UNIT 19 INFERENCE-I

Structure **Page No**

19.1 INTRODUCTION

We shall begin this unit by recalling the multivariate normal distribution. Many of our explanations use the representation of the rows of X as k points in n-dimensions. In section 19.2, we will introduce the assumptions that a random sample is constituted by the observations. We shall confine it for random sampling, for which we shall assume that the traits or the measurements considered for the different trials are independent and the joint distribution of all the variable remains the same. In section 19.3, we shall discuss the maximum likelihood estimation.

Objectives

After studying this unit, you should be able to:

- describe the likelihood function and the role of maximum likelihood estimation in deriving the normally distributed estimators;
- understand the distribution of a normal random vector;
- compute the mean and covariance;
- assess the role of multivariate normal distribution in finance.

19.2 MULTIVARIATE NORMAL DISTRIBUTION AND RANDOM SAMPLING

Let us recall the multivariate normal distribution, A p-dimensional random variable x is said to have a p variate normal distribution if and only if every linear function of x has a univariate normal distribution.

Let $X' = (X_1, X_2, \ldots, X_p)$ represent a p-dimensional random variable. The mean and variance-covariance matrix are denoted by μ and Σ, respectively and are given by

$$\mu' = E(x') = (E(x_1), E(x_2), \ldots, E(x_p)) = (\mu_1, \mu_2, \ldots, \mu_p) \qquad (1)$$

$$\Sigma = E[(x-\mu)\,(x-\mu)] \qquad (2)$$

$$= \begin{bmatrix} var(x_1) & cov(x_1, x_2) & \ldots & cov(x_1, x_2) \\ cov(x_2, x_1) & var(x_2) & \ldots & cov(x_2, x_p) \\ \vdots & \vdots & \ldots & \vdots \\ cov(x_p, x_1) & cov(x_p, x_2) & \ldots & var(x_p) \end{bmatrix}$$

The density of a p-variate normal random variable x is given by

$$N_p(\mu, \Sigma) = (2\pi)^{-1/2} \,|\Sigma|^{-1/2} \exp\left\{\frac{-1}{2}(x-\mu)'\,\Sigma^{-1}(x-\mu)\right\} \qquad (3)$$

where the mean of x is μ, and the variance-covariance matrix of x is Σ.

We now obtain the multivariate normal density by transforming the random vector $z = (z_1, z_2, \ldots, z_p)$ where each z_i has the $N(0,1)$ and the z_is are independent. Thus $E(z) = 0$ and $\text{cov}(z) = I$. When random variables are independently distributed, then their density function is the product of their individual densities. We can write

$$f(z) = f_1(z_1)\, f_2(z_2) \ldots f_p(z_p)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-z_1^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-z_2^2/2} \cdots \frac{1}{\sqrt{2\pi}} e^{-z_p^2/2} \quad \left[\text{since } f_i(z_i) = \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2}\right]$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^p} . e^{\sum_j z_j^2/2}$$

$$= \frac{1}{(\sqrt{2\pi})^p} e^{-z'z/2} \tag{4}$$

If we extend z as a multivariate normal density, then it is with mean vector **0** and covariance matrix **I**.

Here, we wish to obtain the multivariate normal density with arbitrary mean vector μ and covariance matrix Σ which is positive definite. We define the transformation for $y = \sigma z + \mu,\ y = \Sigma^{1/2} z + \mu$.

where $\Sigma^{1/2}$ is the (symmetric) square root matrix. The mean vector and covariance matrix for the transformed random vector y are :

$$E(y) = E\left(\Sigma^{1/2} z + \mu\right) = \Sigma^{1/2} E(z) + E(\mu)$$

$$= \Sigma^{1/2} \cdot 0 + \mu = \mu \tag{5}$$

$$\text{cov}(y) = \text{cov}\left(\Sigma^{1/2} z + \mu\right) = \Sigma^{1/2} \text{cov}(z) \left(\Sigma^{1/2}\right)'$$

$$= \Sigma^{1/2} \cdot I \cdot \Sigma^{1/2} = \Sigma \tag{6}$$

Let us now find the density of $y = \Sigma^{1/2} z + \mu$ from the density of z given in Eqn. (4).

The density of $y = \sigma z + \mu$ is $g(y) = f(z)\frac{dz}{dy} = f(z)|1/\sigma|$.

The analogous expression for

$$y = \Sigma^{1/2} z + \mu \text{ is } g(y) = f(z)\left|\Sigma^{1/2}\right| \tag{7}$$

It may be noted that in Eqn. (7) we use the vertical bars for the determinant of a matrix as opposed to the use of vertical bars to indicate the absolute value of a scalar quantity used as $\left|\frac{dz}{dy}\right|$.

Now Eqn. (7) can be written as

$$g(y) = f(z)\left|\Sigma^{-1/2}\right| \tag{8}$$

$$= f(z)|\Sigma|^{-1/2} \tag{9}$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^p} \cdot \frac{1}{|\Sigma|^{1/2}} e^{-z'z/2} \qquad \text{[using Eqn. (4)]}$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^p |\Sigma|^{1/2}} e^{-\left[\Sigma^{-1/2}(y-\mu)\right]'\left[\Sigma^{-1/2}(y-\mu)\right]/2} \quad \left[\text{since } Z = \frac{y-\mu}{\Sigma^{1/2}}\right]$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^p |\Sigma|^{1/2}} e^{-(y-\mu)'\left(\Sigma^{1/2}\Sigma^{1/2}\right)^{-1}(y-u)/2}$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^p |\Sigma|^{1/2}} e^{-(y-\mu)'\Sigma^{-1}(y-\mu)/2} \qquad (10)$$

Which is the multivariate normal density function with mean vector u and covariance matrix Σ.

Now, we shall discuss random sampling. For this, consider the data matrix X of order $n \times k$, which can be plotted for the n-dimensional scatterplot by representing the columns of X as points. We can write

$$X_{n\times k} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} = \left[x_1, x_2, \cdots x_n\right] = \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{bmatrix}$$

where y_i' is considered as the elements of the rows of the data matrix. Let us consider $\overline{x_i}$ as the sample mean of the i-th observation which is $\overline{x_i} = \frac{1}{n}\sum_{h=1}^{n} x_{hi}$; $i = 1, 2, \ldots, k$

then the deviation vector denoted by e_i is computed as

$$e_i = y_i - \overline{x}_i 1, \text{ where } 1 = [1, 1, \ldots, 1]'$$

$$= [x_{i1} - \overline{x}_i, x_{i2} - \overline{x}_i \ldots x_{ik} - \overline{x}_i]'$$

for two vectors e_i and e_j, we have $e_i'e_j = n\, s_{ij}$, where s_{ij} is the sample variance-covariance matrix. Also, the sample correlation coefficient

$$\hat{R} = \begin{bmatrix} 1 & r_{iy} \\ r_{iy} & 1 \end{bmatrix}, \text{ with } r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}} = \cos\left(\theta_{ij}\right)$$

where θ_{ij} is the angle between e_i and e_j.

Example 1: For $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

With $\sigma_1^2, \sigma_2^2 > 0, -1 < \rho < 1$.

Find the expression for the probability density function of a bivariate normal distribution.

Solution: Since $\det\Sigma = \sigma_1^2\sigma_2^2\left(1-\rho^2\right) > 0$,

Σ^{-1} exists and is given by

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_1^2} \end{bmatrix} \cdot \frac{1}{1-\rho^2}$$

Furthermore, for $X = \left(x_1, x_2\right)' \neq 0$

$$X'\Sigma X = \left(\sigma_1 x_1 + \rho\sigma_2 x_2\right)^2 + \left(1-\rho^2\right)\sigma_2^2 x_2^2 > 0.$$

Hence Σ is positive definite.

With $\mu = (\mu_1, \mu_2)'$

$$(x-\mu)'\Sigma^{-1}(x-\mu) = \frac{1}{1-\sigma^2}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right).$$

The probability density function of a bivariate normal random variable with values in E^2 is

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

Example 2: Let $\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Since, for $X = (x_1, x_2)' \neq 0$

$$X'\Sigma X = 2(x_1 + x_2/2)^2 + \frac{3}{2}x_2^2 > 0,$$

Σ is positive definite. Hence

$$f_x(x) = \frac{1}{2\pi\sqrt{3}}\exp\left\{-\frac{2}{3}\left[\frac{(x_1-2)^2}{2} + \frac{(x_2-3)^2}{2} - \frac{(x_1-2)(x_2-3)}{2}\right]\right\}$$

is the probability density function of a bivariate normal random variable $X = (X_1, X_2)'$ with mean $(2,3)'$ and covariance Σ.

Here $\rho = \frac{1}{2}$.

Example 3 (The equivalence of zero covariance and independence for normal variables): Let $X_{(3\times1)}$ be $N_3(\mu, \Sigma)$ with

$$\Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Check whether X_1 and X_2 are independent or not? Also check the independence of (X_1, X_2) and X_3?

Since X_1 & X_2 have covariance $\sigma_{12} = 1$, they are not independent. However, partitioning X & Σ as

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \cdots \\ X_3 \end{pmatrix}, \Sigma = \begin{bmatrix} 4 & 1 & \vdots & 0 \\ 1 & 3 & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & 2 \end{bmatrix} = \begin{bmatrix} \underset{2\times2}{\Sigma_{11}} & \vdots & \underset{2\times1}{\Sigma_{12}} \\ \cdots & \cdots & \cdots \\ \underset{1\times2}{\Sigma_{21}} & \vdots & \underset{1\times1}{\Sigma_{22}} \end{bmatrix}$$

We see that $X_1 = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and X_3 have covariance matrix $\Sigma_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Therefore, (X_1, X_2) and X_3 are independent. This implies X_3 is independent of X_1 and also of X_2.

Example 4 (The distribution of a linear combination of the components of a normal random vector): Consider the linear combination $a'x$ of a multivariate normal random vector determined by the choice $a'x = [1,0,\cdots,0]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = x_1$.

And

$$a'\mu = [1,0,\cdots,0]\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mu_1.$$

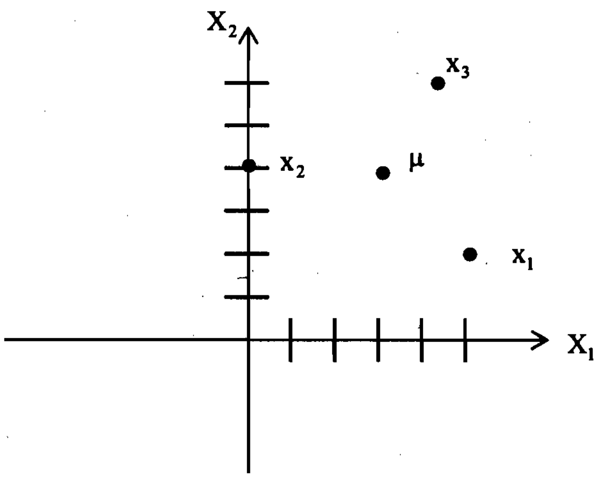
We have

$$a'\Sigma a = [1,0,\cdots,0]\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_{11}$$

and it follows from Result (1) that X_1 is distributed as $N(\mu_1, \sigma_{11})$. More generally, the marginal distribution of any component X_i of X is $N(\mu_i, \sigma_{ii})$.

[**Result 1:** If X is distributed as $N_p(\mu, \Sigma)$, then any linear combination of variables $a'X = a_1X_1 + a_2X_2 + \cdots + a_p \times p$ is distributed as $N(a'\mu, a'\Sigma a)$.]

Example 5: Consider $X = \begin{bmatrix} 5 & 0 & 4 \\ 2 & 4 & 6 \end{bmatrix}$, the scatter plot of X is



Mean vector is $\mu = \begin{bmatrix} \dfrac{5+0+4}{3} \\ \dfrac{2+4+6}{3} \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

Here, $e_1 = y_1 - \mu_1 1 = \begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$

$$e_2 = y_2 - \mu_2 1 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$$

Also, $3s_{11} = e_1'e_1 = 14$

$3s_{12} = e_1'e_2 = -2$

$3s_{22} = e_2'e_2 = 8$

Therefore, $s_n = \begin{bmatrix} \frac{14}{3} & -2/3 \\ -2/3 & 8/3 \end{bmatrix}, r_{12} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}} = -0.189$

and $\hat{R} = \begin{bmatrix} 1 & -0.189 \\ -0.189 & 1 \end{bmatrix}$.

Now, try the following exercises:

E1) For X distributed as $N_3(\mu, \Sigma)$, Find the distribution of $\begin{bmatrix} X_1 & - & X_2 \\ X_2 & - & X_3 \end{bmatrix}$.

E2) If X is distributed as $N_5(\mu, \Sigma)$, find the distribution of $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$.

E3) Prove that if the covariance matrix Σ of a normal random vector $X = (X_1, \cdots, X_p)'$ is a diagonal matrix, then the components of X are independently distributed normal random variables.

E4) Find the sample correlation matrix $\hat{R}$ for the data matrix $X = \begin{bmatrix} -1 & 2 & 5 \\ 3 & 4 & 2 \\ -2 & 2 & 3 \end{bmatrix}$.

E5) Let X be $N_3(\mu, \Sigma)$ with $\mu' = [-3, 1, 4]$ and $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. Which of the following random variables are independent? Give reasons.

a) X_1 and X_2

b) X_2 and X_3

c) (X_1, X_2) and X_3

d) $\left(\frac{X_1 + X_2}{2}\right)$ and X_3

e) X_2 and $\left(X_2 - \frac{5}{2}X_1 - X_3\right)$.

E6) Find the mean and the covariance matrix of the random vector $X = (X_1, X_2)'$ with probability density function

$f_x(X) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(2x_1^2 + x_2^2 + 2x_1x_2 - 22x_1 - 14x_2 + 65\right)\right\}$ and $X \in E^2$.

In the following section, we shall define and maximum likelihood estimation of multivariate normals.

19.3 MAXIMUM LIKELIHOOD ESTIMATION

Suppose $x_1, x_2, \cdots, x_n$ is a random sample of vector observation from a population with a density $h(x_j;\theta)$, where θ is a parameter vector that identifies the particular member of a family being sampled.
The likelihood function, denoted by $L(x;\theta)$, is defined by

$$L(x;\theta) = \prod_{i=1}^{n} h(x_i;\theta)$$

where we view $L(x;\theta)$ as a function of θ for a fixed set of sample observation $x = (x_1, x_2, \cdots, x_n)$.
In the case where the density is $N_p(\mu, \Sigma)$, then

$$L(x,\mu,\Sigma) = \left(\frac{1}{2\pi}\right)^{n/2} |\Sigma|^{-n/2} \exp\left\{\frac{-1}{2}\sum_i (x_i - \mu)\Sigma^{-1}(x_i - \mu)\right\}.$$

Frequently, the logarithm of the likelihood function is used.

The value of a parameter, as a function of the data, that maximizes the likelihood function is called a maximum likelihood estimator (MLE). In many important cases MLEs can be found quite readily by differentiating the likelihood function, or its logarithm, setting the result equal to zero, and solving. In other cases the maximum occurs at a point where the derivative does not exist, and hence other procedures have to be followed. Often numerical methods have to be used. Iterative proportional fitting and the Newton-Raphson method are two such algorithms frequently used to generate MLEs in this case.

Properties of Maximum Likelihood Estimators

1) **Unbiasedness:** An estimate S of a parameter θ is said to be unbiased if $E(S) = \theta$. In the case where S is a vector or matrix of estimators, then we say the estimate is unbiased if the expected values of the constituent components equal the component values of the vector or matrix of parameters. The sample mean $\overline{X}$ is an unbiased estimate of the population mean vector μ, and if the sample sums-of-squares cross-product matrix is divided by the sample size minus 1, then it is an unbiased estimate of the true variance-covariance matrix.

2) **Sufficiency:** Suppose that the data matrix $X = (X_1, X_2, \ldots, X_n)$ where X_i is a p-dimensional vector, forms a realization of a random vector, and that a family F of possible distributions is specified. A statistic S is said to be sufficient for F if the conditioned density $f_{XIS}(XIS)$ is the same for all the distributions in F. The structure of the likelihood function indicates the form that sufficient statistics take on. In particular, a necessary and sufficient condition for S to be a sufficient statistic is that the likelihood function is expressible as $L(\theta;X) = h(X,\theta)g(X)$.

From this result, known as factorization theorem, it is clear that maximum likelihood estimates are often the functions of sufficient statistics. For the multivariate normal family, the factorization theorem shows directly that the mean vector $\overline{X}$ and sample variance-covariance matrix S are sufficient statistics for the population parameters μ and Σ, respectively.

Now, we shall discuss the following theorem.

Theorem1: If $x_1, x_2, \ldots, x_n$ is a random sample from $N_p(\mu, \Sigma)$, then the maximum likelihood estimators of μ and Σ are

$$\hat{\mu} = \bar{x} \tag{11}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)' = \frac{1}{n}W$$

$$\hat{\Sigma} = \frac{n-1}{n}\cdot S = \frac{1}{n}W \tag{12}$$

where $W = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)'$.

Proof: Since the x_is are independent (because they arise from a random sample), the likelihood function (joint density) is the product of the densities of the x_i's :

$$L(\mu, \Sigma) = \prod_{i=1}^{n} f(x_i; \mu, \Sigma)$$

$$= \prod_{i=1}^{n} \frac{1}{\left(\sqrt{2\pi}\right)^p |\Sigma|^{1/2}} e^{-(x_i - \mu)'\Sigma^{-1}(x_i - \mu)/2}$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^{np} |\Sigma|^{n/2}} e^{-\sum_{i=1}^{n}(x_i - \mu)'\Sigma^{-1}(x_i - \mu)/2} \tag{13}$$

We use the notation $L(\mu, \Sigma)$ because we consider $x_1, x_2, \ldots, x_n$ to be known or available from a future sample. For the given values of $x_1, x_2, \ldots, x_n$, we seek the values of μ and Σ that maximize Eqn. (13). We first express Eqn. (13) in a form that will facilitate finding the maximum.

The scalar quantity $(x_i - \mu)'\Sigma^{-1}(x_i - \mu)$ is equal to its trace.
Hence, we have

$$\sum_{i=1}^{n}(x_i - \mu)'\Sigma^{-1}(x_i - \mu) = \sum_{i=1}^{n} tr(x_i - \mu)'\Sigma^{-1}(x_i - \mu)$$

$$= tr\left[\Sigma^{-1}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)'\right] \tag{14}$$

$$\begin{bmatrix} \text{Here}, 1\cdot tr(A+B) = tr(A) + tr(B) \\ 2\cdot tr(AB) = tr(BA) \end{bmatrix}$$

Now by adding and subtracting $\bar{x}$ in the sum in the right side of Eqn. (14), we obtain

$$\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)' = \sum_{i=1}^{n}\left(x_i - \bar{x} + \bar{x} - \mu\right)\left(x_i - \bar{x} + \bar{x} - \mu\right)'$$

$$= \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)' + n\left(\bar{x} - \mu\right)\left(\bar{x} - \mu\right)'$$

$$= W + n\left(\bar{x} - \mu\right)\left(\bar{x} - \mu\right)' \tag{15}$$

The other two terms in the expression preceding Eqn. (15) vanish because $\sum_i \left(x_i - \bar{x} \right) = 0$. Using Eqn. (14) and Eqn. (15), we obtain

$$L(\mu, \Sigma) = \frac{1}{\left(\sqrt{2\pi}\right)^{np} |\Sigma|^{n/2}} e^{-\mathrm{tr}\,\Sigma^{-1}\left[w + n\left(\bar{x}-\mu\right)\left(\bar{x}-\mu\right)'\right]/2}. \tag{16}$$

Because the natural logarithm is an increasing function, the maximum of ln L will occur at the same point as the maximum of L. We prefer to work with ln L because of its simpler form for differentiation:

$$\ln L(\mu, \Sigma) = -np \ln \sqrt{2\pi} - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \mathrm{tr}\Sigma^{-1} \left[w + n\left(\bar{x} - \mu\right)\left(\bar{x} - \mu\right)' \right]$$

$$= -np \ln \sqrt{2\pi} - \frac{n}{2} \ln|\Sigma| - \frac{1}{2} \mathrm{tr}\left(\Sigma^{-1} w\right) - \frac{n}{2}\left(\bar{x} - \mu\right)' \Sigma^{-1} \left(\bar{x} - \mu\right). \tag{17}$$

To find the maximum likelihood estimator for μ, we differentiate $\ln L(\mu, \Sigma)$ given in Eqn. (17) with respect to μ, and set the resulting expression equal to 0:

$$\frac{\partial \ln L\ (\mu,\ \Sigma)}{\partial\ \mu} = -0 - 0 - 0 + n\ \left(\Sigma^{-1}\bar{x} - \Sigma^{-1}\mu\right) = 0 \tag{18}$$

which gives

$$\hat{\mu} = \bar{x}.$$

It is clear that $\hat{\mu} = \bar{x}$ maximizes $\ln L(\mu, \Sigma)$ with respect to μ, because the last term of Eqn. (17) is ≤ 0 and the term vanishes for $\hat{\mu} = \bar{x}$.

Before differentiating $\ln L(\mu, \Sigma)$ to find $\hat{\Sigma}$, we substitute $\hat{\mu} = \bar{x}$ in Eqn. (17) and rewrite $|\Sigma|$ in term of Σ^{-1} to obtain

$$\ln L\left(\bar{\mu}, \Sigma\right) = -np \ln \sqrt{2\pi} + \frac{n}{2} \ln \left|\Sigma^{-1}\right| \frac{-1}{2} \mathrm{tr}\left(\Sigma^{-1} W\right). \tag{19}$$

We now differentiate Eqn. (19) with respect to Σ^{-1} ;

$$\frac{\partial \ln L\left(\hat{\mu}, \Sigma\right)}{\partial \Sigma^{-1}} = -0 + n\Sigma - \frac{n}{2} \mathrm{diag}(\Sigma) - W + \frac{1}{2} \mathrm{diag}(W) = 0 \tag{20}$$

From which we have

$$\hat{\Sigma} - \frac{1}{2} \mathrm{diag}\left(\hat{\Sigma}\right) \frac{1}{n}\left[W - \frac{1}{2} \mathrm{diag}(W)\right] \tag{21}$$

or $\hat{\Sigma} = \frac{1}{n} W$

To see that $\hat{\Sigma} = \frac{W}{n}$ is the solution to Eqn. (21), note that for the off-diagonal elements we have $\hat{\sigma}_{jk} = w_{jk}/n, j \neq k$, and for the diagonal elements,

$\hat{\sigma}_{jj} - \hat{\sigma}_{jj}/2 = w_{jj}/n - w_{jj}/2n$ or $\hat{\sigma}_{jj}/2 = w_{jj}/2n$

[Note that we solved Eqn. (21) for Σ instead of Σ^{-1}, even though we differentiated with respect to Σ^{-1}. Otherwise we would have obtained $(W/n)^{-1}$ as the maximum likelihood estimator for Σ^{-1}. This illustrates the invariance property of maximum likelihood estimators; That is, $\left[\left(W/n\right)^{-1}\right]^{-1} = W/n$ is the maximum likelihood estimator for $\left(\Sigma^{-1}\right)h-1=\Sigma$. Since Σ is positive definite, W is positive definite with probability1.]

Example 6: Consider a portfolio consisting of n securities, and denote the rate of return of security i by X_i. Let C_i denotes the weight of security i in the portfolio. Then, the rate of return of the portfolio is given by the linear combination,

$$R = C_1X_1 + C_2X_2 + \cdots + C_nX_n \tag{22}$$

To see this, let x_i be the current market value of security i, and let Y_i be its future value. The rate of return of security i is given by $X_i = \frac{Y_i - x_i}{x_i}$. Let $Q_0 = \sum_{i=1}^{n} \xi_i x_i$, where ξ_i denotes the number of shares (not weight) of security i, and let Q denotes the random variable that represents the future portfolio value. Assuming that no rebalance will be made the weight C_i in the portfolio is given by,

$$C_i = \frac{\xi_i x_i}{Q_0} \tag{23}$$

It follows that the rate of return, $R = \frac{Q - Q_o}{Q_o}$, of the portfolio is obtained as

$$R = \frac{\sum_{i=1}^{n} \xi_i \left(y_i - x_i\right)}{Q_o} = \sum_{i=1}^{n} C_i X_i \tag{24}$$

As claimed the mean $E(R)$ and the $V(R)$ are given by,

$$E(R) = \sum_{i-1}^{n} c_i \mu_i, V(R) = \sum_{i=1}^{n} \sum_{i=1}^{n} c_i \sigma_{ij} c_j.$$

Respectively, no matter what the random variables X_i are.

Now, of interest is the potential for significant loss in the portfolio. Namely, given a probability level α, what would be the maximum loss in the portfolio ?

The maximum loss is called the value at risk (VaR) with confidence level $100\alpha\%$ and VaR is a prominent tool for risk management. More specifically, VAR with confidence level $100\alpha\%$ is defined by $z_\alpha > 0$ that satisfies

$$P\{Q - Q_0 \geq -z_\alpha\} = \alpha \tag{25}$$

And VAR's popularity is based on aggregation of many components of market risk into the single number z_α. Fig. 1 depicts the meaning of Eqn. (25)

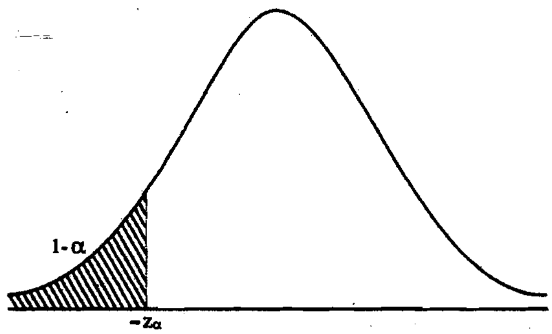


Fig. 1: The density function of $Q - Q_0$ and VaR

Since $Q - Q_o = Q_o R$, we obtain gap $P\{Q_o R \le -Z\alpha\} = 1-\alpha$ (26)
Note that the definition of VAR does not require normality. However, calculation of VAR becomes considerably simpler, if we assume that $(X_1, \cdots, X_n)$ follows an n-variate normal distribution. Then, the rate of return R, in Eqn. (24) is normally distributed with mean $E(R) = \sum_{i=1}^{n} c_i \mu_i$, and variance $V(R) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i \sigma_{ij} c_j$.

The value Z_α satisfying Eqn. (26) can be obtained using a table of the standard normal distribution. In many standard textbooks of statistics, a table of the survival probability

$$L(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \quad x > 0,$$

is given. Then using the standardization $\left(Y = \frac{x-\mu}{\sigma}\right)$ and symmetry of the density function $\left(\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x \in R\right)$ about 0, we can obtain the value Z_α with ease. Namely, letting $r_\alpha = -z_\alpha / Q_0$, it follows from Eqn. (26) that

$$1-\alpha = p\left\{\frac{R-\mu}{\sigma} \le \frac{r_\alpha - \mu}{\sigma}\right\},$$

where $\mu = E(R), \sigma^2 = V(R)$ whence

$$1-\alpha = L\left(-\frac{r_\alpha - \mu}{\sigma}\right).$$

Therefore, letting x_α be such that

$$L(x_\alpha) = 1-\alpha, \text{ we have } z_\alpha = Q_0(x_\alpha \sigma - \mu), \quad (27)$$

$$-x_\alpha = -\frac{r_\alpha - \mu}{\sigma}$$

$$(x_\alpha \sigma - \mu) = -r_\alpha$$

$$x_\alpha \sigma - \mu = \frac{Z_\alpha}{Q_o}$$

$$Z_\alpha = Q_o(x_\alpha \sigma - \mu)$$

The value z_α in Eqn. (27) is Var with confidence level $100\alpha\%$. The value x_α is the $100(1-\alpha)$ percentile of the standard normal distribution given in Table 1.
For example, if $\mu = 0$, then the 99% VaR is given by $2.326\sigma Q_0$.

Note: Since the risk horizon is very short, e.g. one day or one week, in market risk management, the mean rate of return μ is often set to be zero. In this case, VAR with confidence level $100\alpha\%$ is given by $x_\alpha \sigma Q_0$.

Table 1: Percentiles of the standard normal distribution

$100(1-\alpha)$	10	5.0	1.0	0.5	0.1
x_α	1.282	1.645	2.326	2.576	3.090

Now try the following exercises:

E7) Show that $\sum_{i=1}^{n} tr(y_i - \mu)' \Sigma^{-1}(y_i - \mu) = tr\left[\Sigma^{-1} \sum_{i=1}^{n} (y_i - \mu)(y_i - \mu)'\right]$.

E8) Show that $\frac{\partial \ln L(\mu, \Sigma)}{\partial \mu} = n\left(\Sigma^{-1}\bar{y} - \Sigma^{-1}\mu\right)$.

E9) Let $X = \left(X_{(1)}, X_{(2)}\right)^n$ with $X_{(1)} = \left(X_1, \cdots X_q\right)'$ and $X_{(2)} = \left(X_{q+1}, \cdots, X_p\right)'$, and let μ be similarly partitioned as $\mu = \left(\mu_{(1)}, \mu_{(2)}\right)$, and let Σ be partitioned as $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ where Σ_{11} is a submatrix of dimension $q \times q$. Prove that if X has normal distribution with mean μ and covariance matrix Σ, and $\Sigma_{12} = \Sigma_{21} = 0$, then $X_{(1)}$ and $X_{(2)}$ are independently normally distributed with means $\mu_{(1)}, \mu_{(2)}$ and covariance matrices Σ_{11}, Σ_{12}, respectively.

E10) Show that $\sum_{j=1}^{n} (x_j - \bar{x})\,(\bar{x} - \mu)'$ and $\sum_{j=1}^{n} (\bar{x} - \mu)\,(x_j - \bar{x})'$ are both $p \times p$ matrices of zeros. Here $x'_j = [x_{j1}, x_{j2}, \ldots, x_{jp}]$, $j = 1, 2, \ldots, n$ and $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$.

Now let us sum up this unit.

19.4 SUMMARY

In this unit, we have covered the following points.

1. A multivariate normal distribution is a generalization of the one-dimensional normal distribution to higher dimensions. It is also closely related to matrix normal distribution.

2. Estimation of parameters of multivariate normal distribution requires maximum likelihood estimation (MLE), which is a popular statistical methods used for fitting a mathematical model to some data. Maximum likelihood estimation gives a unique and easy to determine solution in the case of normal distribution and many other problems. That is, it has widespread applications in various fields, including structural equation modeling, psychometrics and econometrics, data modeling in nuclear and particle physics, path-choice modeling in transport networks, etc.

19.5 SOLUTIONS/ANSWERS

E1) $$\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = AX \quad \text{(suppose)}$$

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

$$A\Sigma A' = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

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

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

E2) We set $X_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}, \mu_1 = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \Sigma_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$

And note that with this assignment $X, \mu,$ and Σ can respectively be rearranged and partitioned as

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

or,

$$X = \begin{bmatrix} \underset{(2\times1)}{X_1} \\ \cdots \\ \underset{(3\times1)}{X_2} \end{bmatrix}, \mu = \begin{bmatrix} \underset{(2\times1)}{\mu_1} \\ \cdots \\ \underset{(3\times1)}{\mu_2} \end{bmatrix}, \Sigma = \begin{bmatrix} \underset{(2\times2)}{\Sigma_{11}} & \vdots & \underset{(2\times3)}{\Sigma_{12}} \\ \cdots & \vdots & \cdots \\ \underset{(3\times2)}{\Sigma_{21}} & \vdots & \underset{(3\times3)}{\Sigma_{22}} \end{bmatrix}$$

Thus, from the distribution

$$N_2(\mu_1, \Sigma_{11}) = N_2\left(\begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}\right)$$

It is clear from this example that the normal distribution for any subset can be expressed by simply selecting the appropriate means and covariances from the original μ & Σ.

E3) Since Σ is diagonal matrix of order p, let us prove the same for p = 2.

for $p = 2$, $\Sigma = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$

here, $\sigma_{12} = \sigma_{21} = 0$ as Σ is a diagonal matrix.

The density function for bivariate normal distribution can be given by:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}\left(1-\rho_{12}^2\right)}}$$

$$\text{x} \exp\left\{-\frac{1}{2\left(1-\rho_{12}^2\right)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

since $\sigma_{12} = 0 \;\therefore\; \rho_{12} =$ correlation coefficient = 0

this gives,

$$f(x_1,x_2)=\frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}}}\text{ x }\exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]\right\}$$

$$=\left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}}}\cdot\exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2\right\}\right]\cdot\left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{22}}}\cdot\exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right\}\right]$$

$f(x_1,x_2)=f(x_1).f(x_2)$

i.e. if the random variables X_1 and X_2 are uncorrelated, so that $\rho_{12}=0$, then the joint density are the product of two univariate normal density and X_1, X_2 are independent.

This result is true in general.

E4) $\mu=\begin{bmatrix}2\\3\\1\end{bmatrix}$.

$e_1=\begin{bmatrix}-3\\0\\3\end{bmatrix}$, $e_2=\begin{bmatrix}0\\1\\-1\end{bmatrix}$ and $e_3=\begin{bmatrix}-3\\1\\2\end{bmatrix}$

$s_n=\begin{bmatrix}6 & -1 & 5\\-1 & 2/3 & -1/3\\5 & -1/3 & 14/3\end{bmatrix}$.

E5) a) Since x_1 and x_2 have covariance $\sigma_{12}-2$, they are not independent.

b) Since x_1 and x_3 have covariance $\sigma_{13}=0$, they are independent.

c) Partitioning x and Σ in pairs (x_1, x_2) and x_3, we get

$$x=\begin{bmatrix}x_1\\x_2\\\cdots\\x_3\end{bmatrix}, \Sigma=\begin{bmatrix}1 & -2 & \vdots & 0\\-2 & 5 & \vdots & 0\\\cdots & \cdots & \vdots & \cdots\\0 & 0 & \vdots & 2\end{bmatrix}=\begin{bmatrix}\Sigma_{11} & \vdots & \Sigma_{12}\\\cdots & \vdots & \cdots\\\Sigma_{21} & \vdots & \Sigma_{22}\end{bmatrix}.$$

Now here $\begin{bmatrix}x_1\\x_2\end{bmatrix}$ and x_3 have covariance matrix $\Sigma_{12}=\begin{bmatrix}0\\0\end{bmatrix}$, therefore (x_1, x_2) and x_3 are independent.

d) Similarly, partitioning x and Σ in $\frac{x_1+x_2}{2}$ and x_3,

$$x=\begin{bmatrix}\frac{x_1+x_2}{2}\\\cdots\cdots\cdots\\x_3\end{bmatrix}$$

Further, $\begin{bmatrix}\frac{x_1+x_2}{2}\\x_3\end{bmatrix}=\begin{bmatrix}1/2 & 1/2 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$

Thus, $A = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Now,

$$A\Sigma A' = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \vdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{bmatrix}$$

Since the covariance matrix $\Sigma_{12} = 0$ (which is the covariance between $\frac{x_1 + x_2}{2}$ and x_3), therefore, $\frac{x_1 + x_2}{2}$ and x_3 are independent.

E7) $\sum_{i=1}^{n} tr(y_i - \mu)' \Sigma^{-1} (y_i - \mu) = tr\Sigma^{-1} \sum_{i=1}^{n} (y_i - \mu)(y_i - \mu)'$

It follows from the fact that "trace of a sum of matrices is equal to the sum of the traces of the matrices".

Also, $x'Ax = tr(x'Ax) = tr(Axx')$ where A is a $k \times k$ symmetric matrix and x be a $k \times 1$ vector.

E8) Since we know that

$$\ln L(\mu, \Sigma) = -np \ln\sqrt{2\pi} - \frac{n}{2} |n| |\Sigma| - \frac{1}{2} tr(\Sigma^{-1} w) - \frac{n}{2} (\bar{y} - \mu)' \Sigma^{-1} (\bar{y} - \mu)$$

$$\Rightarrow \frac{\partial \ln L(\mu, \Sigma)}{\partial \mu} = 0 - 0 - 0 - n(-\Sigma^{-1} (\bar{y} - \mu))$$

$$= n(\Sigma^{-1}\bar{y} - \Sigma^{-1}\mu)$$

(since $\frac{n}{2}(\bar{y} - \mu)' \Sigma^{-1} (\bar{y} - \mu) = \frac{n}{2} (\bar{y} - \mu)^2 \Sigma^{-1}$ and

$\frac{\partial}{\partial \mu} \left[\frac{n}{2} \Sigma^{-1} (\bar{y} - \mu)^2 \right] = -n\Sigma^{-1} (\bar{y} - \mu)$.

E9) Given, $x = \begin{bmatrix} x_{(1)} \\ x_{(2)} \end{bmatrix}$ $\mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix}$

$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, $\Sigma^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix}$ (Since $\Sigma_{12} = \Sigma_{21} = 0$)

Also, x is $N_p(\mu, \Sigma)$ with $|\Sigma| \neq 0$,

Now, $(x - \mu)'\Sigma^{-1}(x - \mu) = [(x_{(1)} - \mu_{(1)})'(x_{(2)} - \mu_{(2)})']$

$$\begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} x_{(1)} & \mu_{(1)} \\ x_{(2)} & \mu_{(2)} \end{bmatrix}$$

$$= (x_{(1)} - \mu_{(1)})' \Sigma_{11}^{-1} (x_{(1)} - \mu_{(1)}) + (x_{(2)} - \mu_{(2)})' \Sigma_{22}^{-1} (x_{(2)} - \mu_{(2)})$$

i.e., $x_{(1)}$ and $x_{(2)}$ are independently normally distributed with mean $\mu_{(1)}, \mu_{(2)}$ respectively and covariances matrices Σ_{11} and Σ_{22} respectively.

E10) Let us take, $\sum_{j=1}^{n} (x_j - \bar{x})(\bar{x} - \mu)'$

$$= \sum_{j=1}^{n} x_j \bar{x} - \sum_{j=1}^{n} x_j \mu - \sum_{j=1}^{n} \bar{x}^2 + \sum_{j=1}^{n} \bar{x} \mu$$

$$= n \bar{x}^2 - n \bar{x} \mu - n \bar{x}^2 + n \bar{x} \mu = 0$$

Similarly, other can be solved.

19.6 PRACTICAL ASSIGNMENT

Session 4

1. Write a programme is 'C' language to find the ML estimation of mean (μ) and variance (σ^2).