

Applied Multivariate Analysis

Prof. Amit Mitra

Prof. Sharmishtha Mitra

Department of Mathematics and Statistics

Indian Institute of Technology, Kanpur

Lecture No. # 08

Random Sampling from Multivariate Normal Distribution and Wishart Distribution - I

Today, we will start looking at inference problems inference about multivariate normal distribution. So specifically, we have already introduced, what is a multivariate normal distribution, and what are the characterizing parameters of a multivariate normal distribution. So, today we will address the problem of inference about the multivariate mean vector μ and the covariance matrix σ .

(Refer Slide Time: 00:45)

Random sampling from multivariate normal
Multivariate normal popⁿ $N_p(\mu, \Sigma)$; $\Sigma > 0$
 X_1, \dots, X_n be a random sample from $N_p(\mu, \Sigma)$
Jt p.d.f of X_1, \dots, X_n
$$\prod_{i=1}^n f_{X_i}(x_i)$$
$$= \prod_{i=1}^n (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x_i - \mu)' \Sigma^{-1}(x_i - \mu)\right)$$
$$= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)' \Sigma^{-1}(x_i - \mu)\right)$$

So, we have actually the following setup that it is a basically random sampling from multivariate normal distribution. So suppose, we have a population, multivariate normal population $N_p(\mu, \Sigma)$. So, what we have here is this μ to be giving the mean vector and Σ to be the covariance matrix, we assume that Σ is positive definite. Now, the elements in μ ; this μ vector is unknown, and so is the Σ matrix. So, what we do is, we take random samples from this multivariate population, and then try to build up

the inference procedure concerning the unknown mean vector μ and the unknown covariance matrix Σ . So, what we do is, we take a random sample X_1, X_2, \dots, X_n be a random sample from this multivariate normal population and then using this random sample X_1, X_2, \dots, X_n the random sample of size n will be the inference procedure about the mean vector μ and the covariance matrix Σ .

Now, given this random sample the joint pdf of X_1, X_2, \dots, X_n remember that it is a random sample and hence this X_1, X_2, \dots, X_n are independent. So, that the joint distribution the joint pdf of X_1, X_2, \dots, X_n is given by the following that i equal to 1 to n , then the pdf of the respective components X_1, X_2, \dots, X_n . So, that this is $f(x_i)$. So, that this is equal to i equal to 1 to n 2π to the power minus p by 2 determinant of Σ to the power minus half, then exponent minus half $x_i - \mu$ transpose $\Sigma^{-1} x_i - \mu$. So, this is ... what is the joint pdf. So, this can be compactly written as 2π to the power minus n p by 2 determinant of Σ to the power minus n by 2 then exponent minus half summation i equal to 1 to n $x_i - \mu$ transpose $\Sigma^{-1} x_i - \mu$. So, the first thing about this random sample X_1, X_2, \dots, X_n from the multivariate normal distribution, the first question that; we are going to answer is that what is the sufficient statistic that is based on this random sample X_1, X_2, \dots, X_n .

(Refer Slide Time: 04:12)

Sufficient statistic

Case 1: μ and Σ are unknown

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

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

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

$$+ 2 \sum_{i=1}^n (x_i - \bar{x})' \Sigma^{-1} (\bar{x} - \mu) \quad \text{--- (1)}$$

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

$$= n (\bar{x} - \bar{x})' \Sigma^{-1} (\bar{x} - \mu) = 0$$

So, we move on to the sufficiency. So, we will be deriving sufficient statistic under various possibilities. Let us consider the first case, case one which is the most general case μ and Σ are unknown. So, we have both the two parameters characterizing this multivariate normal population, the μ vector and the covariance matrix Σ both

of them are unknown. Now, let us consider this particular term which is present in the exponent this is a term which is present in the exponent, let us see how we can manipulate this particular term this is summation i equal to 1 to n x_i minus μ transpose sigma inverse x_i minus μ . So, we introduce an \bar{x} term.

Now, note that this derivation of sufficient statistic for the multivariate normal population goes along exactly in the same line as to what we usually do when we have a univariate normal population. So, we introduce this \bar{x} vector which is the sample mean vector in this following way sigma inverse x_i minus \bar{x} the same adjustment in the second term as well. So, this now, we will split it into two terms this is one term here and this is the second term here and then take sigma inverse multiplication term by term. So, the first term will be x_i minus \bar{x} transpose sigma inverse x_i minus \bar{x} minus \bar{x} and then this term multiplied by sigma inverse into this particular term now, that is a term which is independent of i and hence we will have n times \bar{x} minus μ transpose sigma inverse \bar{x} minus μ , and then the cross product term of this term multiplied by this and this second term out here and the two terms would be similar the two cross product terms and what will be having is the following that it is summation i equal to 1 to n x_i minus \bar{x} transpose sigma inverse \bar{x} minus μ .

Now, this is if we look at the two factor here it is basically coming from the cross product, because we having two terms which are similar just differing by transpose and the two terms is same. Now, what is going to happen with this particular term is the following. Let me put this as equation number 1. Now, if we look at this expression here, what is that going to give us note that this summation i equal to 1 to n x_i minus \bar{x} transpose sigma inverse \bar{x} minus μ . Now, this term Now, this part here is independent of i and hence it can stay outside the summation.

So, what we well be having is summation i equal to 1 to n x_i minus \bar{x} transpose and then that is post multiplied by this sigma inverse \bar{x} minus μ which is independent of i . Now, what is going to happen when we look at the summation here, it is going to give us 0, because it is the deviation from this mean vector μ \bar{x} vector sample mean and hence this is going to give us n times \bar{x} transpose minus n times \bar{x} transpose. So, this is n times \bar{x} transpose minus this \bar{x} transpose times sigma inverse \bar{x} minus μ and that is equal to 0. So, this cross product term in equation number 1 this specifically. So, this term is going to vanish, because this is equal to 0.

(Refer Slide Time: 08:58)

$$\begin{aligned}
 &\Rightarrow \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \\
 &= \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\
 &= \text{tr} \left(\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \right) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\
 &= \sum_{i=1}^n \text{tr} \left((\mathbf{x}_i - \bar{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \right) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\
 &= \sum_{i=1}^n \text{tr} \left(\boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})' \right) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\
 &= \text{tr} \left(\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})' \right) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\
 &= \text{tr} \left(\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})' \right) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\
 &\text{where } (n-1)\boldsymbol{S} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})'
 \end{aligned}$$

So, we will have thus. So, this will imply that this summation i equal to 1 to n \mathbf{x}_i minus $\boldsymbol{\mu}$ transpose sigma inverse \mathbf{x}_i minus $\boldsymbol{\mu}$, this term is equal to summation i equal to 1 to n \mathbf{x}_i minus $\bar{\mathbf{x}}$ transpose sigma inverse \mathbf{x}_i minus $\bar{\mathbf{x}}$. So, this is basically that term of here, what we have this as a first term and the second term remains as it is and the third term vanishes. So, we will have that n times $\bar{\mathbf{x}}$ minus $\boldsymbol{\mu}$ transpose sigma inverse $\bar{\mathbf{x}}$ minus $\boldsymbol{\mu}$. Now, this is a scalar quantity and hence we can write this as trace of this term here summation i equal to 1 to n \mathbf{x}_i vector minus $\bar{\mathbf{x}}$ vector transpose sigma inverse into \mathbf{x}_i vector minus $\bar{\mathbf{x}}$ vector leave the second term as it is n times $\bar{\mathbf{x}}$ minus $\boldsymbol{\mu}$ transpose sigma inverse $\bar{\mathbf{x}}$ minus $\boldsymbol{\mu}$. Now, trace of $\mathbf{a} + \mathbf{b}$ is trace \mathbf{a} plus trace \mathbf{b} and hence we can take the trace inside.

So, we can write this as a summation i equal to 1 to n trace of this term which is inside the summation. So, this is \mathbf{x}_i vector minus $\bar{\mathbf{x}}$ vector transpose sigma inverse \mathbf{x}_i minus $\bar{\mathbf{x}}$ vector leave the second term as it is; so, $\bar{\mathbf{x}}$ minus $\boldsymbol{\mu}$ vector transpose sigma inverse $\bar{\mathbf{x}}$ minus $\boldsymbol{\mu}$ vector. Now, here we have trace of this particular quantity which is a scalar quantity. So, what we will do is we will use the result that trace of $\mathbf{a} \mathbf{b}$ equal to trace of $\mathbf{b} \mathbf{a}$, where we considered this to be a first term and this to be second term. So, trace of $\mathbf{a} \mathbf{b}$ would be trace of $\mathbf{b} \mathbf{a}$.

So, this would be given by summation i equal to 1 to n trace of sigma inverse \mathbf{x}_i vector minus $\bar{\mathbf{x}}$ vector that multiplied by this particular vector out here. So, that is \mathbf{x}_i minus $\bar{\mathbf{x}}$ transpose second term is left untouched sigma inverse $\bar{\mathbf{x}}$ minus $\boldsymbol{\mu}$. Now, this sigma inverse is something which is actually independent of this i index out

here, we will once again make use of the fact that trace of a plus b is trace of a plus trace of b and hence we can take this trace outside also. So, we will write this as trace of sigma inverse is independent of this i index and hence it can be taken outside the summation. So, what we have is trace of sigma inverse i equal to 1 to n x i vector minus x bar vector that multiplied by x i vector minus, this x bar vector transpose that plus the second term carried forward as it is. So, this is x bar vector transpose sigma inverse x bar minus this mu vector.

Now, what is this particular term, if we look at this particular term here carefully it is nothing but a constant multiplier of the sample variance covariance matrix. So, we can write that as trace of sigma inverse times n minus 1 times s that plus n times x bar vector minus this mu vector transpose sigma inverse x bar minus this mu vector, where we have used the notation that n minus 1 times, s is the sample variance covariance matrix with a divisor n minus 1. So, that; that is equal to summation i equal to 1 to n x i vector minus this x bar vector into x i vector minus x bar vector it is transpose. So, this term here which was actually in the exponent of the joint pdf this was actually if we look at this term here. So, the term which we have simplified and reduced it into that form is nothing but here sitting in the exponent. So, we will go to that term there. So, we have that term as trace of sigma inverse n minus 1 times s plus n times x bar minus mu transpose sigma inverse x bar minus mu. So, we will use this term.

(Refer Slide Time: 14:30)

$$\Rightarrow \text{j.t. p.d.f. of } X_1, \dots, X_n$$

$$(2\pi)^{-np/2} |S|^{-n/2} \exp\left[-\frac{(n-1)}{2} \text{tr } S^{-1} s - \frac{n}{2} (\bar{x} - \mu)' S^{-1} (\bar{x} - \mu)\right] \quad (2)$$

$$= f(x_1, \dots, x_n) g_{\mu, \Sigma}(\bar{x}, s) \quad \checkmark$$
 where,

$$f(x_1, \dots, x_n) = (2\pi)^{-np/2}$$

$$g_{\mu, \Sigma}(\bar{x}, s) = |S|^{-n/2} \exp\left[-\frac{(n-1)}{2} \text{tr } S^{-1} s - \frac{n}{2} (\bar{x} - \mu)' S^{-1} (\bar{x} - \mu)\right]$$
 By Neyman Fisher factorization, (\bar{x}, s) is jointly sufficient for (μ, Σ)

So, this would imply that the joint pdf of X_1, X_2, \dots, X_n the random sample of size n from that multivariate normal distribution with mean vector mu and covariance matrix sigma

is given by the following. That it is 2π to the power minus $n p$ by 2 determinant of σ to the power minus n by 2 , and then the exponent term is what we have done simplification on. So, that is n minus 1 by 2 trace of σ inverse s is that matrix, what we have defined in the last slide. So, this minus n by 2 \bar{x} vector minus this μ vector transpose σ inverse \bar{x} minus μ . Now, when the joint pdf of $X_1 X_2 \dots X_n$ is written in this particular form, it is easy now, to use the Neyman Fisher factorization theorem in order to derive what is the set of sufficient statistic, when we have both the parameters μ vector, and the covariance matrix σ to be unknown.

Now, this is written in this form, I will say what these are basically. So, I am writing it in this way $f(x_1, x_2, \dots, x_n)$, now this let me first complete this and then I will explain what these quantities are now, I am factorizing this particular joint pdf of $X_1 X_2 \dots X_n$ in the form that the first term is a function only of x_1, x_2, \dots, x_n . It does not include any parameter in this particular distribution the parameters are μ vector and the covariance matrix σ and the second part is written as a function it is a function of μ and σ and it is a function of x_1, x_2, \dots, x_n the random sample only through \bar{x} and s . So, this is a function which depends on $\mu, \sigma, x_1, x_2, \dots, x_n$ only through \bar{x} and s . So, what are these functions from the joint pdf that we have written here. So, that let me put a number here equation number 2.

So, we will have this particular quantity here what we have defined. So, this $f(x_1, x_2, \dots, x_n)$ we will have to look for the terms into which depend on x_1, x_2, \dots, x_n only. So, it cannot involve in a μ it cannot involve any element of σ . So, this as we see that it is a constant function which is 2π to the power minus $n p$ by 2 all other terms which are present here depends on μ and σ . So, it cannot be separated from there. And the second function that we have introduced $g(\mu, \sigma, \bar{x}, s)$ is basically the rest which is determinant of σ to the power minus n by 2 and then exponent minus n minus 1 by 2 trace of σ inverse s minus n by 2 \bar{x} minus μ transpose σ inverse \bar{x} minus this μ we close the bracket here. So, this is a term which depends on σ, σ is here and so it is here it depends on μ, μ is present here and it depends on x_1, x_2, \dots, x_n the random sample only through the two statistic which is \bar{x} vector the sample mean vector and s which is a sample variance covariance matrix.

And hence if we use this particular factorization it is in the form of the Neyman Fisher factorization and hence by Neyman Fisher factorization theorem by Neyman Fisher factorization. We have \bar{X} the random vector \bar{X} the sample mean random variable

vector and s is jointly sufficient for the unknown μ and σ . So, this basically is a set of jointly sufficient statistic for the unknown parameters the mean vector μ and the covariance matrix σ . So, if this is a set of sufficient statistic as in any other statistical inference problem instead of having carrying $X_1 X_2 \dots X_n$ the entire set of random samples we can carry this \bar{X} vector and this S the sample variance covariance matrix for all the inference procedure concerning this μ vector and covariance matrix σ . So, this gives us the desired compression of the data without loss of information. So, this contains all the information that is present in $X_1 X_2 \dots X_n$ about the parameter vector μ and the matrix σ the population variance covariance matrix. So, this will be the jointly sufficient statistic, if we have both these two set of parameters rather to be unknown. Now, we move on to the second possibility where σ is known to us.

(Refer Slide Time: 20:37)

Case 2 : Σ is known and μ is unknown

$j \in \{1, 2, \dots, n\}$

$$(2\pi)^{-n/2} |\Sigma|^{-n/2} \exp\left[-\frac{n-1}{2} \text{tr} \Sigma^{-1} S - \frac{n}{2} (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)\right]$$

$$= f^1(x_1, \dots, x_n) g_{\mu}(\bar{x})$$

where $f^1(x_1, \dots, x_n) = (2\pi)^{-n/2} |\Sigma|^{-n/2} \exp\left[-\frac{n-1}{2} \text{tr} \Sigma^{-1} S\right]$

$$g_{\mu}(\bar{x}) = \exp\left[-\frac{n}{2} (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)\right]$$

By Neyman Fisher factorization, \bar{x} is sufficient for μ .

So, we look at case 2 σ is known and this μ vector is unknown. So, it is a theoretical possibility actually for all practical purposes what we usually have is both the mean vector μ and the covariance matrix σ to be unknown, but it is a theoretical possibility that we have σ matrix to be known, and the μ vector to be unknown. Then if this is the only parameter set present in this population we will try to compress the data. So, that it would contain every information that is present in the data about this vector μ . Now, how are we going to do that, now note that we have this as a joint pdf of $X_1 X_2 \dots X_n$. So, we will use this particular form itself and then infer then use the Neyman Fisher factorization in order to get to this official statistic and what is that, the

joint pdf of X_1, X_2, \dots, X_n is once again what is going to be used in order to find the sufficient statistic.

So, this as we have seen is $(2\pi)^{-np/2} |\Sigma|^{-1/2} e^{-\frac{1}{2} (\bar{x} - \mu)^T \Sigma^{-1} (\bar{x} - \mu)}$. Now, if this is this we write it as another function $f_1(x_1, x_2, \dots, x_n)$. So, we are trying to reduce it to the form of the Neyman Fisher factorization. Now, the only unknown parameter in this case is the mean vector. So, we are trying to find out what is that particular function which depends on μ and also depends on x_1, x_2, \dots, x_n through some statistic. Now, we will see that it can be reduced into this particular form, where the first function $f_1(x_1, x_2, \dots, x_n)$ is a function of x_1, x_2, \dots, x_n .

So, that is basically a known function when we have x_1, x_2, \dots, x_n to be given. So, this now takes the form that it is $(2\pi)^{-np/2} |\Sigma|^{-1/2}$. Now, note that Σ is now unknown to us, we have assume then case 2 that Σ is known and hence this determinant of Σ is also known to us. So, this also is a function which is known and if we look at in the exponent the first term is $\frac{1}{2} \text{trace of } \Sigma^{-1} s$. Now, Σ is known to us and so will be Σ^{-1} and s is a matrix which is based on x_1, x_2, \dots, x_n only. So, this also thus would be absorbed in the first function. So, this is exponent to the power minus n by 2 trace of $\Sigma^{-1} s$.

So, this entire function here thus is known to us if x_1, x_2, \dots, x_n is known to us and this μ vector \bar{x} . So, this is $e^{-\frac{1}{2} (\bar{x} - \mu)^T \Sigma^{-1} (\bar{x} - \mu)}$ thus the rest of it which is $e^{-\frac{1}{2} (\bar{x} - \mu)^T \Sigma^{-1} (\bar{x} - \mu)}$. So, what is the characteristics of this particular function, the second function it depends on μ well that is clear and it depends on x_1, x_2, \dots, x_n through this \bar{x} vector. Note that once again Σ is known to us and so will be Σ^{-1} , this is a function which depends on the unknown parameter vector μ and its dependence on x_1, x_2, \dots, x_n . Now, is compressed through it is compressed in terms of this vector \bar{x} .

So, this in this particular form the factorization once again will be using the Neyman Fisher factorization by Neyman Fisher factorization. The Neyman Fisher factorization theorem will tell us that, this \bar{x} vector random variable. So, this \bar{x} vector is now, sufficient for the unknown mean vector. So, if both μ and Σ are unknown then \bar{x} and s is jointly sufficient for μ and Σ if Σ the population variance

covariance matrix is known, then \bar{x} is sufficient for μ . Now, we look at the third possibility which once again is a theoretical possibility.

(Refer Slide Time: 25:54)

Case 3 : Suppose μ is known ($= \mu_0$) and Σ is unknown

$$\begin{aligned}
 \text{Joint p.d.f. of } x_1, \dots, x_n &= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \mu_0)' \Sigma^{-1} (x_i - \mu_0)\right) \\
 &= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \text{tr} \sum_{i=1}^n (x_i - \mu_0)' \Sigma^{-1} (x_i - \mu_0)\right) \\
 &= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \text{tr} \sum_{i=1}^n (x_i - \mu_0)(x_i - \mu_0)'\right) \\
 &= f^2(x_1, \dots, x_n) g_{\Sigma}(t)
 \end{aligned}$$

where $f^2(x_1, \dots, x_n) = (2\pi)^{-np/2}$
 $g_{\Sigma}(t) = |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \text{tr} \sum_{i=1}^n t(x_i - \mu_0)'\right)$

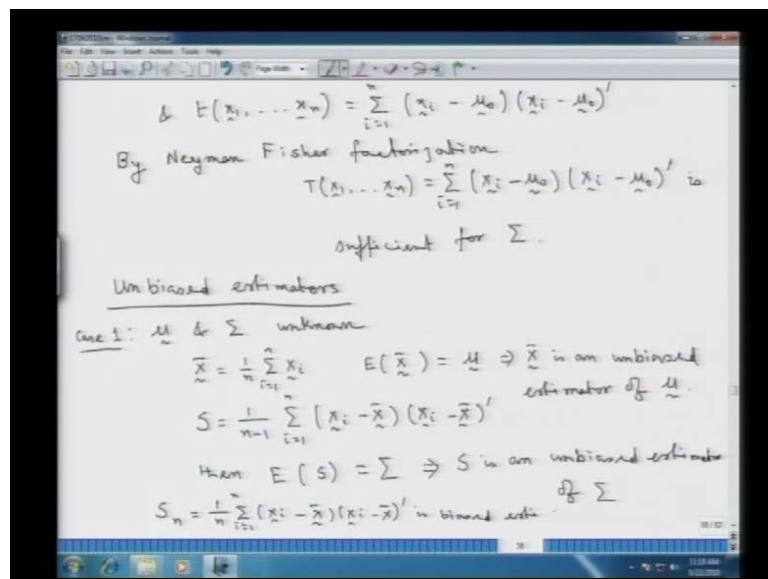
Suppose in the third possibility μ is known say equal to μ_0 vector and the covariance matrix σ is unknown. So, this is the set of ... then we will try to find out what is the sufficient statistic corresponding to the unknown this σ matrix. Now, the joint pdf we do not need to look at the form of the joint pdf that we had derived, we will take the original form the joint pdf of X_1, X_2, \dots, X_n is $(2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \mu_0)' \Sigma^{-1} (x_i - \mu_0)\right)$, because we have assumed that the mean vector μ is known to us and that is equal to μ_0 . So, we plug in this particular value out here $\Sigma^{-1} \sum_{i=1}^n (x_i - \mu_0)'$.

Now, if we have this particular form we are trying to derive remember, we are trying to derive the sufficient statistic under the condition that μ is equal to μ_0 is known to us and σ is unknown. So, we will once again have to use the Neyman Fisher factorization theorem. So, let us write these terms as it is and then we will write this exponent term as now, note that this particular term is scalar quantity and hence we can write trace of that particular quantity. So, it is half trace of $\sum_{i=1}^n (x_i - \mu_0)' \Sigma^{-1} (x_i - \mu_0)$. So, once again we will use this trace of $A + B = \text{tr} A + \text{tr} B$ and $\text{tr} AB = \text{tr} BA$. So, using the two trace results, what we can have is this quantity exponent.

So, we will take first the trace inside and then we will consider this to be one term this to be one term and along with sigma inverse this to be the second term and then flip these two terms are trace of a b will be equal to trace of b a. That this... thus will reduce to half trace of sigma inverse times summation i equal to 1 to n x i minus mu naught into x i minus mu naught this transpose. So, we have taken this trace sigma inverse outside. So, this once again we will write it as a function say f 2 x 1 x 2 x n now, our objective is to use the Neyman Fisher factorization theorem and this the unknown quantity is sigma now, so we will have to write it in terms of sigma and some statistic here. Let us write this as some t we will have to define what it t where the first function f 2 x 1 x 2 x n.

Now, would be a function of x 1 x 2 x n and only known quantities what is that, if we look at this term only this is the term which satisfies that particular criterion. So, this is 2 pi to the power minus n p by 2 and this g sigma t is basically determinant of sigma to the power minus n by 2 and e to the power minus half trace of sigma inverse I denote this entire quantity what we have here as this t matrix. Now, that t matrix is a function of x 1 x 2 x n. So, this is t of x 1 x 2 x n only, because mu naught is known to us. So, that is a known quantity.

(Refer Slide Time: 30:39)



And this t x 1 x 2 x n is nothing but this particular matrix which is summation i equal to 1 to n x i vector minus, this mu naught vector that multiplied by this x i vector minus mu naught vector its transpose. So, see what we have obtained, we have obtained the following that this second factor is a function of the parameter matrix sigma, sigma is appearing here, and so it is present here and the term and the quantity which is... which

we have denoted by T $\times 1 \times 2 \times n$ it depends on $x_1 \times 2 \times n$ through this particular matrix. So, that would imply by Neyman Fisher factorization, once again by Neyman Fisher factorization this $T X_1 X_2 X_n$ which is summation i equal to 1 to n x_i minus μ naught x_i minus μ naught transpose is sufficient for the unknown parameter matrix which is sigma matrix.

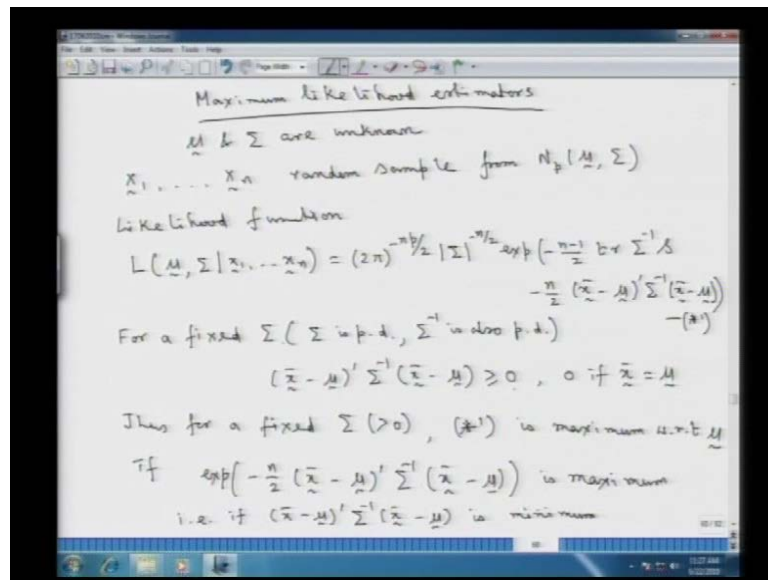
So, this is all the three possible cases when we talk about sufficient see that is going to be derived on the basis of the random sample $x_1 \times 2 \times n$. The first case we had seen that that is the most general case, that is the most practical case actually that both the mean vector μ and the covariance matrix sigma are unknown to us and in such a situation \bar{x} vector and the sample variance covariance matrix s is jointly sufficient for μ and sigma in the second case we had assumed that sigma is known to us μ is unknown. So, if that is the situation then the \bar{x} vector is going to be sufficient for the unknown mean vector μ and in the third case we had the mean vector μ to be known to us say equal to μ naught and in such a situation this matrix T what we have this statistic actually is sufficient for the sigma matrix. That is all about this sufficiency possibilities when we are considering a multivariate normal distribution. Now, second point of interest that one usually is interested in is what is this unbiased estimator corresponding to this unknown quantities. So, let us answer these question unbiased estimators.

Now, we have this μ and sigma are both unknown. So, if both this mean vector μ and the covariance matrix sigma are unknown. So, this is corresponding to the case 1, what we had considered for sufficient statistic. In such a situation we have already derived for a general case of a multi random sampling for the multivariate distribution not necessarily a multivariate normal distribution that, if we have this \bar{X} vector is a sample mean the random vector. So, that is this i equal to 1 to n from here, we have seen earlier that expectation of this \bar{X} vector is the population vector μ . So, this would imply that \bar{X} vector is an unbiased estimator of the population mean vector which is μ . Now, what about the sigma matrix, we have once again in the earlier lectures I had proved that if we consider S to be equal to $\frac{1}{n-1}$ $\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$.

So, this is the sample variance covariance matrix with a divisor $n-1$, then we have already proved this result that expectation of this S matrix is nothing but this sigma matrix that is for a general case that we had proved not necessarily for a multivariate normal distribution the result still holds true, if we have the underline distribution to be a

multivariate normal distribution. This would imply that S is an unbiased estimator of the covariance matrix which is Σ . So, S is going to be an unbiased estimator of Σ and \bar{X} is going to be an unbiased estimator for μ . So, we will have the two, if on the other hand we consider S the sample variance covariance matrix with a divisor n , then that is definitely not going to remain an unbiased estimator. So, that is going to be a biased estimator for the small sample however, for the large sample it is going to be unbiased in the limit. So, this \bar{X} is an unbiased estimator of the unknown mean vector μ and S with a divisor $n - 1$ is an unbiased estimator of Σ on the other hand, if we have defined S_n to be $1/n$ of the same quantity as, what we had considered with this X_i , this is what I just refer to that this quantity is biased estimator.

(Refer Slide Time: 37:03)



However, this S_n is unbiased in the limit. So, these are the two unbiased estimator for more general case that μ and Σ are unknown. The unbiased estimators in the other two situations were say Σ matrix is known to us or μ vector is known to us can similarly be derived. Now, let us move on to another important estimation procedure which is the maximum likelihood estimator. So, we will be deriving a corresponding to this unknown quantities μ and Σ , we take the most general case, where μ and Σ are unknown. The other cases of μ being known or Σ being known can be tackled in a similar way. So, suppose μ and Σ are unknown, we have still this X_1, \dots, X_n random sample from that population $N_p(\mu, \Sigma)$, then we will look at the likelihood function.

So, the Likelihood function is $L(\mu, \Sigma)$ given the observations x_1, x_2, \dots, x_n and that is nothing but the joint pdf actually looked upon as a function of μ and Σ given x_1, x_2, \dots, x_n and in order to find out the maximum likelihood estimators, we will try to see what μ and Σ basically maximizes this particular likelihood function. So, as we had seen earlier the joint pdf which we are now taking as the Likelihood function that is given by the following $(2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)\right\}$. So, this is now, the Likelihood function of μ and Σ given x_1, x_2, \dots, x_n . Now, we note the following that for a fixed Σ positive definite, we will have the quantity which is here, now we are at the first step fixed this Σ to any particular matrix any positive definite matrix and then we will try to see what is the value of μ which would maximize this particular likelihood with respect to that particular new vector. Now, Σ is positive definite, I will just extend at this particular bracket little bit Σ is positive definite Σ^{-1} is also positive definite.

So, for that fixed Σ positive definite matrix Σ^{-1} is also positive definite. So, what will be having is for that fixed Σ $(\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)$, this would be greater than or equal to 0 why, because Σ^{-1} is positive definite actually that would be greater than 0, if we look at that. So, this is greater than equal to 0, an 0 if \bar{x} is equal to μ only, because Σ is positive definite. So, this Σ^{-1} being positive definite will imply that this quadratic form will be equal to 0 only, if we have \bar{x} to be equal to μ that is this vector that we have taken here is just a null vector. So, this would be 0 if we have \bar{x} to be equal to 0.

Now, let me give an equation number star here star 1 say, now for this fixed Σ if we have this quantity, we can say that thus for a fixed Σ positive definite star 1 is maximum with respect to this mean vector μ if we have, now note that in this likelihood function here, μ is present only in the exponent. So, if we are trying to maximize the likelihood function star 1 with respect to μ for a fixed value of the Σ matrix, then this is going to be the slightly hood function is going to be maximized, if the exponent is maximized. Now, in the exponent also you will see that this is a quantity which does not dependent on μ and hence if we are trying to maximize the exponent with respect to μ , then it is maximization corresponding to this particular path of the exponent. So, this for a fixed Σ star 1 is maximum with respect to μ , if e to the

power minus n by 1 \bar{x} minus μ transpose Σ inverse \bar{x} minus μ is maximum. So, we will be looking at, what is the value of that μ which would maximize this particular quantity. Now, thus if we look at this exponent here, this exponent is going to be maximized, if this part what is sitting in the negative part of this exponent is minimized.

That is ; if this \bar{x} minus μ transpose Σ inverse \bar{x} minus μ is this is... if that is minimum if this part is. So, we are trying to find out what is that value of μ which would maximize this particular quantity and that is equivalent to finding what is that value of μ which actually would minimize this particular quantity, because it has got negative sign out here and what is the value of μ that would minimize this particular quantity. We have seen that for a fixed Σ this quantity is greater than or equal to 0 . So, this quantity here is going to be minimize, if we choose μ to be \bar{x} . So, \bar{x} vector is the one that is going to minimize this particular quantity and hence that is what is going to maximize this exponent part here, and that would maximize the likelihood.

(Refer Slide Time: 44:52)

$\Rightarrow \hat{\mu} = \bar{x}$ would maximize $L(\mu^*)$ w.r.t. μ for a fixed Σ
 \Rightarrow MLE of μ is $\hat{\mu}_{MLE} = \bar{x}$
 Likelihood f^n at $\mu = \bar{x}$
 $L(\hat{\mu}, \Sigma | x_1, \dots, x_n) = (2\pi)^{-n/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} (\bar{x} - \hat{\mu})' \Sigma^{-1} (\bar{x} - \hat{\mu})\right)$
 $= (2\pi)^{-n/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} (\bar{x} - \hat{\mu})' A (\bar{x} - \hat{\mu})\right)$
 where $A = (n-1)S = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$
 Maximization of $L(\mu^*)$ is equivalent to maximization of $\log L(\mu^*)$.
 $i.e. \propto \frac{n}{2} \log |A| + \frac{n}{2} \log |\Sigma^{-1}| - \frac{1}{2} (\bar{x} - \hat{\mu})' A (\bar{x} - \hat{\mu})$
 $i.e. \propto \frac{n}{2} \log |\Sigma^{-1}| - \frac{1}{2} (\bar{x} - \hat{\mu})' A (\bar{x} - \hat{\mu})$

So, that what we can thus say is that, this would imply that this μ hat equal to \bar{x} would maximize this star 1 likelihood with respect to this mean vector μ for a fixed Σ matrix. However, we see that this maximizing value of μ does not depend on the particular level of Σ , where we have fixed and hence this μ hat equal to \bar{x} is going to maximize the likelihood with respect to μ for every possible values of Σ . So, that this would imply that the maximum likelihood estimator of μ is μ hat MLE is equal to the sample mean vector \bar{X} which is also incidentally the unbiased

estimator. So, we have got this to be the maximum likelihood estimator of the mean vector μ . Let us now, try to see what is the maximum likelihood estimator of σ in order to derive that let us now, look at the likelihood function at μ equal to \bar{x} .

So, the likelihood function at μ equal to \bar{x} is... let us denote that by this as $\hat{\mu}$ $\sigma^{-1} \times 1 \times 2 \times n$. So, that it is equal to 2π to the power minus n by 2 determinant of σ to the power minus n by 2 and then we have in the exponent. Now, what would that look like, let us look at the likelihood function. So, this is the likelihood function star 1, where we are looking at the likelihood function at the maximum likelihood point $\hat{\mu}$ equal to \bar{x} . So, here this \bar{x} and $\hat{\mu}$, what we will be having will get cancelled out. So, μ will be replaced by $\hat{\mu}$ and that is equal to \bar{x} . So, this would be a null vector.

So, there will not be any contribution coming from here. So, we will have the exponent to be given by minus n minus 1 by 2 trace of $\sigma^{-1} S$. Now, let us write this term as 2π to the power minus n by 2 for convenience, we write it in the μ notation as minus half trace of $\sigma^{-1} A$ say what is that matrix A , where this A is equal to n minus 1 times this S matrix. So, that is nothing but that particular product i equal to 1 to n x_i minus \bar{x} into x transpose. So, this is now, the form of the likelihood which we denote by say star 2 this is the star 2 from of the likelihood at μ equal to \bar{x} the maximize likelihood point.

Now, maximization of star 2 is equivalent to maximization of log of this likelihood maximization of star 2 is equivalent to maximization of log of star 2 natural logarithm that is of minus n by 2 log of 2π then we take σ^{-1} . So, we will take this as positive sign n by 2 log of determinant of σ^{-1} . So, we what we have done is, we have kept this σ^{-1} and taken the positive sign n by 2 and taken the log of that. So, it is plus n by 2 log of determinant of σ^{-1} , this minus half trace of $\sigma^{-1} A$. Now, note that this particular quantity is constant and hence there is nothing to maximize here. So, that is of n by 2 log determinant of σ^{-1} minus half trace of this $\sigma^{-1} A$.

(Refer Slide Time: 50:00)

$$\text{i.e. } \frac{n}{2} \log |\Sigma^{-1} A| - \frac{1}{2} \text{tr} \Sigma^{-1} A - \frac{n}{2} \log |A|$$

$$\text{i.e. } \frac{n}{2} \log |\Sigma^{-1} A| - \frac{1}{2} \text{tr} \Sigma^{-1} A \quad \text{--- } (*^3)$$
 Let λ_j denote the eigen values of $\Sigma^{-1} A$ (λ_j is also $(j=1, \dots, p)$)
 The eigen values of $A^{1/2} \Sigma^{-1} A^{1/2}$

$$(*^3) = \frac{n}{2} \log \prod_{j=1}^p \lambda_j - \frac{1}{2} \sum_{j=1}^p \lambda_j$$

$$= \frac{n}{2} \sum_{j=1}^p \log \lambda_j - \frac{1}{2} \sum_{j=1}^p \lambda_j$$

$$= \frac{1}{2} \sum_{j=1}^p (n \log \lambda_j - \lambda_j) \quad \text{--- } (*^4)$$

$$(*^4) \text{ is maximised if } \lambda_j = n \quad \forall j = 1, \dots, p$$

Now, we will do a bit of adjustment here that is of $n/2 \log$ determinant of sigma inverse A. So, what we have done is, we have introduced an A in order to have a similar expression as to that, what we have here on the trace. So, A is a matrix which does not depend on sigma that is not going to matter much. So, this is the second term which is minus half trace of sigma inverse A. So, this is minus half trace of sigma inverse A then we have actually taken this A matrix here. So, we will have to make adjustment for that. So, determinant of this matrix will be determinant of this plus determinant of this into determinant of A.

So, that we will require a term which is $n/2 \log$ of determinant of A in order to adjust that particular term. Now, once again see that this particular term is independent of sigma and if we are interested in finding the sigma which would maximize the likelihood star 2 that is not going to depend on this particular function, because this is independent of sigma. So, that is of $n/2 \log$ determinant of sigma inverse A minus half trace of sigma inverse A. So, in order to find out thus the maximum likelihood estimator of sigma, we can maximize this particular quantity, what we have here with respect to sigma and that would lead us to maximum likelihood estimator. Now, let us denote by λ_j let λ_j denote the eigenvalues of this matrix sigma inverse A of this sigma inverse A now, this j of course, what is the dimension of this sigma inverse A sigma was coming from the p-dimensional multivariate normal distribution. So, that is p by p and A is a constant multiplier of the sample variance covariance matrix that is also a p by p matrix and hence the order is p by p sigma is positive definite and they probably t 1 A

will also be positive definite and hence this matrix entirely is positive definite and thus this j is from 1 to upto p .

So, suppose λ_j denote the eigenvalue of $\Sigma^{-1}A$ you may also note the following that, this since λ_j s are the eigenvalues of the $\Sigma^{-1}A$ λ_j is also the eigen value of $A^{-1/2}\Sigma^{-1}A^{-1/2}$ matrix that is easy to see, because if you look at the eigen value equation then this would lead us to this. Now, we will replace this let us denote this by $\star 3$. So, this is that equation now, if λ_j s are the eigen values j equal to 1 to p of this $\Sigma^{-1}A$ matrix then this $\star 3$ is equal to n by 2 \log of now, determinant of this particular matrix would be product of the respective eigenvalues. So, it is product this is for j equal to 1 to upto p this minus half. Now, trace of $\Sigma^{-1}A$ would be just sum of these eigenvalues. So, this is λ_j . So, this is equal to n by 2 now, \log of this product is summation of this \log of λ_j terms this minus half summation λ_j .

So, this we can write this these summations are j equal to 1 upto p . So, this is equal to half summation j equal to 1 to p n times $\log \lambda_j$ this minus. So, we have taken this half outside there. So, what remains is this λ_j . So, in order to maximize the likelihood, we can maximize this particular quantity with respect to what now with respect to these λ_j quantities and what is that λ_j with should maximize this likelihood, if each of these λ_j s are equal to n . So, this is $\star 4$. So, $\star 4$ is maximized if λ_j is equal to n , for all j equal to 1 to upto p . So, that is basically the λ_j , now λ_j actually is a the crucial term λ_j is what is now, being involve which is being associated with this unknown, covariance matrix Σ and that λ_j being equal to n , that is λ_j being equal to n for j equal to 1 to p is what is going to give us lead us to rather than maximum likelihood estimator.

(Refer Slide Time: 55:45)

Handwritten mathematical derivations on a whiteboard:

$$\Rightarrow A^{-1} = \sum^{-1} A^{-1/2} = P^{-1} \Lambda^{-1} P^{-1} = n I_p$$

i.e. $\Sigma^{-1} = n A^{-1}$

i.e. $\Sigma = \frac{1}{n} A$ would maximize the likelihood of Σ

$$\Rightarrow \hat{\Sigma}_{MLE} = \frac{1}{n} A = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' = S_n$$

$\hat{\bar{X}}$ is MLE of μ

S_n is MLE of Σ

This would imply, if all lambda js are equal to 1 or lambda j is equal to n, then this would imply that this A half sigma inverse A half that is equal to P and say, what these terms are d lambda n times I p P prime, where P is now holding the auto normalized eigen vectors corresponding to this particular matrix. So, this is what the factorization, what we have now these are orthogonal matrices. So, what we have that is n times I p. Now, using this particular result, it is simple to get to the maximum likelihood estimators. So, we have this particular expression, if the eigen values the maximizing eigen values are all n then this A half sigma inverse A half which we had actually seen here that lambda js are also eigen values of this A half sigma inverse A half matrix and hence this in terms of its spectral decomposition, we will have this P D lambda which is this matrix, now P prime and that is equal to n times I p that is sigma inverse is going to be equal to ... Now this is A half A half. So, pre and post multiplying by A minus half, what will be having here is A inverse term here.

And thus this would imply that this sigma is going to be 1 upon n A matrix would maximize the likelihood function with respect to this sigma matrix and what is that. So, this would imply further that this sigma hat maximum likelihood estimator is going to be 1 upon n A, and if you remember what A is that cross product matrix. So, that is X i minus X bar X i minus this X bar transpose i equal to 1 to n. So, this thus is the maximum likelihood estimator which in our previous notation is nothing but this S n. So, it is a sample variance covariance matrix with a divisor n is ... thus the maximum likelihood estimator. So, thus we have proved the following that X bar vector is the

maximum likelihood estimator of this μ vector and S_n this matrix with the sample variance covariance matrix with a divisor n is the maximum likelihood estimator of the sigma matrix.

So, these two thus \bar{X} vector and S_n are the two maximum likelihood estimators. Now, this also reminds us about the invariant normal maximum likelihood estimator, where \bar{X} was also the maximum likelihood estimator and the sample variance with a divisor n was the maximum likelihood estimator of the sigma square quantity corresponding to any invariant normal distribution. It is just to be noted that this \bar{X} as we had in a seen earlier is an unbiased estimator of μ on the other hand, S_n is not an unbiased estimator of the sigma variance covariance matrix. However, it is going to be unbiased in the limit. In the next lecture we will try to derive the distributions of these maximum likelihood estimators \bar{X} S_n or S_{n-1} .