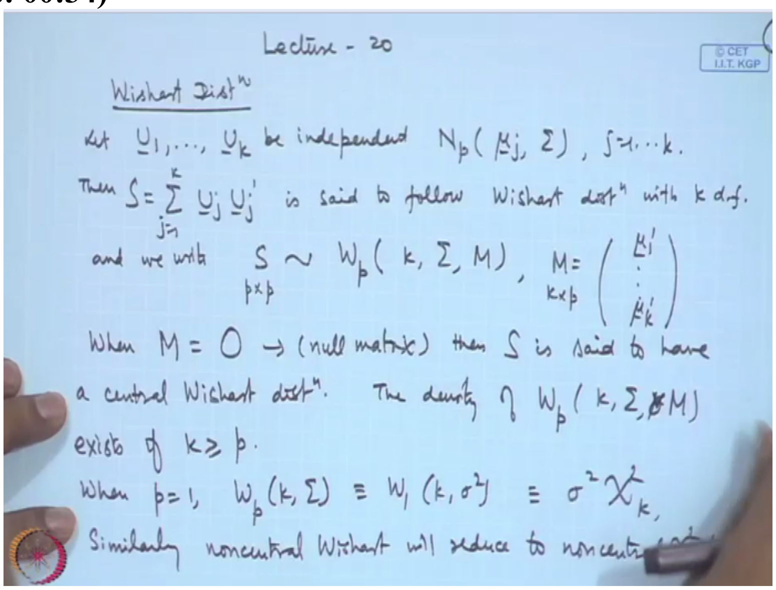
Statistical Methods for Scientists and Engineers
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology – Kharagpur

Lecture - 20
Multivariate Analysis – V

Now, I will discuss the variance, covariance matrix S the sample variance covariance matrix. So for that if we derive the distribution it will be a matrix distribution, it is called Wishart Distribution.

(Refer Slide Time: 00:34)

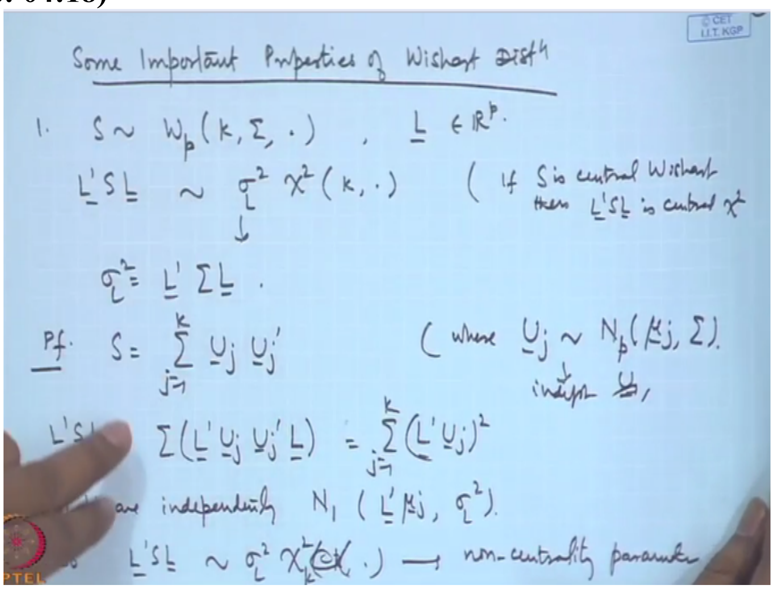


So we can consider this Wishart distribution as a generalization of the chi-square distribution in the Univariate case the sample variance had a chi-square distribution, in fact we wrote it in the form that sigma x/x square/sigma square that follows a chi-square distribution on m-1 degrees of freedom. So now we consider all the components of the dispersion sample dispersion matrix, so we are having sigma x1i-x1 bar square sigma x1i-x1 bar*x2i-x2 bar and so on.

So what is the distribution of that? So let us define the Wishart distribution. So let U1, U2, Uk be independent Np Mu, j, sigma where j=1 to k. Then we say that sigma Uj, Uj prime, j=1 to k, this is said to follow. Wishart distribution with k degrees of freedom and we write as following Wp k sigma, M. This S is p/p, so we let us write this as S here. And here M is the non-centrality matrix. This is of order k/p.

Now, when M= null matrix then S is said to have a central Wishart distribution. And the density function of-- so we write Wp K, sigma Mu M this will exist if k is > or = p. For p=1 Wp k, sigma that is in W1 k, sigma square that is sigma square, chi-square k. Similarly, non-central Wishart will reduce to non-central chi-square for p=1. Before talking about the density function of a Wishart distribution it is quite complicated actually.

(Refer Slide Time: 04:18)



So we firstly look at its properties like the case for multivariate normal distribution. So, some important properties of Wishart distribution. The first properties are that if I have S following Wishart with parameters p, k sigma and here I am not writing that M here because I can consider both the case of central and non-central here. And L is a fixed vector in the p dimensional space. Then L prime SL that will have sigma L square chi-square k.

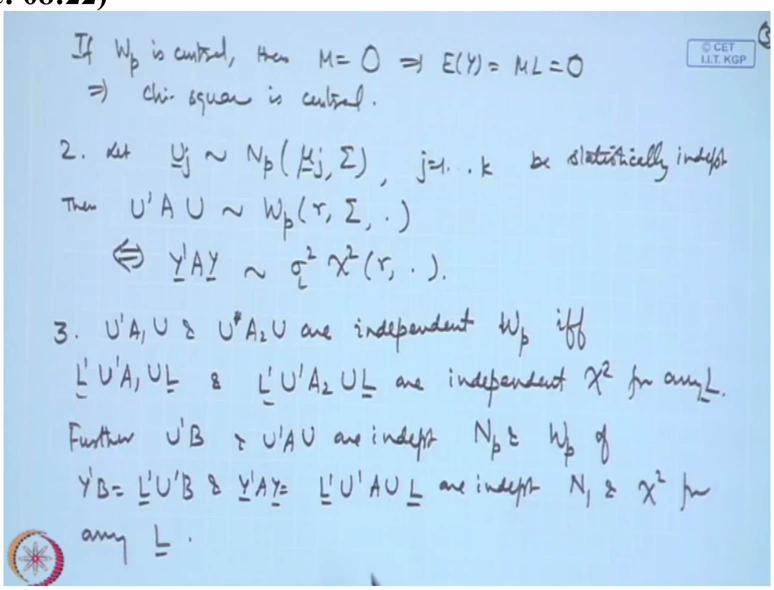
And again if S is central Wishart then L prime SL is central chi-square. And here the sigma L square I am defining to be L prime sigma L. For proof of this now we will make use of the non-central chi-square. So S is written as sigma Uj, Uj, prime j=1 to k where we are actually considering Uj as the multivariate normal. So if I consider and these are independents, okay U1, U2 but because that was a setup that we consider here, U1, U2, Uk are independently distributed.

So let us write here L prime SL that = sigma L prime Uj, Uj prime L that = sigma L prime Uj square j= 1 to k. So L prime Uj these are independent, these are independently distributed normal 1 L prime Mu j and sigma L square. So L prime SL, this will follow sigma L square, chi-square on k degree of freedom and this non-centrality parameter will come there as we discussed in the previous lecture.

That if I am considering x following Np Mu, I, then x prime has an non-central square distribution with p degrees of freedom and non-centrality parameter is given by summation Mu, I square/2. So if you use this then-- because what we are getting here this L prime Mu j they are univariate normal so of course we have put sigma L square if I divide by that then that will come here. So this result follows here.

Now if my original Wishart is central then M will be 0 so expectation of y that will be 0, so chi-square will be central. Let us write that also.

(Refer Slide Time: 08:22)



If Wishart is central then M is 0 null matrix, this implies expectation of y is ML that is 0, this implies chi-square is central. So we have shown a direct correspondence between a Wishart and Chi-square distribution as we have seen in the case of multivariate normal every linear combination is univariate normal. So here in place of linear combination it is quadratic form, so S is a positive (()) (09:07) matrix and considering L prime SL so this is a quadratic form.

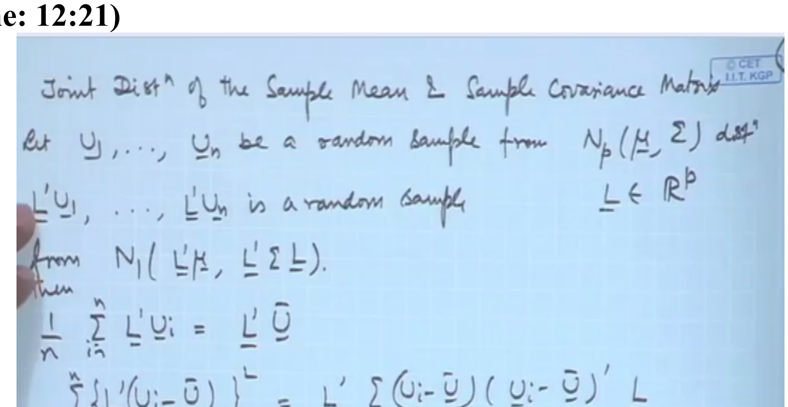
But the quadratic form will have a chi-square distribution. Let us look at the second property. So once again we are considering Uj following Np Mu, j sigma where j=1 to k, suppose they are independent then if I consider. Now in the previous one I defined what is U, I defined the matrix here as U, so if I use this U as components of U1, U2, Un then if I consider U prime A U then this will have Wishart.

This is equivalent to saying Y prime AY; this will be sigma L square chi-square r.. I will skip the proof of this because it involving lot of terms and I do not want to make this course extremely theoretically work. Let us move to further properties of the-- U prime A1, U and U prime A2, U they are independent, independent Wishart iff L prime U prime A1, UL and L prime U prime A2, UL they are independent chi-square for any L.

Further U prime B and U prime AU are independent Np and Wishart P if Y prime B that = L prime U prime B and Y prime A Y that = to L prime U prime A U L they are independent N1 and chi-square for any L. This relation is actually similar to the relation that in the sampling form univariate normal distribution the sample mean and the sample variance are independently distributed. So this is of similar nature.

So now let us talk about the Joint Distribution of the Sample mean and the Sample Covariance matrix.

(Refer Slide Time: 12:21)

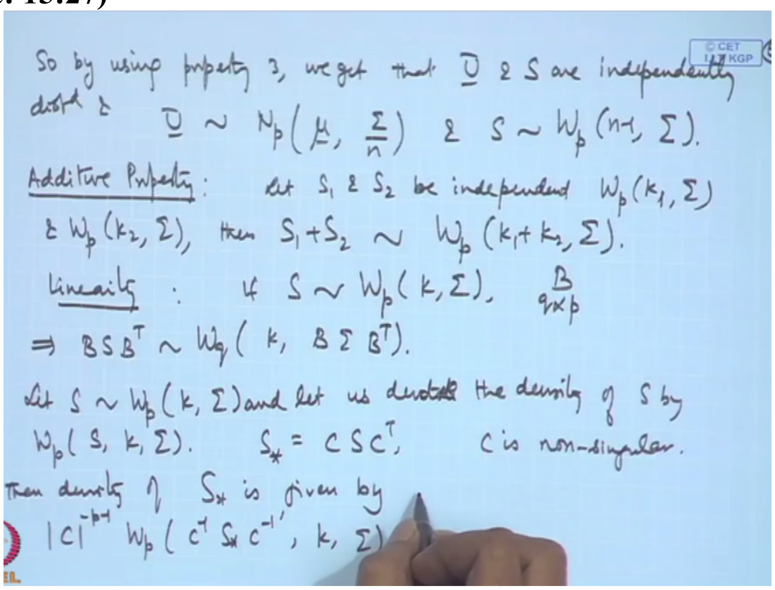


So let U1, U2, Un be a random sample from Np Mu, sigma distribution. Then if I consider L prime U1 and so on L prime Un for any L, L is a p dimensional vector then this is a random sample from N1 L prime Mu, L prime sigma L. So if I consider the sample mean 1/n sigma L prime Ui. And if I consider the sample covariance matrix L prime Ui-U bar square i=1 to n that = L prime sigma Ui-U bar Ui-U bar prime L that = L prime S L.

So from the distribution theory of a univariate normal distribution the sample mean and the sample variance covariance matrix are sample variance, sample covariance is independently distributed. They are independently distributed further L prime U that will be univariate normal L prime Mu, L prime sigma L by n and L prime SL will have L prime sigma L chi-square on n-1.

So now if we use this result this is iff and non-if here so we will get that U bar will have multivariate normal and S will have Wishart.

(Refer Slide Time: 15:27)



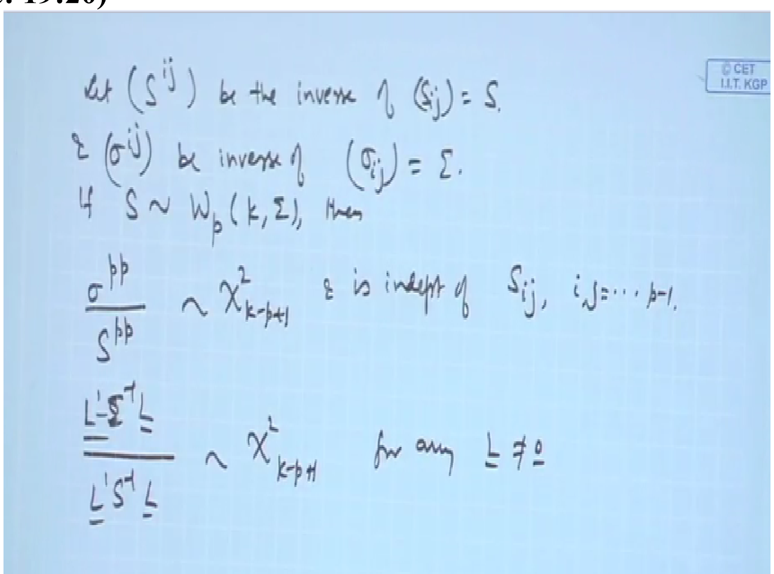
So by using property 3, we get that U bar and S are independently distributed and U bar will follow Np Mu, sigma/n and S will follow Wishart n-1 sigma. So this is a central Wishart distribution. Now like the additive property of chi-square distribution Wishart also has additive property. Let S1 and S2 be independent so Wishart k1, sigma and k2, sigma, then S1+S2 will follow Wishart k1+k2, sigma.

Once again we can prove this result by considering L prime S1 L and L prime S2 L, so there will be central chi-square and then there will be additive so it will become k1+k2. In the case of multivariate normal distribution, we have considered linear combination that means if I consider x as a Np and if I considering B as a q/p matrix then bx will have Nq distribution. Now a similar thing is proved Wishart also.

So this is linearity we can say. If S follow Wp, k, sigma and the B is the q/p matrix then BSB transpose that will follow Wishart q with B sigma B transpose. So the result will follow from definition of the Wishart distribution. Let us talk about the density part here. Let us follow Wishart k, sigma and let us denote the density of S/ say Wp S, k, sigma. Let us define S* that = CSC transpose where C is non-singular.

Then density of S* is given by, so by a transformation of this we get determinant of C to the power –p-1 Wp c inverse S* c inverse prime, k, sigma.

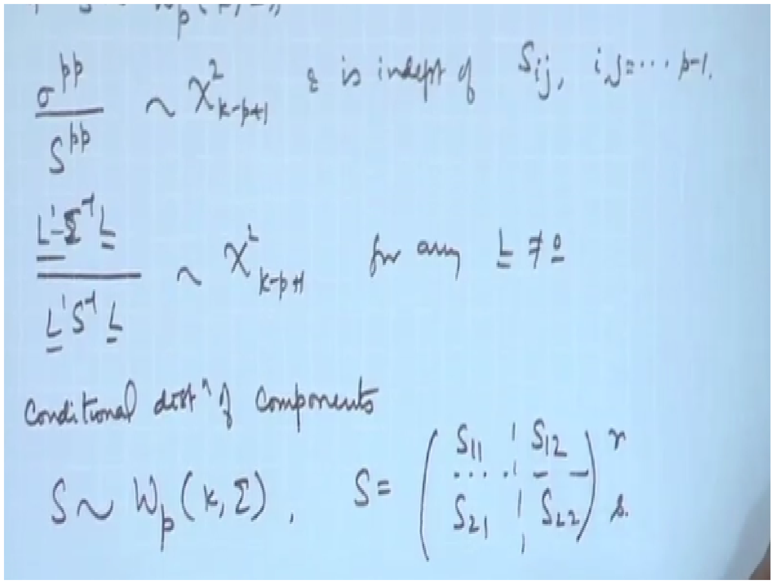
(Refer Slide Time: 19:20)



Let us S ij be the inverse of S ij that is equal to S and sigma ij be inverse of sigma ij that is equal to sigma. Now, if S is having the Wishart distribution then we have sigma pp/S pp that is following chi-square k-p+1and this is independent of S ij ij=1 to p-1. At the same time L prime sigma inverse L/L prime S Inverse L that follow chi-square k-p-p+1 for any L naught 0.

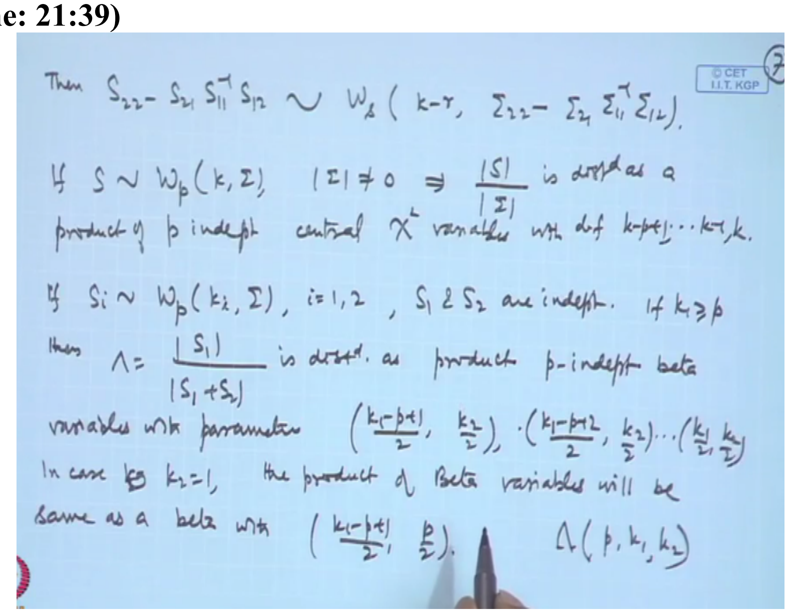
In the case of multivariate normal we have seen the conditional distribution, a similar thing is true for the Wishart also. Now, we have these properties I am stating without proofs because the proofs are quite involved using multivariate parts here.

(Refer Slide Time: 20:56)



So you should know the results here. So Conditional distribution of components. So suppose I assume Wishart distribution with parameter k and sigma and S is partition S11, S12, S21 and S22. Suppose these are r components and these are S components here. Similarly, here this is r components this is S components here.

(Refer Slide Time: 21:39)



Then S22-S21 S11 inverse S12 that has a Wishart on S k-r, sigma22-sigma21-sigma11 inverse sigma12. One more representation of the decomposition of the Wishart determinant is given by the following. If I say S follows Wishart k, sigma determinant of sigma is non-zero then determinant of S/determinant of sigma is distributed as a product of p independent central chi-square variables with degrees of freedom k-p+1 and so on k-1, k.

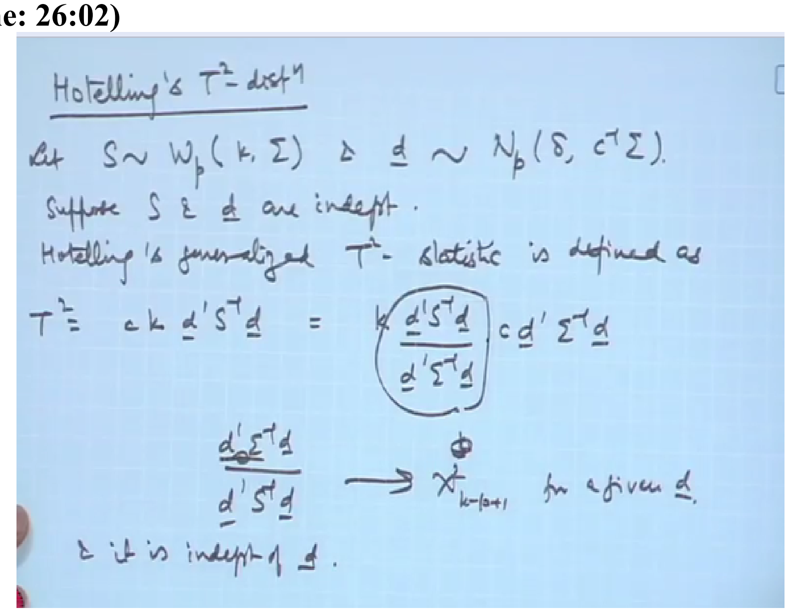
And if Si follow Wishart ki, sigma i=1,2. S1 and S2 are independent, if k1 is > or = p then lambda that = determinant of S1/S1+S2 is distributed as product of p independent beta variables k1-p+1/2, k2/2, k1-p+2/2, k2/2 and so on, k1/2, k2/2. In case, k2=1, the product of beta variables is will be same as a beta with k1-p+1/2, p/2. So this distribution is denoted the lambda p, k1, k2.

So these distributions are use in the study of the covariance coefficient etcetera which I am not paying too much attention at this point here. Now we move to another distribution which is extremely useful. So here we have introduced a Wishart distribution as a generalization of a chi-square distribution and we looked at some of the properties.

So in the testing for the variance, covariance matrix of a multivariate normal distribution you can make use of this and the test are other information will be based Wishart distribution. Let us also consider the concept of t-distribution for the univariate distribution. In the concept of t-distribution came when we are considering the inference on mean but variance is unknown, so we are divided by estimate of sigma that is S there and that was said to have a t-distribution.

Now a similar concept to exist are the when we are considering inference on the mean vector or the multivariate normal distribution. And when the variance covariance matrix is not known. So as a generalization of t-distribution we are considering Hotelling's T square distribution.

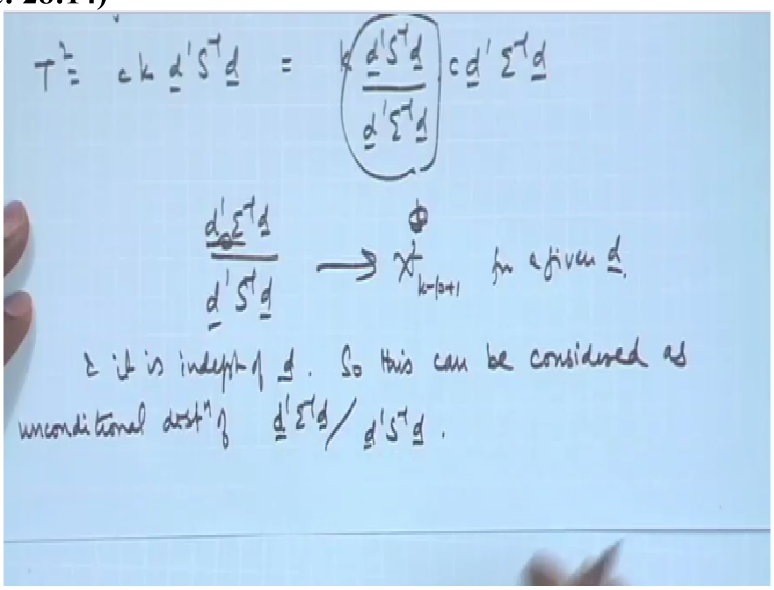
(Refer Slide Time: 26:02)



So let us consider say S following Wp k, sigma and say d follows Np delta c inverse sigma. Suppose S and d are independent, in that case this Hotelling's generalized T square statistics is defined as T square=c k d prime S inverse d. Now this we can interpret as k d prime S inverse d divided by d prime sigma inverse d into d prime sigma inverse d. And this c also we write here. Now if you look at this term here, this is having chi-square k-p+1 for a given d.

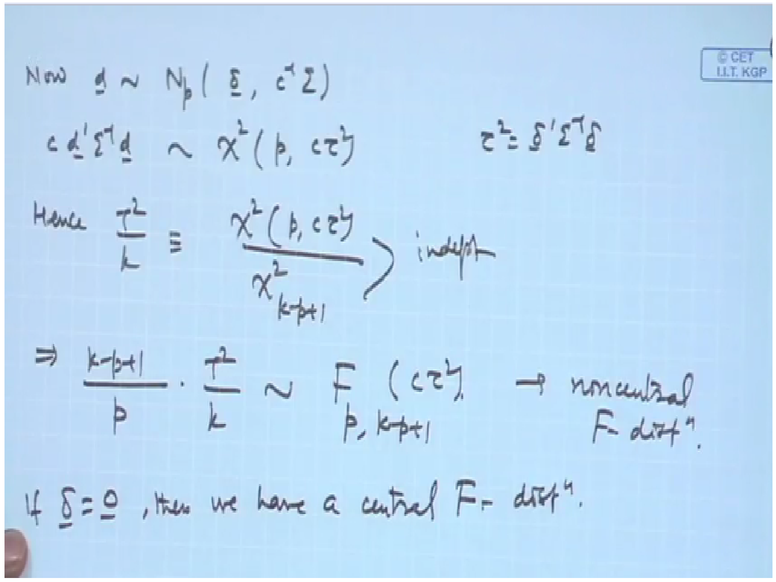
So this property we did at the earlier, this was L prime sigma L/L prime SL. So we have written this property but this will have a chi-square distribution, rather arrears of this, not this d prime sigma inverse d/d prime S inverse d, this is having a chi-square distribution on k-p+1 degrees of freedom. And it is independent of d.

(Refer Slide Time: 28:14)



So this can also be considered as unconditional distribution of d prime sigma inverse d/d prime S inverse d.

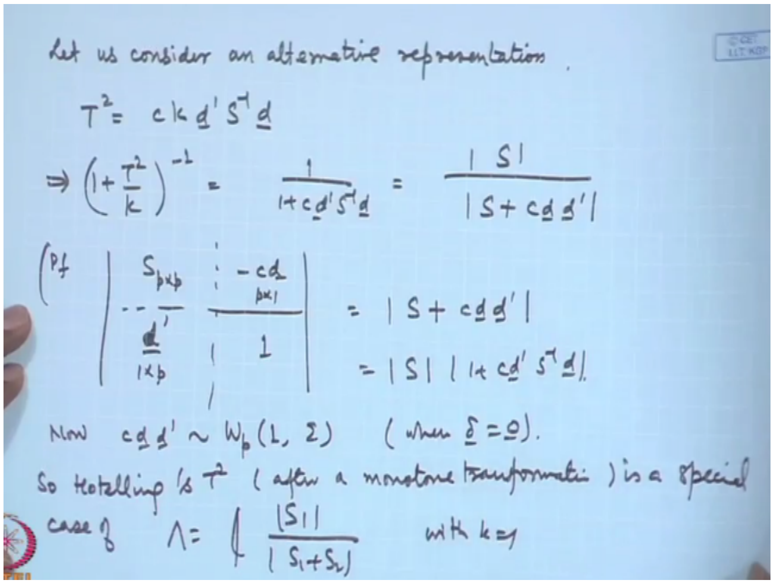
(Refer Slide Time: 28:43)



Now, d is following multivariate normal. So if I consider the c, d prime sigma inverse d that will have chi-square with p and c Tau square where Tau square=delta prime sigma inverse delta. Hence your T square/k is actually chi-square p, c Tau square/chi-square k-p+1. So these are ratio, so this is something like a non-central F distribution which I used in the previous class, that if I consider ratio of--

If I have a central chi-square and in the denominator I have an in the denominator I have central chi-square and in the numerator I have a non-central chi-square, then the ratio is a non-central F. So actually we are able to get come to that situation now that this is and these 2 are independent. Basically we are writing here k-p+1/p T square/k follows F distribution so that is non-central F. If delta = 0 then we have a central F. Let us consider an alternative representation.

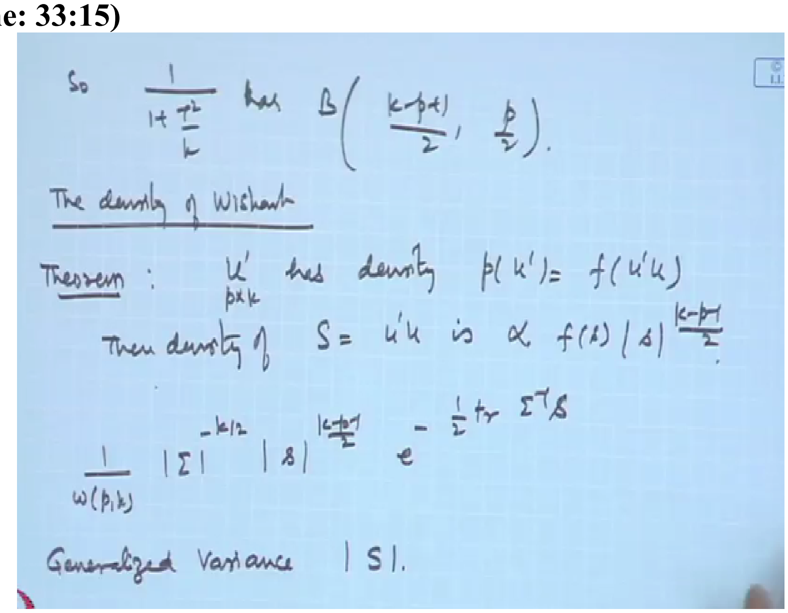
(Refer Slide Time: 31:00)



In the alternative representation let us consider as T square = c k d prime S inverse d. So we can write 1+T square/k inverse = 1/1+c d prime S inverse d this = determinant of S/determinant of S+cd d prime. See to prove this statement we can actually consider c s which is p/p-cd which is p/1, d prime which is of course 1/p and 1. Let us consider the determinant then this can be determinant of s+c d d prime which I can write as determinant of s*1+c d prime s inverse d.

Now c d d prime that will have Wishart 1 sigma when delta = null. So Hotelling's T square after monotone transformation is a special case of lambda=S1/S1+S2 with k=1.

(Refer Slide Time: 33:15)

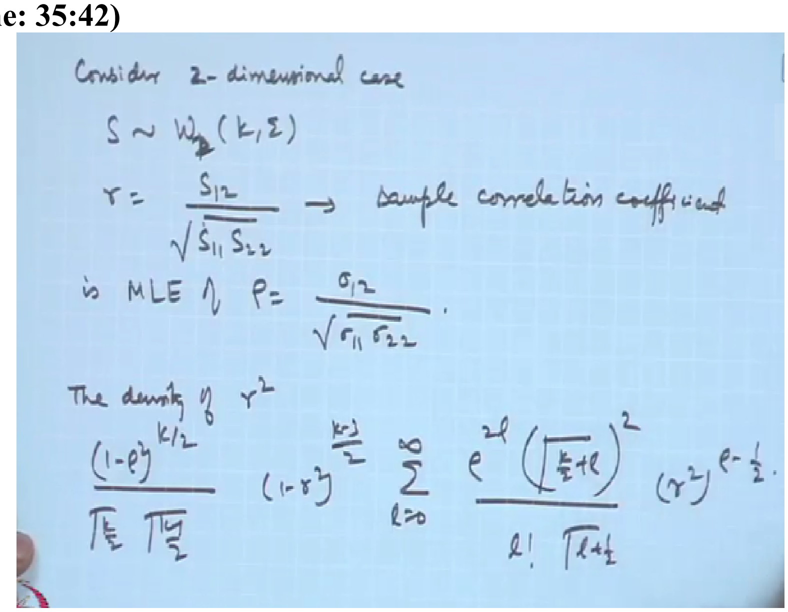


So this we have already proved that this will have--1/1+T square/k has beta distribution with parameter k-p+1/2 and p/2. So we are actually able to divide find out the distribution which is a generalization of the student t-distribution here. I will not be giving the derivation of the density of the Wishart distribution, we simply give the expression here the density of Wishart, so we have the following result. If we U prime p/k has density of the form U prime U then density of S that is equal to U prime U is proportional to F(s) determinant of S to the power k-p-1/2.

So I am considering the density of S as a constant times I will write only the final expression here 1/w(p,k) which is some constant determinant to the sigma-k/2; determinant of S to the power k-p-1/2 e to the power – 1/2 sigma of inverse S. Many times we consider generalized variance that is determinant of S. The distribution of the determinant of S can also be obtained. In terms of this we also define sample correlation coefficient etcetera.

Let me express this terms of Wishart.

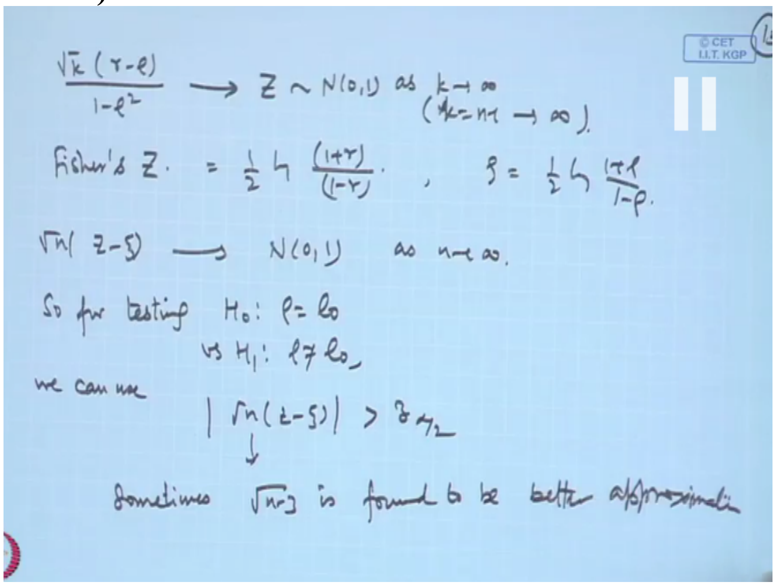
(Refer Slide Time: 35:42)



Consider 2-dimensional case. In the 2-dimensional case S will follow Wishart 2 k sigma. Then r=S12/square root of S1 s22, this is actually the sample correlation coefficient. This is actually maximum likelihood estimator of row that is sigma12/square root sigma11 sigma22.

So the distribution of r, or it function can be determine it is given by the density of r square is given by 1-Rho square to the power k/2/gamma k/2 gamma k-1/2 1/r square to the power k/3/2 sigma Rho to the power 2l (gamma/2+l) square/L factorial gamma L+1/2, L=0 to infinite; r square to the power Rho-1/2. And the density of r can be obtained from here. We also have the asymptotic distribution of r which can be used for the inference purpose.

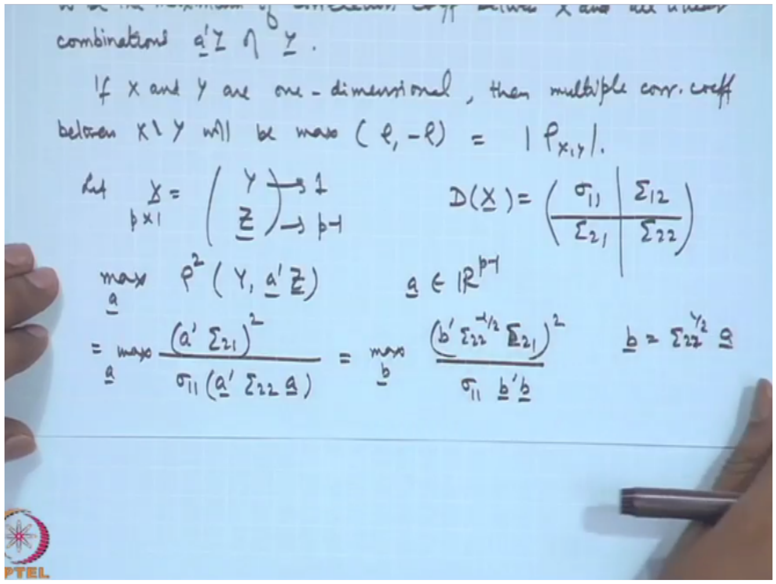
(Refer Slide Time: 37:32)



Square root k r-k/1-Rho square, this converges following normal 0,1 as k times to infinite that is n times to infinite, here k=n-1. And we also define Fisher's Z that is half log 1+r/1-r. And if I consider xi=half log 1+Rho/1-Rho then square root n Z- xi this also converges to normal 0,1 as intense to infinite. So for testing H naught say Rho=Rho naught again say H1 Rho is not equal to Rho naught, we can use root n Z-xi > Z alpha/2.

Sometimes root n-3 is found to be better approximation. Next, we define multiple correlated coefficients which is used in the multivariate analysis. Like in the case of one variable in the case of 2 variables we have discussed the Karl Pearson coefficient of correlation. Similarly, in the several variables we defined multiple correlation coefficients, so let me define that here.

(Refer Slide Time: 39:22)



"Professor to student conversation starts" I have avoided deriving the distributions of various terms in this multivariate portion, those who are interested can look at the book Introduction to Multivariate Analysis by T. W. Anderson, the chapter on multivariate analysis in the book Linear Statistical Inference and its application by C.R. Rao. And there are some other books also for example, M S Srivastava book on Multivariate analysis they consider this distribution theory. Not considering here to save the time here. "**Professor to student conversation ends"**

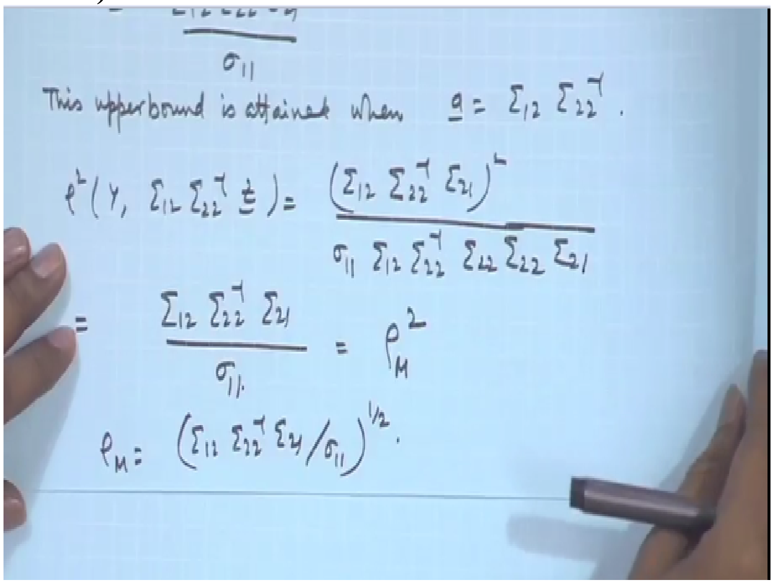
So let us consider say, let x be a random variable and let y be a random vector. Then the multiple correlation coefficient between x and y is defined to be the maximum of correlation coefficient between x and all linear combinations a prime of Rho. If x and y are one dimensional then maximum of Rho and -Rho multiple correlation coefficient between x and y be maximum of Rho and –Rho that is equal to modulus of Rho x, Rho y. Now let us consider say x=y and z.

So this is one dimensional and this is say p-1 dimensional. So x is a p/1 vector, and we want to define the multiple correlation coefficient here. And we partition the special matrix as sigma 11, sigma 12, sigma 21, sigma 22. So this is here 1 dimensional scalar and this is p-1 dimensional. So let us consider say maximum of correlation coefficient let us put this square here between y and a prime z, where a is a p-1 dimensional vector.

So we want to maximize this with respect to a. So this = a prime sigma 21 square/sigma 11 a prime sigma 22 a. We are considering the maximum of this with respect to a. so this we can write as we can substitute b as sigma 22 1/2 a. So if you put that we will get this as maximum with respect to b, b prime sigma 22 to the power-1/2 sigma 21 square/sigma 11 b prime b.

Now here we can apply (()) (43:50).

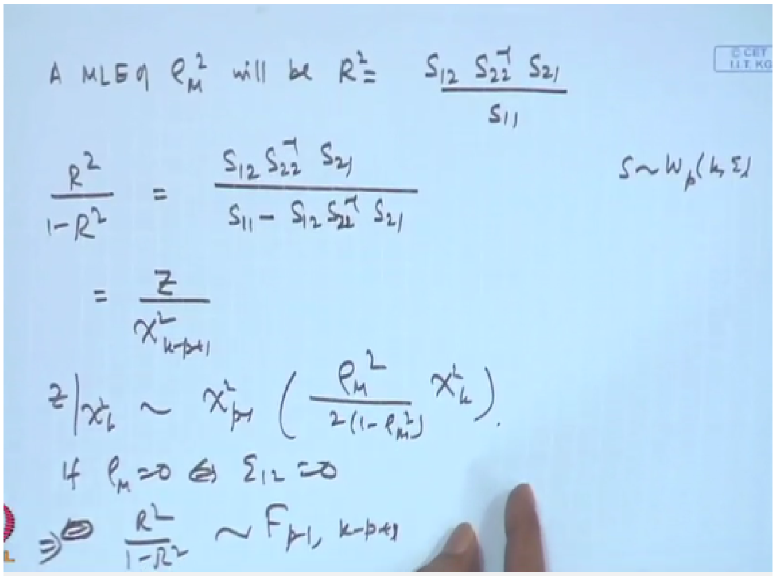
(Refer Slide Time: 43:52)



So this quantity will be < or = maximum of b prime b sigma 12, sigma 22 inverse sigma 21 divided by sigma 11 b prime b. But this term here is canceled out so here is quantity become free from b. This is upper bound is actually obtained, this is attempt when a=sigma 12 sigma 22 inverse.

So we are getting here Rho square y sigma 12 sigma 22 inverse Z that is = sigma 12 sigma 22 inverse sigma 21 square divided by sigma 11 sigma 22 inverse sigma 23 sigma 22 sigma 21, so this becomes identity and we will get this term can canceled out so you get simply sigma 12 sigma 22 sigma 21/sigma 11. So this we call Rho m square. So Rho m is actually = sigma 12 sigma 22 inverse sigma 21/sigma 11 to the power half.

(Refer Slide Time: 45:36)

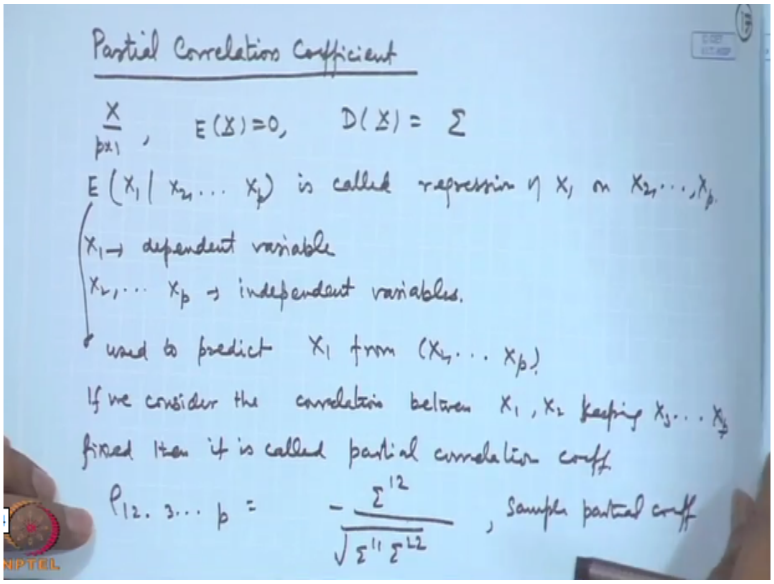


Now a maximum likelihood estimator of Rho m square, this will become simply r square which is calculated simply from the sample analog of this. Later on we will show that in the multiplication analysis, we use this r square as the coefficient of determination and it is an important indicator of the goodness of the regression model that is fitted there.

Now the distribution of r square can also be obtained by distribution theory that I have discussed earlier but I will not be giving the final results here. Actually we can see here that if I consider r square/1-r square then this term is actually = S12, S22 inverse S21/S11-S12 S22 inverse S21. So if I am considering S following Wishart k sigma then this will follow this can be written as Z/chi-square k-p+1.

And the general distribution of Z given chi-square k is chi-square p-1 Rho m square by twice 1-Rho m square chi-square k. So if Rho m is 0 then sigma 12 is 0 and this will imply that r square/1-r square has F distribution on p-1 k-p+1degrees of freedom. So the distribution of the multiple correlation coefficient after a transformation is shown to be F when Rho m = 0. So this is used for the testing of hypothesis regarding multiple correlation coefficients.

(Refer Slide Time: 47:36)

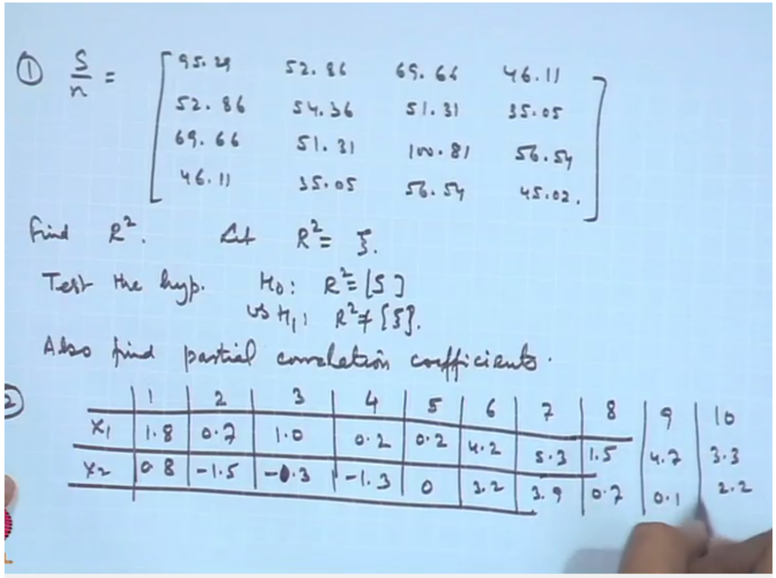


Likewise, we can also talk about Partial Correlation Coefficient. Suppose x is a p/1 vector with expectation x=0 and dispersion matrix = sigma. Then expectation of x1 given x2 to xp this is called regression of x1 on x2 to xp. So here x1 is known as dependent variable, we will discuss this in detail in when we do the regression but right now let me just introduce for the purpose of definition here. And x2 to xp these are called independent variables.

So this is use to predict x1 from x2 to xp. If we consider the correlation between x1 x2 keeping x3 to xp fixed, then it is called partial correlation coefficient. So we have for example Rho12.3 up to p that is equal to –sigma 12 divided by square root sigma 11 sigma 22. And one can obtain sample partial correlation coefficient. From here by considering –S12/square S11 S22.

"Professor to student conversation starts" I will conclude today's lecture by giving some exercise here for calculation of this coefficient and also for testing here. **"Professor to student conversation ends"**

(Refer Slide Time: 50:41)



So S/n is given by say 95.29 52.85 69.66 46.11 52.86 54.36 51.31 35.05 69.66 51.31 100.81 56.54 46.11 35.05 56.54 45.02. Fine R square. Let R square be = say xi. Then test the hypothesis that H naught r square is equal to xi integer versus Rho square is not equal to integer part of zie. Also find partial correlation coefficients.

So this is one exercise another exercise I am asking. Let us consider the data on the performance of student on 2 test, so some performance majors are given here 1 1.8 0.8 2 0.7 -1.5 3 1.0 -1.3 sorry -0.3 4 0.2 -1.3 5 0.2 and 0 6 4.2 3.2 7 5.3 3.9 8 1.5 and 0.7 9 4.7 and 0.1 10 3.3 and 2.2. Here you find MLE's of Mu sigma assuming x1, x2 follow N2 Mu sigma. Find Rho and test H naught Rho=0.8 against H1 Rho is not equal to 0.8 using asymptotic test for Rho.

In the next lecture, I will introduce the use of this Hotelling's T square etcetera for testing for the mean of the multivariate normal distribution or comparing the means of 2 multivariate normal distribution. We will also consider the problems of classification of observations. So this thing I will be covering in the next lecture.