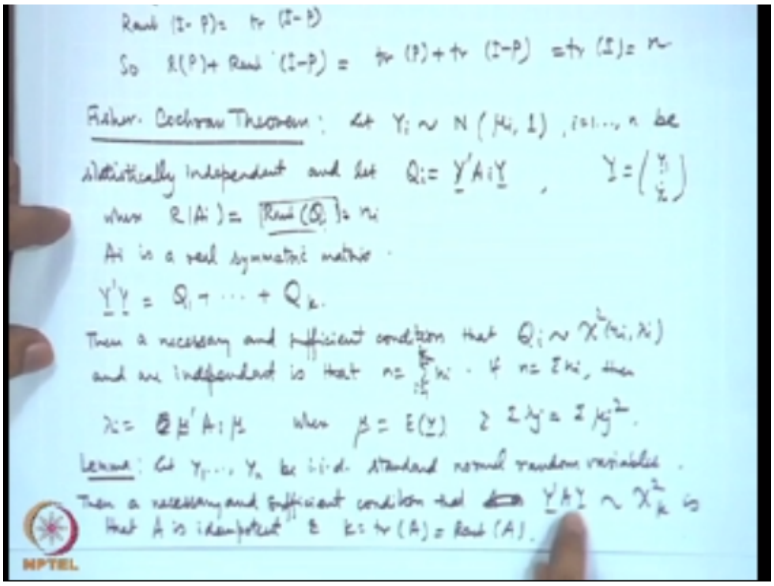
Statistical Methods of Scientists and Engineers
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology – Kharagpur

Lecture - 18
Multivariate Analysis – III

So, we continue our discussion on the multivariate normal distribution and its properties. We have seen various characterizing properties, which also helped us in giving an alternative definition of the multivariate normal distribution. Now, we are trying to see its connections with chi square distribution as in the case of univariate normal distribution.

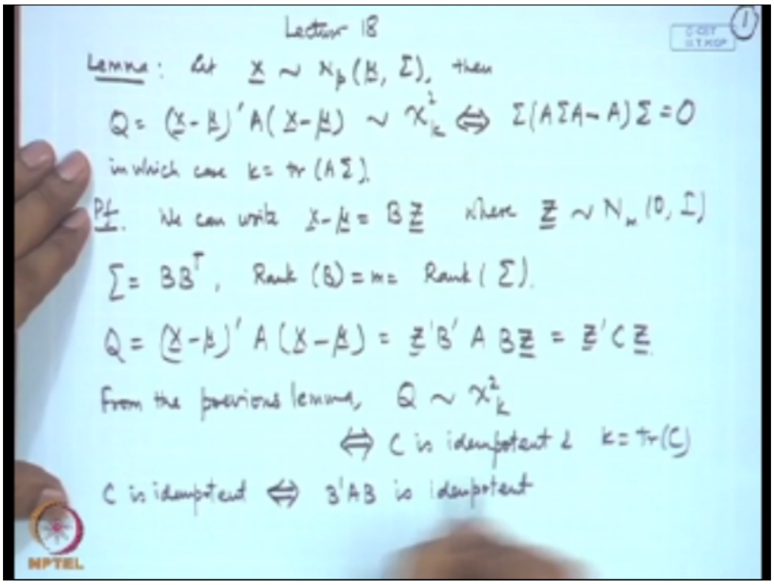
(Refer Slide Time: 00:43)



For that purpose, I stated Fisher-Cochran Theorem and another Lemma, which is saying that Y prime A Y will have a chi square distribution. So, this is giving a necessary information condition that if are having standard random variables, then if I considered Y prime A Y that is a quadratic form. This will have chi square k. We know that Y prime A Y has a chi square k.

But, if I consider any A here, then for idempotent matrix this will be true. Now, let us consider further results on this. The next result is that if X has a Np mu sigma distribution, then let us consider say Q that = X-mu transpose A X-mu, then that follows chi square k, this is if and only if sigma A sigma A - A sigma is null and in this case, you will have k = trace of A sigma.

(Refer Slide Time: 01:17)

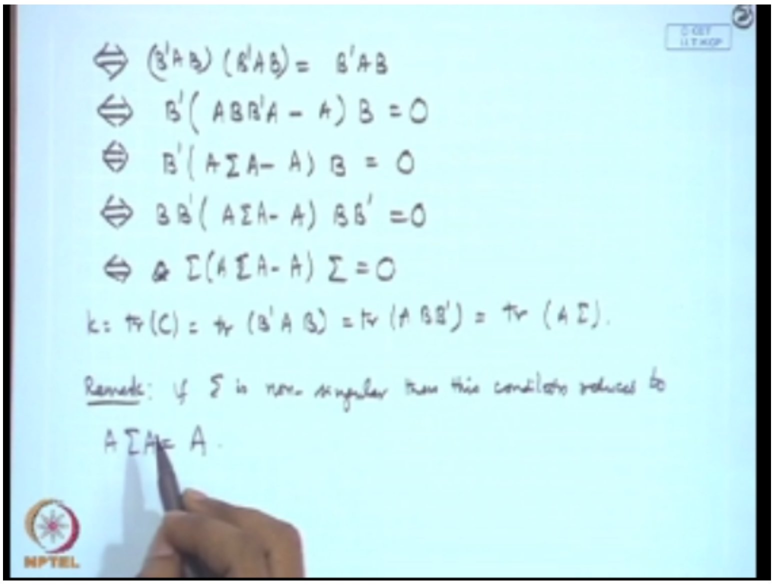


Let us look at the proof of this, so you can write X-mu = some BZ, if you remember the representation that I obtained for necessary and sufficient condition for the multivariate normal distribution, we were able to write a multivariate normal as mu+BZ, where Z is a vector consisting of the standard normal independent random variables of dimension N. So, let us consider the decomposition of sigma as B B transpose, rank of B is m, which is also the rank of sigma.

And the quadratic form Q that is X-mu prime A X-mu. So, since X - mu is B Z, this becomes Z prime B prime A B Z that we can write as Z prime and this B prime A B, we can write as sum matrix C. Now, if we implement this result that if I am having a collection of a standard normal random variables, then Y prime A Y has a chi square k, if and only if A is idempotent.

So that condition will be applied to C and also the trace and rank of A will be = k here. So, if we apply this result, Q will follow chi square k if and only if C is idempotent and k = trace of C that is rank of C. Now, C is idempotent this condition is equivalent to, so C = B prime A B. So, B prime A B is idempotent so this you can write as B prime A B * B prime A B = B prime A B.

(Refer Slide Time: 04:40)

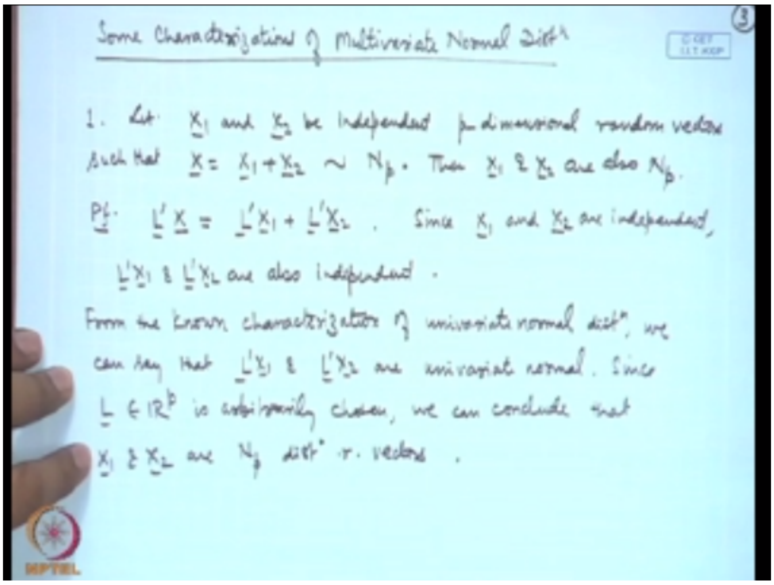


So I bring it to the left hand side, so we can write as A B B prime A - A B = a null matrix. Now, B B prime is sigma, so this becomes B prime A sigma A - A B = null matrix. Again, this is equivalent to I can pre-multiply by B and I can post-multiply by B prime, this is equivalent. Now, a question is that why is this equivalent because if I am having this, I can consider here a transformation from here to get this thing here.

So, this will be implying C sigma A sigma A - A sigma = null. Now, k = trace of C that = trace of B prime A B that = trace A B B prime because of trace of some matrix C*D is same as trace of D*C, so trace of A sigma. Now as a remark, let me mention here if sigma is non-singular, then I can multiply by sigma inverse and sigma inverse here, then this condition is A sigma A = A.

In that way, actually you can say that sigma is a generalized inverse of A that condition will be there. Now before going to, we will also discuss in detailed the noncentral chi square distribution, however, let me talk about certain characterizations of the multivariate normal distribution now. Some characterizations of multivariate normal distribution, let us consider let X1 and X2 be independent p-dimensional random vectors such that X = X1 + X2 follows Np.

(Refer Slide Time: 07:16)

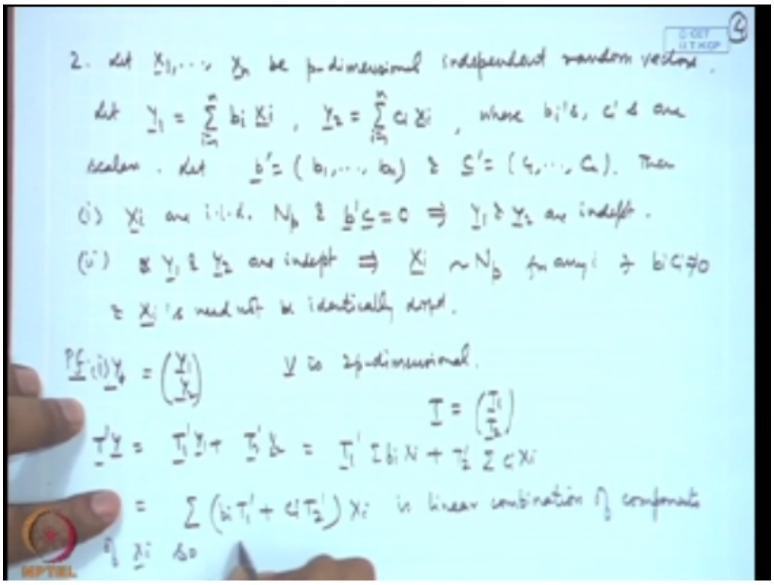


Then X1 and X2 are also Np. Let us look at the proof of this, let us consider say a linear combination of the components of X. So that is becoming L prime X1 + L prime X2. Now, since X1 and X2 are independent, L prime X1 and L prime X2 are also independent. Now, there is a characterization of the univariate normal distribution in terms of the decomposed terms that means if I say X1 and X2 are univariate normal, such that X1 + X2 follows univariate normal, then each of X1 and X2 will be univariate normal.

So from this, we conclude that from the known characterization of, we can say that L prime X1 and L prime X2 are univariate normal. Now this L, I chose arbitrarily of p-dimension since L belongs to Rp is arbitrarily chosen, we can conclude that X1 and X2 are Np distributed random vectors. A second characterization is generalization of this, which let me state in the full form here.

Let X1, X2, Xn be p-dimensional independent random vectors. Let us consider say Y1 = a linear combination of say bi Xi, i = 1 to n and Y2 as an another linear combination of the same where bi's and ci's, they are scalars. Let us consider say b as b1, b2, bn and c as say c1, c2, cn. Then, we will have the following that is Xi's are IID Np and b prime c = 0 implies Y1 and Y2 are independent.

(Refer Slide Time: 10:59)

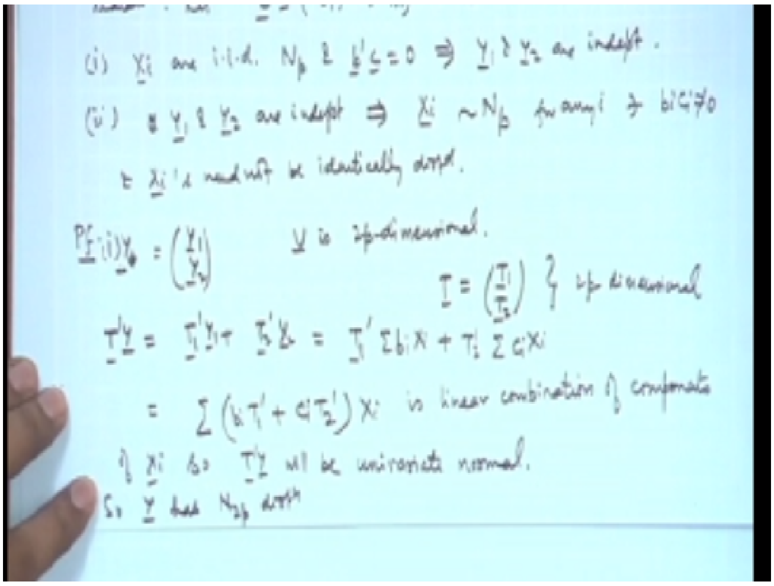


And secondly Y1 and Y2 are independent implies that Xi will follow Np for any i such that Bi Ci is not 0 and Xi’s need not be identically distributed. Let us look at the proof of this, so we can consider the vector Y1, Y2 let us call it say Y, I put them in the 2 dimensional form here. So this is now 2p-dimensional, so V is 2p-dimensional. If I consider linear combination of say T prime Y, then that will become say T1 prime Y1 + T2 prime Y2, where T = T1, T2.

If I am assuming that Xi’s are independent random vectors, so in the first part if I am assuming that Xi’s are multivariate normal, then these are linear combinations of the, because what I have done here, Y1 is a linear combination of Xi's, so this is becoming T1 prime sigma biXi + T2 prime sigma ciXi that = Sigma bi T1 prime + ci T2 prime Xi. So, this is linear combination of components of Xi.

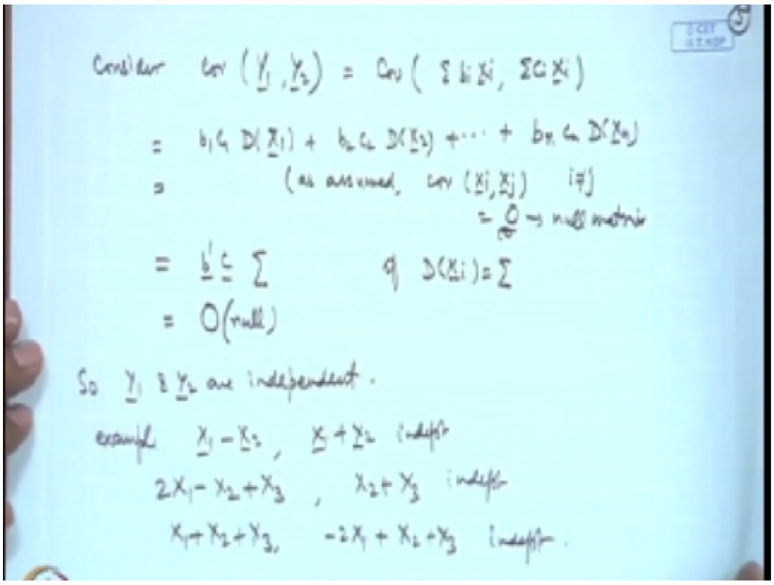
So, T prime Y will be univariate normal. So this is for any T, this is 2p-dimensional, so y has N2p distribution that is 2p-dimensional multivariate normal distribution. Now, let us consider covariance matrix between Y1 and Y2, now that will be = because I have written this as b prime X, see basically what we are getting here is covariance between Y1 and Y2 will become covariance between sigma biXi and sigma ciXi.

(Refer Slide Time: 15:11)



That will consist of, since X_i's are independent, this will reduce to $b_1 c_1$ dispersion matrix of $X_1 + b_2 c_2$ dispersion matrix of $X_2 + b$, I am taking n here, $b_n c_n$ dispersion matrix of X_n. As we have assumed covariance terms between X_1, X_i, X_j for $i \neq j$, they will be null. So, this is nothing but b prime c sigma. If we are writing dispersion matrix of X_i = sigma, then this = this.

(Refer Slide Time: 15:53)

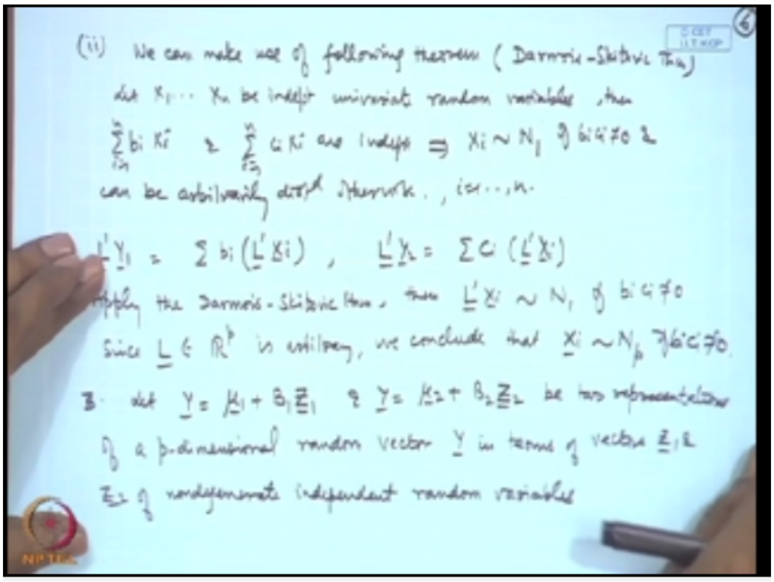


Now, if I am assuming here the b prime c = 0, then this is simply = null. So, we will get Y_1 and Y_2 are independent. So, this result is proved that if X_i's are independent and identically distributed p-dimensional multivariate normal distributions where $b_1c_1+b_2c_2+b_nc_n$, they are 0, then this Y_1 and Y_2 are independent. In particular, you may consider something like this.

For example, I take say X1-X2 and X1+X2, so then they will be independent. Suppose, I consider say 2X1-X2+X3 and say I take X2+X3, then they are also independent because if I consider here, 2*0-1+1 1*1, so that is going to be 0. If I consider say X1+X2+X3 and I consider say -2X1+X2+X3, then here the product is -2+1+1, so they are also independent so like that we can construct independent linear combinations here.

Let us look at the part B of this, in the second part, what we are saying is that if Y1, Y2 are independent, then Xi's must be Np here for any distinct. So let us look at this, so we can make use of, this is called actually Darmois-Skitovic theorem. Let X1, X2, Xn be independent univariate random variables, then sigma biXi, i = 1 to n and sigma ciXi, i = 1 to n are independent, it implies that Xi's will follow normal 1, if bici is not 0.

(Refer Slide Time: 19:31)

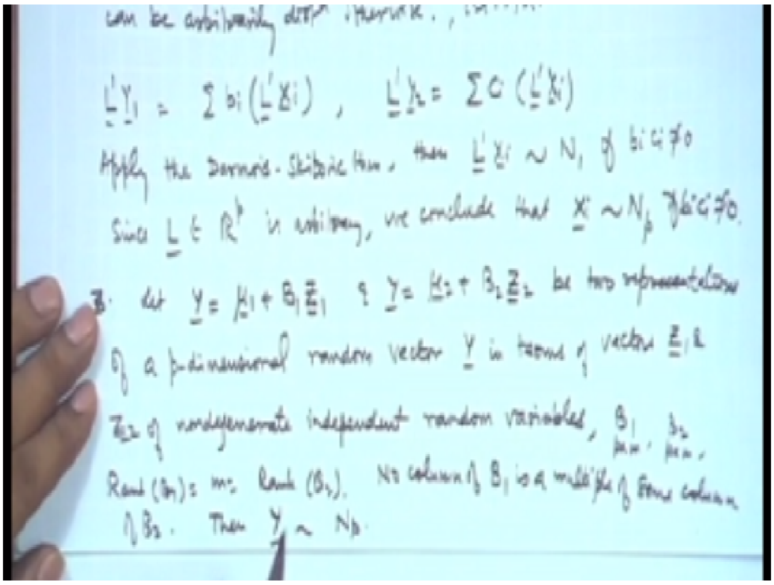


And can be arbitrarily distributed otherwise for i = 1 to n. So, let us consider say L prime Y1, so that = sigma bi L prime Xi and similarly, L prime Y2 = sigma ci L prime Xi. On this, apply the Darmois-Skitovic theorem, then L prime Xi this will follow N1, if bici is not 0. So, L is arbitrary vector in p-dimensional space, we conclude that Xi's will follow Np, if bici is not 0.

Now, if you look at the statement this is again very powerful statement. What we are saying is that if I construct linear combinations of p-dimensional random vectors and if they are independent, then each of the terms in the linear combination will have a p-dimensional normal distribution. Of course, we are putting a condition here that bici must be nonzero that means the corresponding term should be there.

A third characterization is based on the decomposition that I obtained and that we gave as an alternative definition of the multivariate normal distribution also. So, let us consider say Y = mu 1 + say B1Z, let us call it Z1 and say Y = mu 2+B2Z2. Suppose, these be 2 representations of a p-dimensional random vector in terms of say vectors Z1 and Z2 of nondegenerate independent random variables.

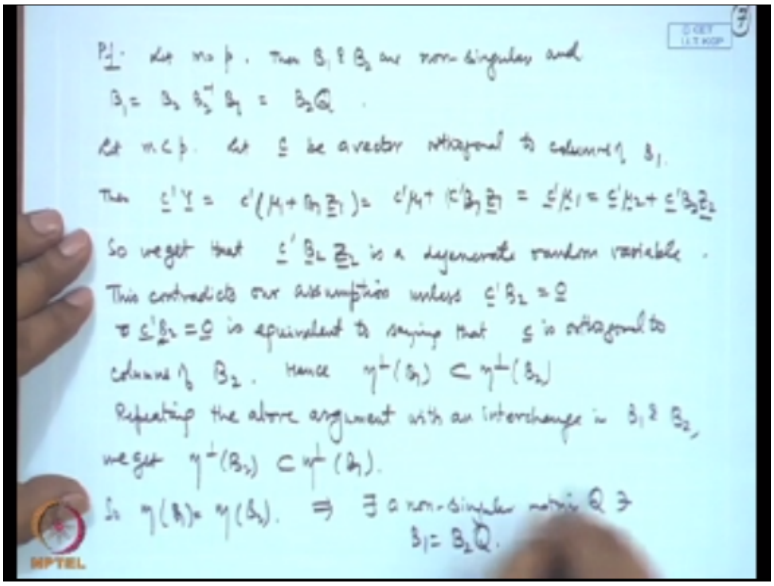
(Refer Slide Time: 24:30)



And this B1 is a p by m matrix, B2 is p by m matrix, rank of B1 is m and rank of B2 is also m. We also assume that no column of B1 is a multiple of some column of B2. Then, Y follows Np, so now you see here. I am actually using the representation that I gave as an alternative definition of the multivariate normal distribution, but in that one, Z1 and Z2 are vector of IID standard normal variables.

Here, I am saying is that this is the vector of simply nondegenerate independent random variables and then, just by putting a condition on B1 and B2, we are getting that Y must have a multivariate normal distribution. So, this is also very powerful characterization of a multivariate normal distribution. Let us consider say m = p, then B1 and B2 are nonsingular and then we can write B1 as B2 B2 inverse B1.

(Refer Slide Time: 25:47)



That we can write as a B2 and some matrix, this term we can write as some Q. Let us assume say m is < than p. Let C be a vector which is orthogonal to columns of B1 and we write here C prime Y that = C prime mu 1 + B1Z1, then that is becoming C prime mu 1+ C prime B1Z1, now this will become 0. So, this is simply C prime mu 1 here. Now, that = C prime mu 2 + C prime B2Z2.

Now, what I am getting here C prime B2Z2 = now this is a scalar, so we are getting that C prime B2Z2 is a degenerate random variable. Now, we assumed that this Z1 and Z2 are vectors of nondegenerate independent random variables. So, here I am getting this as a degenerate random variable. So, this is contradicting our assumption unless we have C prime B2 = 0.

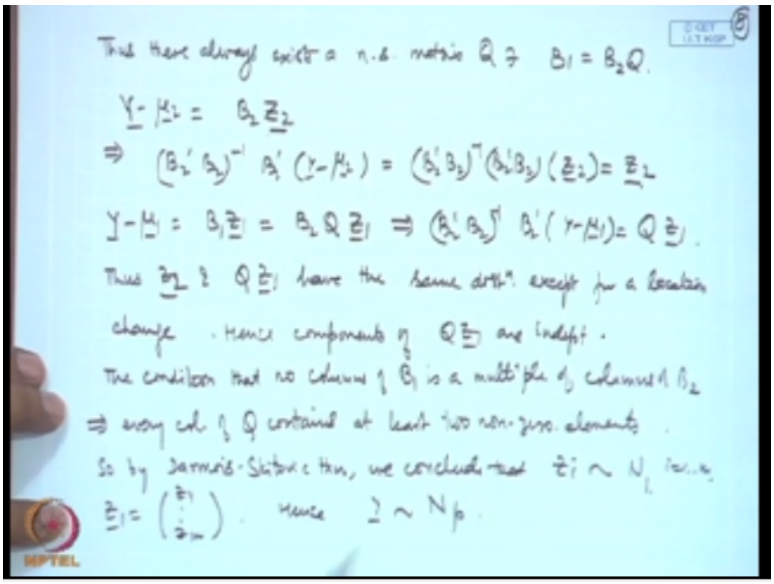
Now, if C prime B2 = 0 is equivalent to saying that C is orthogonal to columns of B2. Now let us look at this, I started with C to be a vector which is orthogonal to the columns of B1 and I am able to prove that C is now orthogonal to the columns of B2. So, this means that the orthogonal column space of say B1 is a subspace of orthogonal column space of B2.

Now in this derivation, I have taken B1 B2, now I started with C to be a vector orthogonal to the columns of B1 in place of that, suppose I write B2 here, then this statement will change here, here I will get C prime mu 2 and here I will get B2Z2, so his will become C prime mu 2 and here then, I can write C prime mu 1 + C prime B1Z1. In that case, I will get the same statement in the reverse way.

So, repeating the argument with an interchange in B1 and B2, we get orthogonal space of B2 is a subspace of orthogonal space of B1, so that means they are same. Basically, column space are B1 and column space are B2 are same. This means that there exists a nonsingular matrix Q such that B1 = B2Q. So, I have written here if m = p, then I am able to write to that B1=B2Q and if m < p then also I am able to obtain a nonsingular matrix Q such that B1=B2Q.

So, this one and 2 give that all the time there will be a nonsingular matrix. Thus there always exist a nonsingular matrix Q such that B1 = B2Q. Now we make use of this, so let us write say Y-mu 2 that = B2Z2. So, this implies B2 prime B2 inverse B2 prime Y-mu 2 that = B2 prime B2 inverse B2 prime B2Z2 that = Z2. So, if I consider now Y-mu 1 that = B1Z1 that = B2QZ1 this implies that B2 prime B2 inverse B2 prime Y-mu 1 that will be = QZ1.

(Refer Slide Time: 31:06)



So, what we are getting is that Z2 and QZ1 they have the same distribution except for a location change. Because both I am able to represent in terms of, see this is Y-mu 2, so mu 2 is the translation here and here I am getting QZ1 that is Y-mu 1 here. So, components of QZ1 they are independent. Now, the condition that no columns of B1 is a multiple of columns of B2, then this implies that every column of Q contains at least 2 nonzero elements.

So by Darmois-Skitovic theorem, then we conclude that Zi's follows normal N1, i = 1 to n. So now Z1 = your components of this, let us call it as Z11, Z12, Z1n. So, what you are getting here is that Y follows Np. So, these are the 3 characterizations, Now, we move over to the actual density function.

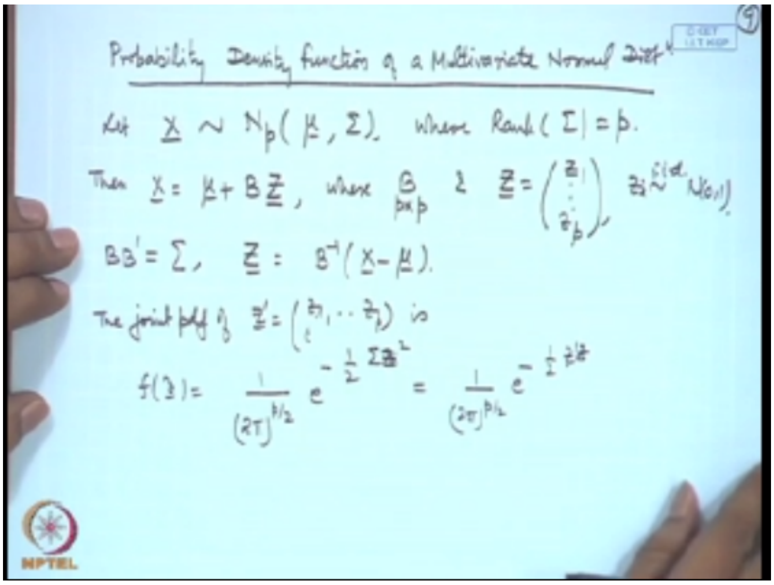
If you remember here in the case of one dimension and 2-dimension distribution, we always define a distribution and we talked about its probability and mass function and the probability density function. In the case of p-dimensional normal distribution, I have not yet actually defined the density function. So, one major reason is that when we talk about higher dimensions.

And if there is, for example here I mentioning sigma as a variance-covariance matrix is positive semidefnite. So if it is a positive definite matrix, then it will have full rank, but if it is not a full rank that means the rank say p-1 or p-2 or in general I am saying mn < p that means there will be some linear relationships among the variables there. If there are complete relationships there, in that case the density will exist on a subspace.

It will not exist on the fully space that was on p-dimensional space. So that is the reason that I gave the definition of the multivariate normal distribution in terms of its linear combinations and then in terms of an alternative representation like mu + BZ where Z is a collection of m independent univariate normal random variables. So, there m was the rank. So that means I am able to actually define in terms of alternative you can say characterization of the multivariate distribution.

I do not necessarily have to write the density function, but now I will write the density function for the full space that means when I considered the full rank, then we talk about the density function and actually, the representation that I have given, it will be exactly used for deriving the density function. So, we talk about probability density function of a multivariate normal distribution.

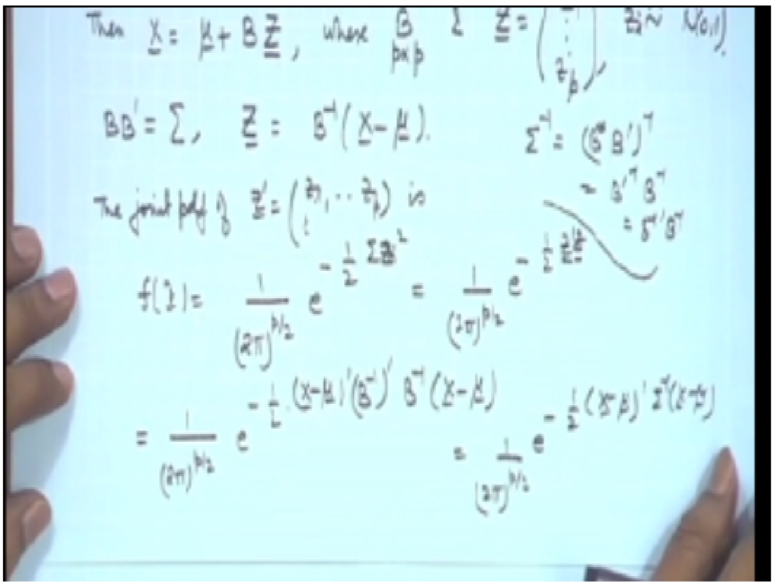
(Refer Slide Time: 36:43)



So, let us consider X following Np mu sigma and I consider full rank, rank of sigma = say p. if rank of sigma = p, then we can write X = mu+BZ, where B is p by p and Z is a vector of independent, these are IID normal 0, 1. So, if that is happening and also this B B prime = sigma and this Z = actually B Inverse X-mu. Now, if I have independent normal random variables, I can write down the density function.

So, the joint pdf of Z = this Z1, Z2, Zp, so Z prime = Z1, Z2, Zp that is nothing but let me use a notation f Z. So that = 1/2 pi to the power p/2 e to the power - 1/2 sigma Zi square. So that will be = 1/2 pi to the power p/2 e to the power - 1/2 Z prime Z. Let me use capital letters here, usually we write small letters for denoting the value of the random variable, but here for the sake of convenience, I am using the capital letters here.

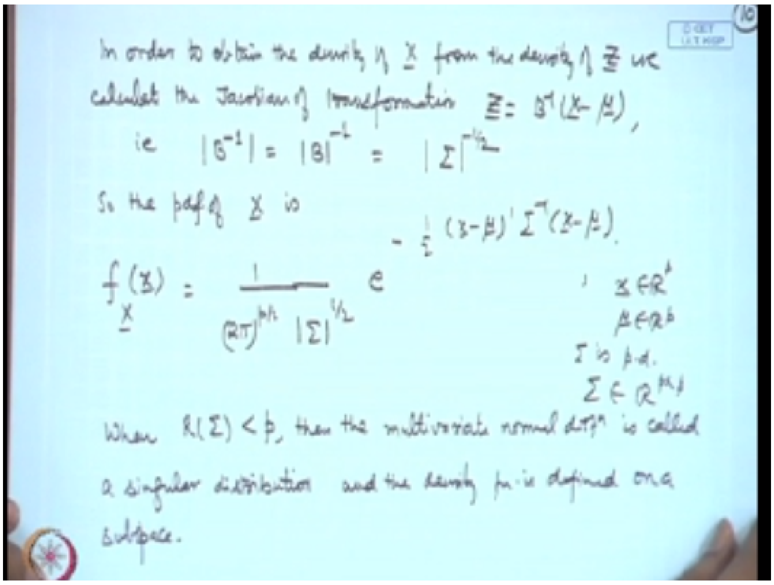
(Refer Slide Time: 39:07)



Now, this Z is given in terms of this, so we write it here that = 1/2 pi to the power p/2 e to the power - 1/2, now Z = this term here, so it is becoming X-mu prime B inverse prime B Inverse X-mu. Now, if am assuming this B B prime = sigma, then sigma inverse = B B prime inverse that = B prime inverse B inverse that will be = B inverse prime B inverse. So we can use this, so this is simply becoming 1/2 pi to the power p/2 e to the power-1/2 X-mu prime sigma inverse X-mu.

Now, if I am obtaining the distribution of X from here, then I have to calculate the Jacobian here. So, what will be the Jacobian term here? In order to obtain the density of X from the density of Z, we calculate the Jacobian of transformation that is Z = B inverse X-mu. So that is given by determinant of B inverse which is same as determinant of B inverse, which is also the determinant of sigma to the power -1/2.

(Refer Slide Time: 40:22)

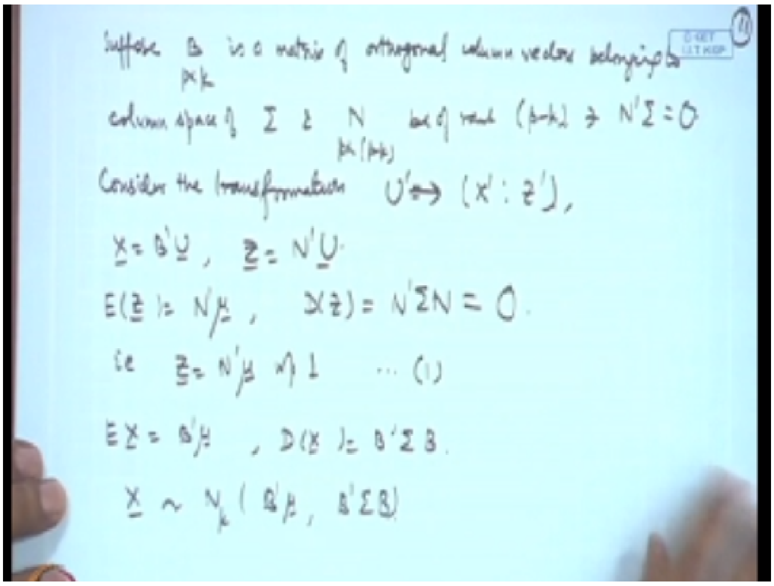


So, the pdf of X is given by that = 1/2 pi to the power p/2 determinant of sigma to the power 1/2 e to the power - 1/2 X-mu prime sigma inverse X-mu. Here, X belongs to Rp, mu belongs to Rp and sigma is positive definite matrix. Sigma is Rp/p that is p/p positive definite matrix. When rank of sigma is < p, then the multivariate normal distribution is called a singular distribution and the density function is defined on a subspace.

Suppose B that is p/k is a matrix of orthogonal column vectors belonging to column space of sigma and N that is p/p-k be of rank say p-k such that N prime sigma is null matrix. So, let us consider the transformation, U prime going to X Z prime, where X is B prime U, Z = N prime

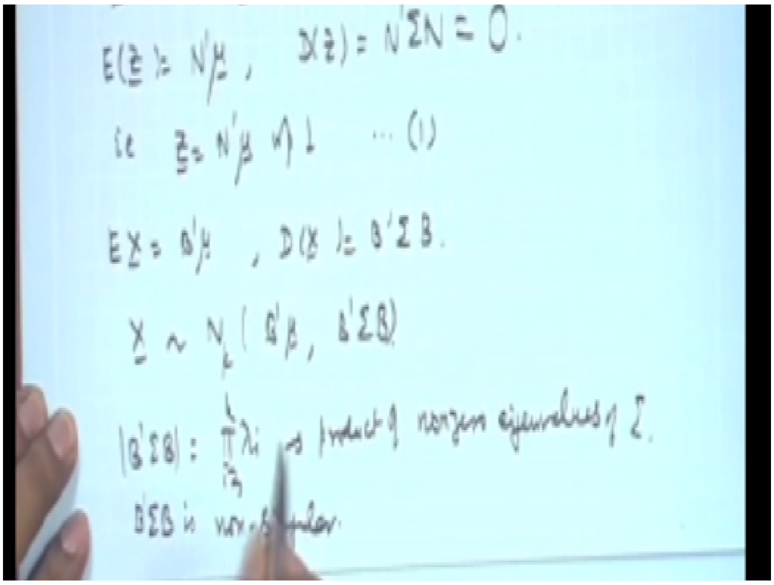
U. Then, expectation of Z = N prime mu, dispersion matrix of Z = N prime sigma N that is becoming null.

(Refer Slide Time: 42:57)



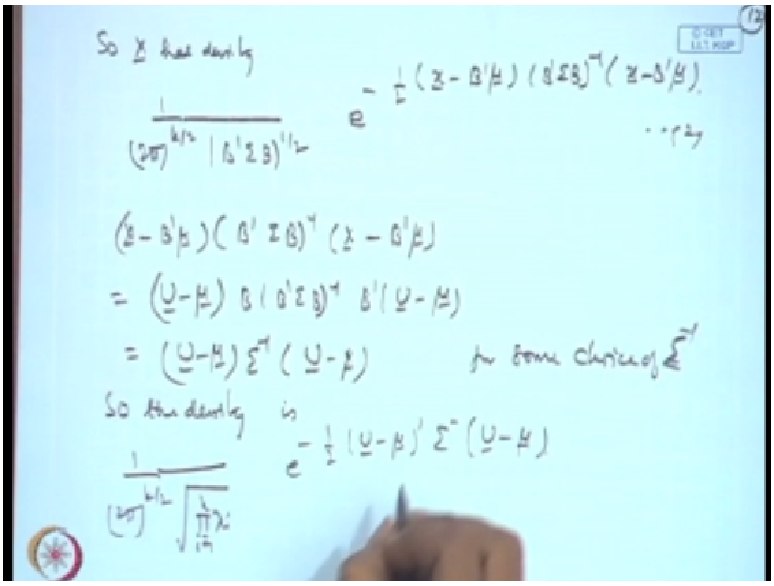
That means Z = N prime mu with probability 1 that is degenerate (()) (44:23) and expectation of X = B prime mu, dispersion matrix of X = B prime sigma B. So, X follows Nk B prime mu B prime sigma B. So, we can write actually B prime sigma B can be written as a product of nonzero Eigen values of sigma. So, B prime sigma B is nonsingular.

(Refer Slide Time: 44:48)



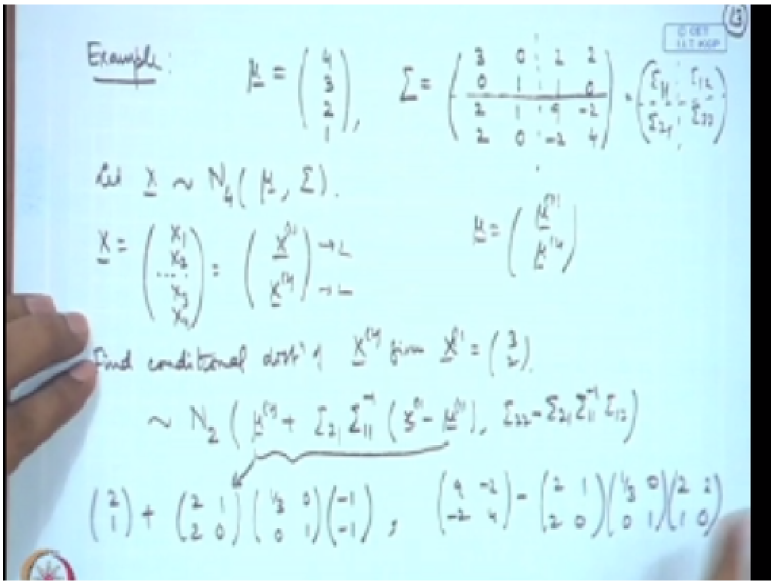
So, X will have density 1/2 pi to the power k/p B prime sigma B to the power 1/2 e to the power -1//2 X-B prime mu B prime sigma B to the power -1 X–B prime mu, so this description 1 and 2 that describes the density. If you consider say X–B prime mu, B prime sigma B Inverse X–B prime mu then that is U-mu B B prime sigma B Inverse B prime U-mu = U-mu, a generalized inverse of this U-mu for some twice of sigma g inverse.

(Refer Slide Time: 45:18)



So the density is actually 1/2 pi to the power k/2 product of the determinant the Eigen values i = 1 to k, e to the power -. So, this is actually density of on a subspace. It is not a density on the full space when the rank of sigma is not full. Now before going to the estimation, let us consider one or 2 applications of this conditional distribution or linear combinations etc. One example of a multivariate normal distribution let me write here.

(Refer Slide Time: 47:21)



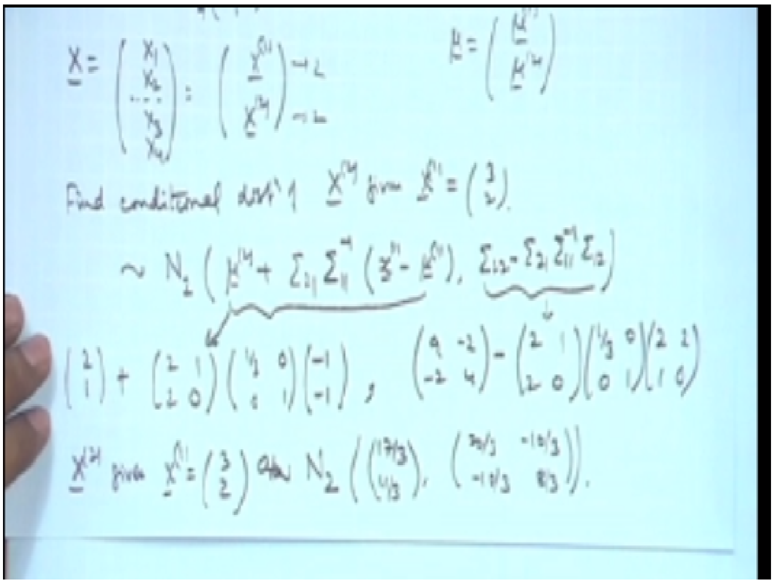
Let us consider say mu = 4, 3, 2, 1 and I consider sigma as 3, 0, 2, 2, 0, 1, 1, 0, 2, 1, 9, -2, 2, 0, -2, 4. So let us take say X following N4 mu sigma, let us consider some partitioning of this, say it = X1, X2, X3, X4, which I am actually writing as a X1and X2, okay. That is this is 2 dimensional and this is 2 dimensional here. Let us define say conditional distribution of say X2 given X1 = say 3, 2.

Now, we have discussed the conditional distribution of one component giving the second component. So this follow N2 and if I considered the corresponding decomposition of mu as a mu 1 and mu 2, then this will become mu 2 + sigma 21, so I am partitioning this as sigma 11, sigma 12, sigma 21, sigma 22. So, if I considered this, then this term is sigma 11, this is sigma 12, this is sigma 21, and this is sigma 22.

So this will become, so let us calculate these terms here. So this one is now 2, 1 + sigma 21 is this term here, 2. 1, 2, 0, sigma 11 inverse is the inverse of this that is 1/3, 1, 0, 0 and then you have X1-mu 1, so 3, 2-mu 1, so that will become -1, -1 and here I will get 9, -2, -2, 4, -sigma 21 that is 2, 1, 2, 0, sigma 11 inverse*sigma 12 that is 2, 2, 1, 0 that is the dispersion term here.

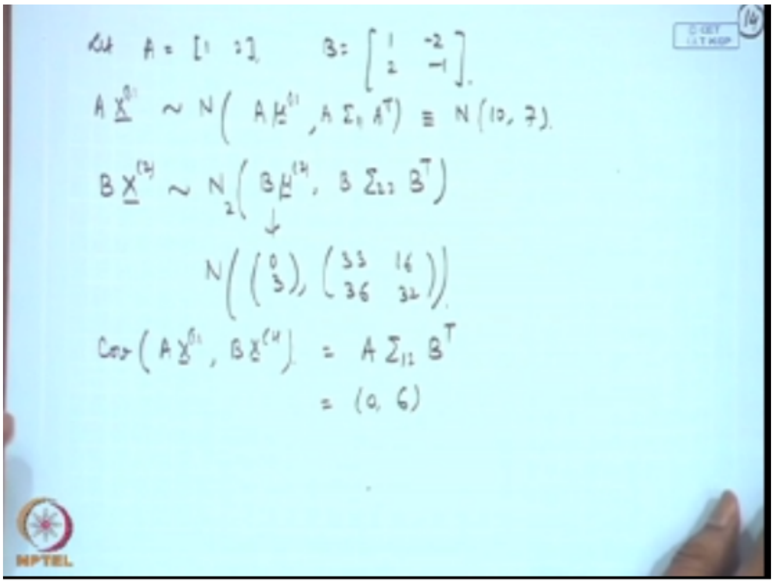
So, I will get here X2 given X1 = 3, 2 as N2 17/3, 11/3, 20/3, -10/3, -10/3, 8/3. So, I am able to obtain the conditional distribution of X2 given X1 = a certain number. So, this is quite interesting here, you can obtain and you can actually look at this, this is 4, 3, 2, 1 and here you see X2 given some value of X1. So, here you can see that there is a dramatic change here, this is 17/2 which is bigger than 5 itself, this is around 4.

(Refer Slide Time: 50:50)



And whereas the original means of X2 was only 2, 1. So, if X1 is given 3, 2 then it has increase the means of X2 and similarly, there is substantial change in the value of the variance-covariance terms here. Let us also define in the same, A = say 1, 2 and let us consider say B = 1, -2, 2, -1. What is the distribution of say AX1? So, according to AX1 will have normal with mean.

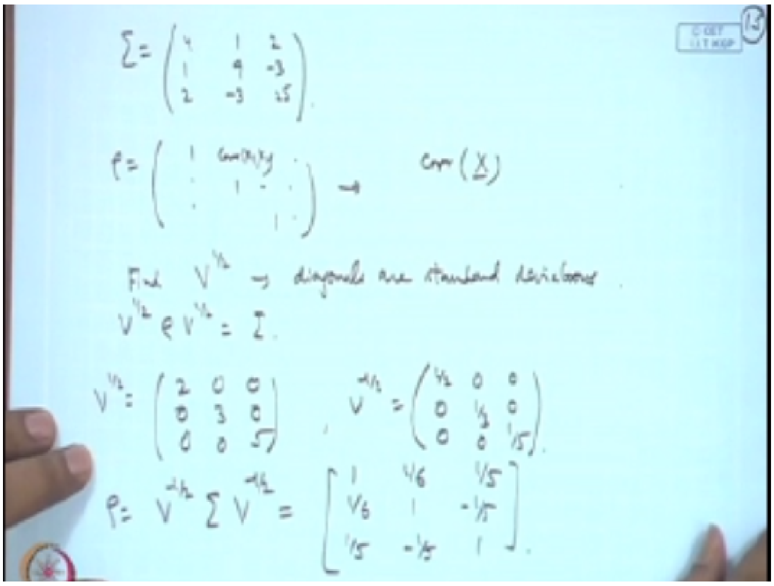
(Refer Slide Time: 52:14)



So A mu 1 because A is a scalar here, A is a row vector here. So this will become a scalar and then you will have A sigma 11 A transpose. So, you can calculate this, this value is simply 10 and this is 7. Similarly, suppose I consider BX2, so BX2 is actually = 2 dimensional vector here that is following normal B mu 2, B sigma 22, B transpose. So, if you can calculate this, these terms have to be 0, 3, 33, 16, 36, 32.

Let us also consider say covariance between AX1 and BX2, then this will become = A sigma 12 B transpose, so that = 0, 6. So, in this example I have shown you a direct application of the distribution theory of the multivariate normal distribution and various things were considered here. Let me give one more exercise here. Let us consider say sigma = 4, 1, 2, 1, 9, -3, 2, -3, 25, okay.

(Refer Slide Time: 54:23)



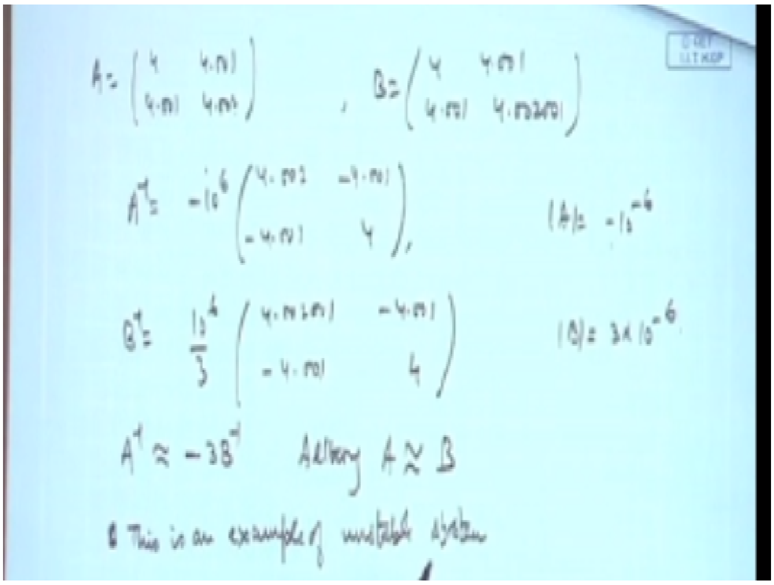
So, here I considered rho as say 1, 1, 1 and here I will consider correlation between, so this is actually correlation matrix, correlation matrix of X, okay. So that means these term will be denoting correlation between X1, X2 not covariance, it is correlation terms here. So, consider find V1/2, where these diagonals are standard deviations and find rho and also show that V1/2 rho V1/2 = sigma for this particular case.

See this is an interesting thing because you can do the manipulation with the variance-covariance matrix because of the positive semidefiniteness of the matrix, this is very important because it has a spectral decomposition, you can have a gram decomposition as V V transpose etc. So, lots of nice properties are coming here. Let us consider V1/2 here will become = 2, 3, 5 that is the standard deviation matrix here.

Let us consider V -1/2 so that will be 1/2, 1/3, 1/5, 0, 0, 0, 0, 0, 0 and rho = V to the power -1/2 sigma V to the power -1/2 that = 1, 1/6, 1/5, 1/6, 1, -1/5, 1/5, -1/5 and 1. I mentioned about the uniqueness of the sigma 21, 11 inverse term actually. So see there is a problem, suppose we are calculating the inverses, then sometimes the inverses will not exist or with the little variation, the inverse may vary too much.

So that means it is an example of an unstable matrix. Let us take one case here at least, let me give you example of one such problem. Let us take say A = 4, 4.001, 4.001, 4.002 and B = say 4, 4.001, 4.001, 4.002001, you can see here that in A and B, 3 terms are exactly the same, the 4th term I have modified only by adding 0.000001 okay only that much difference is there.

(Refer Slide Time: 57:25)



Let us look at say A inverse, A inverse = -10 to the power 6, 4.002, -4.001, -4.001, 4 and if I look at B inverse then that = 10 to the power 6/3, 4.002001, -4.001, -4.001, 4. So you can see that there is a dramatic change in the value here. Actually, determinant of A is turning out to be -10 to the power -6, whereas determinant of B = 3*10 to the power -6. So, there is substantial change in the value.

So, we are getting that A inverse is approximately -3B inverse. If you look at A and B, there is hardly any difference here. In fact, the 3 terms are exactly the same in the 4th term, I am considering the change after 5 decimal places. In the 6th decimal place, there is a minor change by 0.000001, but if you look at the inverses here, A is almost same as B, but if you look at the inverse, so then you are getting substantial change.

So, this is an example of unstable system. The reason is that if I look at this that they are almost linearly dependent here, see A if you look at this is dependent, so this is almost dependent here because there is a small change. In that case, a small change in the value of one term makes a huge change in the value of B inverse. In the next class, I will be talking about the estimation of the parameters of multivariate normal distribution.

We will also discuss the noncentral chi square distribution etc because that concept is coming here and it will be also used in finding out the distributions of the statistics there, so that I will be taking up in the next lecture.