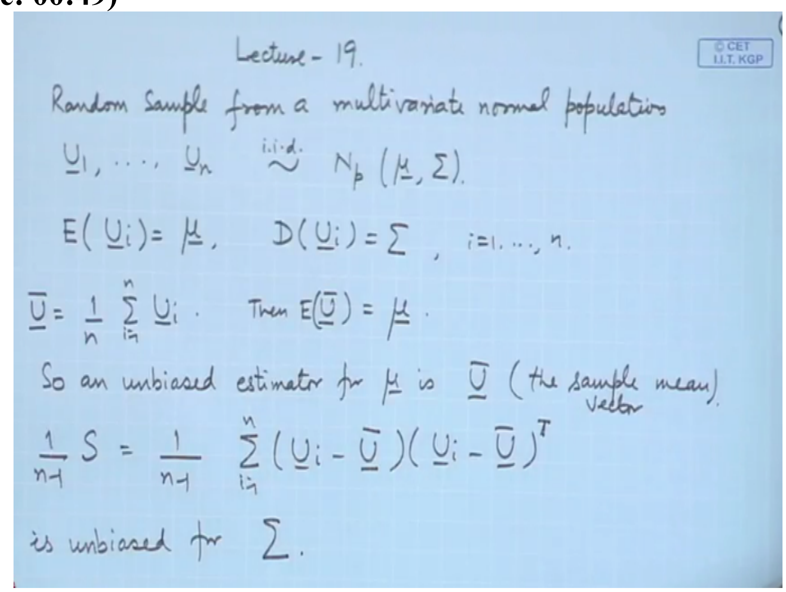
Statistical Methods for Scientists and Engineers

Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Science – Kharagpur

Lecture - 19
Multivariate Analysis – IV

So we have considered the density of the Multivariate Normal Distribution in the previous class. Now if we are doing random sampling from Multivariate Normal Distribution then we want to do the estimation of parameters or we want to do the test on the parameters, so in general inferences on the parameters of a multivariate normal distribution. Firstly, I will discuss the part here.

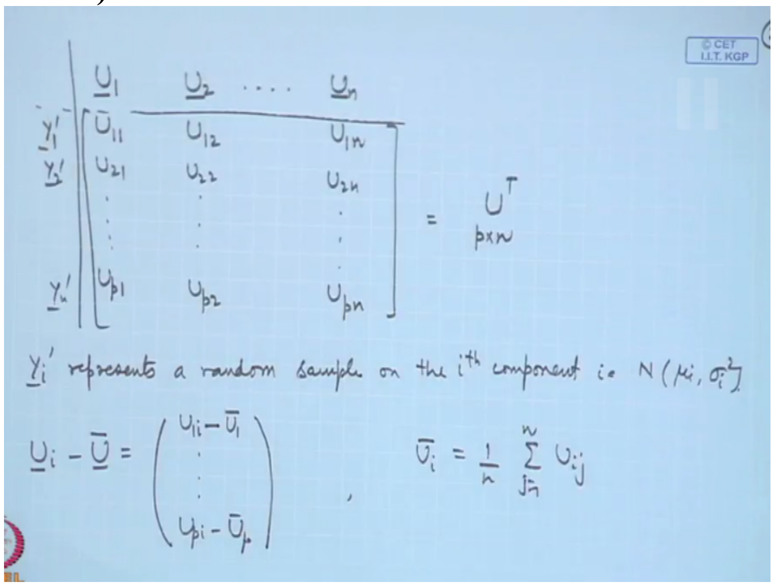
(Refer Slide Time: 00:49)



So let us consider say Random Sample from a Multivariate normal population. So we can consider say U1, U2, Un let me use this notation so these are independent and identically distributed and Np Mu, sigma random variables, so that means these are the observations from a multivariate normal distribution. So we have basically expectation of each Ui that = Mu and the dispersion matrix of Ui = sigma for i=1 to n.

So clearly we can see that if I define U bar = 1/n sigma Ui i=1 to n, then expectation of U bar that will be = Mu. So an unbiased estimator for Mu is the sample mean, sample mean vector. Similarly, we consider, say 1/n-1 S that = 1/n-1 sigma Ui-U bar*Ui-U bar transpose i=1 to n. This is unbiased for Sigma. Let me give the interpretation of this year.

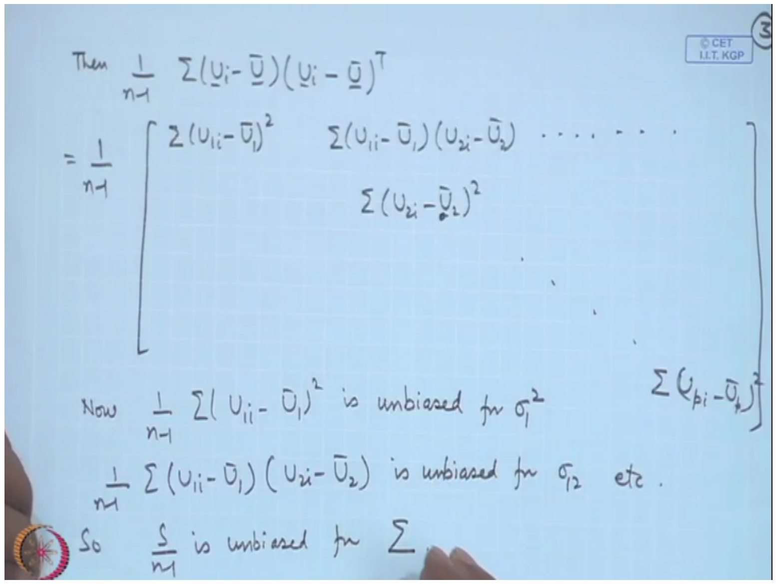
(Refer Slide Time: 03:04)



Let us consider the sample in this fashion, U1, U2, Un, I will write it in this fashion. The components of this is U11, U21 and so on Up1. Similarly, the components of U2 are U12, U22, Up2, and components of Un are U1n, U2n and so on Upn. Let us consider say – I consider this row vector which say Y1 prime, Y2 prime, Yn prime.

And this entire matrix we can use the notation say U transpose which is of order p/n. So we can consider say for example Yi prime represents a random sample on the ith component that is n Mu i sigma i square. Now let us also use the notation Ui–U bar that is equal to U1i–U1 bar and so on, Upi-Up bar. Here individual Ui bars are denoting 1/n sigma Uij j=1 to n. Therefore, this term, that is 1/n-1 sigma Ui-U bar*2Ui-U bar transpose that will represent.

(Refer Slide Time: 05:12)



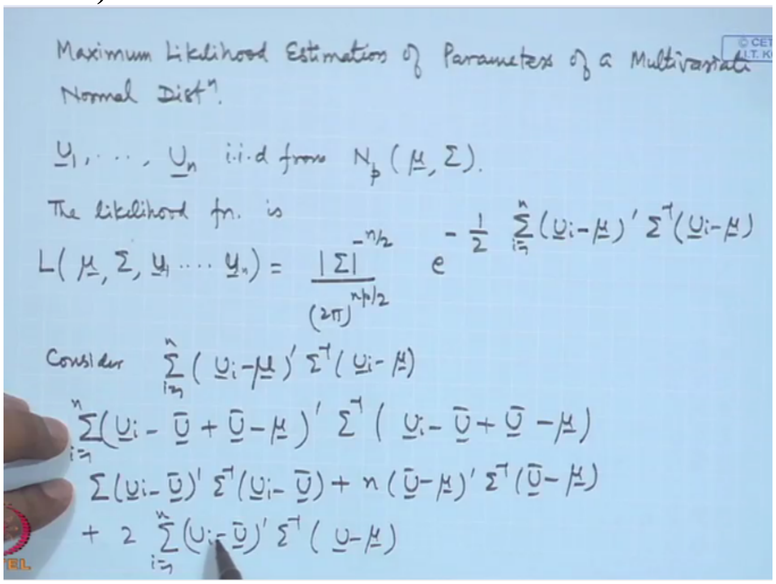
So we can write it here 1/n-1 sigma Ui–U bar*Ui–U bar transpose this will represent 1/n-1, the first component will be sigma U1i–U1 bar, from here we can write here because if I am considering this multiplied by the transpose of this then the first term will become sigma U1i–U1 bar square similarly, in the second diagonal it will become sigma U2i-U2 bar square and in the half diagonal it will become sigma U1i–U1 bar*U2i–U2 bar etcetera.

Here it will be sigma U2i–U2 bar square and so on; sigma Upi–Up bar is square. Now you can see that 1/n-1 sigma U1i–U1 bar square, this is unbiased for sigma 1 square and so on. 1/n-1 sigma U1i–U1 bar*U2i–U2 bar, this is unbiased for sigma12 etcetera. So S/n-1 is unbiased for. So we are able to consider the unbiased estimation for Mu and sigma. We had the concept of minimum variance unbiased estimation in the case of a scalar parameter.

Since we are dealing with the vector parameter here, that concept is no longer is valid here of course we can consider component wise minimum variance unbiased estimation here. Now in the case of one variable we have seen the like for example in normal Mu sigma square we have also looked at the maximum likelihood estimators.

In the one dimensional case the maximum likelihood estimator for Mu was the sample mean and for the sigma square it was 1/n sigma x-x bar square. So here we can consider analog to that and we will get, here we have to find S, so for the variance covariance matrix sigma we will get x/n and for Mu we will get U bar. Let us prove this here.

(Refer Slide Time: 08:18)

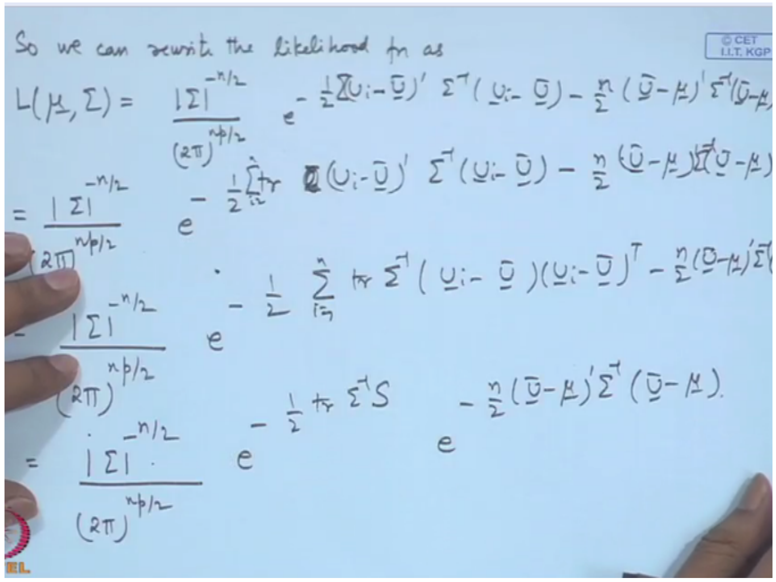


So Maximum Likelihood Estimation of Parameters of a Multivariate Normal Distribution. So as before we have U1, U2, Un is a random sample from normal Np Mu sigma. Now let us go back to the density function of Ui. **"Professor to Student conversation starts"** Yesterday we have seen the density function of Multivariate Normal Distribution when the rank of sigma is full it is given by 1/2 to the power 9/2 determinant of sigma to the power half, e to the power – half x-Mu prime sigma inverse, x-Mu. Now we write this density for U1, U2, Un. **"Professor to Student conversation ends"**

So the likelihood function that is we write L Mu sigma and then of course your U1, U2, Un. I continue using capital letters here just for convenience so it will become determinant sigma to the power–n/2, 2 pi to the power np/2, e to the power-1/2 sigma Ui–Mu prime sigma inverse Ui–Mu, i=1 to n. So firstly let us simplify this expression, sigma Ui–Mu transpose sigma inverse Ui–Mu.

Here we add and subtract sample mean vector so this becomes sigma i=1 to n, Ui–U bar +U bar–Mu prime sigma inverse and we expand this so sigma Ui–U bar prime sigma inverse Ui–U bar+n times U bar–Mu prime sigma inverse U bar–Mu + twice sigma Ui–U bar prime sigma inverse U bar–Mu i=1 to n. Now if I consider here summation and apply on this I get U bar–U bar, so this term is actually 0; this term vanishes. So we are getting only this part here.

(Refer Slide time: 11:54)



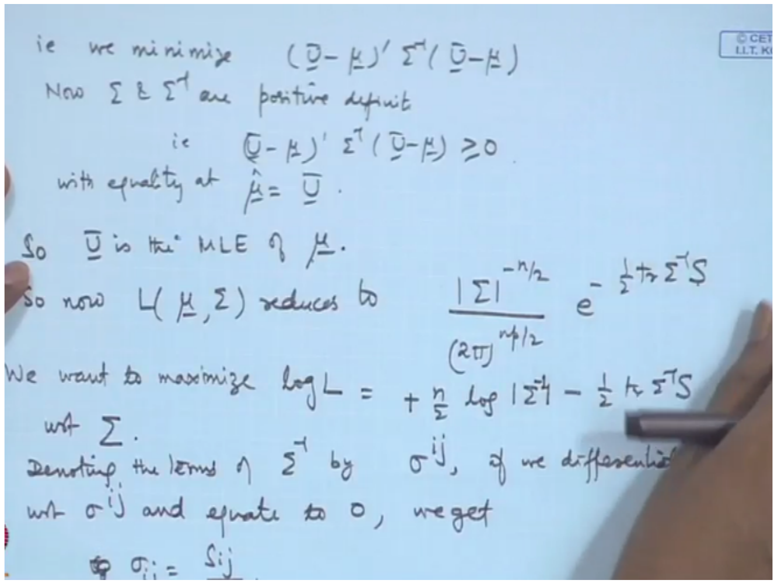
So we can rewrite the likelihood function as L Mu sigma that = determinant of sigma to the power–n/2/2 pi power Np/2, e to the power–1/2. Now if you look at this term, this is actually scalar, so if it is a scalar term I can also write it as trace of this, now trace of this can also be written as trace of this I interchange the order here, I multiply it on the side. So let me write it here = sigma Ui– sorry.

This is sigma Ui-U bar prime sigma inverse Ui-U bar-half n U bar-Mu bar prime sigma inverse U bar – Mu bar. Okay. Now this term I write as e to the power-half trace of sigma Ui-U bar prime, sigma inverse Ui-U bar. Now in the trace, so this will become summation here. I can take the summation outside. So this will become determinant of sigma to the power n/2 divided by 2 pi power Np/2, e to the power-1/2 trace of Sigma inverse Ui-U bar, Ui-U bar transpose-n/2 U bar-Mu sigma inverse U-Mu.

Now I take this summation sign inside then this will become summation here that is becoming S here, so this is determinant of sigma to the power n/2, 2 pi to the power Np/2, e to the power – half trace of sigma inverse S, e to the power–n/2 U bar-Mu prime sigma inverse U bar-Mu. Now we want to maximize this with respect to Mu. Let us consider firstly the maximization with respect to Mu, we first maximum with respect to Mu.

Now there is no Mu term appearing here so that means basically we have to minimize this term.

(Refer Slide Time: 15:19)

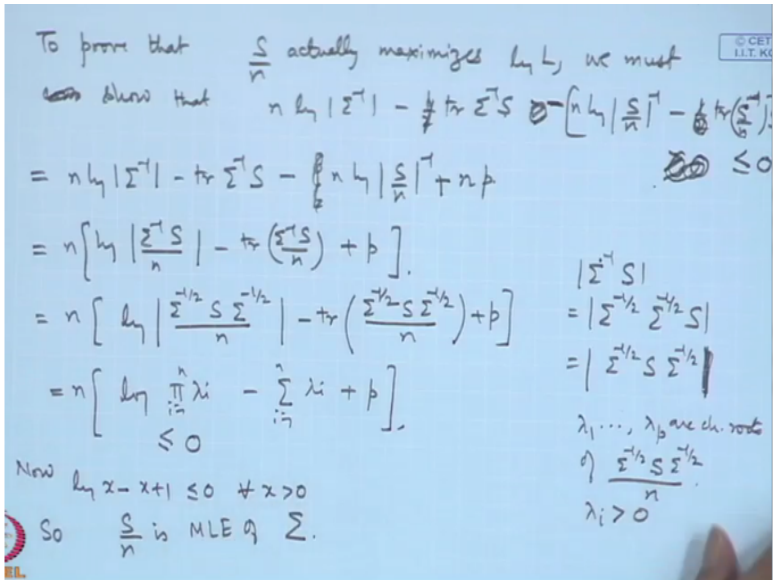


So i.e. we minimize U bar – Mu bar, Mu prime sigma inverse U bar–Mu. Now sigma and sigma inverse they are positive definite. So U bar-Mu prime sigma inverse U-Mu is always > or = 0 with equality at Mu had is equal to U bar. So U bar is the maximum likelihood estimator of Mu. So if I have reduced the second term in the likelihood function to 0 so my likelihood function is now reducing to this term alone.

So now let us consider the maximization of this with respect to sigma. So now, L Mu sigma reduces to determinant of sigma to the power-n/2/2 pi to the power Np/2, e to the power – half trace are sigma inverse S. And we want to maximize. So let us consider maximization of log of L that = –n/2 log of determinant of sigma-half trace of sigma inverse S. So this we want to maximize with respect to sigma.

So if I—here, actually you can see I can put + here then this will become -. So if considering terms of sigma inverse and we can denote, denoting the terms of sigma inverse by say sigma ij, if we differentiate log L with respect to sigma ij and equate to 0 we get sigma ij=Sij/n. Now in order to prove that this is maximum likelihood estimator we should show that actually this is maximizing, that means we must show that--

(Refer Slide Time: 18:24)



To prove that S/n actually maximizes log of L, we must consider we must show that n log determinant sigma inverse – half trace sigma inverse S is always > or = n log S/n inverse – half trace of S inverse n S. Or we can say that this difference should be > or = 0. Now this difference if you consider this is n log determinant of sigma inverse minus trace of half I can remove here because in this terms 2/2 was and here also divided by 2 was here, so I can take it out so this I can remove and this term also I can remove.

So this will become trace of sigma inverse S–n log, so now it is S/n inverse and here S inverse S will become i, so i of P dimension so trace will become = P, so the term will become P here - and there is a n in the denominator so it will become Np and this will become + here. Now this term we can write as n times log of sigma inverse S/n determinant, I am combining this term with this minus trace of sigma inverse S/n, I am taking out n here +P.

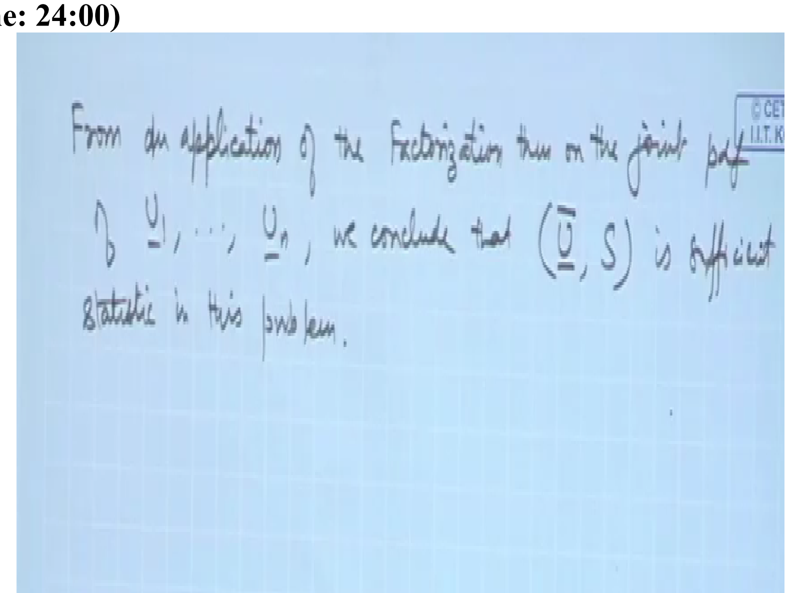
That we can write as n time log of, now here we do some manipulation. This sigma is actually sigma inverse S so we can write it as sigma to the power – half; sigma to the power – half S. Now determinant of AB = determinant of BA. Now these type of breakup was allowed provided we assume sigma to be a positive definite matrix that is why sigma inverse is existing, it is a positive definite matrix we were having a decomposition and from that decomposition the inverse was also possible and then the half matrix was also allowed.

If you remember the calculations that we did in it in one of the previous lectures so we will use that thing here. So this I can write as log of determinant of sigma to the power – half S sigma to the power half/n–trace. Now in the trace also same argument can be used because trace of AB is also = trace of BA. Now this is equal to n times log of product of lambda i i=1 to n–sigma lambda i i=1 to n +p.

Here lambda 1, lambda 2, lambda p they are characteristic roots of sigma to the power minus half S, sigma to the power-1/2/n. And since this is we are starting with the positive definite these are all positive. Now if we consider log of x–x+1 this is always < or or = 0 if I take x=0. So if you look at this term this is always going to be < or = 0. So actually we wanted to prove that S/n maximizes.

So in the log L I substituted sigma=S/n that is why I got n log S/n inverse-trace of sigma inverse that was becoming S/n inverse S. So I should show this is < or = 0 not > or = 0. So this is what we are able to prove. So S/n is maximum likelihood estimator of sigma. Also if you look at the likelihood function which is actually the joint density function form, that we have written here. So from here we can also conclude that U bar and S it is sufficient statistics for this problem.

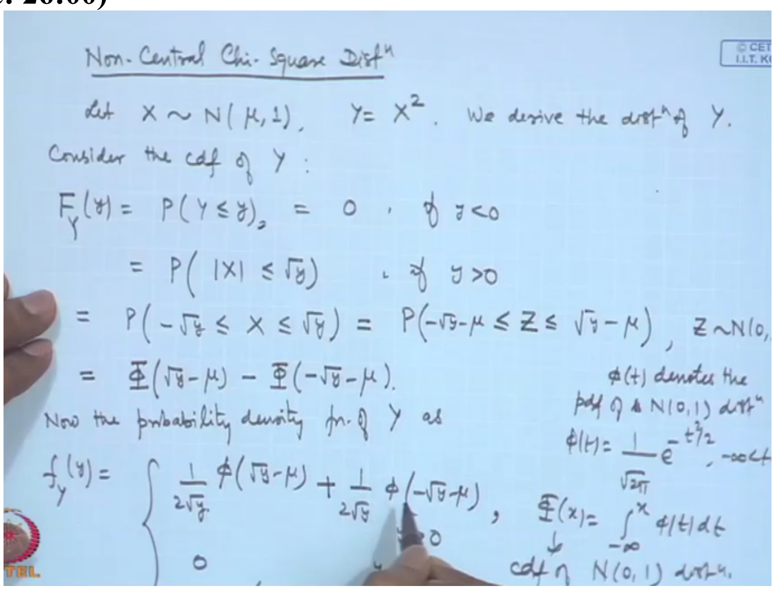
(Refer Slide Time: 24:00)



From an application of the Factorization theorem on the joint pdf of U1, U2 we conclude that U bar and S is sufficient. So this fact will be further useful in the inference problems. So let us summarize, we have considered the Multivariate Normal Distribution and we have discussed several properties of the multivariate normal distribution. Now one or two important points that we saw was use of a non-central chi-square distribution because we have seen that some of the squares of independent normal random variable is a central chi-square.

So if we are considering normal distribution with some non-zero mean, and then if I consider some of the squares then we will get a non-central chi-square. So I will now introduce non-central distribution, they are extremely useful in the multivariate theory. So let me start with the non-central chi-square and then gradually we will talk about non-central T and non-central F distributions also.

(Refer Slide Time: 26:00)



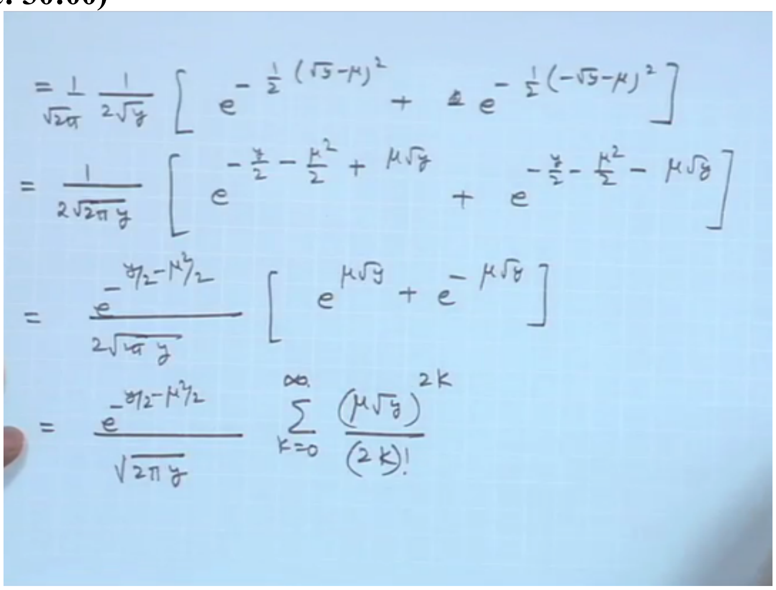
So we talk about Non-Central Chi-square distribution. So let us consider say x following normal Mu, 1 distribution. We have seen that if x follows normal 0,1 then y=x square as a chi-square 1 distribution. Now if x is normal Mu, 1 then x-Mu square will be chi-square 1, but what about x square itself, so let us drive the distribution here. To derive the distribution, we consider, we derive the distribution of y.

So let us consider the cdf of y, so naturally this is going to be 0 if y is < 0. So this = modulus x < or = root y if y is 0. So let us consider this portion here. So this = probability of – root y < or = x < or = root y. Now that = we transform it to standard normal then this is becoming Z is < or = root y-Mu and here - here z follows normal 0,1. So this in terms of capital phi function which is a cdf of standard normal distribution, we can write it as.

So we have derived the cumulative distribution function of x square, so we can also find out the probability density function, so derivative of capital phi will be small phi. So let us revise the definition. This is small phi t denotes the pdf of standard normal distribution that is phi t=1/root 2 pi, e to the power–t square/2 and capital phi x that is nothing but the cumulative distribution function that is cdf of normal 0, 1 distribution.

So if I differential capital phi I will get a small phi the root y–Mu and I will get 1/2 root y. And here I will get, there is a - here and there will be - here, so it will become +1/2 root y small phi – y–Mu. This is for y > 0, it is 0 for y. So of course equality at 0 we may include at one of the points that does not make any difference here.

(Refer Slide Time: 30:00)



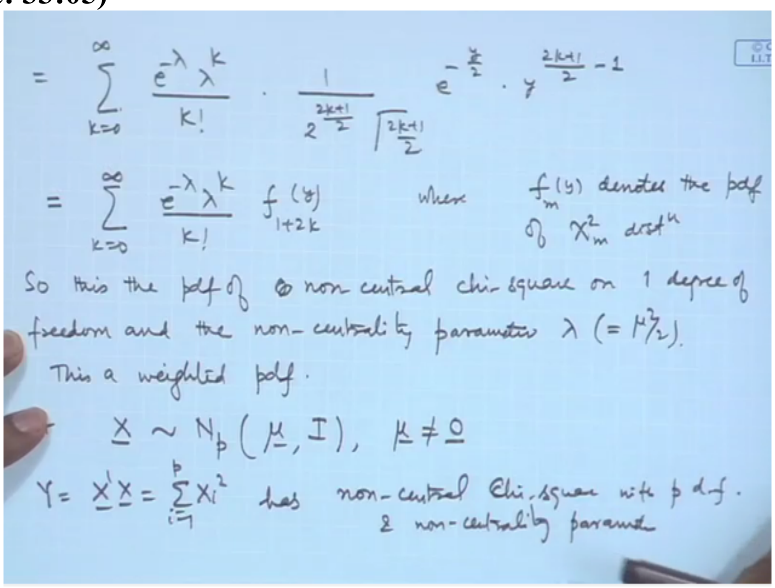
Now this we simplify. We can write it as 1/2 root y and 1/ root 2 pi will also come out and I will get e to the power half root y–Mu square+1, e to the power +1/2–root y–Mu square. I am writing the part where the density is positive, in the 0 part I am not writing here. Let us simplify this portion. So this is becoming = 1/2 root 2 pi y and this term here I can write e to the power–y/2–Mu square/2+Mu root y.

Similarly, in the second part it will become–y/2-Mu square/2 and this one will give me the - sign, -Mu root y. So this term I can keep outside, it will be become e to the power–y/2-Mu 2 square/2 root 2 pi y, I get e to the power Mu root y + e to the power–Mu root y. If we consider the expansion of e to the power Mu root y and e to the power-root y–Mu root y, so this become simply e to the power– y/2-Mu square/2 divided by.

Here in terms will be + and – so they will get canceled out and the event terms will get added up. And if you then you will get two times so this 2 will go away, I will get divided by root 2/y sigma Mu root y to the power 2k/2k factorial k=0 to infinite. Now let us substitute here, say Mu square/2= say lambda, then this will become e to the power-lambda-y/2/root 2 pi y sigma lambda to the power k, 2 to the power k, y to the power k/2k factorial.

I multiply and divide by k factorial here. So let us simplify this here.

(Refer Slide Time: 33:05)



This will become = -- I combine the terms in the particular way, e to the power–lambda, lambda to the power k/k factorial; 1/2 to the power 2k+1/2 gamma 2k+1/2; e to the power– y/2; y to the power 2k+1/2-1. So I have combined all the terms in a very particular way, let us see how it is coming. So e to the power-lambda, lambda to the power k/k factorial I am writing here. Then there is another term that is 2k factorial and then there will be a k factorial here.

So here I can actually cancel the terms like here I will get 2k, so I will cancel with the first term here then I will get 2k-2 that I cancel with k-1 and so on. Now what is remaining is 2k-1; 2k-2 and so on-- no 2k-3 and so on, so that terms I combined and it can be written as a gamma 2k+1/2 because there is a divided by 2 terms coming here. Now the terms which I left here, that is for example 2k so there was a 2 here, then there was another 2k-1 and so on, that will be again coming here, and then there is a square root 2 here, so that I put together as 2 to the power k+ 1/2.

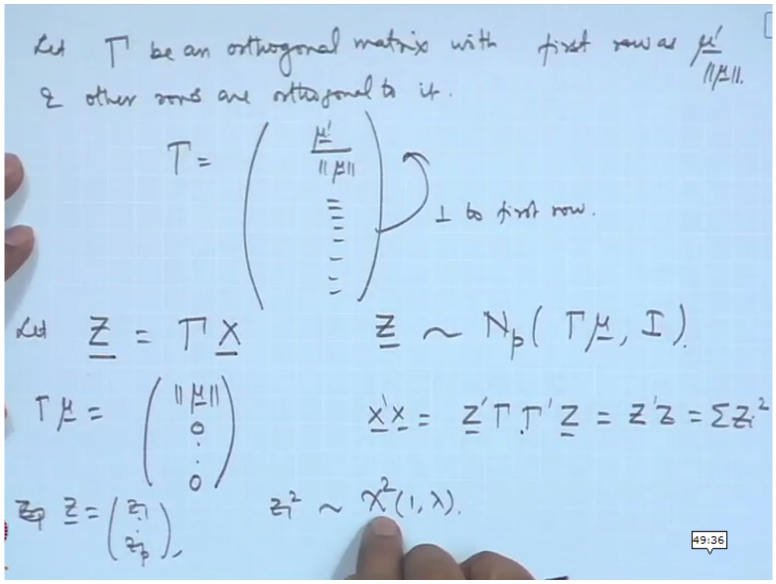
Then there is a e to the power–y2 term that I write here and the power of y to the power 2k/2 and then here we have –1/2 so that I write as half -1. So this particular way of writing down this gives it as interpretation that it is = sigma e to the power–lambda, lambda to the power k/k factorial f of 1+2ky where this fm(y) denotes the density of chi-square m distribution. So the interpretation for the density function of a y which is = x square.

Here y=x square and the interpretation for the density of that is, it is a weighted, because these are Poisson's weights of central chi-square. So this is the pdf of non-central chi-square on one degree of freedom and the non-centrality parameter lambda, lambda is actually Mu square/2. This is a weighted pdf; the weights are actually the Poisson weights here. So now, let us consider x to be Np Mu, i.

That means I am considering p components so x1, x2, xp are independent normal which means Mu1, Mu2, Mu p and variances are unity and they are independent and in general I am assuming Mu to be non-zero because at 0 it will simply give me chi-square central chi-square. So now I am looking at y=x prime x that is sigma xi square i= 1 to p. So then this has non-central chi-square with p degrees of freedom and non-centrality parameter.

Lambda that is half Mu prime, that is sigma Mu y square/2 lambda that is 1/2 Mu prime Mu that is sigma Mu y square/2. Let us look at this.

(Refer Slide Time: 38:14)



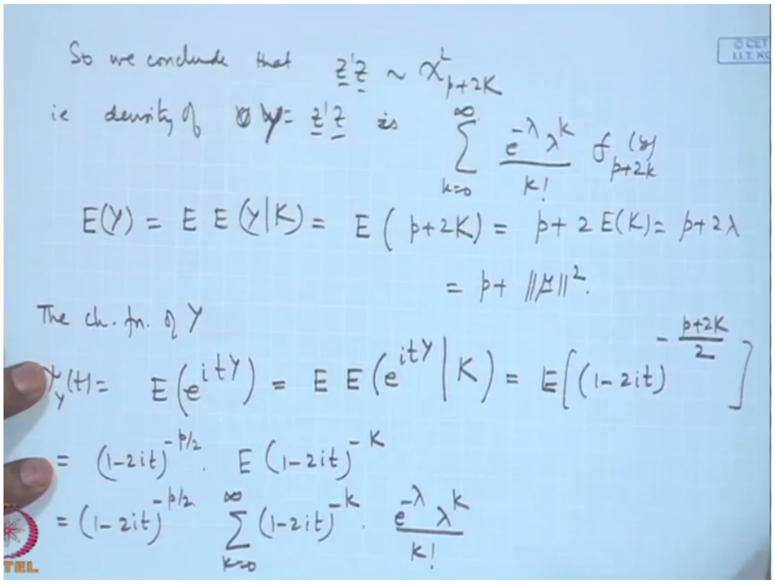
Let us define gamma to be an orthogonal matrix with first row as say new prime/ norm of Mu. And other rows are orthogonal to it. That is, I am writing gamma = something like Mu prime by norm of Mu and other rows these are orthogonal okay, orthogonal to first row. Let us consider say z=gamma x then z will follow Np gamma Mu, i. But what is gamma Mu? Gamma Mu= because this Mu, so Mu prime Mu you will get.

So that is norm of Mu square so you will get norm of Mu and other terms will become 0 because the other rows are orthogonal to the first row. So if I consider x prime x that=z prime gamma, gamma prime z that=z prime Z because gamma is orthogonal that = simply sigma Zi square. So if I consider the first component that is Z=Z1, Z2, Zp then Z1 square will follow chi-square 1 lambda.

That is non-central chi-square distribution with one degree of freedom and lambda as the non-centrality parameter. So this we will write as chi-square p lambda and this one we are writing as chi-square one lambda, sometimes we write as chi-square one lambda like this also, that is chi-square p lambda. So these are various forms of this notation here. So Z1 square is chi-square 1 lambda and what you are getting is Z prime Z that = to Z1 square+Z2+Zk square.

So these are central chi-square, so what you are getting chi-square 1+2k where k is Poisson lambda + chi-square 1+chi-square 1, these are central and these are all independent.

(Refer Slide Time: 41:20)



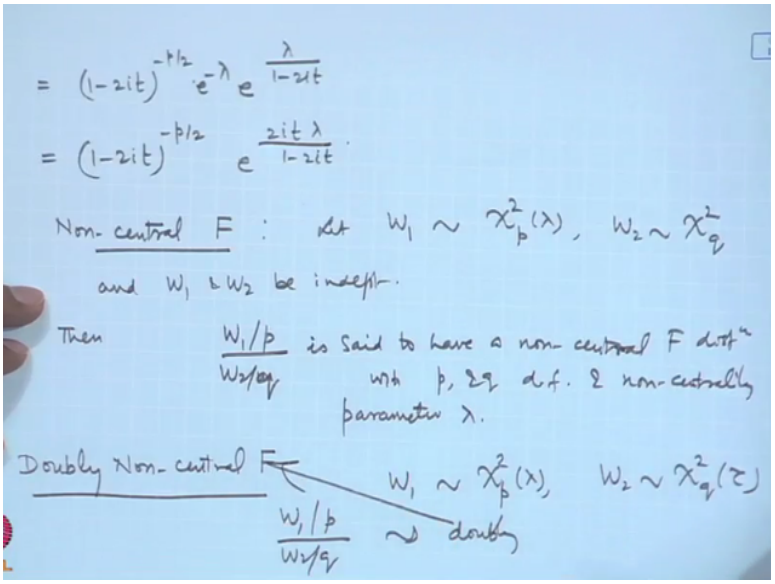
So we conclude that Z prime Z follows chi-square p+2k that is pdf of say V or say W=Z prime Z that will be = e to the power - lambda, lambda to the power k/k factorial f of p+2ky where k=0 to infinite. We can look at some elementary properties for example, we have written actually y=x prime x, now x prime z=z prime z so this is actually y the density of y.

So if we consider expectation of y we can write it as expectation of expectation y/K, that is equal to equal to expectation of p+2k because for a chi-square distribution if k is given then it become central and it = to the number of degrees of freedom, that = p+2 expectation of k that = p+2 lambda, that = p+ norm of Mu square. We can also consider the characteristic function of y that = chi y of t that is expectation of e to the power i ty.

So that is equal to expectation of expectation e to the power i ty given k. So that = -- given k it is chi-square so we know it is equal to expectation of 1-2it to the power-p+2k/2. Now to consider the expectation of this with respect to k we consider the k following Poisson lambda here, so it = 1/2it to the power p/2 expectation of 1-2it to the power–k, that = 1-2it to the power–p/2, because k is following Poisson lambda.

Now this term can be combined with this so you get it as simply 1-2it to the power–p/2 sigma e to the power–lambda by k factorial lambda by 1-2it to the power k.

(Refer Slide Time: 44:49)

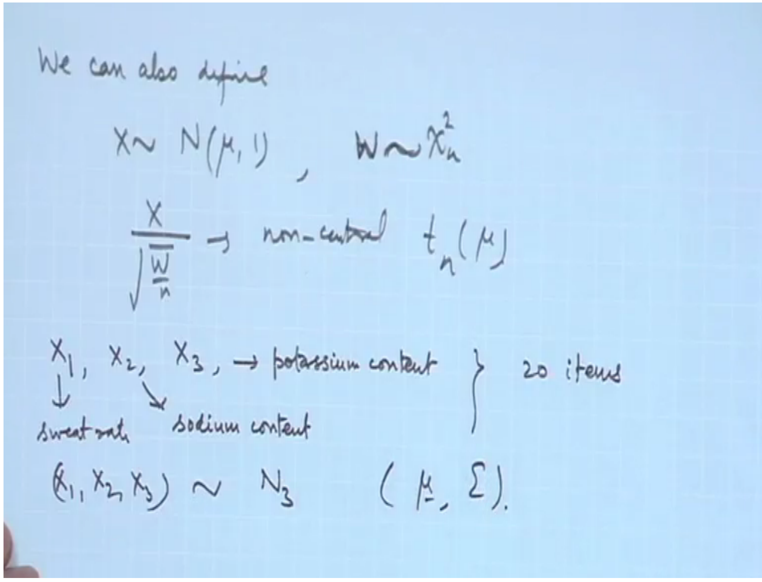


So you can get that = 1-2it to the power –p/2 e to the power lambda/1-2it e to the power lambda 1-. So after combining the terms, now once we are aware to determine the characteristic function of the non-central chi-square distribution other characteristics like and other things can also be found out easily. So I am leaving this discussion at this point. Now, if you remember the definition of T-distribution the definition of F-distribution we have made use of the chi-square.

So now if that chi-square is replaced by a non-central chi-square the similar changes will occur. So let me define, non-central F, so let say W1 follow chi-square p lambda and W2 follow chi square say q. And W1 and W2 be independent then W1/p divided by W2/n is said to have a non-central F with say—sorry this is q here with p and q degrees of freedom and non-centrality parameter lambda.

Now there can be possibility that the denominator chi-square is also non-central so we call it Doubly non-central F, say W1 follows chi-square p lambda and W2 follows chi-square Q Tou. So in that case if we consider W1/p/W2/q, then this is called Doubly Non-Central F. Of course they should be independent.

(Refer Slide Time: 47:53)



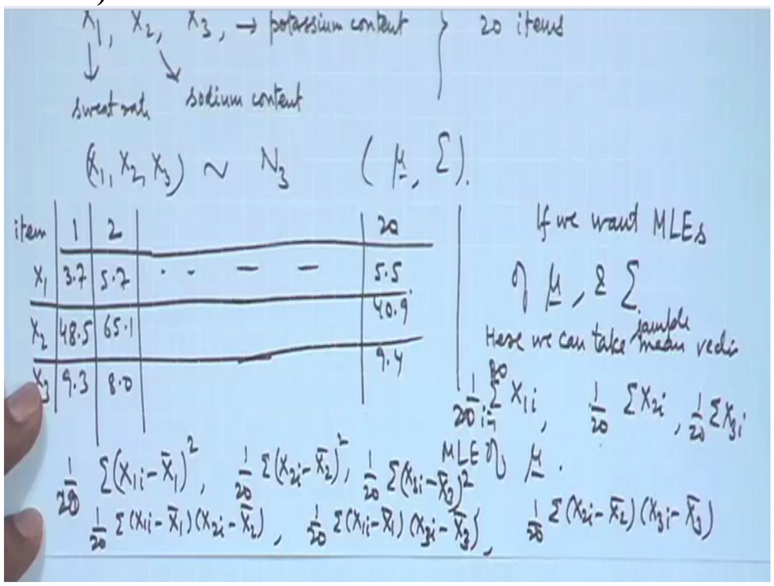
We can also define X following normal Mu,1 and say W follow chi-square n. Then if I consider x/root W/n and this is called non-central t. So these are some of the distribution that are used in when we deal with general multivariate normal distribution and these quantities will be appearing in the distributions of the tested statistics which we use for constructing the test for the parameters of a multivariate normal distribution or for constructing the confidence intervals etcetera.

Now in the case of Univariate normal distribution we had that x1, x2. xn, if they are it is a random sample and if we consider the sample variance sigma xi-x bar whole square/-1 and we called it as S square. We have obtained distribution of that that is n-x square/sigma square it follows chi-square distribution n-1 that is the freedom. Now what could be a possible generalization of this to multi-dimension.

Because in the multi-dimension for variance covariance matrix sigma we are getting the sample variance covariance matrix S, so we are considering S/n-1 by as estimator, so what could be the distribution of that. So we need the concept of a matrix distribution. So in the next lecture I would be covering this matrix distribution for this. Let us look at one or two applications of the sampling from multivariate distribution.

Suppose I am considering say x1, x2, x3 so these are considered as sweat rate; x2 is considered as the sodium content; x3 is considered as the potassium content. So this is data on 20 times. So it is assumed that x1, x2, x3 this is having N3 distribution with say Mu and sigma and data is calculated data is observed in the following fashion.

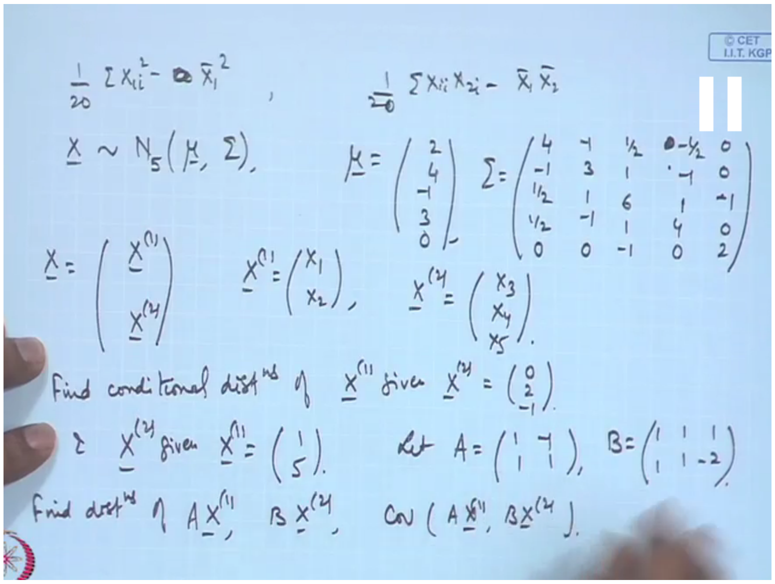
(Refer Slide Time: 50:59)



For item number 1, so for each item we are writing down the values of x1, x2, x3. So it is 3.7, 48.5, 9.3. For the second item it is 5.7, 65.1, 8.0 like that up to 20. We are having for 20th item 5.5, 40.9, 9.4. So if we want say maximum likelihood estimators of Mu and sigma then here we can consider mean vector sample mean vector we can consider here sigma say x1i i=1 to n 1/n, so 1/20 here 1/20 sigma x2i 1/20; sigma x3i etcetera.

So this will give the MLE of Mu. For calculation of the MLE or sigma then I need to consider 1/19-- 1/20 sigma x1i-x1 bar square 1/20; sigma x2i-x2 bar square 1/20; sigma x3i-x3 bar square. We also need to calculate the cross product terms like 1/20 sigma x1i-x1 bar*x2i-x2 bar 1/20 sigma x1i-x1 bar*x2i-x2 x3i-x3 bar 1/20; sigma x2i-x2 bar*x3i-x3 bar.

(Refer Slide Time: 53:33)



For calculation purpose the simplifications can be done, for example one may use like 1/20 sigma x1 say x1i square–20 x1 bar square so this will get canceled out. And similarly, for the cross product term you may consider 1/20 sigma x1i, x2i–x1 bar, x2 bar etcetera, so one may consider these quantities also for simplification. I end up with some exercise. Let us consider say x following N5 Mu, sigma and I define Mu as the 24-130 and sigma matrix as 4-1 1/2 1/2 -1 3 1 -1 0; 1/2 1 6 1 -1; 0 sorry –1/2 -1 1 4 0; 0 0 -1 0 2. So this is a 5/5 matrix.

Let us consider say x=x1 and x2, where x1 I am taking to be x1, x2 and x2 I am taking to be x3, x4, x5. So find conditional distributions of say x1 given x2=0 2 -1. And x2 given say x1=1 5. Let us also take A=say 1-1 1 1 and b= say 1 1 1; 1 1 -2. Find the distributions of A x1, B x2, covariance between A x1, B x2. Also find P that is 2/2 matrix and Q 3/2 such that Px1 and Qx2 are independently distributed. So I am leaving it as an exercise, you can try.

"Professor to Student conversation starts". So in the next lecture I will consider a matrix distribution for the sample dispersion matrix S. So it is called Wishart distribution in the next lecture I will introduce this thing.**" Professor to Student conversation ends."**