Applied Multivariate Analysis

prof. Amit Mitra

prof. Shramishtha Mitra

Department of Mathematics and Statistics

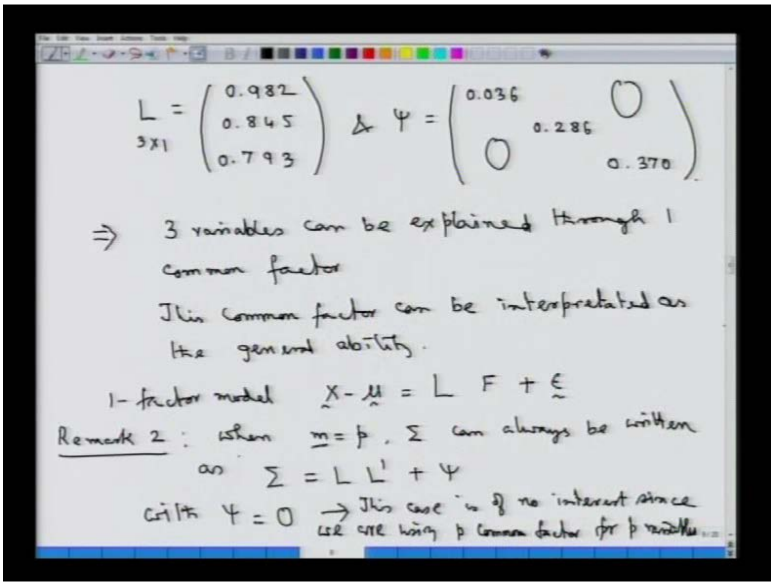
Indian Institute of Technology, Kanpur

Lecture No. # 38

Factor Analysis

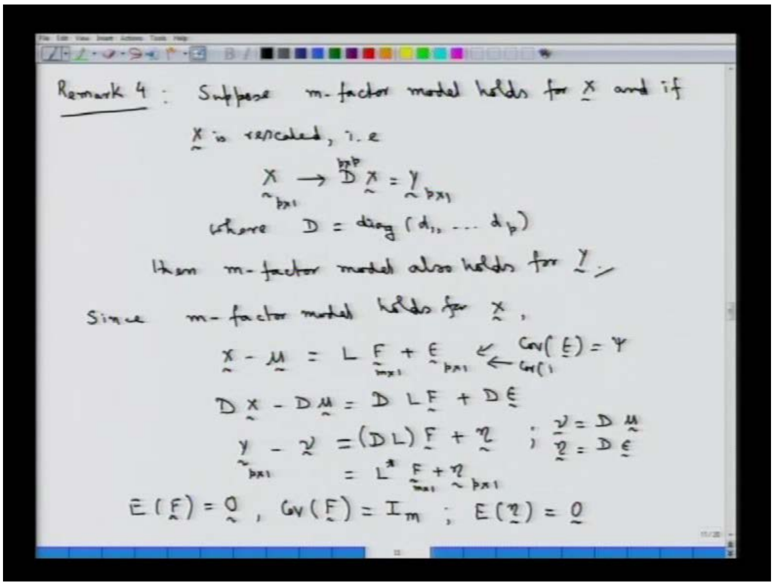
In this lecture, we will continue or discussion on factor analysis. In the last lecture, we had the given some preliminary introduction about factor analysis. We had also looked at as an example, if we have a covariance matrix sigma; how to verify, whether a particular m order factor model holds for such a sigma matrix or not. We had also seen some important results concerning factor analysis. Specifically we had these remarks at the end of the example.

(Refer Slide Time: 00:44)



That remark two had the when we had said that if we take m equal to p, then sigma can always be return as sigma equal to L L dash plus psi. Thus an m factor model will always hold in such a situation. In the next remark, we had seen how the reduction in the number of parameters is affected, when we have in a factor model.

(Refer Slide Time: 01:09)



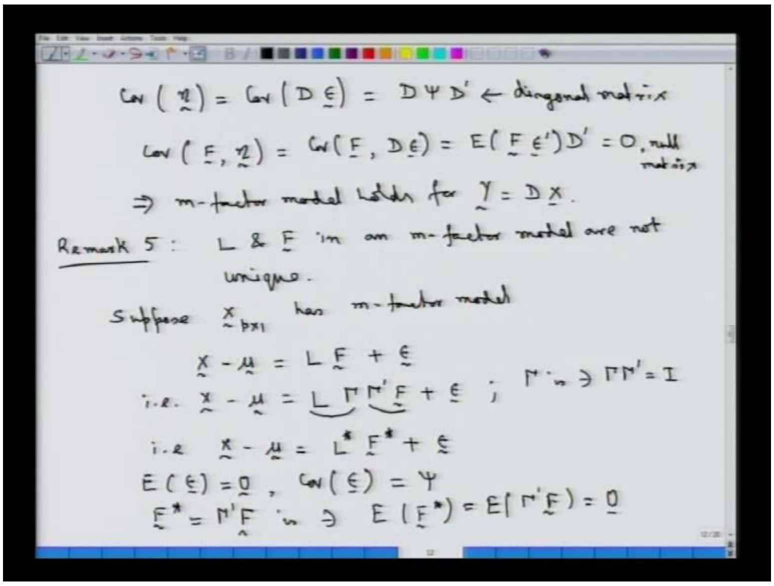
Now, let us look at the next important thing which, goes as remark four. (No audio from 01:13 to 01:19) Suppose m factor model holds for X, and if X is rescale, that is if X is transformed to D X, wherein this D is a diagonal matrix. Remember this is p by 1. So, this is diagonal matrix D 1, D 2, D p. Then, m factor model also holds for the rescale variable - that is Y, so we have got this D to be the diagonal matrix which is p by p order. So, the m factor model will also hold for Y.

Now, let us see Y is that. So now, this is what the remark says, now since m factor model holds for X, holds for X. We can write this X as X minus mu; where mu is mean vector this is equal to L F plus epsilon - where L is loading matrix, F is a vector of m specific factors. I am sorry F is a vector of m common factors, and epsilon is a vector of p specific factors. So, this is what is the setup for the factor analysis.

Now, if we pre multiply this equation by this diagonal matrix D, what we get is that this D X minus D mu that is equal to D L F plus D epsilon. Now, this D X we had earlier denoted by Y. So, let the p equal to Y, and let us denote this D mu by nu that is equal to… Let us write this as D L F this write this as eta, wherein what we have used in this mu equal to D times mu, and this eta vector is D times epsilon right.

Now, this particular form here will represent now, this is a p by one-dimensional random vector here. Now, this we can denote as some L star say that F plus eta. So, this looks as if, it is an m factor model for this random vector Y, provided the assumptions that we had a usually in mind for the factor analysis model holds. Now, the vector of the common factors remains exactly the same. So, this is m by 1 vector. Now, what is order this eta - eta as it is define, it is p by 1. Now, expectation of F of course, nothing as changed from the previous equation, this will be equal to a null vector. Then the covariance matrix of this F vector, that would be an identity matrix - identity matrix of order m, because this is the m factor model for the random vector X.

(Refer Slide Time: 05:14)



Now, concerning this eta, expectation of this eta vector will be expectation of D time's epsilon. So, that will be equal to a null vector. And furthermore this covariance matrix of eta, this is equal to the covariance matrix of D times, this epsilon vector. Now, epsilon as it is given here, this epsilon will have a covariance structure, covariance of epsilon equal to psi matrix which is diagonal matrix, that is what is assumption for a name factor model.

So, that we will have here as D psi matrix times D transpose. So, what would be the characteristic of this particular matrix. This matrix will also be diagonal matrix, as we have this D, and also D prime which is exactly the same. So, they are diagonal matrix, psi the starting matrix is a diagonal matrix. And also, we will have the covariance between F and eta, this would be covariance between F- F is unchanged.

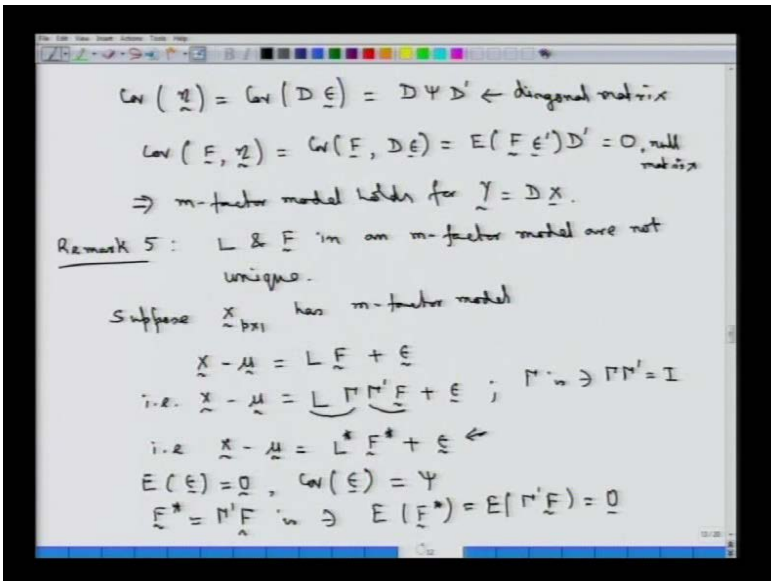
So, this is D epsilon. Now, expectation of F is equal to 0. So, this is equal to expectation of F epsilon transpose times, this D transpose. Because this F an epsilon are coming from the original m factor model, we will have in this particular model, further mode that covariance between F and epsilon. This would be equal to a null matrix, and hence this is what we will also have a null matrix. Thus we see that, if we are having and m factor model to hold for X, then F X is rescale that is X is transform to D X with D a diagonal matrix. We have been able to write this Y minus nu - nu is a expectation vector of this Y vector, which is equal to L star F times eta. Where in this F an eta satisfies require conditions for an m factor model two hold.

So, this will imply that m factor model wholes for Y equal to D X right. Now, we will look at the next important remark, this would be remark number five. Which will say that L and F in and m factor model are not unique. That is it sticks that if we have a random vector X, and we are looking at expressing that random vector in terms of a m factor model. Then this L is what is matrix of factor loadings, and F is the vector of our m common factors, the choice of L N F are not unique. Now, Y do is say so, let us try to understand what we are trying to achieve. Suppose X this is p by 1, has m factor model or an m factor model wholes for X will be able to right, X minus mu to be equal to L F plus epsilon with the corresponding assumption on F N epsilon to hold.

Now, on the right hand side, if we introduce and orthogonal matrix gamma - gamma transpose the nothing will change as such, where in this gamma is such that, it is a orthogonal matrix. So, that gamma - gamma transpose is equal to an identity matrix. Now, if we have this, we have this X minus mu return in terms of this, that is in other words we can remains this X minus mu equal to L star. Wherein L star is L times gamma, and this is say and F star vector - where F star is this gamma prime F, this plus epsilon.

So, we have a new loading here. And this F star. The new vector this is an m dimensional vector, it needs to a satisfy the conditions in order to say, in order that we can say that this is an m factor factor model for this random vector X. Now, epsilon there is no change in epsilon. So, this expectation of epsilon is still a null vector, and the covariance matrix of a epsilon is psi, the diagonal matrix which is coming from the previous formulation. Now, this F star is such that - F star is equal to gamma prime F vector. This is such that, expectation of this F star will be equal to expectation of this gamma prime F this will be a null vector, because this F has got a expectation as null vector.

(Refer Slide Time: 10:26)



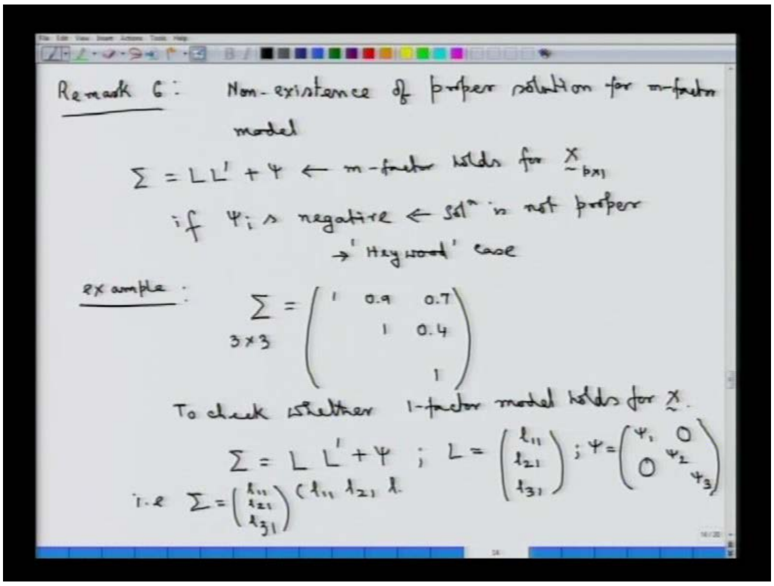
Then the covariance matrix of this F star will be equal to the covariance matrix of what we have to find is gamma prime F. So, this is gamma prime F, this will be equal to gamma prime this will be equal to gamma prime. Then covariance matrix of F times gamma. Now covariance matrix of F, because F is the vector of common factors coming from the m factor model. So, this is an identity matrix. So, this will be gamma prime gamma this will be an identity matrix right.

So, this is what is concerning the covariance matrix of F, and furthermore the covariance between this epsilon, because epsilon is unchanged here. So, we need to look at the covariance matrix of epsilon, and F star. So, covariance matrix of epsilon and F star - this is equal to the covariance matrix of epsilon, and this gamma prime F. This is equal to expectation of epsilon F prime, this would be a F prime F prime times this gamma matrix.

Now, the relationship between epsilon and F. F is the vector of common factors in the original m factor model, and hence we will have the covariance between epsilon and F to be equal to 0. And null matrix that multiplied by this gamma is also in a null matrix. So that, if we have return this particular model as in here, we are having F star such that expectation of F star is equal to 0, covariance matrix of a F star is an identity matrix of order m. And the covariance matrix of epsilon, and F star that is equal to a null matrix. Epsilon of course, is unchanged and hence that is got expectation equal to a null vector, and covariance matrix diagonal psi matrix - this will imply that this X minus mu equal to L star F star plus epsilon is an m factor model for X.

So, what we have we what are we trying to see, we are trying to see that this is m factor model for X with the loading matrix as L, and the vector of common factors as F. Now, the same can be expressed in terms of another L star, where L star is just equal to L times gamma matrix, where gamma is orthogonal matrix. So, this also has this representation. So, we are a different loading matrix L star, then the original starting L. And we have a different a vector of common factors F star, which is different from the starting F. So, the choice of L, and F is definitely not unique. No in order to make this particular choice of L N F unique, some additional conditions are sometimes impost a like the following condition - some conditions are imposed. So, as to have the m factor model unique. For example, one such condition is that L prime psi inverse L to be a diagonal matrix. So, such additional conditions may be imposed on the model. So, as to have the choice of the L, and the corresponding F vector to be unique.

(Refer Slide Time: 14:53)



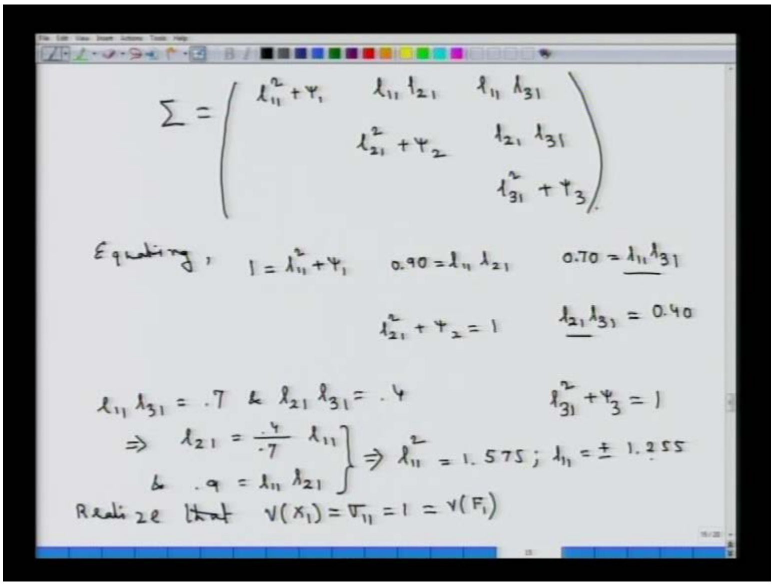
Now, in the next remark which is remark number 6, which remark number 6 talks about non-existence non-existence of proper solution for m factor models. Now, in some situation, suppose we starts from a variance, covariance matrix as in sigma. We might get after the solution, now let me just write it. Sigma equal to L L dash plus psi is what would lead us to believing that an m factor model wholes for the original set of random variables, p dimensional X. Now, in some situations starting from a sigma matrix, we might still be able to solve this particular equation, but we might be getting psi i is… So, if we have psi i's negative. Then, the solution is not a proper solution. Why ((

)) so, what are psi i's - psi i's other specific variances. So, those are the variances of specific factors.

Now, they cannot be negative, and hence if in some situation by solving such an equation in order to verify, whether and in factor model holds for X. If we get in the solution that psi i is an negative. Then the solution is not a proper solution. Now, such a situation is refer to as the Heywood case. So, the Heywood case basically tells us that, in order to the get this solution if we get psi i is negative. That solution is not a proper solution, and the terminology that is use in order to a actually say such a case, you will say that it has some property like what is call the Heywood case right. Let us look at an example of such a Heywood case, where the proper solution will not exist.

So, we take a sigma matrix which is 3 by 3 matrix, which is having one in the diagonal, so it is basically variance covariance matrix of standardize variables. And we take the following values 0 .9, 0 .7 and 0.4. So, this is a starting covariance matrix. We are trying to c, to check whether one factor model holds for the random vector X, which has this as the covariance matrix right. Now, in order to do that we need to frame the following equation, which says a sigma equal to L L dash plus psi; where this L is going to be equal to, because we have seen that whether a one factor model holds. So, this is an l 1 1, l 2 1, l 3 1, and this psi is the diagonal matrix with psi 1, psi 2, and psi 3 as the three diagonal entries. Now, if we plug in this particular value. We will have this sigma equal to our l 1 1, l 2 1, l 3 1, that into its transpose. So, its l 1 1, l 2 1, l 3 1 this plus this psi matrix.

(Refer Slide Time: 18:53)

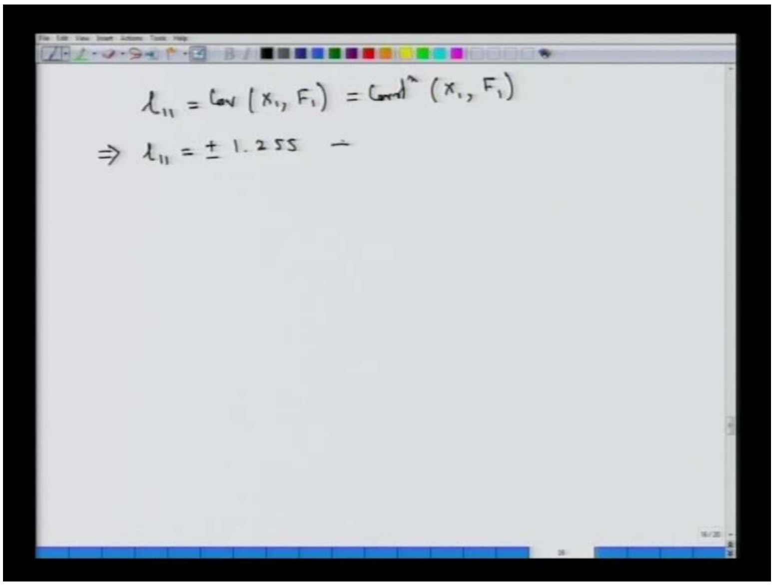


So, we will have this sigma to be equal to if we look at this particular multiplication, and then at the psi matrix to that multiplied vector, what will be getting is l 1 1 square plus psi 1 on the 1 1 th element. l 1 1, l 2 1, l 1 1, l 3 1 this is l 2 1 square plus psi 2, and this is l 2 1, l 3 1 and the 3 3 th element is l 3 1 square plus psi 3. Now, we know what this particular sigma matrix. So, equating what we get is the following 1 equal to l 1 1 square plus psi 1, because the 1 1 th entry of this sigma matrix is equal to 1. The other values also gives us the following that 0.9 0 is equal to your l 1 1, l 2 1. Then we have the value as 0.7 0 that is equal to l 11, l 3 1. This entry l 2 1 square plus psi 2 - this is equal to 1, and l 2 1, l 3 1 that is equal to the given value which is 0.40. And we have this l 3 1 square plus psi 3, that is also equal to 1.

So, we need to solve this particular set here, and then come up with the values of l 1 1, l 2 1, l 3 1 and psi 1, psi 2, and psi 3, if we use first these two equations. This l 1 1, l 3 1 that is equal to 0.7, and l 2 1, l 3 1 that is equal to 0.4. So, this will simply, because l 3 1 is common out here, we will have this l 2 1 equal to 0.4 by 0.7 times l 1 1 right. And furthermore, what we have from this equation is 0.9 equal to l 1 1 times l 2 1. So, these two collectively would imply or rather give us the solutions, this will lead us to l 1 1 square that is equal to 1.575. That is l 1 1, it will be equal to plus or minus this square root of this particular number which turns out to be 1.255 right.

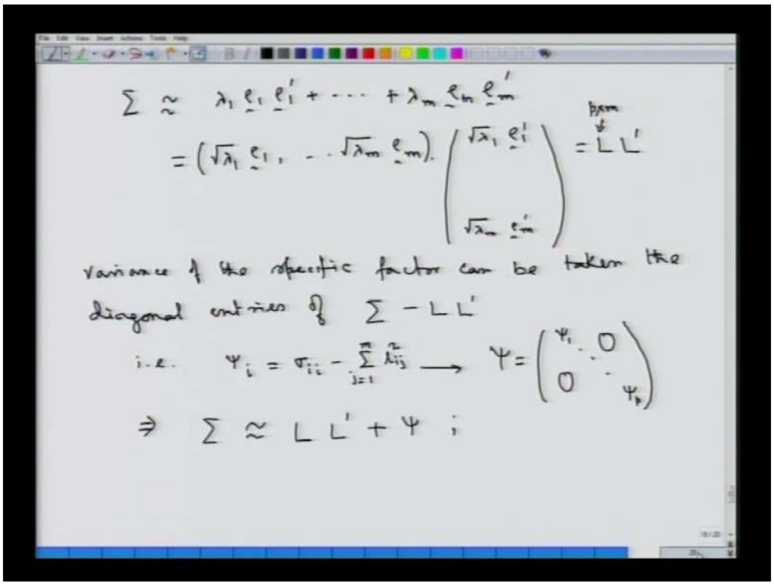
So, we have a solution l l l equal to this. Now, we will see why this is not a feasible solution, now realize that this variance of X 1 is equal to sigma 1 1. What is that equal to from the given sigma matrix that is equal to 1. So, this is equal to 1, which is also equal to when we are looking at this variance of the first common factor. l 1 1 is a loading of X 1 on F 1. Now, the two component X 1, and F 1 both of them have a variance equal to one. And what we have seen earlier is that, this L ij is a covariance between X i and F j.

(Refer Slide Time: 22:15)



So, this l 1 1 is nothing, but covariance between X 1 and F 1. Since the variance of a X 1 and F 1 are both equal to 1. We will have this as also the correlation between X 1 and F 1. Now, the solution what we have got is l 1 1 equal to plus or minus 1.22, which is an absorb value. So, this will imply that l 1 1 equal to plus or minus 1.255 is an absent. So, if we have this lambda m plus 1 to up to lambda p close to 0. We can neglect the contribution of these eigen values lambda m plus 1 to lambda p to sigma, that is in the spectral decomposition as in here. We have from lambda one to up to lambda p, we are assuming that beyond the certain point m - lambda m plus 1 to up to lambda p are negligible. They are close to 0. And hence, we can neglect the contribution of these terms - the last p minus m terms.

(Refer Slide Time: 23:37)



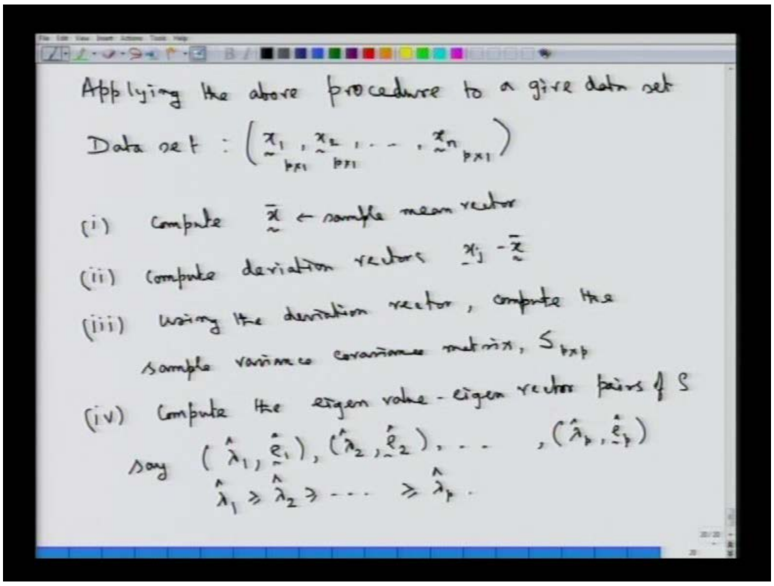
And then, we can say that let in such a situations sigma is approximately equal to our first m terms, that is lambda 1 e 1 e 1 prime plus lambda m e m prime. Now, we can write this particular expression here, up to m terms. This is a approximate, because we have chopped off from lambda m plus 1 to up to lambda p those contributions. So, we can right similar that previous set up, that this is equal to root over this - root over lambda m e m; this is the transpose of it - lambda 1 e 1 prime root over lambda m e m prime right.

So, if sigma is approximately equal to this. We can take the variance of the specific factors, variance of the specific factors can be taken as diagonal entries of sigma. Now, we will write this as L L dash, where this L matrix is this particular matrix which is p by m. So, this matrix is what we have writing a p by m, this is m by p it is transpose, by choosing the diagonal entries of sigma minus L L dash. That is what we are having is this psi i equal to sigma i i minus this L I ij square for j equal to 1 to up to m.

So, we will look at this sigma minus this L, L transpose and then from that different matrix; if we pick up this the diagonal elements, and then say that our psi i as going to be that sigma i i. This sigma i is diagonal entry of this. And this quantities L L dash is diagonal quantity - i th diagonal quantity. And this will imply that we will have this sigma, approximately equal to L L dash plus the psi matrix wherein using this psi is here.

We will form the psi matrix, which is psi 1, psi 2, and psi p - the specific variances. All this quantities are zero's right. Now, in this approximation, note that the diagonal entries of sigma would exactly match with the diagonal entries of L L dash plus psi, because L L dash is up to this particular a term e m or lambda m terms. And psi is what we are taking as the diagonal entries of this particular difference matrix, and hence this sigma being approximated by L L dash plus psi. This approximation - in this approximation, the diagonal entries will be exactly equal to 0, and the half diagonal entries of sigma, and L L dash plus psi will differ. Now, we will use this particular concept in order to estimate L, and the corresponding psi matrix from the data.

(Refer Slide Time: 27:23)



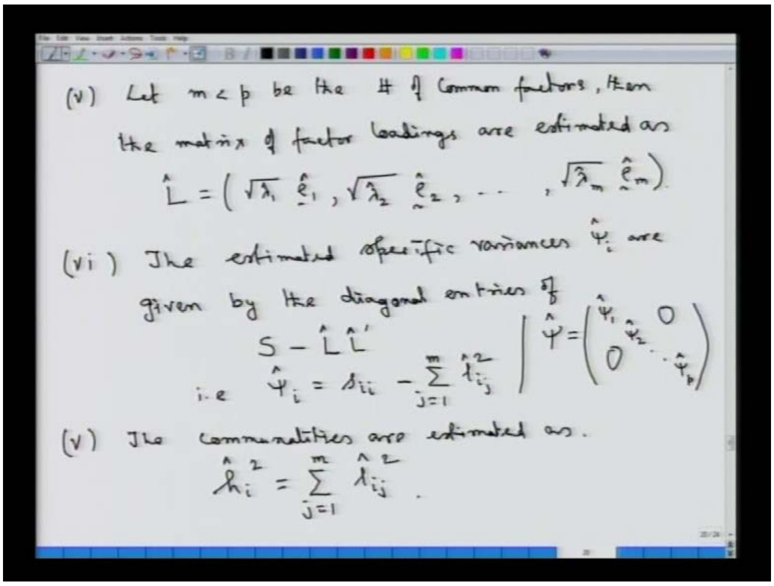
So, what we will now, look at is applying the above procedure above procedure to a given data set. Now, the data set is comprising of x 1. So, that is the first p dimensional realization, x 2 this is the second p dimensional realization, and this is say x n which is the n th p dimensional realization. So, these are the realizations which we have as the data set. So, for any practical purpose as such, where we do not have any idea about what is the covariance matrix of the underling random variables. We will just be having this x 1, x 2, x n as the given data set.

Now, given this particular data set, we will apply the previous concept when we have looked at that is sigma matrix, and then give this algorithm for actually estimation of the loading matrix. And the psi the matrix are specific factors variances. So, at the first step from this given data, we will compute, I will just give this step by step procedure. What will first compute, it is the sample mean vector - the observed sample mean vector. Given this is calculated - we will calculate the deviation vectors, deviation vectors are given by this x j minus x bar quantities. Now, using this deviation vectors or otherwise; using the deviation vectors compute the sample variance covariance matrix, the sample variance covariance matrix say is given by this capital S.

Now, once we have this now this is going be the estimate, as such of this sigma the population variance covariance matrix. Now, we will compute the eigen value - eigen vector pairs, compute the eigen value - eigen vector pairs of S. Say those are given by lambda 1 hat, e 1 hat - lambda 2 hat, e 2 hat, and now this is a p dimensional observations p dimensional observations and of them. So, the variance covariance matrix is also p dimensional, and we will have these as the corresponding eigen values, and eigen vector pairs. These are given caps, because we had a estimated sigma by S and we look at lambda one hat as an estimate of lambda 1, which was the eigen value corresponding to the, the largest eigen value corresponding to the sigma matrix.

Now here, we will have similar relationship between the lambda i hats. So, lambda 1 hat is greater than or equal to lambda 2 hat is greater than or equal to lambda p hat.

(Refer Slide Time: 31:30)



Now, we will use this estimated eigen values, and the S corresponding estimated eigen vectors in order to. So, in the fifth step here, let us say that m less than p be the number of common factors, that we are going to choose. Then the matrix of factor loadings are estimated as this L hat, which is going to be equal to root over lambda 1 hat times this e 1 hat root over lambda 2 hat times e 2 hat and so on. This is root over lambda m hat, where m is the number of common factor is that we are choosing, and this is going to be this e m hat right.

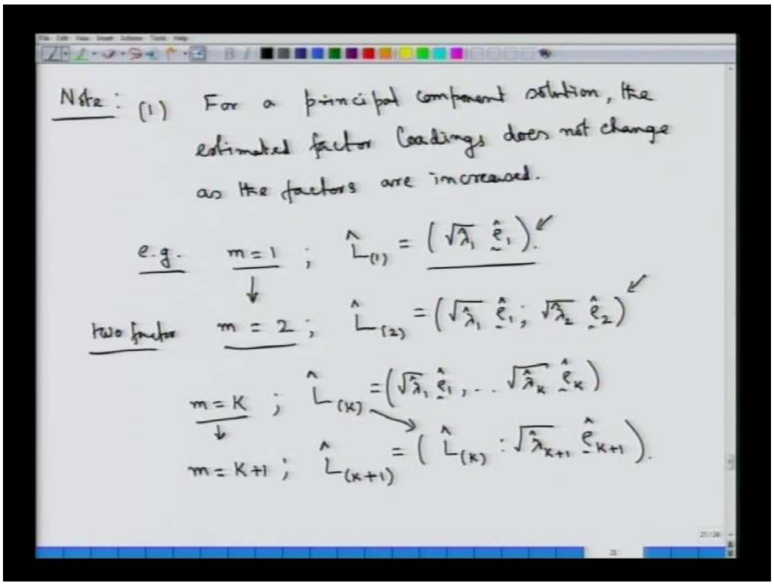
Now, why is this so, because if we look at this formulation here, what we had chosen was this matrix truncated up to the m th point. And since, we are going to have the estimates from the sample estimated sample variance covariance matrix, from just the sample variance covariance matrix. And the eigen value eigen vector decomposition, if we have choosen m less than p to be the number of common factors. Then the matrix of factor loading are estimated by this. Now, once we have this factor loading matrix as this, the next step would be two estimate the specific variances.

So, what we will have is a following the estimated specific variances. The estimated specific variances psi i hats are given by the diagonal entries of, diagonal entries of which matrix. Now, it would be S minus L hat L hat transpose, why is that so, because in relationship with this particular relationship we had chosen, the variance of this specific factors as sigma minus L L dash. Now, sigma is a estimated by S, L is a estimated by L hat.

And hence, we will be using this S minus L hat L hat transpose, it is diagonal entries will be chosen as the specific, as the estimates of this specific variances psi i. That is we will have this psi i, psi i hat will be equal to s i i, where s i i is the diagonal entry of this S matrix this minus j equal to 1 to up to m l ij hat squares. Where l ij hat is the i,j th element of this L hat matrix. Now, lastly now that is what is the estimation, because L has been estimated as L hat corresponding to the m factor model, and this after this we will have this psi matrix to be estimated as psi hat matrix. Which will have the entries as psi 1 hat, psi 2 hat, and psi p hat; rest of the elements are zeros, which is given by this.

Last thing as result remark the communalities. The communalities are estimated as your are h i hat square which is just particular term, because s i i is equal to communality plus the specific variance, and hence this communalities for the different common factors are going to be given by l i hat square terms of this right. So, this is how from a given data vector, we will be able to estimate all all the things which are require, in order to an actually have an m factor model for the given data. So, this is a step wise procedure.

(Refer Slide Time: 36:25)



Now, under such a situation, we make note of the following facts. The first note says that for a principal component solution, for a principal component solution, the estimated factor loadings estimated factor loadings, does not change as the number of factors are increased.

What it is trying to convey is the following message, that suppose we have an m factor model - we have estimated the factor loadings, and the specific variances for an N factor model. If we one to go from N factor model, two N plus 1 factor model. Now, the previous factor loadings - the factor loadings for the first m factors will not change. If we are actually estimating the factor loading and specific variances, under this particular principal component solution approach. Now, why is that so. For example, this look at simple situation. Suppose we have an m factor model, if we have an m factor model then this L hat matrix say L 1 hat matrix, which this one signifies that it is a 1 factor model, this will be equal to what? This will be equal to root over lambda 1 hat times e1 hat.

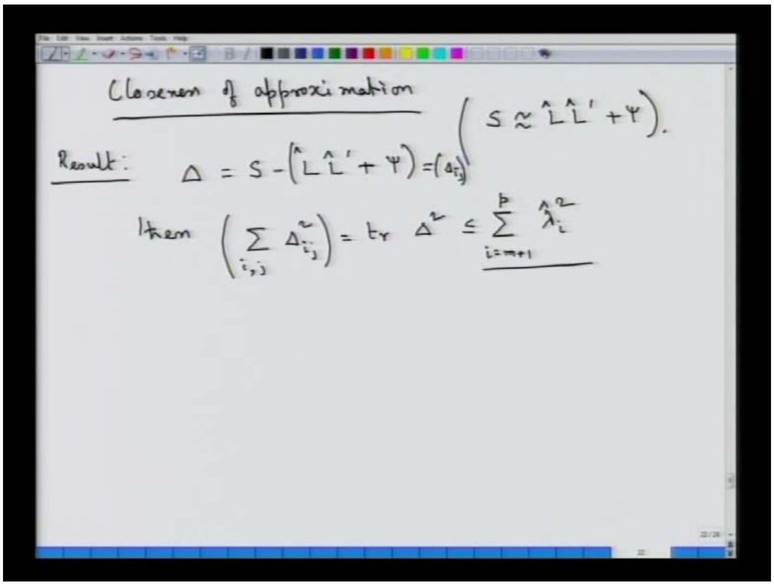
Now, if we one to go from the first factor one factor model to a two factor model. So, this is a two factor model, that is m equal to 2. Now, the loading matrix for m equal to 2 will be given by root over lambda one hat, e 1 hat, still its first column. And the second column is just augmented that you will have this as lambda 2 hat be 2 hat. So, what we observe is that, this was the factor loading matrix for a one component model. And this is the factor loading matrix for a two component model that is m equal to 2. So, the two component model has two columns – this is the first column is the factor loadings corresponding to a one component model, and hence it does not change when we are moving from a one factor model to a two factor model. In general, if we have m equal to k say; then this L k hat matrix will have it is entries as l 1 hat, e1 hat root over l lambda k hat e k hat.

Now, if from m equal to k, we want to move and move had and have a k plus 1 factor model for some reason. Then what we will be having as this factor loading matrix for this k plus one-dimensional factor model, to will just p this factor loading matrix corresponding to this k factor model. This will be augmented by one more column which is lambda k plus 1 hat times e k plus one hat.

So, as such as we are seen here that the factor - the previous factor loadings are not going to change. If we are moving from a lower order factor model say k th order factor model, to a k plus 1 th order factor model. Now, when we are using this principal component approach - principal component method for a estimation of L and psi. We are making some approximation, as we have seen in here, that we are going to approximate this L - this S. In terms of L L hat this one, just write it here. This S is being a approximated by L hat L hat transpose plus this psi hat matrix.

This is similar to that approximation that we had use for sigma matrix. So, this approximation is off what nature? If we look at the diagonal entries of S, they are going to match with the diagonal entries of this L hat L hat transpose plus this i hat matrix. Only the non-off diagonal entries of the two matrices S and L hat L hat transpose, and psi are going to differ.

(Refer Slide Time: 41:16)



So, the following result gives us a measure of closeness of approximation. Now, what is this approximation? This approximation is that S, we are approximating by L hat L hat transpose plus psi. Now, this is an important result which gives us the measure of closeness, some idea about the closeness of this approximation. If we denote by delta, this difference matrix which is S minus L hat L hat transpose plus psi. So, this is this delta matrix is going to measure the degree of closeness of this approximation. So, let us denote this, these elements as small delta i j's. Then this summation of the delta i j square some over i j, which is also going to be equal to trace of this delta square matrix, this is symmetric matrix. This is going to be less than or equal to summation lambda i terms, lambda i hat squares i equal to m plus 1 to up to p.

So, what it is says is that, this delta matrix - the matrix of differences comprising of delta i j as the i j th element of this delta matrix. This some of square of all these deviation matrix deviation matrix elements delta i j. So, this is the sum of square of all the deviation is going to be bounded, by it is less than or equal to summation i equal to m plus 1 to up to p lambda i hat squares. Now, this summation is what? The summation is the contribution of lambda i hat squares for the remaining for for the last p minus n eigen values. So, we had at the starting point said that, this type of method is going to what well, if we have the last p minus m eigen values to be negligible. And hence, in such a situation, this approximation would be very close. In the next lecture, we will look at proving this particular result.