Applied Multivariate Analysis

Prof. Amit Mitra

Prof. Shramishtha Mitra

Department of Mathematics and Statistics

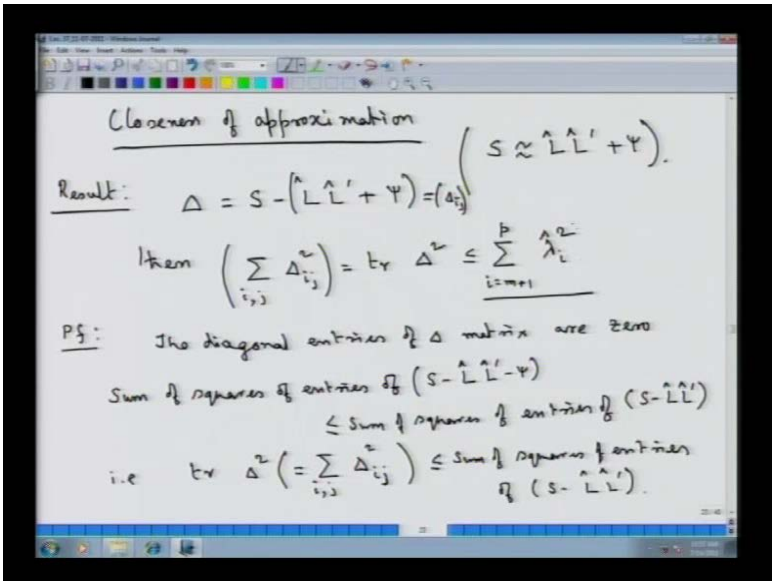
Indian Institute of Technology, Kanpur

Lecture No. # 39

Factor Analysis

In this lecture, we continue our discussion on methods of estimation of L and psi of a factor model. Now in the last lecture, we had discussed about the principal component method for estimation of L and psi; and we had come up to the point of looking at the closeness of comparison of approximating S by L hat L hat transpose plus psi.

(Refer Slide Time: 00:42)



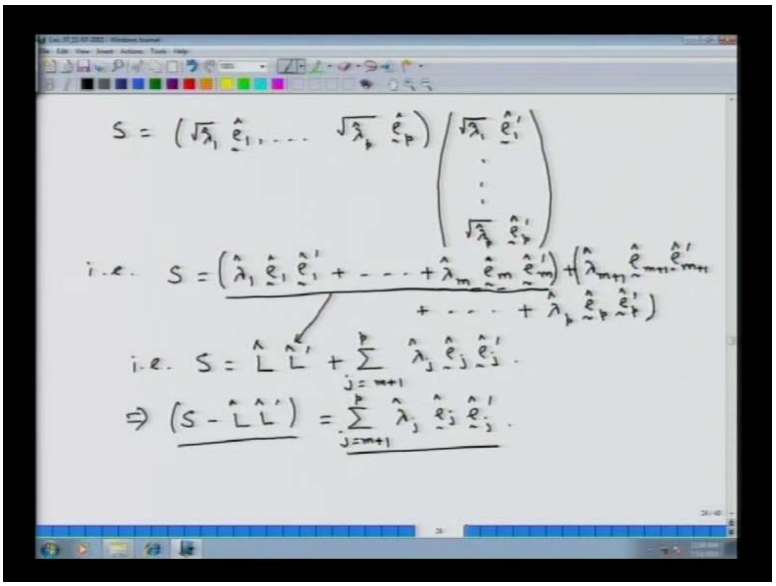
So, we had stated this result in the last lecture that denoting by delta, this matrix which is S minus L hat L hat transpose plus psi delta i j being the i j eth element of this delta matrix. Now, as we had discussed in the last lecture that, the way that this L hat and psi is formed, the diagonal entries of S minus L hat L hat transpose plus psi this element, the diagonal entries will be equal to 0; and only the off diagonal entries of this delta matrix is non-zero. So, since we had approximated S by L hat L hat transpose plus psi, we have that particular observation. Now under such a situation, we will have this sum of squares

of the entries of this delta matrix bounded by the sum of squares of the Eigen values, from m plus 1 to up to p, the smallest p minus m Eigen values to be the quantity, which actually bounds this term here, which is the sum of squares of all the entries of this delta matrix.

So, the results that states that summation over i j of delta i j square, this is lesser or equal to i equal to m plus 1 to p lambda i hat square. So, the first thing that we will do in this lecture is to prove this particular result, which gives us a bound on the approximation, which we get from the factor, the principal component method. Now, first we realize that the diagonal entries of the diagonal entries of delta matrix are 0, and thus the sum of squares of entries of S minus L hat L hat transpose minus psi, this is lesser or equal to the sum of squares of entries of this S minus L hat L hat transpose.

Because we are going to have the diagonal entries of these to be equal to 0, the diagonal entries of S minus L hat L hat transpose are non-zero, and hence the sum of squares of the of all the entries of S minus L hat L hat transpose minus psi will be lesser or equal to the sum of square entries sum of squares of entries of S minus L hat L hat transpose. Now, if we look at this left hand side here that is nothing but trace of delta square which of course, as we have denoted that is equal to double summation over i j this delta i j square. So, this is less than or equal to sum of squares of entries of this S minus L hat L hat transpose matrix. Now, we will look at what this sum of squares entries of S minus L hat L hat transpose is, and then we will have that quantity to be equal to this upper bound.

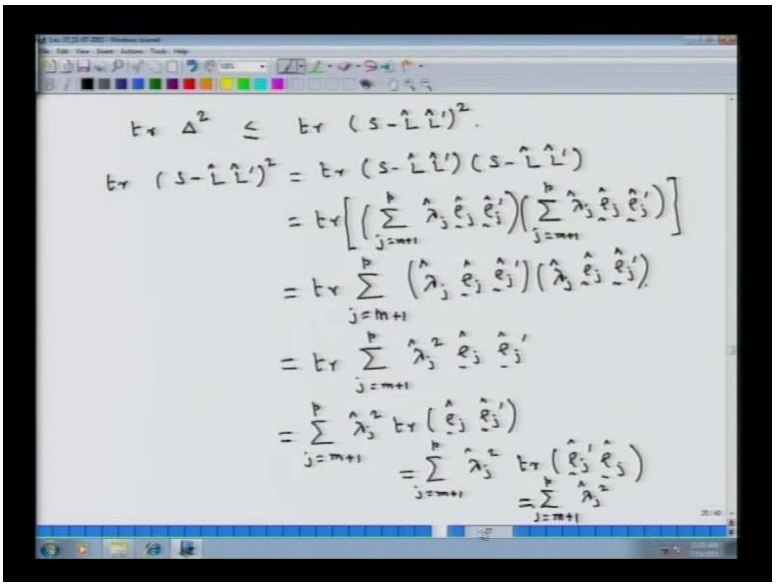
(Refer Slide Time: 04:18)



Now, let us look at how we had defined S through the Eigen value Eigen vector pairs. So, S as it was defined, its lambda 1 hat, e 1 hat, then this was of order p. So, this is lambda p hat, e p hat, this in to the transpose matrix, which is lambda 1 hat, e 1 hat transpose, and the last entry being lambda p hat times e p hat transpose vector right; that is this S is equal to lambda 1 hat, e 1 hat, e 1 hat transpose plus I will just write this m eth term here, the m eth term is lambda m hat e m hat e m hat transpose, this plus lambda m plus 1 hat e m plus 1 hat e m plus 1 hat transpose to up to the last term, which is lambda p hat e p hat e p hat transpose right. Now, if you look at this part here, there is a reason why I have written all these terms here up to m and beyond m.

So, if we look at the first m terms, these first m terms are going to come, if we look at L hat L hat transpose matrix. In other words, this S can be written as this is the first m terms contribution to S is basically L hat L hat transpose, this plus I will just put it as a sum summation j equal to m plus 1 to up to p, this is lambda j hat e j hat in to e j hat transpose. So, this will imply that this S minus L hat L hat transpose, this matrix is equal to the terms, which we have neglected; assuming that lambda m plus 1 up to lambda p they are having a negligible contribution, this is lambda j hat e j hat e j hat transpose. So, we have this S minus L hat L hat transpose to be given by this right hand side here. Now, we will use this form of S minus L hat L hat transpose in the previous equation, sum of the squares square entries of S minus L hat L hat transpose; now this of course, is a symmetric matrix.

(Refer Slide Time: 07:28)



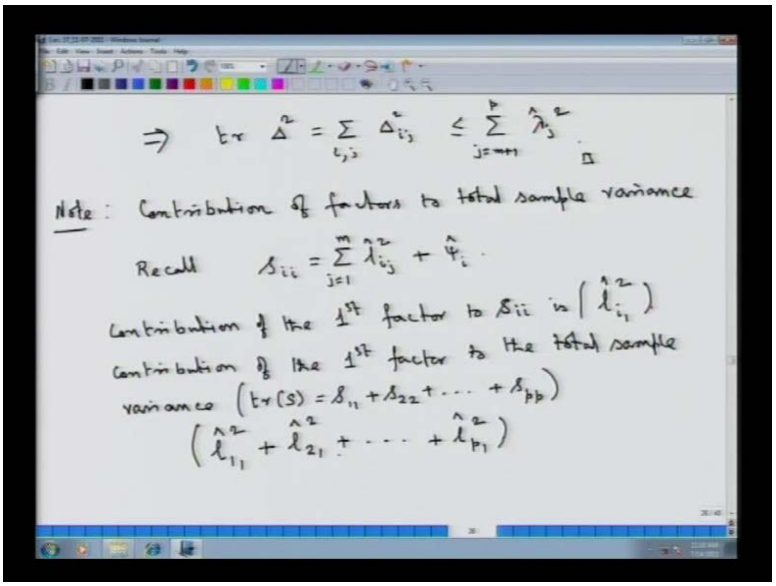
So, what we are going to have is that this trace of delta square from the previous equation is less than or equal to trace of this S minus L hat L hat transpose this square why is that so? Because we are looking at sum of squares entries of this symmetric matrix, and hence we are writing that in terms of the trace of this S minus L hat L hat transpose square. Now, what is this equal to let us look at what is the trace of S minus L hat L hat transpose square matrix, this is trace of I just write the two matrices L hat L hat transpose this into S minus L hat L hat transpose. Now, we will plug in the value of S minus L hat minus S minus L hat L hat transpose, as in this, in this expression here, and then we will able to write it as this is summation over j equal to m plus 1 to up to p, this is lambda j hat e j hat times e j hat transpose. So, that is this matrix, and then product by the same thing here j equal to m plus 1 to up to p lambda j hat e j hat times e j hat transpose right.

Now, if we look at this particular product here, inside the trace, what is going to happen is that we recall that this e 1 hat e 2 hat up to e p hat, they are ortho normalized Eigen vectors corresponding to the corresponding corresponding to the Eigen values. And hence, if we are looking at a particular j here, and a j prime in this summation, then the product will make that product will be equal to 0, because we will be looking at multiplication of e j hat prime times, if we look at another j prime here, then it will be e j hat prime multiplied by e j prime and that would be equal to 0, because the vectors are ortho normal.

And hence, only the terms, for which the same index in this sum, and the same index in this sum are multiplied, and that would lead us to the following expression, which is summation j equal to m plus 1 to up to p lambda j hat times this e j hat e j hat transpose times the same quantity from the other summation say in index; so, that is lambda j hat times e j hat times e j hat transpose. Now all other terms, for which the indices in these two summations differ, they will be equal to 0, because of the orthogonal properties of this e j vectors.

So, what do we get after this multiplication? This lambda j hat anyways is scalar, so this e j hat transpose e j hat that would be equal to 1, because we have e j hat vectors to be ortho normal. So, this is nothing but trace of the matrix, which is j equal to m plus 1 to up to p lambda j hat square e j hat e j hat transpose. Now, we will (( )) take the trace some term by term, because trace of the sum is sum of the traces. So, lambda j hat square anyway is a scalar quantity. So, we take the trace straight away inside trace of e j hat times e j hat transpose; and furthermore the trace of a b equal to trace of b a. So, what we are going to have is summation j equal to m plus 1 to up to p lambda j hat square trace of e j hat transpose e j hat and that is equal to 1. So, what we have is this summation to be just equal to j equal to m plus 1 to up to p lambda j hat square.

(Refer Slide Time: 12:18)



So, this would imply the desired result that trace of delta square, which is summation, double summation i j delta i j square that is less than or equal to the term that we have

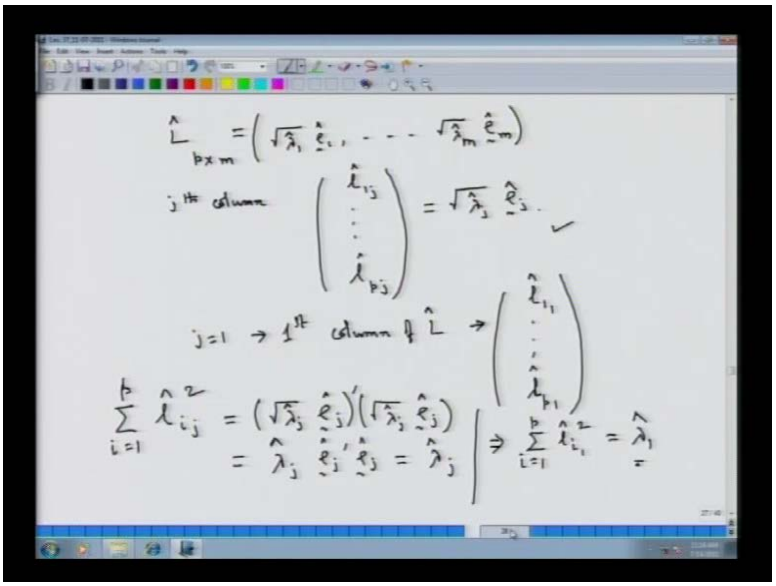
got in here. So, that is summation j equal to m plus 1 to up to p lambda j hat square. So, this proves the result that this is quantifying the closeness of approximation that this matrix, which is the difference between S and the matrix, which approximates S in this principle component method for estimation of L and psi that this is bounded by this quantity on the right hand side. So, after proving this particular result, we move on further, and look at another aspect of this principle component method for estimation of L and psi, which will give us idea about the contribution contribution of factors to total sample variance.

Now, in order to see the type of contributions that a particular factor would be having contribution of factors to total sample variance; we recall that in this principle component based method, we have this S i to be equal to summation L i j square, S i j equal to summation L i j square, j equal to 1 to up to m, this plus L i j hat square actually, this plus psi i hat. So, this is what is the expression for approximation that we are going to approximate S i i, by this particular term. And from here, we can say that the contribution contribution of the first factor to S i i is going to be given by…

Now, if you look at this particular term here, this is L i 1 hat square L i 2 hat square and so on up to L i m hat square. So, the contribution of the first factor is related to the term for which j is equal to 1 in this expression. So, what we are going to have is that it is L i 1 hat square. Now, if we look at the total sample variance, and then the contribution of this first factor to the total sample variance to the total sample variance; now what is total sample variance? Total sample variance is trace of the S matrix, which is equal to S 11 plus S 1 S 22 plus up to S pp, this is a p dimensional random vector that we are considering.

So, the contribution of the first factor to the total sample variance would be from these expressions. So, we will have that contribution to be equal to L 1 1 hat square, so this is the contribution of the first factor in S 1 1. Similarly, the contribution of the first factor to S 2 2 would be given by this expression this is L 2 1 hat square and so on, this up to the contribution of the first factor on the p eth component that is S p p would be given by L hat this particular term. So, this is the total contribution of the first factor to the total sample variance, which is given by trace of S which is summation of S i i is i equal to 1 to up to p is this particular quantity.

(Refer Slide Time: 17:00)



Now, it is interesting to see what this term is actually equal to if we look at the L hat matrix in this principal component based methods; this L hat, which is of the dimension of L by p I am sorry p by m, this is given by root over lambda 1 hat e 1 hat and so on; this is truncated up to the m term the number of factors, so this is lambda m hat times e m hat right. So, if we look at this, then the j eth column here is this the j column is say I write that as L 1 j hat L p j hat; now that is equal to root over lambda j hat times this e j hat vector. So, we have this as the j eth column.

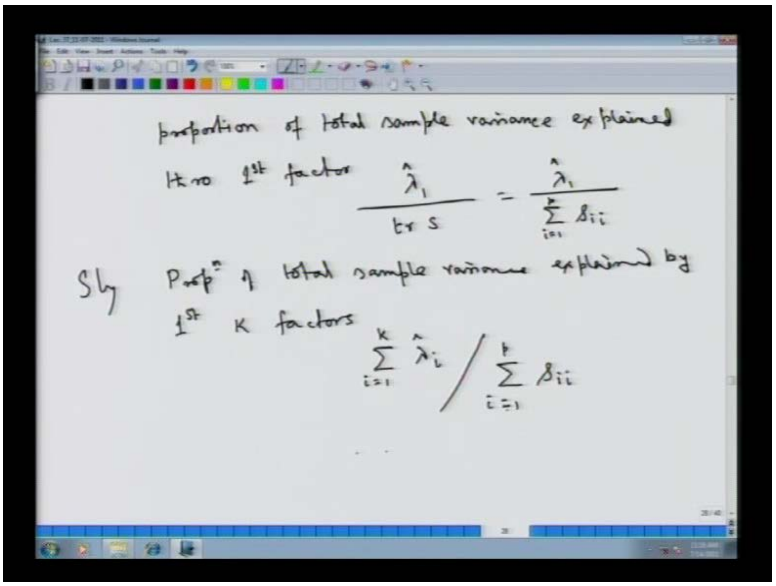
Now if you look back at this contribution term what we have got this was the contribution of the first factor to the total sample variance, which was L 1 1 hat square L 2 hat square L p 1 hat square. So, for j equal to 1 what we are going to have is the first column of this L hat matrix, and that is nothing but L 1 1 hat up to L p 1 hat; why are we looking at this particular expression? Just to identify that this contribution of the first factor to the total sample variance is nothing but sum of squares of the entries of the first column of this L hat matrix, this is the j column.

So in general, if we consider any j as in here, this summation which is L i j summation i equal to 1 to up to p hat square that is this the norm of this particular vector is going to be given by root over lambda j hat e j hat, this multiplied by its transpose lambda j hat e j hat transpose. So, what is this is, this is equal to because this is e j hat e j hat transpose; so, we will have I am sorry this transpose is on the other side; so, its root over lambda j

hat e j hat transpose into root over lambda j hat e j hat. So, this is going to give us this lambda j hat, and this is e j hat transpose times e j hat, they are ortho normal; and then this is going to be just equal to 1, so that this particular term is equal to just lambda j hat.

Now, this will imply that the expression that we previously got as the contribution of the first factor to the total sample variance is this term is just going to be given by lambda 1 hat, because its concerns this first row here. So, for the particular value that this is L i 1 hat square for i equal to 1 to up to p, this term is nothing but lambda 1 hat. So, this is the way, in which actually the contribution of the respective factors to the total sample variance can be calculated.

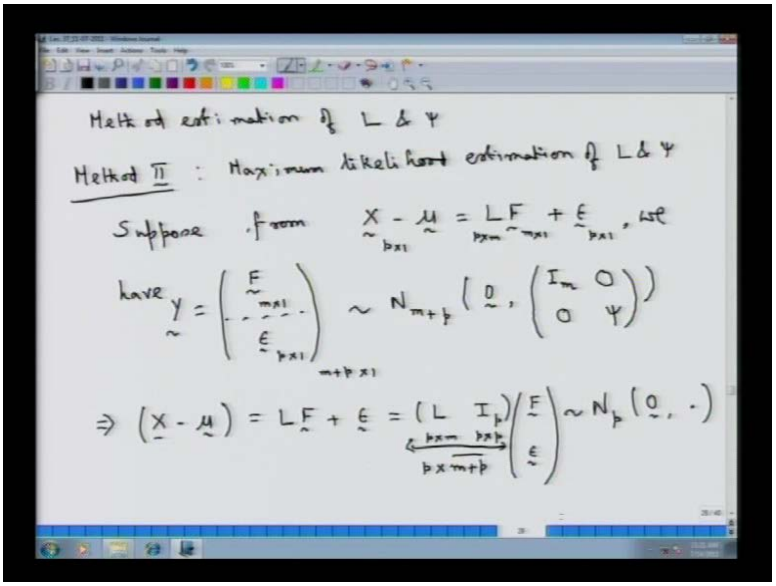
(Refer Slide Time: 21:00)



Now, using that we can further say that the proportion of total sample variance total sample variance explained or captured through first factor is going to be given by lambda 1 hat divided by this trace of S, which is going to be that summation, which is lambda 1 hat divided by summation of S i i terms, i equal to 1 to up to p. In a similar way, the contribution of a say k factors, the first k factors can be defined, so this is, this also can be extended. So the proportion of total sample variance explained by first K factors would be given by the first k factors, would be associated with the first k Eigen values. So, this is i equal to 1 to up to k this divided by summation i equal to 1 to p S i i and so on right.

So, this gives us a way actually from the principle component based method that how we can actually quantify the contribution of different factors in explaining the total sample variance, which was the basic objective actually to capture the variance covariance structure, and as a byproduct we are looking at these outputs; that will conclude this method of principal components.

(Refer Slide Time: 23:03)



Now, we will look at another method of estimation; method of estimation of L and psi; this is the first method that we had discussed was the principle component method. This is method number II; this is maximum likelihood estimation maximum likelihood estimation of L and psi. Now, once we say that it is maximum likelihood method of estimation, we would require certain distributional assumption on this, the random variables concerning the system; and we will have to impose certain conditions, certain distributional assumptions. Now, suppose from this m factor model x minus mu equal to L F plus epsilon, we have the joint distribution of the random vectors involved on the right hand side; this we callethe dimensions this is p by 1, this L is a factor loading matrix p by n, this is the vector of m common factors, and this is the vector of p specific factors.

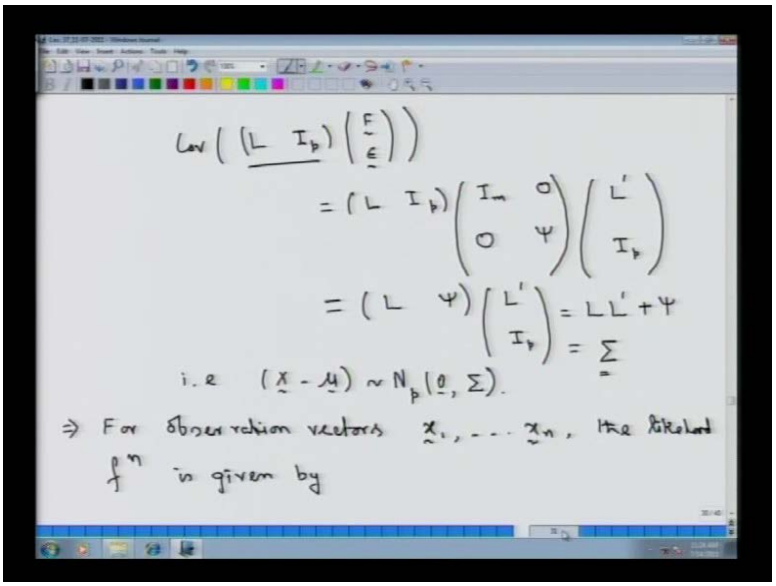
And suppose, we have the joint distribution of F m dimensional and that augmented with epsilon vector, which is p dimensional; so, this entire vector here, which is m plus p dimensional random vector; suppose the joint distribution of this is a N p dimensional

multivariate normal, with a mean vector as a null vector; and the covariance matrix naturally has to satisfy this structure of the factor model. Now in the factor model, this F vector, which is a vector of the common factors; it has got a covariance matrix to be an identity matrix; and then the covariance matrix of this epsilon vector the vector of specific factors is having a diagonal structure, which is psi matrix, and the covariance between F and epsilon in such a factor model needs to be a null matrix.

And hence, we assume that the joint distribution of F and epsilon has got this m plus p dimensional multivariate normal with a mean vector 0, and this as the covariance matrix right. Now this would imply that our X minus mu, which is L F plus epsilon; this can be written as L times an identity matrix of order p, this multiplied by this random vector, this F and epsilon. So, if you multiply this what we are going to get is L F plus this epsilon vector. Now, if we denote this m plus p dimensional random vector by Y, we have a result from multivariate distribution theory that if Y follows a multivariate normal; now this is a matrix of constants. So, this will, the distribution of this will also be a multivariate normal of the order that is determined through this what is the order - this is p by m and this is a p by p right.

So, this has got p rows and m plus p columns. So, the order of this matrix, which is L augmented with i p is p rows and m plus p columns; and hence the dimension of this X minus mu vector, multivariate normal, it will be a multivariate normal distribution N p with a mean vector given by this multiplied by the expectation vector of F and epsilon what is that that is a null vector. So, this is a null vector and a covariance matrix, which we are going to calculate.

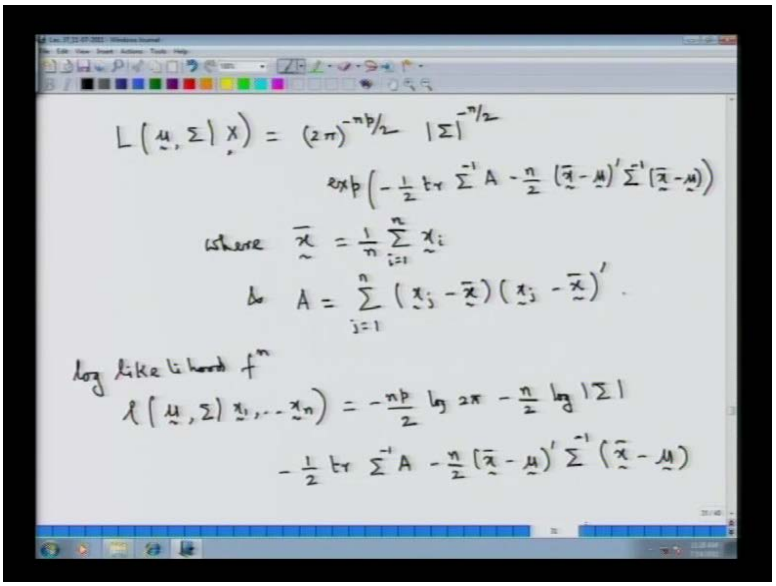
(Refer Slide Time: 27:36)



Now, the covariance matrix this covariance matrix of L i p that multiplied by F epsilon, this is going to be given by this is a matrix of constants. So, this is going to be given by L i p multiplied by the covariance matrix of this. Now, the covariance matrix of this is I m null null psi; so what we are going to get this is I m null null matrix psi matrix; and then the transpose of this is going to come here; so that is L transpose I p right. Now if we take the multiplications, what we are going to get is that this multiplied by this is going to lead us to L, and L I p multiplied by this is going to give us psi, and then this is L transpose I p. So, what do we get we get L l transpose plus psi now what is L L transpose plus psi in a factor model that is nothing but the sigma matrix.

So, we can fill in this particular dot here, and say that our x minus mu has got multivariate normal p dimension with mean vector 0, and a covariance matrix equal to sigma. Well, you have the covariance matrix of sigma straighter, covariance matrix of x straightaway equal to sigma anyway; but we had derived that covariance matrix of this L F plus epsilon through the multivariate normality distribution of this augmented F with epsilon vector. Under the assumption of joint multivariate normality of F and epsilon vector, what we have realized is that this X minus mu has got a p dimensional multivariate normal with a mean vector 0 and a covariance matrix as sigma. Now, we are in a position to frame the likelihood estimation, because we can now write the form of the likelihood function for observations observation vectors x 1, x 2, x n; for observation vectors x 1 x 2 x n; the likelihood function is given by…

(Refer Slide Time: 30:22)



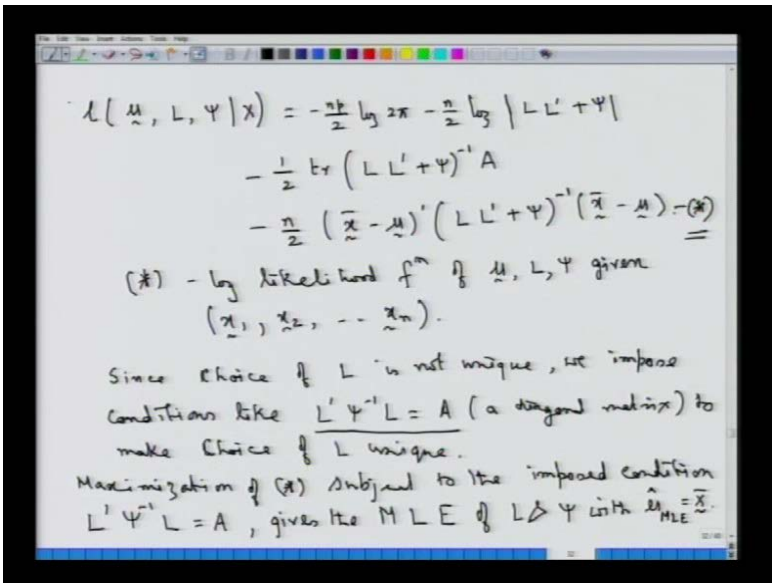
Let us denote that likelihood function as L; now this is going to be written as a function of mu and sigma, given the data x 1, x 2, x n; so given x, this is going to be given by 2 pi to the power minus n p by 2 why it is that, because we have random sample of size n, and we have each of the random samples a multivariate normal with dimension p. So, for each of these random samples, we will have a 2 pi to the power minus p by 2 that is the dimension of the multivariate normality; and we have n such terms, and hence we have this factor as 2 pi to the power minus n by 2, and then we will have a determinant sigma to the power minus n by 2, from where does this come; if we once again look at the structure of multivariate normal in the density of each of these, we will have determinant of sigma to the power minus half; and since we have n observations, we will have this actually leading us to determinant of sigma to the power minus n by 2 that multiplied by the exponent.

Now, we will straight away write the form of the exponent that we usually use, which is minus half trace of sigma inverse a this minus n by 2 x bar minus mu transpose sigma inverse x bar minus mu right, where this x bar vector is 1 upon n summation x i vectors. So, it is the sample mean vector obtained from the observations x 1 vector x 2 and up to x n; and this A matrix, this is actually the formulation of calculating the maximum likelihood estimators of multivariate normal distribution. So, that this A is j equal to 1 to up to n x j minus x bar vector into x j minus x bar vector its transpose.

So, this is the standard form that we will have this as the exponent of the joint distribution of x 1, x 2, x n; each of them having a multivariate normal distribution, having a multivariate normal distribution with the order of the multivariate normality as p. Now if this is the likelihood function, we can also write the log likelihood, log likelihood function; let us denote that by small l, this is l mu sigma given this x 1, x 2, x n or in short just to as we had written as x here. So, that is going to be equal to minus n p by 2 log 2 pi this minus n by 2 log determinant of sigma this minus half trace of sigma inverse A minus n by 2 x bar vector minus this mu vector transpose sigma inverse x bar minus this mean vector mu right.

Now, from this expression here, we what is the basic objective is to get the maximum likelihood estimators of L and psi right that is what we are aiming at. So, in order to get the maximum likelihood estimators of L and psi, we will write this log likelihood function in terms of those; note that although we have written the log likelihood function of mu and sigma, it is in this particular sigma that we have L and psi.

(Refer Slide Time: 34:54)



And hence this can be written in terms of log likelihood function in terms of L and psi given x 1, x 2, x n, this is going to be minus n p by 2 log of 2 pi minus n by 2 log of determinant. Now, here we are going to use the fact that sigma here we had determinant of sigma here; in place of sigma, we will write L L dash plus psi; so, this is L L dash plus psi determinant of this matrix this minus minus half trace of sigma inverse A. So, this is

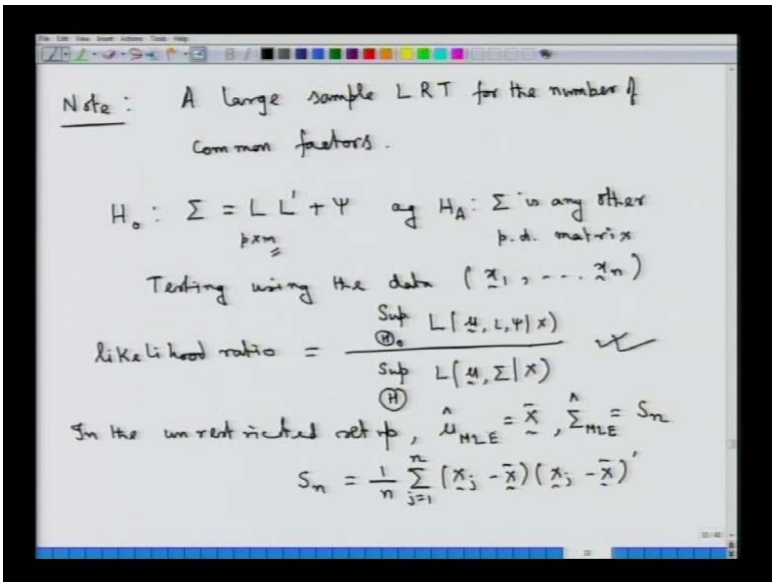
L L dash plus psi whole inverse times that A matrix this minus n by 2 x bar minus mu transpose sigma inverse that is L L dash plus psi whole inverse times this x bar minus mu. So, what we have done is to look at this log likelihood function, this was as a function of mu and sigma, we transformed that actually using the factor model, just by replacing sigma by the corresponding L L dash plus psi quantity.

So, we have this expression, which is star as the log likelihood, is the log likelihood function log likelihood function of this mu L psi, given the observations set that is x 1, x 2, x n; and the choice of L is not unique; as we have seen in the n factor model that the choice of L is not unique, and as such we need to impose certain additional conditions in order to make L unique. And that is a type of condition, which is used in order to maximize this subject to the conditions, which would make L to be unique

So, we write this that since choice of L is not unique, we impose conditions like L dash psi inverse L equal to A - a diagonal matrix to make L choice of L unique; and then maximization of star maximization of star, subject to the imposed condition, subject to the imposed condition L dash psi inverse L equal to A diagonal matrix gives the maximum likelihood estimators of L and psi; with of course, mu hat the maximum likelihood estimator to be given by the usual maximum likelihood estimator, which is x bar. So, using the form star as in here to be the likelihood function of mu L and psi, and using a condition like this to be A a given diagonal matrix; we can maximize the star with respect to this condition and arrive at the maximum likelihood estimators of L and psi with mu hat maximum likelihood estimator to be given by x bar.

Now, there are other methods of estimation of L and psi, when we are talking about the factor analysis model; but we have in this course look at two such important most widely used methods of estimation of L and psi. The first one that we had discussed was the principal component based methods, and the second one in a more general setup, when we are looking at maximum likelihood method of estimation.

(Refer Slide Time: 40:14)



So, we will conclude this estimation part here, and to conclude actually, we will just look at 1 small note, which is important, which gives us a large sample test large sample L R T test for the number of common factors. So, what we are going to address is the following fact that the number of common factors as such is not known to us, and what is the best way to judge, what is the number of common factors that should be used in a particular x vector model. We look at a large sample likelihood ratio test; now we are doing this, because we are just now actually obtained the maximum likelihood estimators of L and psi; and this large sample likelihood ratio test is based on that maximum likelihood method of estimation. Now what we are doing here is that we are saying that a null hypothesis is of the form that sigma is equal to L L dash plus psi with the order of L to be p by n with a chosen p here.

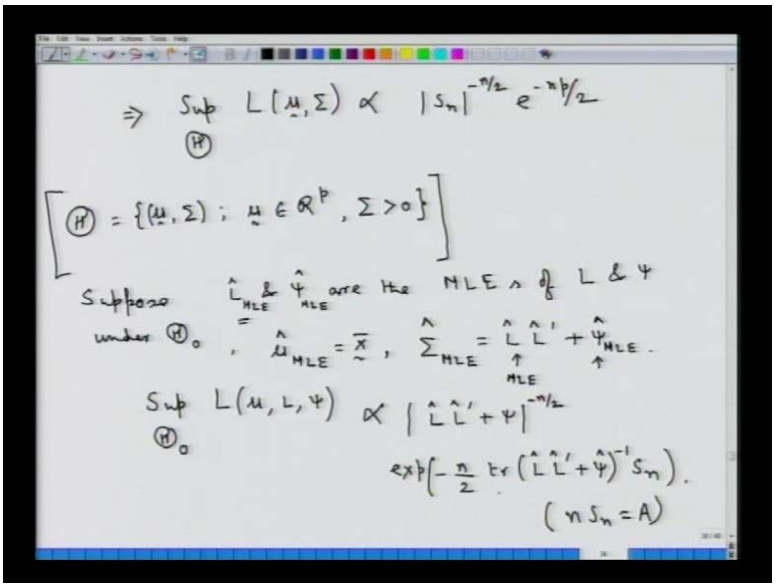
Suppose, our interest is to test this null hypothesis against the alternate hypothesis that sigma is any other is any other positive definite matrix any other positive definite matrix. So, we are going to test this, if null hypothesis is accepted, then we take this m to be the number of common factors as in here; now for certain m, as we will see that this might actually get rejected, and then we will accept that this that particular choice of m does not hold good for the given observation vectors x 1, x 2, x n. Now this testing is to be carried out using the data as in x 1, x 2, x n; so based on these n data vectors, we are going to test this null hypothesis against the alternate hypothesis H A using the L R T philosophy

Now, the likelihood ratio as we know, likelihood ratio is going to be given by this is supremum over script theta naught of the likelihood function; now this likelihood function will be under the null hypothesis, another null hypothesis meaning thereby we will have mu L psi, given this x, the entire data vector that divided by supremum over script theta; now when we talk about script theta, it is with respect to not this factor model, it is with respect to the mean vector mu and a sigma being any any positive definite matrix.

So, we will have to compute this particular term, which is going to be called the likelihood ratio, when we are looking at testing of this null hypothesis; and then we will use large sample standard large sample theory of likelihood ratio in order to formulate this testing of H naught against H A. Now in order to do so, what we would require is these two supremum quantities. Now in the unrestricted setup in the unrestricted setup mu hat maximum likelihood estimator is going to be given by x bar that we have already seen time and again.

And this sigma hat M L E is going to be S n, where S n is nothing but one upon n summation x j minus x bar into x j minus x bar transpose j is equal to 1 to up to n. So, that is the maximum likelihood estimators in the unrestricted setup. So, this will actually look at the denominator part, and then in the likelihood function, we will have to plug in these values of mu as x bar, and sigma as S n, which is of this particular form.

(Refer Slide Time: 45:16)



And then, we will be able to write this supremum over script theta, the entire parameter space; now what is entire parameter space? The entire parameter space just to recall this script theta is the set of all mu and sigma, wherein this mu belongs to R to the power p, and sigma is positive definite; so that is my script theta. So, supremum over script theta of L mu sigma, I will just drop this x, so this is going to be proportional to I am I am ignoring the constants. So, it is supremum over theta L mu sigma that is going to be given by determinant of S n to the power minus n by 2 e to the power minus n p by 2. This we had of course, done when we are looking at maximum likelihood method of estimation in multivariate normal and testing; but if one wants to recall why it is so? You will have to look at this expression here. So, this is basically the likelihood function.

You plug in mu hat equal to x bar, so this term would be equal to 0. So, plugging in mu hat equal to x bar, this will be a null vector multiplied by whatever sigma hat you plug in this is going to be that; and this A is this matrix, which is n times. So, this

A is nothing but n times S n; so we will have in place of sigma inverse n times S n inverse being plugged in, and what we are going to get is the expression that written here. So, supremum over theta L mu sigma is going to be proportional to determinant of S n to the power minus n by 2 e to the power minus n p by 2.

Now, suppose L hat and psi hat are the maximum likelihood estimators of L and psi under script theta naught. Now how we are going to obtain this L hat and psi hat that is using the method of maximum likelihood estimation that we have just now touched upon. So, using those that technique in order to get to this L hat M L E, and psi hat M L E under H naught; under H naught, how it is going to matter? Under H naught, we will have a fixed m here; now for that fixed m, we will actually look at the m factor model; and then that is going to determine that this L hat M L E and psi hat M L E are going to have the dimensions as what is specified through the null hypothesis that m factor.

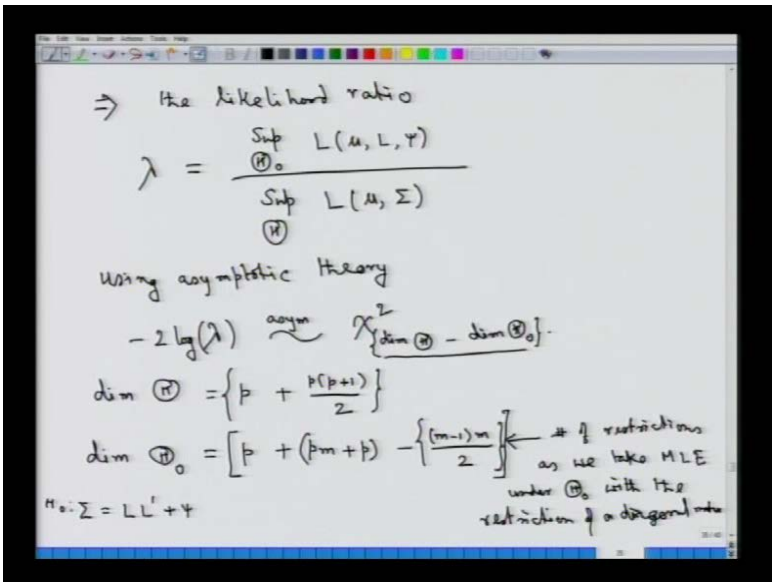
Now along with that, we will have mu hat M L E, irrespective of this L hat psi hat that is always going to be equal to this x bar term. Now if L hat is a maximum likelihood estimator of L, and if psi hat is the maximum likelihood estimator of psi, we will have the maximum likelihood estimator of sigma hat by invariance property of the maximum likelihood estimator. We will have this as L hat L hat transpose plus psi hat, where this is

the maximum likelihood estimator, as we have given in here; this also is the maximum likelihood estimator right.

Now, we are in a position to write this supremum over script theta naught of the likelihood functions. Now this likelihood function will be in terms of mu L psi, this given x, I will just drop it; now this one is going to be proportional to terms similar to this; now this is determinant of sigma in the likelihood function; now sigma is denoted by sigma hat, which is L hat L hat transpose plus psi hat. So, this is going to be proportional to L hat L hat transpose plus psi whole to the power minus n by 2, and here the exponents terms do not cancel out and give us a nice form like the previous one; this is going to be bit complicated, because terms do not cancel out.

The second term of course, will be equal to 0, because its x bar minus mu. So, mu hat being equal to x bar will lead the second term to be equal to 0. However, the first term is going to be trace of L hat L hat transpose plus psi hat this is sigma. So, trace of sigma inverse is this times that I will just write this as A, because it is better to write that as no if you write that as A, then this term will be given by this S hat right. So, we have this particular term here, so n times S n is basically A, and we have the supremum over script theta of the likelihood function given by this.

(Refer Slide Time: 50:57)



So, we can formally thus write the likelihood ratio. So, this will imply that the likelihood ratio lambda is supremum over script theta naught of L mu L I am sorry L mu L and psi;

this L is corresponding to the factor (( )) supremum over script theta of L mu sigma so that this term, we can look at what terms we had got earlier, use this form here, and use this form here, in order to write this particular likelihood; and so, we can use this form, and this form here in order to write the final form of this likelihood ratio using asymptotic theory; using asymptotic theory, we know that this lambda or minus 2 log lambda this follows asymptotically a central chi square on the degrees of freedom, which is dimension of script theta minus the dimension of script theta naught.

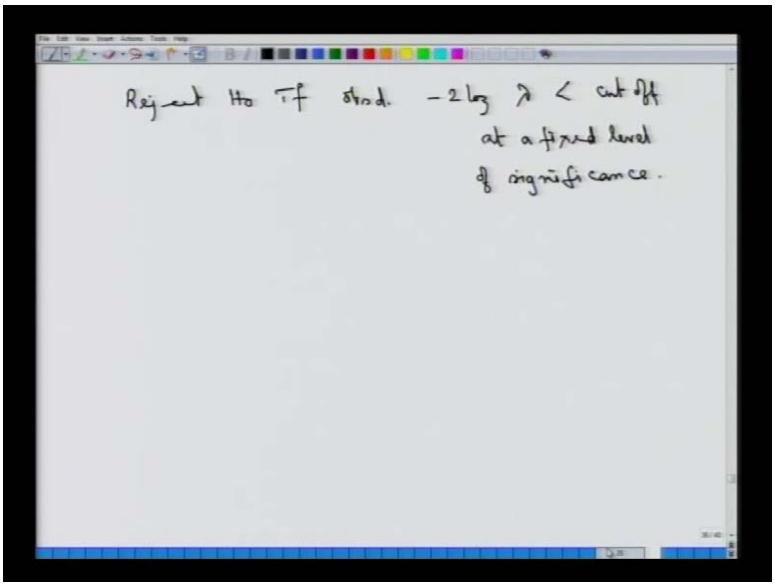
So, we will look at this likelihood ratio, and then using the asymptotic theory, we have minus 2 log lambda following asymptotically chi square central with degrees of freedom as dimension of theta minus dimension of theta naught; from a given data, we will obtain what is observed value of lambda, what is observed value of minus 2 log lambda, and looking at the appropriate degrees of freedom of that central chi square, we will actually look at acceptance or rejection of null hypothesis. So, what we are going to we are going to require this particular number here.

So, what we see is the dimension of script theta; now script theta was this space here. So, this has got p entries and the number of distinct elements of sigma is going to be the dimension of this script theta; so that this is equal to p for mu and p into p plus 1 by 2 for the sigma matrix. Now if you look at the dimension of script theta naught; script theta naught says that this sigma, so this H naught is what we have giving us L L dash plus psi. Now, the dimension of script theta naught, it still has though the components of mean vector; so, that is p of them. And then we will have the elements here, and the elements here, which are going to give us, this is p m; p m is the number of terms corresponding to the factor loading matrix L, this plus the p diagonal entries of psi are going to give us this.

Now, the dimension of script theta naught is something less, because of the type of restriction that we had imposed; now this is thus going to be the number of restrictions, which is this. Now, what is or how is this one coming? This basically is coming that from the number of restrictions, as we take M L E under script theta naught with the restriction of a diagonal matrix. So, since we have the diagonal matrix of that m by m matrix, which is you can just go back a little bit, this is basically the restriction that we plug in. So, this L transpose psi inverse L, this matrix is the diagonal matrix; and hence, these are the number of restrictions that we have in computing the maximum likelihood

estimators. And hence, we will have the dimension of script theta as it is given by this and the dimensional script theta naught to be equal to this. So, there is no problem in computing the degrees of freedom of this central chi square, which is a asymptotic distribution of minus 2 log lambda.

(Refer Slide Time: 55:47)



So, we will thus reject null hypothesis; now if we look at this likelihood ratio, this is the likelihood ratio; so when we are going to reject the null hypothesis, if this particular contribution the numerator terms supremum over script theta naught is too small with respect to supremum over script theta, then we usually reject the null hypothesis. So, we will reject null hypothesis, if observed minus 2 log lambda is less than the cut off at a fixed level of significance. So, that concludes actually this testing procedure, which was based on the maximum based on the large sample theory of the likelihood ratio test for the number of number of common factors that we will be choosing in an m factor model. So, that concludes our discussion with about the factor analysis, there are other concepts in factor analysis like factor rotation, which also occupies an important place, but we will end the concept of factor analysis here from the next lecture, we will look at the canonical correlation analysis.