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

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

$$F(x_1, x_2, \cdots, x_p) = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(u_1, u_2, \cdots, u_p) \, du_1 \, du_2 \cdots du_p.$$

Marginal CDF

Let *F(x, y)* be the cumulative distribution function of two random variables *X* and *Y*, the marginal cumulative distribution function of *X* is

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

Therefore,

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

Let us consider,

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

Where *f(u)* is called the marginal density function of *X* is

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

Similarly we define G(y), the marginal cumulative distribution function of Y, and g(y), the marginal density function of Y.

Let $F(x_1, x_2, ..., x_p)$ be the cumulative distribution function of the random variables $X_1, . . . X_p$. The marginal cumulative distribution function of some of $X_1, . . . X_p$, say $X_1, ..., X_r$, (r < p) is

$$P(X_1 \le x_1, \cdots\cdots, X_r \le x_r) = P(X_1 \le x_1, \cdots\cdots, X_r \le x_r, X_{r+1} \le \infty, \cdots, X_p \le \infty)$$

$$= F(x_1, x_2, \cdots, x_r, \infty, \cdots, \infty)$$

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

and the marginal density function of $X_1, . . . X_r$ 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 $X_1, . . . X_p$ are obtained in the obvious similar fashion.

Marginal and Conditional distribution of Multivariate normal distribution

Let p dimensional random vector $x = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ has a normal distribution $N(x, \mu, \Sigma)$ with $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ where x_1 *and* x_2 are two subvectors of respective dimensions p and q with $p + q = n$.

Note: that $\Sigma = \Sigma^T$ and $\Sigma_{21} = \Sigma_{21}^T$.

Statement

- The marginal distributions of x_1 *and* x_2 are also normal with mean vector μ_i and covariance matrix Σ_{ii} $(i = 1,2)$ respectively.

- The conditional distribution of x_i given x_j is also normal with mean vector

$$\mu_{i|j} = \mu_i + \Sigma_{ij} \Sigma_{jj}^T (x_j - \mu_j)$$

and covariance matrix

$$\Sigma_{i|j} = \Sigma_{ij} - \Sigma_{ij}^T \Sigma_{ii}^{-1} \Sigma_{ij}$$

Proof

The joint density of is

$$f(x)=f(x_1,x_2)=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

$$=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}exp\left[-\frac{1}{2}Q(x_1,x_2)\right]$$

where Q is defined as

$$Q(x_1,x_2)=(x-\mu)^T\Sigma^{-1}(x-\mu)$$

$$=\left[(x_1-\mu_1)^T,(x_2-\mu_2)^T\right]\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\begin{bmatrix}x_1-\mu_1\\ x_2-\mu_2\end{bmatrix}$$

$$=(x_1-\mu_1)^T\Sigma^{11}(x_1-\mu_1)+2(x_1-\mu_1)^T\Sigma^{12}(x_2-\mu_2)+(x_2-\mu_2)^T\Sigma^{22}(x_2-\mu_2)$$

Here we have assumed

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

we have

$$\Sigma^{11}=\left(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T\right)^{-1}=\Sigma_{11}^{-1}+\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22}-A_{12}^T\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}$$

$$\Sigma^{22}=\left(\Sigma_{22}-\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}=\Sigma_{22}^{-1}+\Sigma_{22}^{-1}\Sigma_{12}\left(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T\right)^{-1}\Sigma_{12}^T\Sigma_{22}^{-1}$$

$$\Sigma^{12}=-\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22}-\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}=\left(\Sigma^{21}\right)^T$$

Substituting the second expression for Σ^{11}, first expression for Σ^{22} and Σ^{12} into $Q(x_1,x_2)$ to get

$$Q(x_1,x_2)=(x_1-\mu_1)^T\left[\Sigma_{11}^{-1}+\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22}-A_{12}^T\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\right](x_1-\mu_1)$$

$$-2(x_1-\mu_1)^T\left[\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22}-\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}\right](x_2-\mu_2)$$

$$+(x_2-\mu_2)^T\left(\Sigma_{22}-\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}(x_2-\mu_2)$$

$$=(x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1)+(x_1-\mu_1)^T\left[\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22}-A_{12}^T\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\right](x_1-\mu_1)$$

$$-2(x_1-\mu_1)^T\left[\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22}-\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}\right](x_2-\mu_2)$$

$$+(x_2-\mu_2)^T\left(\Sigma_{22}-\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}(x_2-\mu_2)$$

$$=(x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1)+\left[(x_2-\mu_2)-\Sigma_{12}^T\Sigma_{11}^{-1}(x_1-\mu_1)\right]^T\left(\Sigma_{22}-\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1}$$

$$\left[(x_2-\mu_2)-\Sigma_{12}^T\Sigma_{11}^{-1}(x_1-\mu_1)\right]$$

The last equal sign is due to the following equations for any vectors u and v and a symmetric matrix $A = A^T$

$$u^T Au - 2u^T Av + v^T Av = u^T Au - u^T Av - u^T Av + v^T Av$$

$$\Rightarrow u^T A(u - v) - (u - v)Av = u^T Au - u^T Av - u^T Av + v^T Av$$

$$\Rightarrow (u - v)^T A(u - v) = (u - v)^T A(u - v)$$

We define

$$b \stackrel{\Delta}{=} \mu_2 + \Sigma_{12}^T \Sigma_{11}^{-1} (x_1 - \mu_1)$$

$$A \stackrel{\Delta}{=} \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}$$

and

$$Q_1(x_1) \stackrel{\Delta}{=} (x_1 - \mu_1)^T \Sigma_{11}^{-1} (x_1 - \mu_1)$$

$$Q_2(x_1, x_2) \stackrel{\Delta}{=} \left[(x_2 - \mu_2) - \Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1)\right]^T \left(\Sigma_{22} - \Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}^T\right)^{-1} \left[(x_2 - \mu_2) - \Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1)\right]$$

$$= (x_2 - b)^T A^{-1} (x_2 - b)$$

and get

$$Q(x_1, x_2) = Q_1(x_1) + Q_2(x_1, x_2)$$

Now the joint distribution can be written as

$$f(x) = f(x_1, x_2) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} exp\left[-\frac{1}{2} Q(x_1, x_2) \right]$$

$$= \frac{1}{(2\pi)^{n/2} |\Sigma_{11}|^{1/2} \left|\Sigma_{22} - \Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}^T\right|^{1/2}} exp\left[-\frac{1}{2} Q(x_1, x_2) \right]$$

$$= \frac{1}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2}} exp\left[-\frac{1}{2}(x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1) \right] \frac{1}{(2\pi)^{q/2} |A|^{1/2}} exp\left[-\frac{1}{2}(x_2 - b)^T A^{-1}(x_2 - b) \right]$$

$$= N(x_1, \mu_1, \Sigma_{11}) N(x_2, b, A)$$

The third equal sign is due to theorem

$$|\Sigma| = |\Sigma_{11}| \left|\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}\right|$$

The marginal distribution of x_1 is

$$f_1(x_1) = \int f(x_1, x_2) dx_2 = \frac{1}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2}} exp\left[-\frac{1}{2}(x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1) \right]$$

and the conditional distribution of x_2 given x_1 is

$$f_{2|1} = \frac{f(x_1,x_2)}{f(x_1)} = \frac{1}{(2\pi)^{q/2}|A|^{1/2}} exp\left[-\frac{1}{2}(x_2 - b)^T A^{-1}(x_2 - b) \right]$$

with

$$b = \mu_2 + \Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1)$$

$$A = \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}$$

Hence the proof.

Characteristic function

The characteristic function of a random vector X is defined as $\phi_X(t) = E\left[e^{it'X}\right]$, where t is a vector of reals, $i = \sqrt{-1}$

Theorem

Let $X = (X_1, X_2, \ldots, X_p)'$ be normally distributed random vector with mean μ and positive definite covariance matrix Σ, then the characteristic function of X is given by

$\phi_X(t) = e^{it'\mu - \frac{1}{2}t'\Sigma t}$, where $t = (t_1, t_2, \ldots, t_p)'$ is a real vector of order p ×1.

Proof

We have

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

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

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

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

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

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

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

Now the characteristic function of Y is

$$\phi_Y(u) = E\left[e^{iu'Y}\right] = Ee^{i(u_1Y_1+\cdots+u_pY_p)} = Ee^{iu_1Y_1}\cdots Ee^{iu_{p1}Y_p} \quad \textbf{- - - (1)}$$

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

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

$$= e^{-\frac{1}{2}\left(u_1^2+\cdots+u_p^2\right)} = e^{-\frac{1}{2}u'u}. \quad \textbf{- - - (2)}$$

Thus,

$$\phi_X(t) = E\left[e^{it'X}\right] = E\left[e^{it'(CY+\mu)}\right] = e^{it'\mu}\, E\left[e^{it'(CY)}\right]$$

$$= e^{it'\mu}\, Ee^{i(C't)'Y} = e^{it'\mu}\, e^{-\frac{1}{2}(C't)'(C't)}$$ from equation (1) and (2)

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

Hence the proof.

Distribution of linear combinations of multivariate normal vector

Theorem

If every linear combination of the components of a vector X is normally distributed, then X has normal distribution.

Proof

Consider a vector X of p-components with density function $f(x)$ and characteristic function $\phi_X(t) = E\left[e^{iu'X}\right]$ and suppose the mean of X is μ and the covariance matrix is Σ.

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

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

$$Ee^{i(u'X)} = e^{iu'\mu-\frac{1}{2}u'\Sigma u}$$

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

Moment generating function

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

Proof

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

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

Make the nonsingular transformation

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

Therefore, the density function of *Y* is

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

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

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

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

Now the moment generating function of *Y* is

$$M_Y(u) = Ee^{u'Y} = Ee^{(u_1Y_1 + \cdots + u_pY_p)} = Ee^{(u_1Y_1)} \cdots Ee^{(u_pY_p)}$$

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

Thus we can say

$$\phi_X(t) = E\left[e^{t'X}\right] = E\left[e^{t'(CY+\mu)}\right] = e^{t'\mu}\, E\left[e^{t'(CY)}\right]$$

$$= e^{t'\mu}\, E\left[e^{(C't)'Y}\right] = e^{t'\mu}\, e^{\frac{1}{2}(C't)'(C't)} = e^{t'\mu + \frac{1}{2}t'CC't} = e^{t'\mu + \frac{1}{2}t'\Sigma t}$$

Note

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

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

$$=1+t\,\mu_1'+\frac{t^2}{2!}\mu_2'+\cdots+\frac{t^r}{r!}\mu_r'+\cdots \qquad ----- (1)$$

Differentiating r times equation *(1)* with respect to t and then putting $t = 0$, we get

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

Raw Moments of multivariate normal distribution

First Moment

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

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

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

$$=M\left\{\mu_n+\frac{1}{2}\frac{\partial}{\partial t_n}\begin{pmatrix}t_1\left(t_1\sigma_{11}+t_2\sigma_{12}+\cdots+t_n\sigma_{1n}+\cdots+t_p\sigma_{1p}\right)\\+\cdots+t_n\left(t_1\sigma_{n1}+\cdots+t_n\sigma_{nn}+\cdots+t_p\sigma_{np}\right)\\+\cdots+t_p\left(t_1\sigma_{p1}+\cdots+t_n\sigma_{pn}+\cdots+t_p\sigma_{pp}\right)\end{pmatrix}\right\}|_{t=0}$$

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

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

Second Moment

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

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

Therefore,

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

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

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

Third Moment

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

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

$$= \left[\begin{array}{l} M\left\{\mu_r + \sum_{j=1}^{p} t_j \sigma_{rj}\right\}\left\{\mu_l + \sum_{j=1}^{p} t_j \sigma_{lj}\right\}\left\{\mu_n + \sum_{j=1}^{p} t_j \sigma_{nj}\right\} + M\left\{\mu_n + \sum_{j=1}^{p} t_j \sigma_{nj}\right\}\sigma_{lr} \\ + M\left\{\mu_l + \sum_{j=1}^{p} t_j \sigma_{lj}\right\}\sigma_{nr} + M\left\{\mu_r + \sum_{j=1}^{p} t_j \sigma_{rj}\right\}\sigma_{nl} \end{array}\right]_{t=0}$$

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

Fourth Moment

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

$$= \frac{\partial}{\partial t_m}\left[\begin{array}{l} M\left\{\mu_r + \sum_{j=1}^{p} t_j \sigma_{rj}\right\}\left\{\mu_l + \sum_{j=1}^{p} t_j \sigma_{lj}\right\}\left\{\mu_n + \sum_{j=1}^{p} t_j \sigma_{nj}\right\} + M\left\{\mu_n + \sum_{j=1}^{p} t_j \sigma_{nj}\right\}\sigma_{lr} \\ + M\left\{\mu_l + \sum_{j=1}^{p} t_j \sigma_{lj}\right\}\sigma_{nr} + M\left\{\mu_r + \sum_{j=1}^{p} t_j \sigma_{rj}\right\}\sigma_{nl} \end{array}\right]_{t=0}$$

$$
=\left[\begin{array}{l}
M\left\{\mu_m+\sum_{j=1}^{p}t_j\sigma_{mj}\right\}\left\{\mu_r+\sum_{j=1}^{p}t_j\sigma_{rj}\right\}\left\{\mu_l+\sum_{j=1}^{p}t_j\sigma_{lj}\right\}\left\{\mu_n+\sum_{j=1}^{p}t_j\sigma_{nj}\right\} \\
+M\left(\begin{array}{l}\sigma_{rm}\left\{\mu_l+\sum_{j=1}^{p}t_j\sigma_{lj}\right\}\left\{\mu_n+\sum_{j=1}^{p}t_j\sigma_{nj}\right\}+\sigma_{lm}\left\{\mu_r+\sum_{j=1}^{p}t_j\sigma_{rj}\right\}\left\{\mu_n+\sum_{j=1}^{p}t_j\sigma_{nj}\right\} \\ +\sigma_{nm}\left\{\mu_r+\sum_{j=1}^{p}t_j\sigma_{rj}\right\}\left\{\mu_l+\sum_{j=1}^{p}t_j\sigma_{lj}\right\}\end{array}\right) \\
+M\left\{\mu_m+\sum_{j=1}^{p}t_j\sigma_{mj}\right\}\left\{\mu_n+\sum_{j=1}^{p}t_j\sigma_{nj}\right\}\sigma_{lr}+M\,\sigma_{nm}\,\sigma_{lr} \\
+M\left\{\mu_m+\sum_{j=1}^{p}t_j\sigma_{mj}\right\}\left\{\mu_l+\sum_{j=1}^{p}t_j\sigma_{lj}\right\}\sigma_{nr}+M\,\sigma_{lm}\,\sigma_{nr} \\
+M\left\{\mu_m+\sum_{j=1}^{p}t_j\sigma_{mj}\right\}\left\{\mu_r+\sum_{j=1}^{p}t_j\sigma_{rj}\right\}\sigma_{nl}+M\,\sigma_{rm}\,\sigma_{nl}
\end{array}\right]_{t=0}
$$

$$
=\mu_m\,\mu_r\,\mu_l\,\mu_n+\mu_l\mu_n\,\sigma_{rm}+\mu_r\,\mu_n\,\sigma_{lm}+\mu_r\,\mu_l\,\sigma_{nm}+\mu_m\,\mu_n\,\sigma_{lr}+\sigma_{nm}\,\sigma_{lr}+\mu_m\,\mu_l\,\sigma_{nr}
+\sigma_{lm}\,\sigma_{nr}+\mu_m\,\mu_r\,\sigma_{nl}+\sigma_{rm}\,\sigma_{nl}.
$$

Determination of mean and variance

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

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

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

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

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

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

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

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

$$=\frac{\partial}{\partial t_m}\left\{\begin{array}{l} M\left[\sum_{j=1}^{p} t_j\,\sigma_{rj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{lj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{nj}\right]+M\left[\sum_{j=1}^{p} t_j\,\sigma_{nj}\right]\sigma_{lr} \\ +M\left[\sum_{j=1}^{p} t_j\,\sigma_{lj}\right]\sigma_{nr}+M\left[\sum_{j=1}^{p} t_j\,\sigma_{rj}\right]\sigma_{nl} \end{array}\right\}_{t=0}$$

$$=\left\{\begin{array}{l} M\left[\sum_{j=1}^{p} t_j\,\sigma_{mj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{rj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{lj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{nj}\right] \\ +M\left\{\sigma_{rm}\left[\sum_{j=1}^{p} t_j\,\sigma_{lj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{nj}\right]+\sigma_{lm}\left[\sum_{j=1}^{p} t_j\,\sigma_{rj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{nj}\right]+\sigma_{nm}\left[\sum_{j=1}^{p} t_j\,\sigma_{rj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{lj}\right]\right\} \\ +M\left[\sum_{j=1}^{p} t_j\,\sigma_{mj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{nj}\right]\sigma_{lr}+M\sigma_{nm}\,\sigma_{lr}+M\left[\sum_{j=1}^{p} t_j\,\sigma_{mj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{lj}\right]\sigma_{nr}+M\sigma_{lm}\,\sigma_{nr} \\ +M\left[\sum_{j=1}^{p} t_j\,\sigma_{mj}\right]\left[\sum_{j=1}^{p} t_j\,\sigma_{rj}\right]\sigma_{nl}+M\sigma_{rm}\,\sigma_{nl} \end{array}\right\}_{t=0}$$

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

Covariance matrix of multivariate normal distribution

Theorem

Let *x* follow a multivariate normal distribution

$$x \sim N(\mu, \Sigma) \qquad \text{- - - - (1)}$$

Then, the covariance matrix of *x* is

$$Cov(x) = \Sigma \qquad \text{- - - - (2)}$$

Proof

Consider a set of independent and standard normally distributed random variables

$$z_i \sim N(0, 1),\ i = 1, 2, \ldots, n. \qquad \text{- - - - (3)}$$

Then, these variables together form a multivariate normally distributed random vector

$$z \sim N(0_n, I_n) \quad \text{- - - - (4)}$$

Because the covariance is zero for independent random variables, we have

$$Cov(z_i, z_j) = 0, \text{ for all } i \neq j \quad \text{- - - - (5)}$$

Moreover, as the variance of all entries of the vector is one, we have

$$Var(z_i) = 1, \text{ for all } i = 1, 2, \ldots, n. \quad \text{- - - - (6)}$$

Taking (5) and (6) together, the covariance matrix of z is

$$Cov(z) = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} = I_n \quad \text{- - - - (7)}$$

Next, consider an $n \times n$ matrix A solving the equation $AA^T = \Sigma$. Such a matrix exists, because Σ is defined to be positive definite. Then, x can be represented as a linear transformation of z

$$x = Az + \mu \sim N(A0_n + \mu, AI_nA^T) = N(\mu, \Sigma). \quad \text{- - - - (8)}$$

Thus, the covariance of x can be written as

$$Cov(x) = Cov(Az + \mu). \quad \text{- - - - (9)}$$

With the invariance of the covariance matrix under addition

$$Cov(x + a) = Cov(x). \quad \text{- - - - (10)}$$

and the scaling of the covariance matrix upon multiplication

$$Cov(Ax) = A \, Cov(x) \, A^T, \quad \text{- - - - (11)}$$

this becomes

$$\begin{aligned} Cov(x) &= Cov(Az + \mu) \\ &= Cov(Az) = A \, Cov(z) \, A^T \\ &= AI_nA^T = AA^T = \Sigma. \end{aligned}$$

Hence Proved.

SCL PROBLEMS

1. The equivalence of zero covariance and independence for normal variables

Let X be distributed as $N_3(\mu, \Sigma)$, where $\mu^T = (1, -1, 2)$ and $\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$.

Which of the following random variables are independent? Explain.

(a) X_1 and X_2, (b) X_1 and X_3, (c) (X_1, X_3) and X_2, (d) X_1 and $X_1 + 3X_2 - 2X_3$, (e) $\frac{X_1 + X_3}{2}$ and X_2, and (f) X_2 and $X_2 - \frac{5}{2}X_1 - X_3$.

Procedure

- To identify the covariance of X_1 and X_2.
- To identify the covariance of X_1 and X_3.
- To calculate the covariance of (X_1, X_3) and X_2 using $A\Sigma A^T$.
- To calculate the covariance of X_1 and $X_1 + 3X_2 - 2X_3$ using $A\Sigma A^T$.
- To calculate the covariance of $\frac{X_1 + X_3}{2}$ and X_2 using $A\Sigma A^T$.
- To calculate the covariance of X_2 and $X_2 - \frac{5}{2}X_1 - X_3$ using $A\Sigma A^T$.

Calculation

(a) X_1 and X_2

Given $\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$

Since X_1 and X_2 have covariance, $\sigma_{12} = \sigma_{21} = 0$. Therefore the random variables X_1 and X_2 are independent

(b) X_1 and X_3

Since X_1 and X_3 have covariance, $\sigma_{13} = \sigma_{31} = -1$. Therefore the random variables X_1 and X_3 are not independent

(c) (X_1, X_3) and X_2

Rearrange the covariance matrix and partition it. The new covariance matrix is as following:

$$X = \begin{bmatrix} X_1 \\ X_3 \\ \cdots \\ X_2 \end{bmatrix} \text{ and } \Sigma^* = \begin{bmatrix} 4 & -1 & \vdots & 0 \\ -1 & 2 & \vdots & 0 \\ \cdots & \cdots & \vdots & \cdots \\ 0 & 0 & \vdots & 5 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \vdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{bmatrix}$$

Now here, $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ and X_2 have covariance matrix $\Sigma_{13} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Therefore (X_1, X_3) and X_2 are independent. This implies X_2 is independent of X_1 and X_3.

(d) X_1 and $X_1 + 3X_2 - 2X_3$

Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & -2 \end{bmatrix}$, then $AX = \begin{bmatrix} X_1 \\ X_1 + 3X_2 - 2X_3 \end{bmatrix}$ and $AX \sim N(A\mu,\ A\Sigma A^T)$,

Where

$$A\Sigma A^T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 3 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 4 & 0 & -1 \\ 6 & 15 & -5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 3 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 61 \end{bmatrix}$$

Therefore X_1 and $X_1 + 3X_2 - 2X_3$ are not independent.

(e) $\frac{X_1 + X_3}{2}$ and X_2

Let partitioning X and Σ in $\frac{X_1 + X_3}{2}$ and X_2, $X = \begin{bmatrix} \frac{X_1 + X_3}{2} \\ \cdots \\ X_2 \end{bmatrix}$

then

$$A = \begin{bmatrix} \frac{X_1 + X_3}{2} \\ \cdots \\ X_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$$

$$A\Sigma A^T = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \\ \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & 0 & \frac{1}{2} \\ 0 & 5 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \\ \frac{1}{2} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 5 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \vdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{bmatrix}$$

Since the covariance matrix $\Sigma_{13} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ (which is the covariance between $\frac{X_1 + X_3}{2}$ and X_2). Therefore $\frac{X_1 + X_3}{2}$ and X_2 are independent.

(f) X_2 and $X_2 - \frac{5}{2}X_1 - X_3$

Let partitioning X and Σ in $X_2 - \frac{5}{2}X_1 - X_3$ and X_2, $X = \begin{bmatrix} X_2 - \frac{5}{2}X_1 - X_3 \\ \cdots \\ X_2 \end{bmatrix}$

Then

$$A = \begin{bmatrix} X_2 - \frac{5}{2}X_1 - X_3 \\ \cdots \\ X_2 \end{bmatrix} = \begin{bmatrix} -\frac{5}{2} & 1 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$A\Sigma A^T = \begin{bmatrix} -\frac{5}{2} & 1 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} -\frac{5}{2} & 0 \\ 1 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -9 & 5 & \frac{1}{2} \\ 0 & 5 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{2} & 0 \\ 1 & 1 \\ -1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 27 & \vdots & 5 \\ \cdots & \vdots & \cdots \\ 5 & \vdots & 5 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \vdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{bmatrix}$$

Therefore X_2 and $X_2 - \frac{5}{2}X_1 - X_3$ are not independent.

Result

(a) X_1 and X_2 are independent.

(b) X_1 and X_3 are not independent

(c) (X_1, X_3) and X_2 are independent

(d) X_1 and $X_1 + 3X_2 - 2X_3$ are not independent

(e) $\frac{X_1 + X_3}{2}$ and X_2 are independent.

(f) X_2 and $X_2 - \frac{5}{2}X_1 - X_3$ are not independent.

2. **Linear combinations of random vectors**

Let X_1, X_2, X_3 and X_4 be independent and identically distributed 3×1 random vectors with

$$\mu = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

(a) Find the mean and variance of the linear combination $a'X_1$ of the three components of X_1 where $a = [a_1\ a_2\ a_3]'$.

(b) Consider two linear combinations of random vectors

(i) $\frac{1}{2}X_1 + \frac{1}{2}X_2 + \frac{1}{2}X_3 + \frac{1}{2}X_4$ and

(ii) $X_1 + X_2 + X_3 - 3X_4$

Find the mean vector and covariance matrix for each linear combination of vectors and also the covariance between them.

Procedure

- To calculate the Mean vector is $a'\mu$.
- To calculate the covariance matrix is $a\Sigma a'$.

Calculation

(a) Mean and variance of the linear combination $a'X_1$ of the three components of X_1

Given $\mu = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$, $a = [a_1\ a_2\ a_3]'$ and $\Sigma = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$

Let us consider a linear combination $a'X_1$ of the three components of X_1. This is a random variable with mean

$$a'\mu = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = 3a_1 - a_2 + a_3$$

and Variance

$$a\Sigma a' = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} = 3a_1^2 + a_2^2 + 2a_3^2 - 2a_1a_2 + 2a_1a_3$$

That is, a linear combination $a'X_1$ of the components of a random vector is a single random variable consisting of a sum of terms that are each constant times a variable. This is very different from a linear combination of random vectors,

$$c_1X_1 + c_2X_2 + c_3X_3 + c_4X_4$$

This is a random vector. Here each term in the sum is a constant times a random vector.

(b) Mean vector and covariance matrix for each linear combination of vectors and also the covariance between them

Now consider two linear combinations of random vectors

(i) $\frac{1}{2}X_1 + \frac{1}{2}X_2 + \frac{1}{2}X_3 + \frac{1}{2}X_4$

By the result $V_1 = c_1X_1 + c_2X_2 + ... + c_nX_n$ with $c_1 = c_2 = c_3 = c_4 = 1/2$, the linear combination has mean vector

$(c_1 + c_2 + c_3 + c_4)\mu = 2\mu$

$$2\mu = \begin{bmatrix} 6 \\ -2 \\ 2 \end{bmatrix}$$

and covariance matrix is

$$(c_1^2 + c_2^2 + c_3^2 + c_4^2)\Sigma = 1 \times \Sigma$$

$$\Sigma = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

(ii) $X_1 + X_2 + X_3 - 3X_4$

The linear combination of random vectors, we apply $V_1 = c_1X_1 + c_2X_2 + ... + c_nX_n$ with $b_l = b_2 = b_3 = 1$ and $b_4 = -3$ to get mean vector

$(b_1 + b_2 + b_3 + b_4)\mu = 0\mu$

$$0\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and covariance matrix is

$$(b_1^2 + b_2^2 + b_3^2 + b_4^2)\Sigma = 12 \times \Sigma$$

$$12 \times \Sigma = \begin{bmatrix} 36 & -12 & 12 \\ -12 & 12 & 0 \\ 12 & 0 & 24 \end{bmatrix}$$

Finally, the covariance matrix for the two linear combinations of random vectors is

$(c_1b_1 + c_2b_2 + c_3b_3 + c_4b_4)\,\Sigma = 0\,\Sigma$

$$0 \times \Sigma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Result

(a) Mean of the linear combination $a'X_1$ of the three components of X_1 is $3a_1 - a_2 + a_3$ and variance is $3a_1^2 + a_2^2 + 2a_3^2 - 2a_1a_2 + 2a_1a_3$

(b) Mean vector and covariance matrix for each linear combination of vectors and also the covariance between them

(i) $\frac{1}{2}X_1 + \frac{1}{2}X_2 + \frac{1}{2}X_3 + \frac{1}{2}X_4$: Mean vector is $\begin{bmatrix} 6 \\ -2 \\ 2 \end{bmatrix}$ and covariance matrix is $\begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$.

(ii) $X_1 + X_2 + X_3 - 3X_4$: Mean vector is $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ and covariance matrix $\begin{bmatrix} 36 & -12 & 12 \\ -12 & 12 & 0 \\ 12 & 0 & 24 \end{bmatrix}$