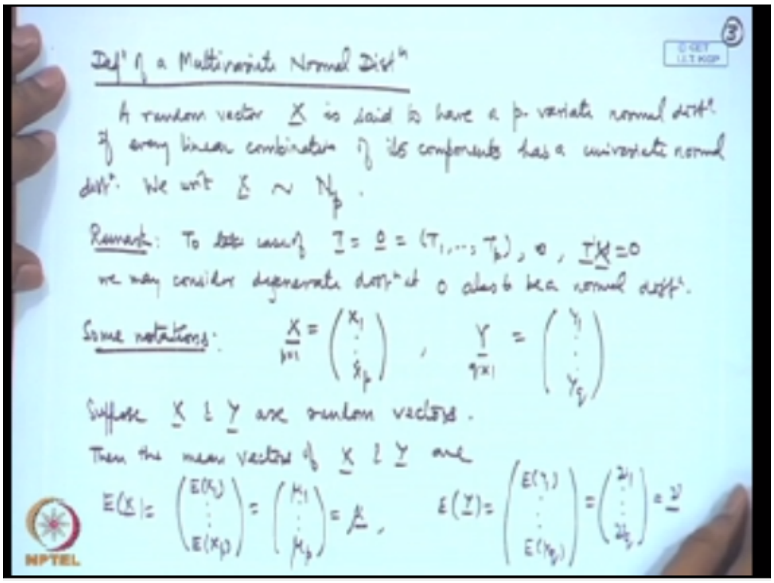
Statistical Methods of Scientists and Engineers
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology – Kharagpur

Lecture - 17
Multivariate Analysis - II

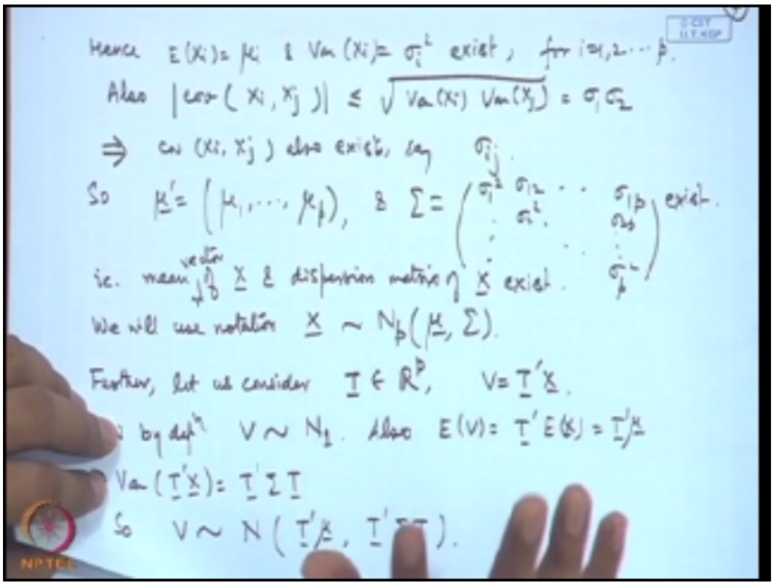
Yesterday, we have introduced multivariate normal distribution. So, this is a p dimensional distribution and let me recall the definition of the multivariate normal distribution. The definition was in terms of its linear combination. So, we say that a random vector X is having a p-variate normal distribution if every linear combination of it is component has a univariate normal distribution and the notational form was X follows Np.

(Refer Slide Time: 00:39)



Now, as a consequence of this definition, we proved certain properties. For example, we showed that if X has a multivariate normal distribution, then its mean vector and variance-covariance matrix will exist. So for example, we showed here that the mean vector mu and the variance-covariance matrix sigma will exist and therefore, we modified our notation to X following Np mu sigma.

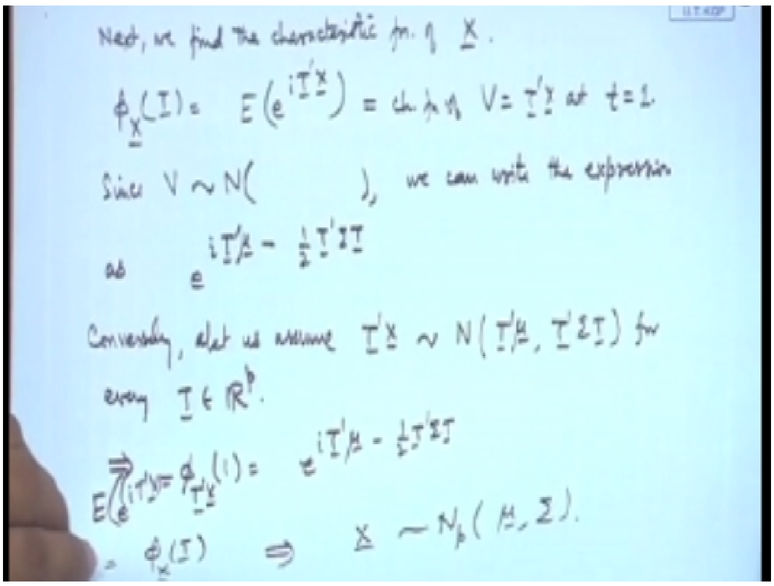
(Refer Slide Time: 01:18)



So that means, when I make a statement, then X is a multivariate normal distribution, then at the same time, we will have it is a mean vector mu which is have p dimensional vector in the Rp and sigma which will be a p/p matrix. Now, the nature of this matrix is it is a real symmetric matrix. But, at the same time because it is a variance-covariance matrix, it will also be positive semidefinite.

We actually showed this statement through the definition of positive semidefiniteness that means we consider A prime sigma A and we are able to show that it is actually non-negative and at the same time, we were also able to find out the characteristic function of this and the characteristic function is of the form e to the power i T prime mu–1/2 prime sigma T. So, using this we can also derive the distribution of the linear combinations.

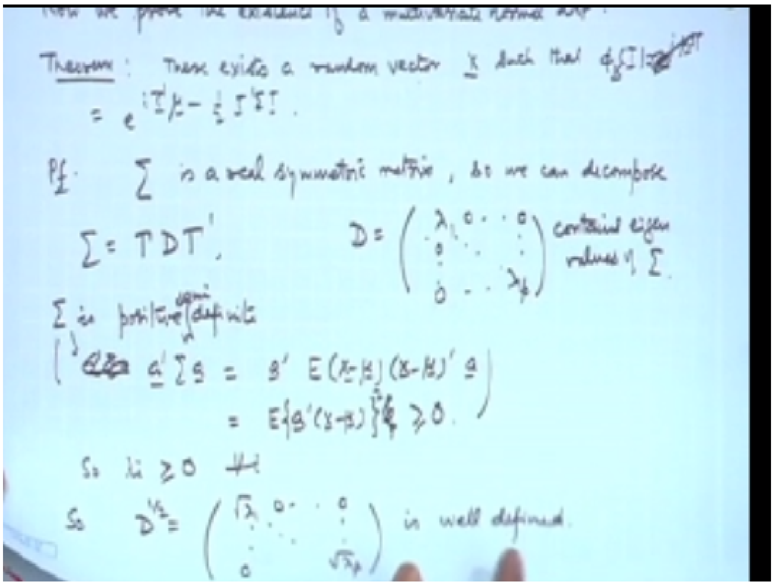
(Refer Slide Time: 02:09)



We were also able to prove the independent result that if the matrices diagonal, sigma is a diagonal matrix, then the components will be independent. We also considered a decomposition of the full random vector in terms of that also we proved. Finally, we proved that given mu and sigma, we can always find out a random vector whose distribution will be Np with mean vector mu and variance-covariance matrix sigma.

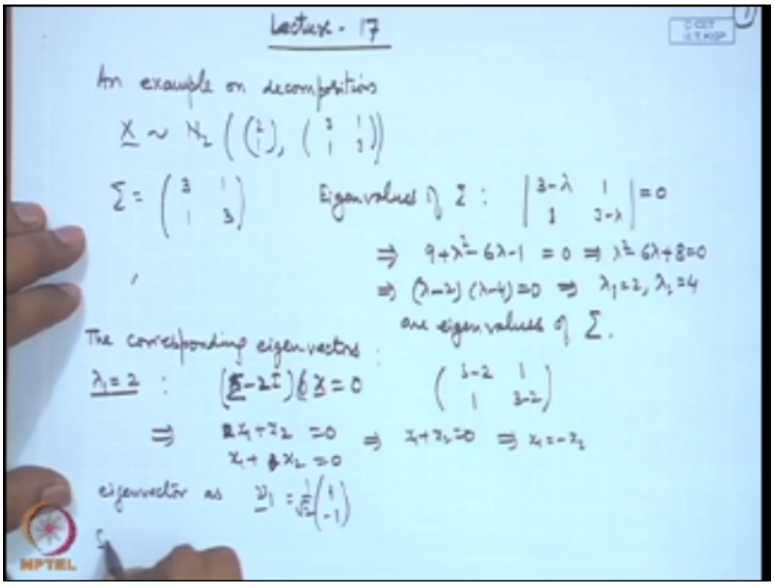
For this, we applied decomposition approach. We said that if there is a real symmetric matrix sigma, then we can have decomposed in the form of gamma D gamma prime, where D is the diagonal matrix consisting of Eigen values of sigma. The proof of positive definiteness was done. Now, first we let me start with an example which will show this decomposition. So, let me start with one problem here.

(Refer Slide Time: 03:00)



Let us consider an example here for the decomposition. An example on decomposition let us consider say, X follows N 2 2, 1, 3, 1, 1, 3. So, in fact it is a positive definite matrix here. This is the mean vector. So, this is the basically bivariate normal distribution. Now, let us consider this sigma. Sigma is 3, 1, 1, 3. Let us consider the Eigen values of sigma. So, we applied a standard procedure.

(Refer Slide Time: 03:38)

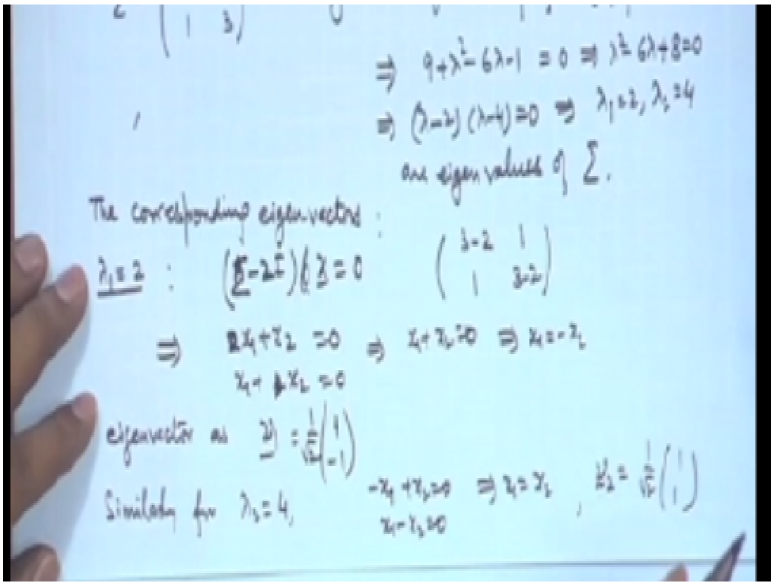


We consider 3 - lambda 1, 1, 3 – lambda, the determinant is = 0. This implies you will have 9 + lambda square - 6 lambda - 1 = 0. So, we will have lambda – 2 * lambda - 4 is = 0. This implies lambda 1 is = 2 and lambda 2 is = 4. These are the Eigen values of sigma. Let us consider the Eigen vectors, the corresponding, so let us consider say for lambda 1 is = 2, for this we will have to consider sigma - 2i * say a vector x is = 0.

So, that will give me 2X1+X2 and X1, if I add these 2, I get X1+X2, so here we get 9 + lambda square - 6 lambda - 1. So, this is giving me lambda square - 6 lambda - 8 is = 0. So, I get lambda - 2. Sorry, this is + 8 that is why we are getting the wrong answer, so lambda – 2 * lambda - 4 is = 0. So that is lambda is = 2 and lambda is = 4 are the 2 Eigen values here. If I considered here 3-2, 1, 1, and 3-2, then I will get here X1+X2 is = 0.

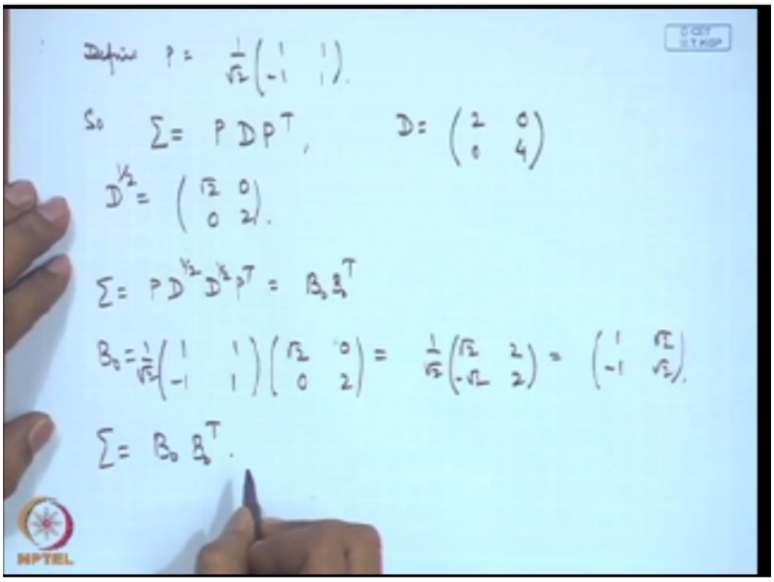
The second equation is also X1+X2 is = 0. So, this implies X1 is = - X2. So, if we normalize we get the Eigen vector as a nu 1, 1-1 and we can normalize it by dividing by a square root 2. Similarly, for lambda 2 is = 4, we will get -X1+X2 is = 0, X1-X2 is = 0. That means X1 is = X2, so if I normalize, I can consider the Eigen vector as 1/root 2, 1, 1. So, based on this we can consider P to be 1/root 2, 1, -1, 1, 1.

(Refer Slide Time: 07:50)



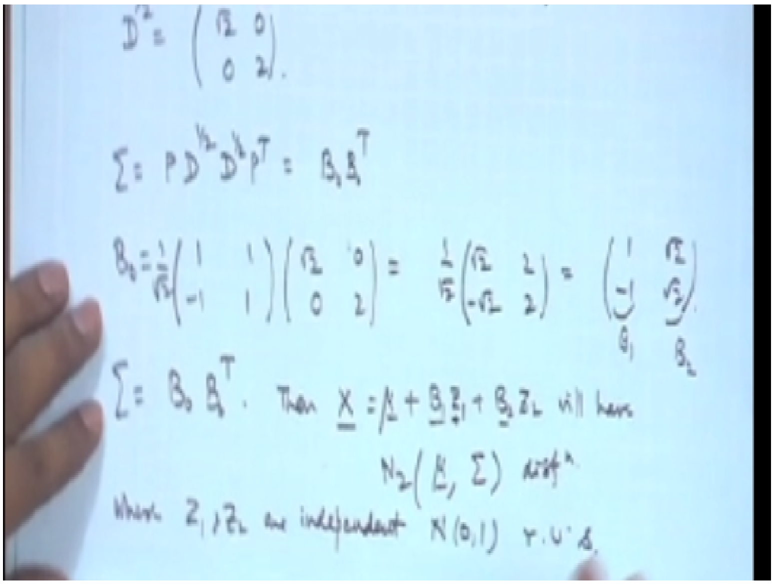
Then, we can check that sigma is = P D P transpose, where D is the diagonal matrix consisting of Eigen values 2 and 4. So, if I consider say D to the power 1/2 then that will become = square root 2 and 2. So, this is the way actually the calculation for the B matrix was done which I showed yesterday for the existence proof here. We considered that decomposition sigma as P D 1/2 D 1/2 and P transpose which is called B0 B0 transpose.

(Refer Slide Time: 08:19)



So, here B0 will become = 1/root 2, 1, -1, 1, 1, root 2, 0, 0, 2. We can calculate it, it is = 1/root 2, root 2, -root 2, 2 and 2 that is = 1, -1, root 2, root 2. So, B0 matrix is coming like this that means sigma can be written as B0 B0 transpose and using this, we can define if I am considering this as B1, this as B2, then I am having B1 Z1 + B2 Z2 + mu. Now, in this particular case, I took mu to be 2, 1.

(Refer Slide Time: 10:16)



So, if I considered this vector, let us call this as mu 2, 1, then I can put it here mu + this. So, this is = x. This will have N 2 mu and this sigma distribution, where your Z1 and Z2 are independent normal 0, 1 random variables. So, here what I have shown here that if I am given a mean vector and variance-covariance matrix which is a positive definite here, in fact it can be positive semidefinite also.

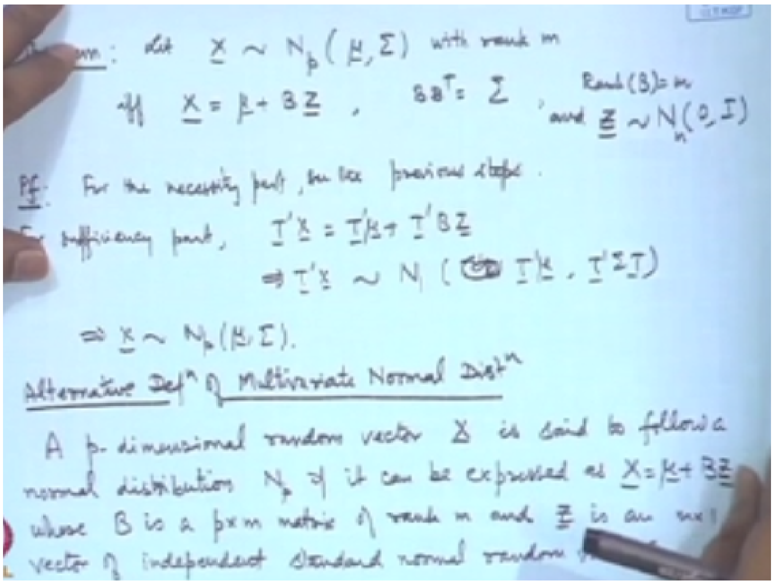
Using this, I considered the decomposition of this; I find out the Eigen value, which are 2 and 4 respectively corresponding to that I find out the Eigen vectors. So that is = 1, -1 multiplied by a constant and similarly, 1, 1 for the second one multiplied by a constant, I considered the normalized one. So that if I consider the matrix of this Eigen vectors, this is actually an orthogonal matrix here.

So, the decomposition of sigma is now P D P transpose, where P is given by this and D is the diagonal matrix consisting of the Eigen values of this in the diagonals. Based on this, I defined D to the power 1/2 that is root 2, 0, 0, 2 that is a square root of the diagonal entries and based on this, I considered B0, B0 is then actually P*D 1/2. So, that is = 1/root 2, 1, -1, 1, 1 and then root 2, 0, 0, 2 and this matrix can be written like this.

And if the columns of B0 are written as B1, B2 and I consider now B1 Z1+B2 Z2, then they are normal distributions, standard normal random variables. So, this is a vector now and I add here a mu vector here and I define X as this. Then, this X will have 2 dimensional normal distributions with mean vector mu given by this and variance-covariance given by this.

So, this is the application here of the theorem that I proved yesterday that there can be always defined a normal distribution with the given mean vector and variance-covariance matrix. Let us proceed further here, for some further properties of the multivariate normal distribution. This result I state in the form of a theorem. Let X follows a multivariate normal distribution with mean vector mu and variance-covariance matrix sigma and the rank of sigma is m.

(Refer Slide Time: 13:47)



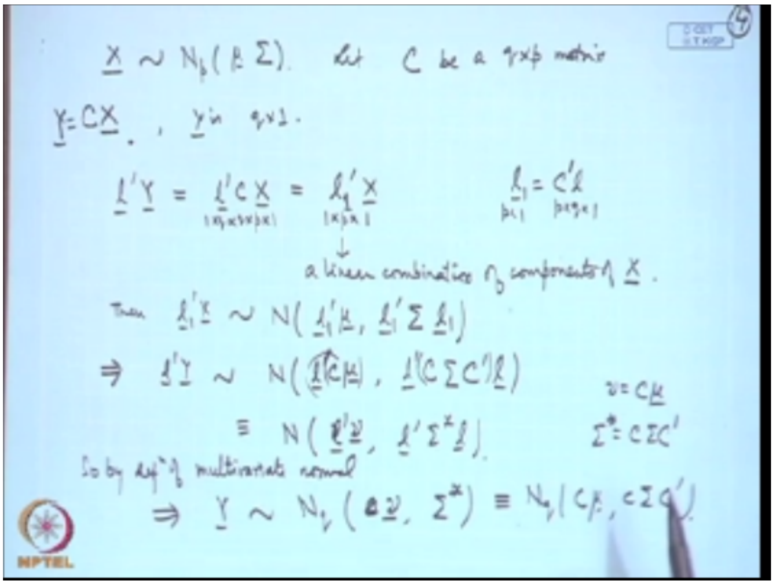
So, this is if and only X can be written as mu + BZ, where B, B transpose = sigma. Rank of B is = m and Z is a collection of independent standard normal random variables. So, the proof of the necessity part that means if I am writing this, then for the necessity part, see the previous steps. Let us consider say T prime X, T prime X is = T prime mu + T prime BZ. So, this implies that T prime X, this will follow N 1 T prime mu T prime sigma T.

So, this implies that X will follow Np mu sigma. So, since this is a necessary and sufficient condition for the multivariate normal distribution, we can give an alternative definition of the multivariate normal distribution in terms of this characterization. If you remember the original definition that I gave for the multivariate normal distribution, that was in terms of the linear combinations only.

The original definition if I recall here, a random vector X is to have a p-variate normal distribution if every linear combination of its components has a univariate normal distribution. But, now by these results that we have proved here, we are now able to give an alternative definition, an alternative definition of the multivariate normal distribution.

So, a p dimensional random vector X is, B is a p/m matrix of rank m and Z is an m/1 vector of independent standard normal random variables. So, this definition is actually used in this representation that I have proved here. So based on this, this is an alternative way of defining a multivariate normal distribution. So, basically again you considered, you can think of this as linear transformations obtained from univariate normal distributions.

(Refer Slide Time: 18:52)



And from there, we are actually (()) (18:27) definition. A previous definition was a sort of characterization. Then, we can again obtain let us consider, suppose I say X follows Np mu sigma and I consider say let C be a q/p matrix and let us consider say C X here and I defined it as Y. So, naturally then Y is q/1. Now, consider say linear combinations of Y. So linear combinations of Y, let us consider something like L prime Y.

So, L prime Y is = L prime C X. So, this is nothing but, I can call it l 1 prime X where l 1 is = C prime l. If I look at the dimension here, C is q/P, so this should be 1/q, this is q/P and this is p/1. So, here l 1 will be having dimension p/1 because this is p/q, this is p/1. So, this is becoming p/1. So, this is 1/p/1. So, if I look at l prime Y I have written it as l 1 prime X. This is a linear combination of components of X.

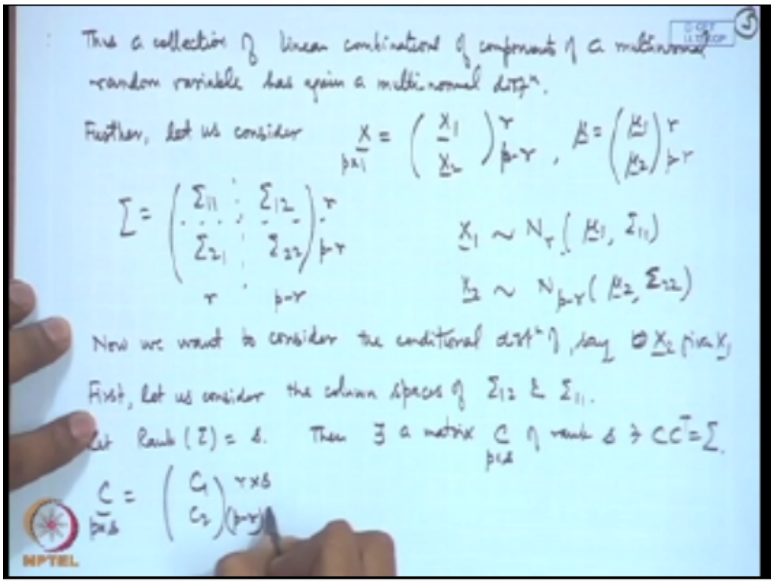
Now, if we remember yesterday's working out after we defined the multivariate normal distribution, let me show you the results once again. We talked about the distribution of the linear combination. If I am considering X following Np mu sigma and T is any P dimensional vector, then T prime X has a univariate normal distribution. I am defining V is = T prime X. It has a univariate normal distribution, T prime mu T prime sigma T.

So if I look at this, then what we are getting l 1 prime X, this will follow normal l 1 prime mu l 1 prime sigma l 1 as the variance term. Now in place of l 1, let us substitute C prime l everywhere, what does it will mean? It will mean l prime Y that will follow normal l prime C mu, l prime C Sigma C prime l. What I have done? I have substituted l 1 is = C prime l everywhere. So, this is a linear combination of C mu.

Let us define say mu is = C mu and let us define say sigma * is = C sigma C prime, then this is nothing but normal distribution with mean vector l prime nu and l prime sigma * l, where l is a q/1 vector. So by definition of multivariate normal, this will imply that Y will follow q dimensional with V vector nu and variance-covariance matrix sigma, which is actuallYn q C mu C sigma C prime.

That means a collection of the linear combination of multivariate normal distribution (()) (23:28) as a multivariate normal distribution with the required number of components. We can write it in the terminology, thus a collection of linear combinations of components of a multinormal random variable has again a multinormal distribution. Next, let us consider the conditional distributions.

(Refer Slide Time: 23:45)



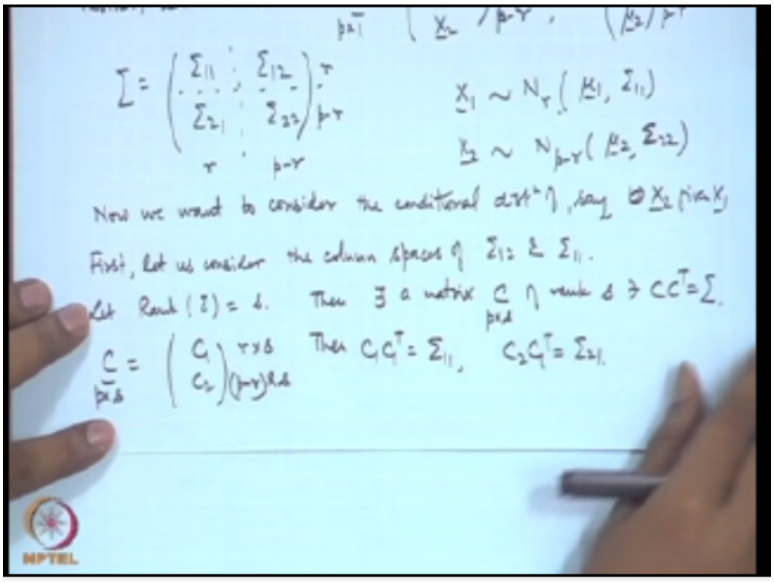
If we recall, if I have XY following a bivariate normal distribution, then the conditional distributions of X given Y and Y given X are univariate normal distribution. So, if I look at X given Y then it is univariate because one dimension. Now, I can consider the decomposition if random vector into 2 parts, each of them maybe random vectors. So, let us consider say X is = X1 with r components and X2 as p-r components, this is p/1 vector.

So, simultaneously I decomposed mu as mu 1, mu 2 in r and p-r components and variance-covariance matrix also I decomposed, here you have r, here you have p-r, here you have r, here you have p-r components. So, basically you are same that X1 will follow N r mu 1 sigma 11, X2 will follow Np-r mu 2 sigma 22. Now, we want to consider the conditional distributions of say X2 given X1, similarly X1 given X2.

So in this one, we will need certain inverses. Let us prove a result for that. First, let us consider the column spaces of sigma 12 and sigma 11. Let us assume say rank of sigma is = say s, then there exists a matrix C, say p/s of rank s. Such that C C transpose is = sigma. This existence I have shown you earlier in the previous discussion and if I decompose the C as C1 C2, so this is p/s.

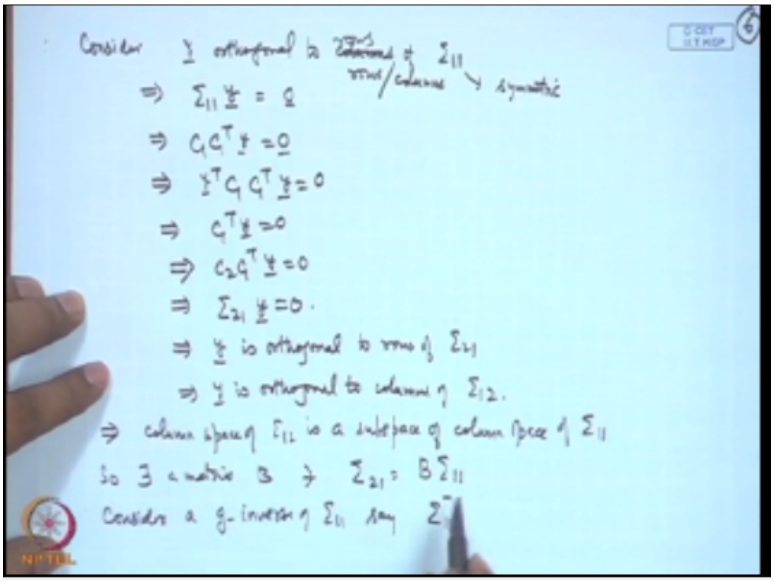
So, this will be some r/s and this will be p-r/s. If we consider this decomposition, then we will have C1 C1 transpose is = sigma 11 and C2 C1 transpose will become = sigma 21. You can write C C transpose, so this will become C1 C2*C1 transpose C2 transpose, so that will give me this. Now, let us consider say a vector Y which is say orthogonal to columns of say sigma 11 that means sigma 11*Y is = 0.

(Refer Slide Time: 27:42)



This will imply that C1 C1 transpose Y is = 0. This will imply if I remultiply by Y transpose, I will get Y transpose C1 C1 transpose Y is = 0. This will imply C1 transpose Y is = 0. This will imply that C2 C1 transpose Y is = 0. This implies that sigma 21 Y is = 0. So, this will imply that Y is orthogonal to columns of, if I am writing it is orthogonal to the rows basically, sorry. It is not rows actually.

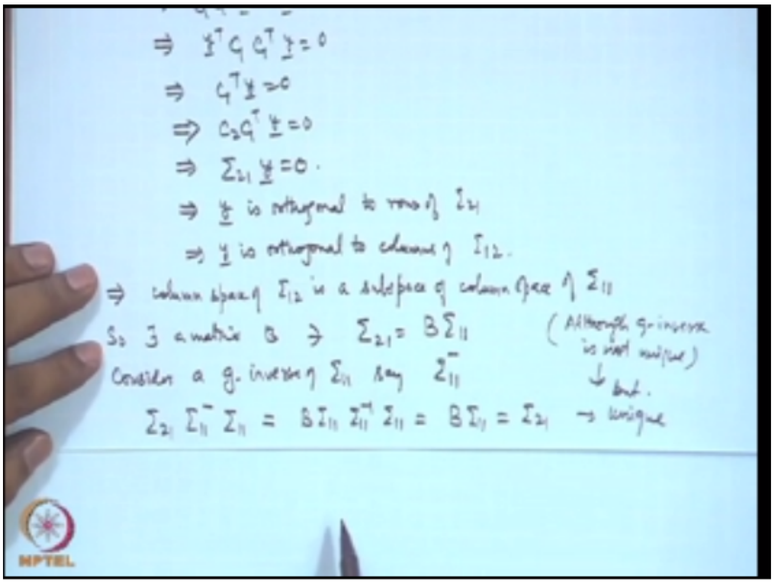
(Refer Slide Time: 28:05)



Because if have considered sigma 1, 1. So if write here, I will be multiplying this column vector into the rows of this. So if they are 0, that means rows of sigma 1, 1 or orthogonal to this one, but again because sigma 11 is a symmetric matrix, so rows and columns both are same, so rows or columns because sigma 11 is a symmetric matrix, so the statement will be same.

So, if Y is orthogonal to rows of sigma 21, this implies Y is orthogonal to columns of sigma 12. Because rows of sigma 21 will be columns of sigma 12 because sigma 21 is transpose of sigma 12. So, this will imply that column space of sigma 12 is a subspace of column space of sigma 11. So, there exists a matrix say B such that sigma 21 is = B sigma 11. So, if I consider a g inverse of sigma 11 say, let us we use the notation say sigma 11 inverse.

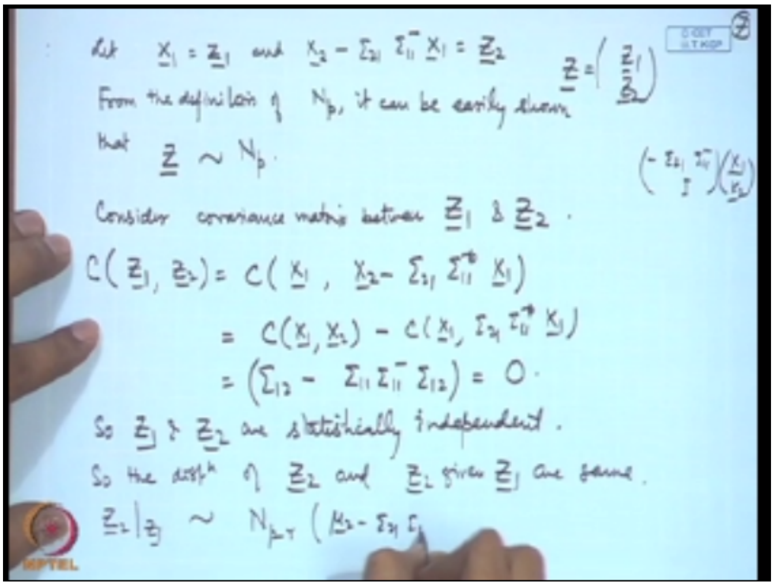
(Refer Slide Time: 31:10)



So, here this I am putting as a generalized inverse here. So, if I consider say sigma 21 sigma 11 g inverse sigma 11, then this will become = B sigma 11 sigma 11 inverse sigma 11 that is = B sigma 11 that is = sigma 21, so this is unique. So, although g inverse is not unique, but this term is unique. So, this term can be utilized for derivation of the conditional distribution which I will be using now.

Let us define X1 is = some Z1 and X2 - sigma 21 sigma 11 g inverse X1 = say Z2. If I am considering X as a multivariate normal distribution and the components X1, X2 are also multivariate normal then naturally this is linear combinations. So, Z1 and Z2 will also be multivariate normal distributions and also, if we look at the dimension, this is r dimensional and here, we are having the dimension of X2 has p-r.

(Refer Slide Time: 32:14)



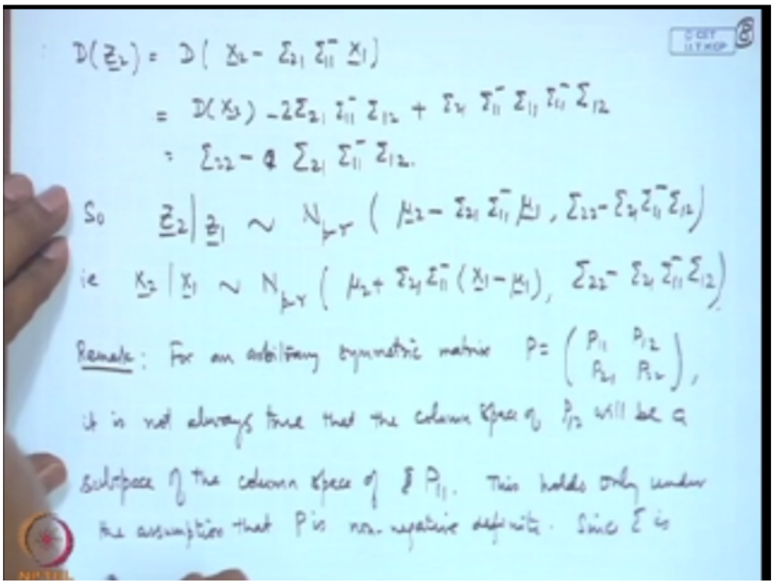
So, if I put say Z1, Z2 let us call it is a Z, then this will have p dimension. So, from definition of Np, it can be easily shown that Z follows Np, this Z1, Z2. Now, let us consider covariance matrix between say Z1 vector and Z2 vector, so let us write it as C Z1 Z2 that is = C of X1 X2 - sigma 21 sigma 11 inverse X1 that is = C of X1 X2 - C of X1 sigma 21 sigma 11 inverse X1.

So, this is = sigma 12 - now, you look at this one, this will give me sigma 11 sigma 11 inverse, so this actually generalized inverse here. Sigma 11 inverse sigma 12, now this is = 0 that is the null matrix. So, this Z1 and Z2 are statistically independent. So, the distribution of Z2 and Z2 given Z1 are same because if Z1 and Z2 are independent, then the conditioning on Z2/Z1 has no effect.

So, the distribution of Z2 and Z2 given Z1 is the same. So, if I write it in terminology, it will turn out to be Z2 given Z1 this follows Np-r and the distribution of Z2 will be coming from the linear combination of X1, X2 because this is the linear combination define Y - sigma 21 sigma 11 inverse and I of X1 X2. So, this linear combination is given this. So, if I consider this then I get a straight forwardly mu 2 - sigma 21 sigma 11 inverse mu 1 as the mean vector.

For the dispersion matrix, it will come as this * sigma 11 sigma 12 sigma 21 sigma 22 and the transpose of this on the other side. So, that gives me straight forwardly, let me write it as the dispersion matrix of Z2. Let us derive this, the dispersion matrix of Z2 can be derived, the dispersion matrix of Z2, so that is = dispersion matrix of X2 - sigma 21 sigma 11 inverse X1. So, this will become = dispersion matrix of X2 - sigma 21 sigma 11 inverse sigma 12.

(Refer Slide Time: 36:32)



And, the same term will come 2 times, I will put 2 times here + sigma 21 sigma 11 inverse sigma 11 sigma 11 inverse sigma 12. So, this is = sigma 22 - twice. Now, this term you see here sigma 21 sigma 11 inverse sigma 11 is again sigma 21, so this term and this term are the same. So, this one of them gets cancelled out. We left with, so we are saying Z2 given Z1 this follows Np-r mu 2 - sigma 21 sigma 11 inverse mu 1 sigma 22 - sigma 21 sigma 11 inverse sigma 12.

Now, we substitute here Z2 is in terms of X2 here, so we put it there. So, X2 given X1 that will follow Np-r mu 2 + sigma 21 sigma 11 inverse and here X1 will get added up because I have brought this term to the other side – mu 1. The variance term will not change. Now, the fact that we have used that is the column space of sigma 12 is a subspace of the column space of sigma 11.

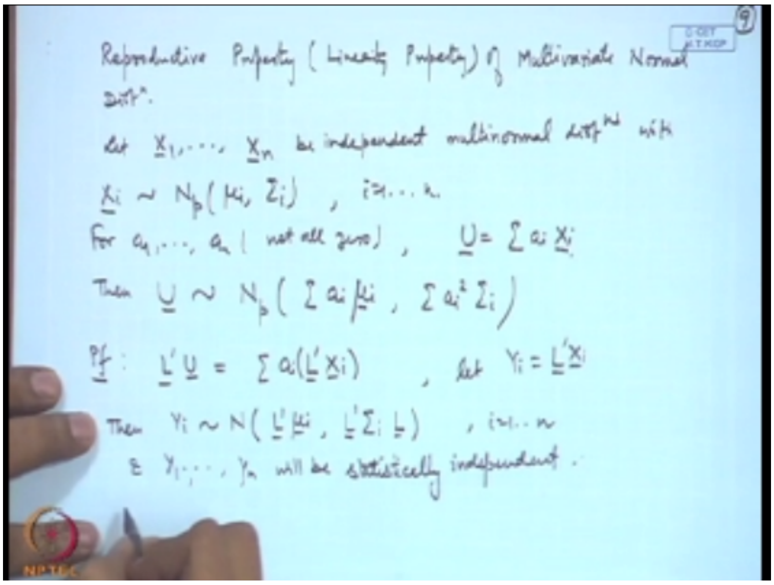
This is following a decomposition that we have used for a positive definite matrix. Because this term is coming here or positive semidefinite also it will be true. If it is not positive semidefinite then this decomposition will not be able to us and therefore, this statement that column space of sigma 12 is a subspace of the column space of sigma 11 need not always be true.

So, we are in fact making full use of the positive semidefiniteness of the variance-covariance matrix. Now, using this property we are able to write down sigma 21 s B times sigma 11 and due to that I can have a unique definition of sigma 21 sigma 11 inverse sigma 11. Now, why this was required? Because it is appearing in the ultimate expression here for the variance-covariance matrix of the conditional distribution.

So, although the g inverse has many representations where the term that we will get here by this calculation it will be unique here. So, as a remark let me write for an arbitrary symmetric matrix say P is = P 11, P 12, P 21, P 22, it is not always true that the column space P 12 will be a subspace of the column space of P 11. This holds only under the assumption that P is non-negative definite. Since sigma is dispersion matrix, this fact holds here.

Now, next we prove the reproductive property of multivariate normal distribution. If we remember that if we are considering independent univariate normal distributions, then the linear combinations of independent univariate normal distributions again have a univariate normal distribution and the means and variances are defined accordingly. Now, this type of property can be generalized to multivariate normal distribution also.

(Refer Slide Time: 42:33)

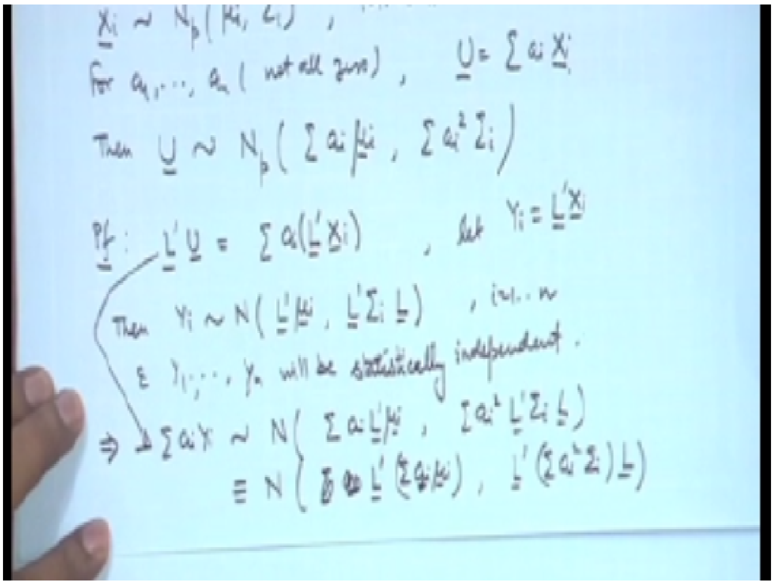


So, let us consider this propertYnow, basically we say linearity property. So, let us consider X1, X2, Xn be independent multinormal distributions with Xi say following Np mu i sigma i, for i is = 1 to n. For a 1, a 2, a n, let us saYnot all 0, let us define say U is = sigma a i X i. Then, U will follow a multivariate normal with mean vector sigma a i mu i and variance-covariance matrix sigma a i square sigma i.

So, you can see this is a straight forward generalization of the result which is label for the univariate normal distribution, there mu i is where the scalars and sigma i square where the variance terms. So, here it has become a matrix here. So, the proof is actually based on the definition that is a linear function we can use. So for example, if I write say L prime U, so what is L prime U, L prime U become sigma a i L prime X i.

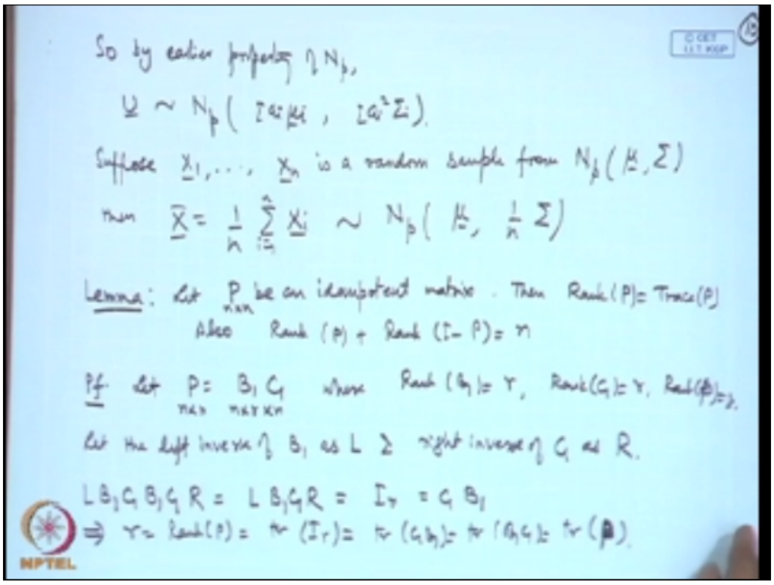
Now, if I define say Yi is = L prime Xi, then Yi will follow univariate normal with L prime mu i and L prime sigma i L. For i is = 1 to n and Y1, Y2, Yn will be a statistically independent. So, now this implies that sigma a i Y i that will follow univariate normal with sigma a i L prime mu i and sigma a i square L prime sigma i L. So, this you can write as normal with sigma L prime sigma a i mu i and L prime sigma a i square sigma i L.

(Refer Slide Time: 45:34)



And this sigma a i Y i is nothing but L prime U, so by the definition of the multivariate normal distribution, we have U following Np sigma a i mu i and sigma a i square sigma i. Now, we can consider the sampling, suppose X1, X2, Xn is a random sample from Np mu sigma distribution, then if I considered X bar vector as 1/n sigma Xi, i is = 1 to n, then that will follow Np mu 1/n sigma.

(Refer Slide Time: 46:37)



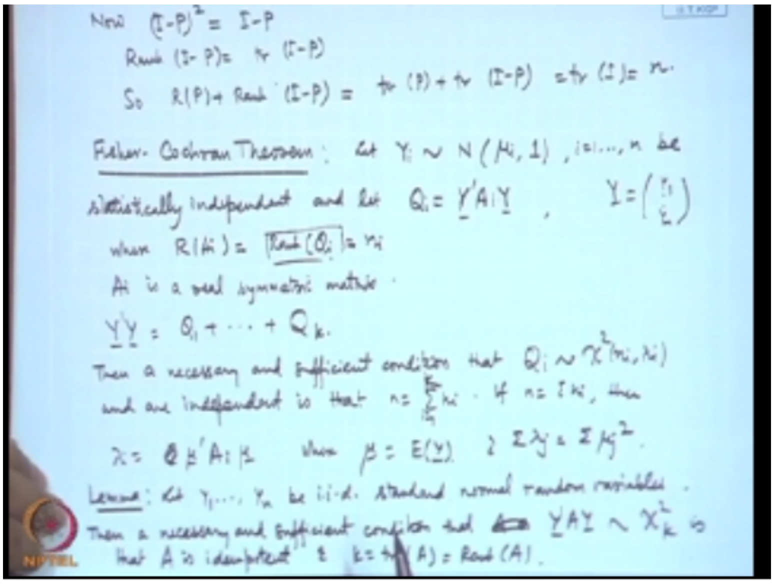
So, we are able to obtain a distribution of sample mean in sampling from a multivariate normal distribution. There are some other results which are related to the multivariate normal distribution especially they will be useful for deriving, for example if you remember in the univariate normal distribution if I considered the sum of squares of the independent normal random variables, then it is having a chi square distribution.

So, similarly if I considered some squares etc which are quadratic forms which are related to the multinormal distribution, then they are also having chi square distributions under certain conditions. So, we have some results which I will just mention here. For example, let P be an idempotent matrix, then rank of P is = trace of P and also rank of P + rank of I-P that will be = dimension, so this is n/n here.

So, that I mention will be = the rank of P + rank of I-P. Let us consider a simple illustration of this. Let us take say P is = say B1 C1, so this is n/n. This is saYn/r*n, where rank of B1 is r, where rank of C1 is r and rank of P is = r. Let us define the left inverse of B1 as L and right inverse of C1 as R. Let us consider say L B1 C1 B1 C1 R that will be = L B1 C1 R that is = I r that is = C1 B1.

So, this will imply that r is = rank of P that is = trace of I r that is = trace of C1 B1 that is = trace of B1 C1 that is = trace of B. Recall the definition of idempotent matrix that is P square must be = P. Also, if I considered I-P square that is = I-P, so that means rank of I-P will be = trace of I-P. So, rank of P + rank of I-P that will be = trace of P + trace of I-P that is = trace of I that is = n.

(Refer Slide Time: 51:40)



This result is useful in proving certain properties and the most important result in this direction is actually known as Fisher-Cochran theorem. Let me give the theorem in its full form. Let us consider say Y i following normal mu i, 1, i is = 1 to n. Suppose (()) (52:44) independent and let us may fine say Q i is = Y prime A i Y. Now, I am considering here Y to be the vector Y1, Y2, Yn and rank of Ai that is = rank of Q i that is = n i.

Actually, rank of a quadratic form is actually the rank of the matrix which is given there. Otherwise, it has no significance as such and I am assuming A i is a real symmetric matrix. So, if I consider Y prime Y is = say Q 1 + Q 2 + Q k, then necessary and sufficient that Q i follows chi square n i lambda i and are independent is that, so this is actually non-central chi square distribution.

I will spend some time on the discussion of noncentral chi square distribution also, because whatever chi square distribution we have done so far are actually central chi square distribution, but if I consider normal distribution with mean mu i then if I consider the square of that. See Y i – mu i square if I consider that will be chi square on one degrees of freedom.

But if I consider Y i square itself, then it will have a noncentral chi square distribution with one degree of freedom and non-centrality parameter mu i square/2. So, here I am getting this quantity here. So, and they are independent is that n is = sigma n i, i is = 1 to k and if n is = sigma n i, then lambda i that is = sigma mu prime A i mu, where mu is the expectation of Y and sigma lambda j that is = sigma mu j square.

As a corollary of this, you have the following result, let Y1, Y2, Yn be independent and identically distributed, standard normal random variables, when a necessary and sufficient condition that Y prime AY follows chi square k is that is idempotent and k is = trace of A that is a rank of A. I will follow up this theorem by some further results on the connection of the multivariate normal distribution with chi square distribution.

And also we will introduce the noncentral chi square distribution because in the discussion, we have used that thing. So, I will briefly discuss the noncentral chi square distribution also. We also have some further characterizing properties of the multivariate normal distribution, so I may briefly describe those things also in the next lecture.