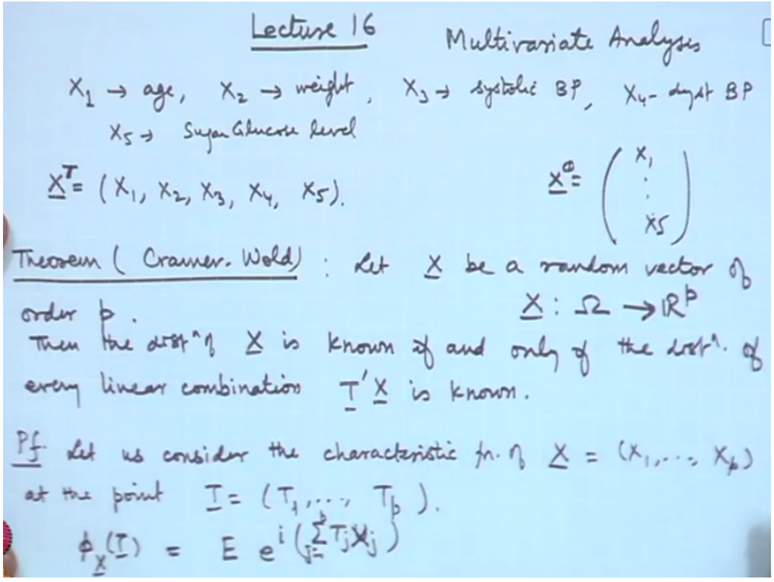
**Statistical Methods for Scientists and Engineers**
**Prof. Somesh Kumar**
**Department of Mathematics**
**Indian Institute of Technology – Kharagpur**

**Lecture - 16**
**Multivariate Analysis - I**

In this section of this course, we will introduce methods of multivariate statistics. Now we have seen in the first section on random variables, that we can also discuss bivariate and multivariate distributions. That means, for example you consider a data on the patient, when a patient goes for some diagnostic test, so concerned physician or the doctor, he may record his various characteristics.

**(Refer Slide Time: 00:54)**




For example, his age is recorded, X2 for example is weight, and X3 could be his systolic blood pressure, X4 could be his diastolic blood pressure, X5 could be his sugar/glucose level, etc. So in general then for a particular patient you have the data on 5 variables. We can use it as a row vector then I will put a transpose here otherwise we can consider column vector, then it will be X1, X2, ... X5.

In particular, we have discussed bivariate normal distribution and also some specific problems on normal distribution in the course on probability and statistics. So the primary thing that we have to notice here is that there may exist some correlatedness among these variables. And now like in the case of univariate distributions, we stabilised normal distribution as one of the important distribution, or you can say more frequently used distributions.

The reason was the application of the centre limit theorem, that means whenever we are considering averages or the summations then the data can be approximated at the distribution of the sums or the means can be approximated by the normal distribution. In a similar way, we also have a multivariate centre limit theorem, which I may mention briefly and therefore that brings into the focus a multivariate normal distribution.

Now a multivariate normal distribution can be considered as an extension of a bivariate normal distribution. And we will introduce the concept. So we will first study multivariate normal distribution and then in particular certain distributions which are used for inference. So for example, in the univariate case you had Chi square distribution, T distribution, etc., which were related to the normal distribution.

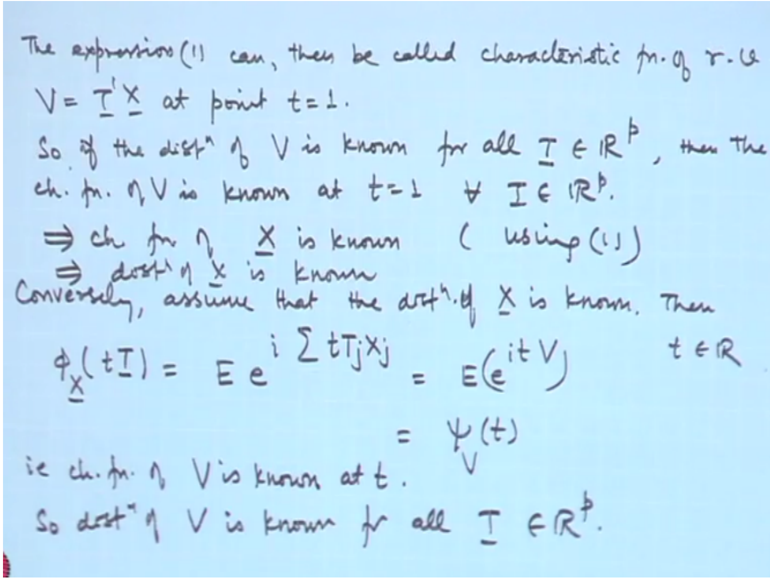
Similarly, in the multivariate case there will be certain distribution such as Wishart distribution or Hotelling's T-squared distribution distribution, etc., which will be used for the inference purposes. So in this particular section of this course, we will introduce various multivariate distribution which are related to multivariate normal distribution. So let me start with the theory of multivariate distributions.

So the first theorem or the first result which is actually attributed to Cramer-Wold. Let X be a random vector of order p. So that means we are assuming X is a mapping from the sample space into Rp. Then the distribution of X is known if and only if the distribution of every linear combination say T prime X is known. So basically characterisation of a multivariate distribution can be done in terms of its linear combinations.

But of course we have to consider every linear combination. The proof is based on a characteristic function approach. Let us consider the characteristic function of say X = X1, X2, Xp at the point say T = T1, T2, Tp. So let us use some notation phi X T, then it is equal to expectation of e to the power i summation Tj Xj, j=1 to p, which also can be written as expectation of e to the power i T prime X.

Now if I give this name as V, then this is = expectation of e to the power iV. So this can be then considered as, let us call this expression as say 1.

**(Refer Slide Time: 07:04)**



The expression (1) can, then be called characteristic fn. of r.v.
$V = \underline{T}'\underline{X}$ at point $t = 1$.
So if the distn of $V$ is known for all $\underline{T} \in \mathbb{R}^p$, then the ch. fn. of $V$ is known at $t = 1$ $\forall\ \underline{T} \in \mathbb{R}^p$.
$\Rightarrow$ ch fn of $\underline{X}$ is known (using (1))
$\Rightarrow$ distn of $\underline{X}$ is known
Conversely, assume that the distn of $\underline{X}$ is known. Then

$$\phi_{\underline{X}}(t\underline{T}) = E e^{i \sum t T_j X_j} = E(e^{itV}) = \psi_V(t) \qquad t \in \mathbb{R}$$

ie ch. fn. of $V$ is known at $t$.
So distn of $V$ is known for all $\underline{T} \in \mathbb{R}^p$.

The expression 1 can then be called characteristic function of random variable V that is = T prime X at point, say t = 1. So if the distribution of V is known for all T belonging to p dimensional equally to any space then the characteristics function of V is known at T = 1 for all T belonging to Rp. This implies characteristics function of x is known using 1. Now this implies that distribution of X is known.

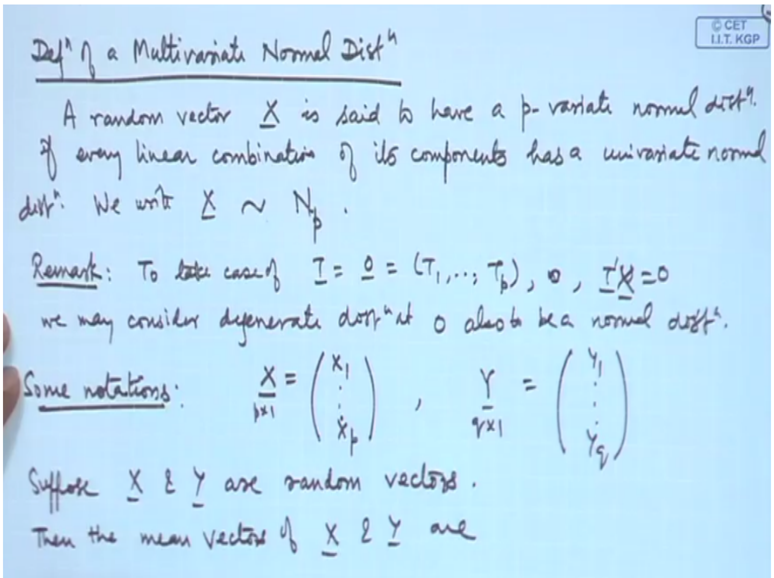
Conversely assume that the distribution of X is known. Then phi X at the point say t of T where T is a real number. Then this is becoming the expectation of e to the power i sigma t Tj Xj that is = expectation of e to the power itV. That is the characteristic function of V at the point t that is characteristic function of V is known at t. So the distribution of V is known for all. So this theorem actually it is a characterisation theorem.

That means it says that the distribution of a random vector can be described in terms of the linear combination provided all the linear combination distributions are known. Conversely if the distribution of random vector is known then all its linear combination will have a known distribution. Now in fact we use this definition in the first case for introducing a multivariate normal distribution.

Later on we will see equivalent versions but we will find this quite convenient to introduce a multivariate normal distribution through this. If you remember a property for the bivariate normal distribution in the course of probability and statistics that we proved that if xy is a bivariate normal distribution, then every linear combination ax+by is univariate normal distribution.

Conversely if for every a and b, ax + by has a univariate normal distribution, then xy has a bivariate normal distribution. So basically you can see that this theorem Cramer-Wold theorem is a most general version of this result. So basically we use the same definition for a multivariate normal distribution.

**(Refer Slide Time: 11:14)**

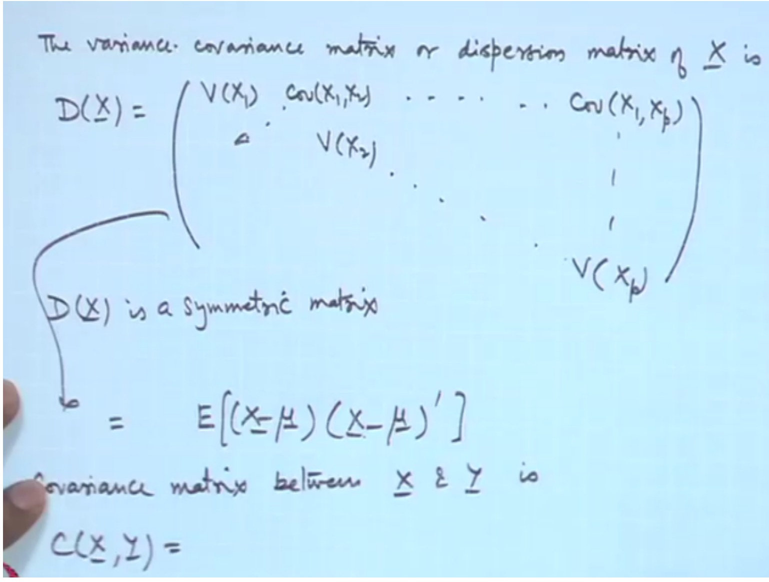



So we define the definition of a multivariate normal random, so a random vector X is said to have a p variate normal distribution if every linear combination of its components has a univariate normal distribution, and we write X follows Np. As a remark, let me mention here that to take the case of T = 0 vector, that means basically we are saying T prime X is 0, then we may consider degenerate distribution at 0 also to be a normal distribution.

Now let me introduce some multivariate notations, suppose I consider, let us consider the p dimensional vector X1, X2, Xp, say this is p by 1 vector another vector is say y which is may be of a q dimension say Y1, Y2, Yq. Suppose X and Y they are random vectors, then the mean vectors are defined by, so here you will have all the components coming here, expectation of X1 and so on, expectation of Xp, which we may call say mu1, mu2, mu p that is = to mu vector.

And similarly say for the Y which let us give some notation say nu1, nu2, nu q, that is a nu vector. Now in the case of one variable we have the variants. So for a multivariate, that means for a random vector, we will have a variance-covariance matrix.

**(Refer Slide Time: 14:55)**

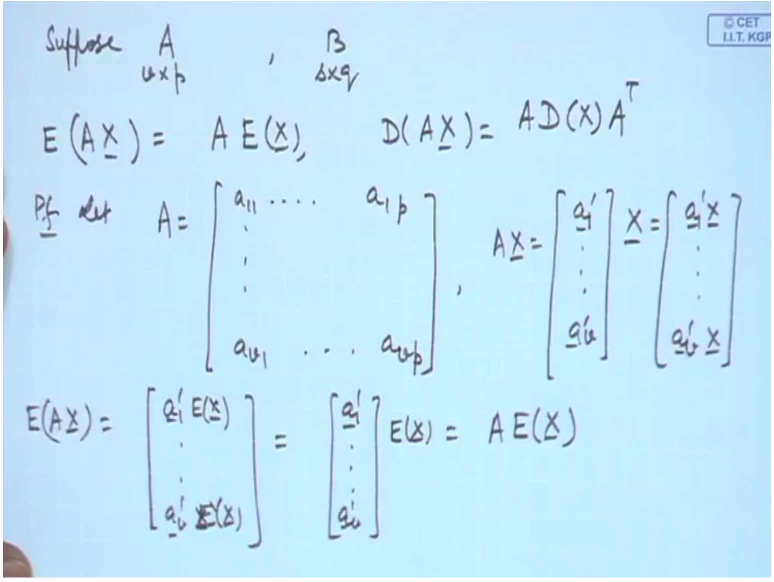



The variance-covariance matrix or say dispersion matrix, it is defined as D X say, that is = variance of X1, variance of X2, so in the diagonal we will have the variance of the components, and in the off diagonal terms we will have covariance between X1, X2 and so on, covariance between X1, Xp. Now these terms will be same, this is a symmetric matrix, Dx is a symmetric matrix. And if we consider, so this is also another interpretation.

So this is equal to expectation of X - mu, X - mu prime. This is a column vector and this is a row vector. So if we multiply we get this matrix here. Now covariance matrix between 2 vectors say X and Y, this is called say CXY, that is consist of all the covariances, that is covariance between X1, Y1, covariance between X1, Y2 and so on, covariance between X1, Yq and so on.

Here you will have covariance between X2, Y1 and so on and here you will have covariance between Xp, Yq. This is a p/q matrix, this is actually if you consider CYX, then it is transpose of this.

**(Refer Slide Time: 17:25)**

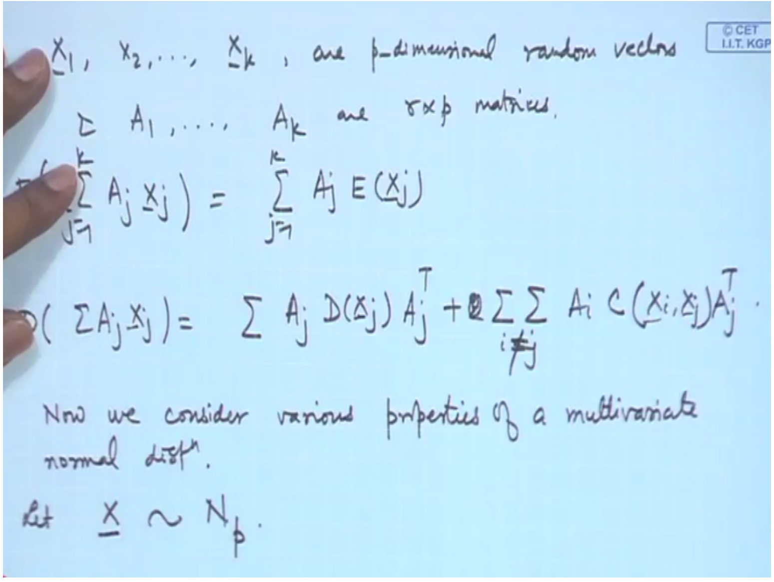



If we consider linear transformation of a vector, suppose A is a v/p matrix and say B is a s/q matrix, then AX if I consider this, then this will have A * expectation of X and that is dispersion matrix of AX that will become A times this dispersion matrix into A transpose. This can be proved easily, let us consider say, let us take A to be A11, and so on A1p and so on Av1, Avp. So Ax then will become = we can consider these as the vectors here.

See this is actually a1 prime and so on av prime vector multiplied by X here. So I can consider it as a1 prime X and so on av prime X, see if I consider expectation of AX, so that will become component wise expectation. That will become a1 prime expectation of X and so on, av prime expectation of X. So that is = a1 prime and so on av prime expectation of X, so that is = A times expectation of X.

In a similar way, we can consider the dispersion matrix of AX, the dispersion matrix of AX it is = expectation of AX - A mu * AX - A mu transpose. So that is = A * expectation of X - mu *X -mu transpose *A transpose. So that is = A times dispersion matrix of X into A transpose. If I use the notation DX = sigma, then we can say that D (AX) = A * sigma A transpose.

**(Refer Slide Time: 20:24)**

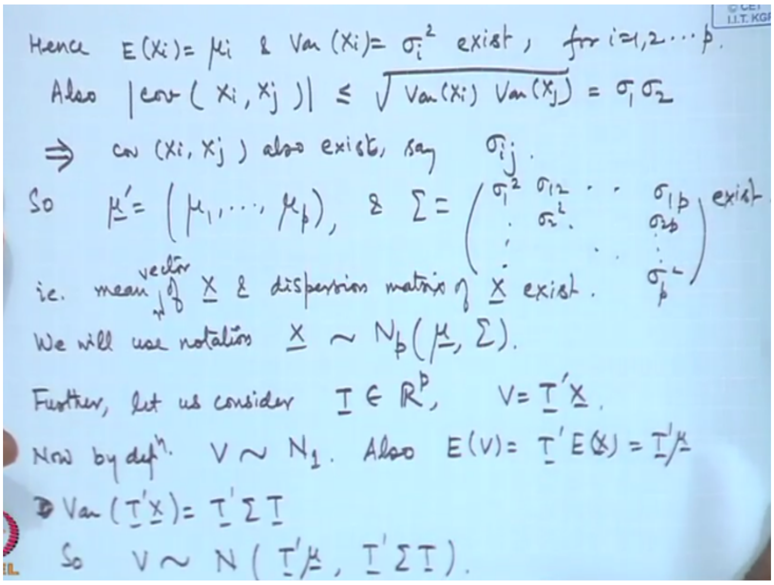



If we are considering say X1, X2 and so on Xk, these are some, they are p dimensional random vectors, and A1, A2, Ap, Ak, these are say r/p matrices. Then let us consider Aj, Xj sigma j = 1 to k, then expectation of this will become, sigma Aj expectation of Xj and also if I consider dispersion matrix of sigma Aj Xj, then that is = sum of Aj, dispersion matrices of Xj Aj transpose + twice, double summation.

So we may put not = I think that will be better, we put Ai C of Xi, Xj Aj transpose. That is the covariance matrix between Xi and Xj. So these are the certain you can say linearity properties of the random vectors. Now we go back to our definition of the multivariate normal distribution. So if you remember we define that if every linear combination has a univariate normal distribution, then we say that X as a p dimensional of p variate, multivariate normal distribution.

So now let us look at properties of the multivariate normal distribution. Let us assume that X follows Np. Then the first thing is that this will imply that, if I consider the components X1, X2, Xp, then Xi will follow N1 for i = 1 to p, by definition of multivariate normal, because every linear combination of Xi, since each Xi is a linear combination by choosing eiX, where ei = 0, 0 and so on, 1 at the i-th place and 0, 0, 0.

**(Refer Slide Time: 24:13)**



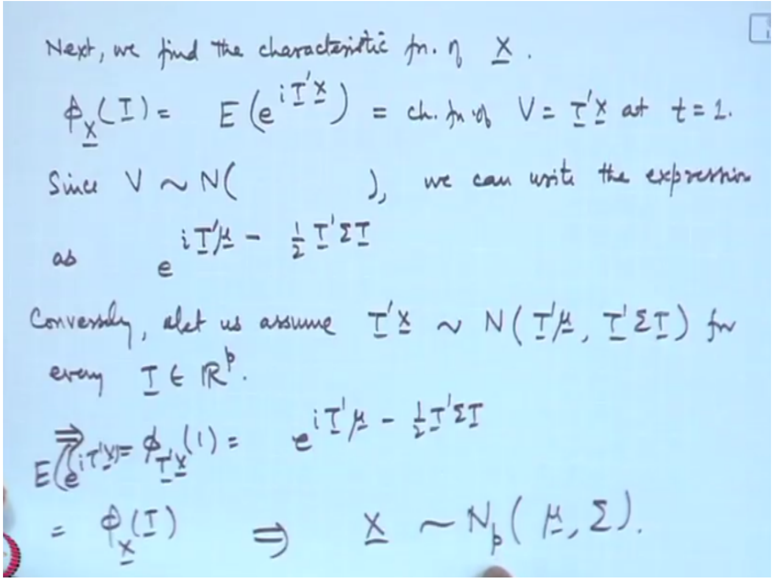

Now if Xi has a univariate normal distribution it will have some mean and variance. Hence expectation of Xi = mu i and variance of Xi that is = sigma i square exist. This will exist for i = 1 to p and also if I consider covariance between Xi and Xj then it is </= square root of variance of Xi and variance of Xj. So that is = sigma 1, sigma 2. So this will imply that if I consider absolute value, this will imply that covariance between Xi, Xj also exists, sigma ij.

So let me write mu1, mu2, mu p and sigma that is = sigma1 square, sigma2 square, sigma p square, sigma12 and so on sigma 1p and so on sigma2p, etc., this exist. That is mean of X and mean vector of X and dispersion matrix of X exist. That means I started with the assumption that if X has a multivariate normal distribution, then certainly it is mean vector and that is dispersion matrix are well defined. So we will use the notation X follows Np, mu sigma.

Now further let us consider say T belonging to Rp, and let us say V = T prime X. Now by definition V follows univariate normal also expectation of V will become = T prime expectation of X that is = T prime mu and the dispersion matrix or you can say variance of T prime X, that will be = T prime sigma T. So what we are proving here is that V follows normal distribution with mean T prime mu and variance T prime sigma T.

So if X as a multivariate normal distribution we are able to identify its mean and variance and at the same time we are able to identify completely the distribution of any linear combination here. Now in terms of the linear combination one can write down the characteristic function that can be used here.

**(Refer Slide Time: 28:24)**



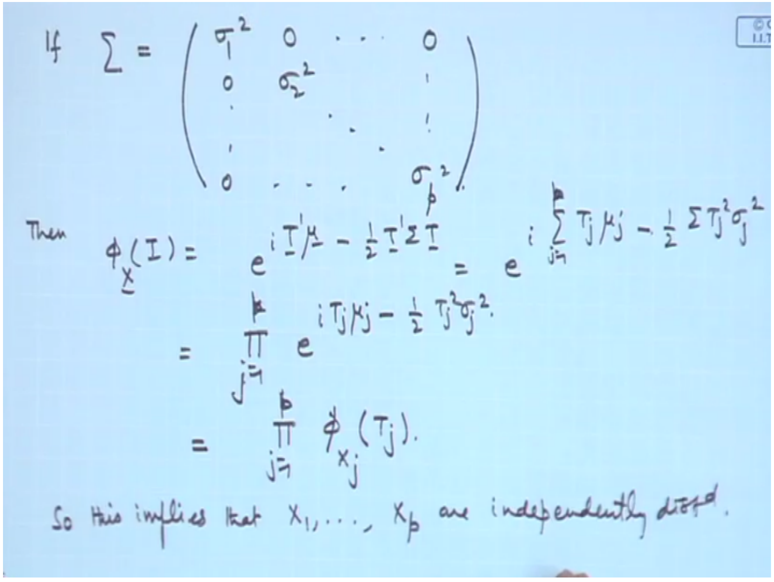
So next we find the characteristic function of X. So phi XT that is = expectation of e to the power o T prime X. So this can be considered as the characteristic function of V that is = T prime X at T = 1. Now since, V has normal distribution we can write down the expression as e to the power i mu T. So T prime mu – 1/2 T prime sigma T, 1/2 sigma square, so this is sigma square, T square T is 1. So this is the characteristic function of X at a general point T.

So if we go by the definition, the definition of a multivariate was characterized in terms of its linear combination. Now assuming a multivariate normal distribution, we are able to identify its mean, vector, its dispersion matrix, at the same time we are also able to identify completely the distribution of its linear combination and also we have found the characteristic function.

Now we will do the converse, if we assume the distribution of the linear combination, let us look at the distribution of the random vector itself. Conversely let us assume T prime X follows normal with T prime mu, T prime sigma T for every T in the p dimensional Euclidean space. Then this will imply that the characteristic function of this T prime X at the point 1 that is = e to the power i T prime mu – 1/2 T prime sigma T.

But according to the definition it is = expectation of e to the power i T prime X. So this is nothing but the characteristic function of X at T, this will imply that X follows Np mu sigma. So the converse result is also now established, that if I know that every linear combination has a univariate normal distribution then it implies the exact form of the distribution of multivariate normal with the mean mu and variance, covariance matrix as sigma.

**(Refer Slide Time: 32:13)**

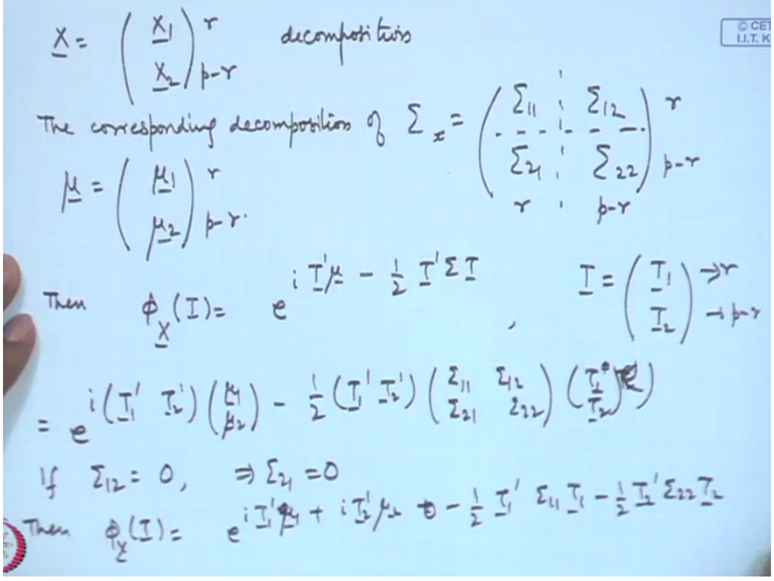



Now what we do we consider the independence criteria, if the variance, covariance matrix is only diagonal, then let us write down phi XT that is = e to the power i T prime mu – 1/2 T prime sigma T. So that is = this will become sigma Tj mu j, so this is e to the power i sigma Tj mu j j = 1 to k – 1/2 sigma Tj square sigma j square. So this I can write as product of i = 1 to k e to the power, let us put j here, i Tj mu j - half Tj square sigma j square.

But this I can consider as the product of the characteristic functions of Xj at the point Tj. So this implies that X1, X2, Xp, this will go up to p actually not k here, they are independently distributed. So as in the case of bivariate normal distribution we have seen that independence condition is equivalent to the covariance between the 2 variables being 0. In the multivariate case also like wise you have a generalisation.

The independence condition is equivalent to all the covariances term being = 0, consequently all the co-relation between the components will be 0. This is equivalent condition, that means it is if and only if. If the random variables are the components are independent all the co-relation between the components will be 0, conversely if all the co-relations are 0 then the random variables X1, X2, Xp they will be independent normal random variables.

**(Refer Slide Time: 35:09)**



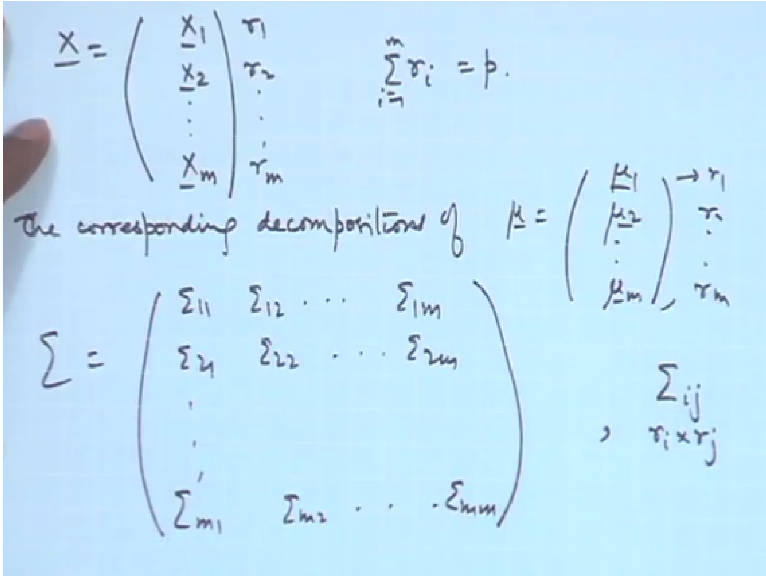

Now this result can be generalised to consider decomposed vector. For example, if I am considering say X = X1, X2, that means I am putting some r terms here and p - r terms here. So we are considering decomposition, so the corresponding decomposition of sigma, let us consider it as sigma11, sigma 12, sigma21, sigma22. So this is r, this is p - r, this is r, this is p - r, so similarly you will have decomposition of mu as mu1, mu2.

Then let us consider characteristic function of X e to the power i T prime mu – 1/2 T prime sigma T. What I consider, I split T also as T1, T2, where this is r components, this is p - r components. If I have this, then I can write it as e to the power i T1 prime T2 prime mu1 mu2 – 1/2 T1 prime T2 prime T1T2. Now if I take sigma12 = 0, this implies sigma21 is also 0, because this is transpose of this.

Then this is = e to the power i T1 prime mu1 + i T2 prime mu2 – 1/2 T1 prime sigma11 T1 – 1/2 T2 prime sigma22 T2, which I can consider as the characteristic function of X1 at T1 into characteristic function of X2 at T2, that is X1 and X2, they are independent. So the correlations being 0 implying independence, this is true for multivariate situation also, that means I consider multivariate components of the p-dimensional vector.

I am considering here 1 as a r dimensional vector and another as a p - r dimensional vector, and then if I put that all the covariances between the components of X1 with the components of X2, they are here in sigma12 and sigma 21. If they are vanishing, then X1 and X2 will be independent, so this result is also true in general.

**(Refer Slide Time: 39:00)**

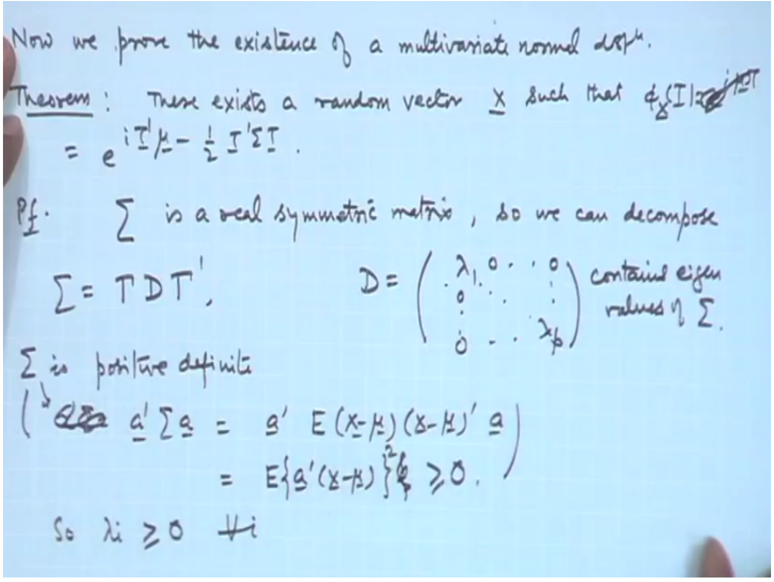



Now I can consider any subsets in place of 2 subsets, if I consider in general any number of subsets, let us consider say, this X as X1, X2 and so on say Xm, where this as r1 component, this as r2 component, this as rm component such that sigma ri, i = 1 to n, = p. So the corresponding decompositions of mu, so that will be mu1, mu2 and so on mu m and for sigma it will be = sigma11, sigma12, sigma1m, sigma21, sigma22, sigma2m, sigmam1, sigmam2, sigma mm.

Here sigma ij matrix, this will be of the order ri * rj, this is r1, r2, rm, etc., so here also you will have the result that if sigma ij for i not = j vanishes, then X1, X2, this is for all i not = j. then X1, X2, Xn they are independently distributed multivariate normal distributions. So I am not explaining the proof here. It is again following on the same line, that if I write the characteristic function of X.

I decompose in place of this i decompose into n terms here T1, T2, Tn and the corresponding decomposition I consider here, then if I apply the independent condition, then it will become the product of the characteristics functions of the X1, X2 and Xm at the corresponding terms.

**(Refer Slide Time: 41:35)**

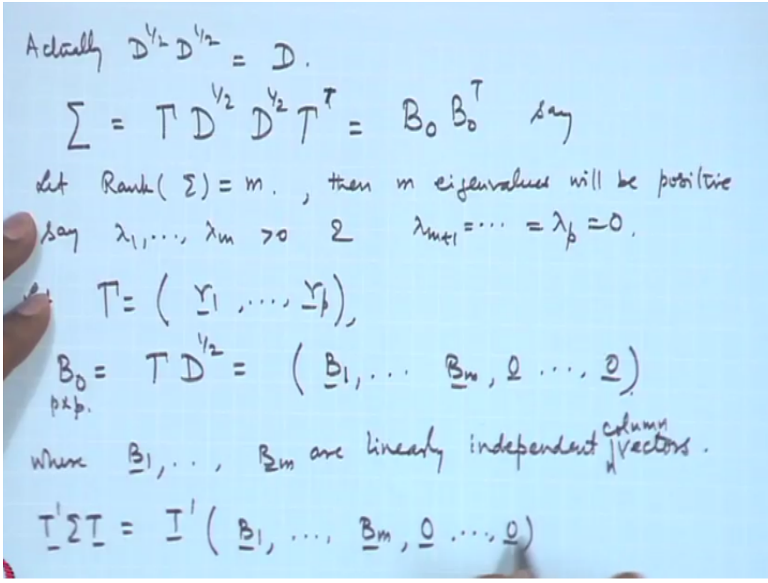



So now we can say that, now we prove the existence of a multivariate normal distribution, let me call it in a form of a theorem. There exists a random vector say X such that phi XT = e to the power i T prime mu – 1/2 T prime sigma T. Now this sigma is a real symmetric matrix, so we decompose it. So sigma = some gamma D gamma prime, where D is actually the diagonal matrix consisting of the Eigen values.

This contains Eigen values of sigma and gamma is an orthogonal matrix which will consist of Eigen vectors corresponding to lambda1, lambda2, lambda p. Now sigma is actually positive definite, the proof is quite simple actually, because sigma how I define sigma was let us consider say X prime or some a prime sigma a, then that will be = to a prime expectation of X - mu X - mu prime a.

So this I can write as expectation of a prime X - mu and this I can write as because X - mu prime, I can write as square here, so this is >/= 0, so this is a positive definite matrix. So lambda i is that is the Eigen values will be >/= 0. So D and square root D, let us define D 1/2, that is = square root of lambda1 and so on, square root of lambda p, this is well defined.

**(Refer Slide Time: 45:14)**

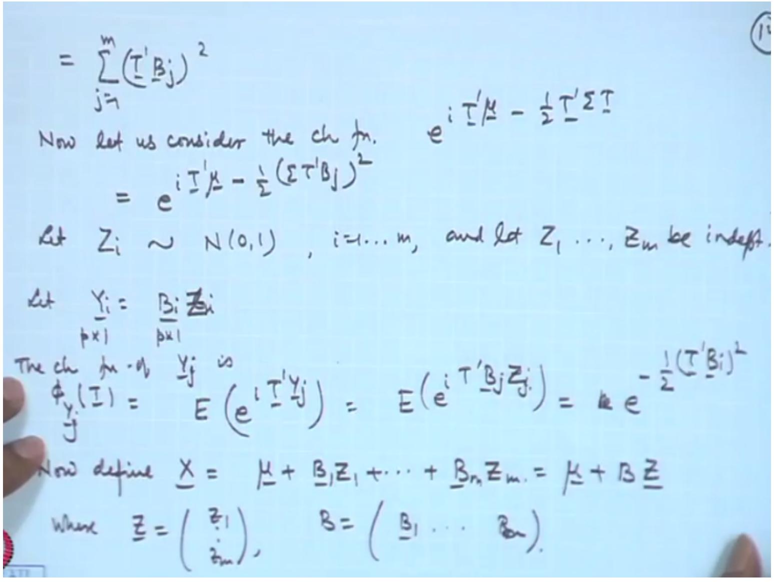



Actually this D to the power 1/2 into D to the power 1/2, this is satisfying that it is = D, that is why we can call it square root matrix. So we can consider sigma that is = gamma D to the power 1/2, D to the power 1/2 gamma transpose that we can write as B0 B0 transpose. Let us consider say rank of sigma to be = m, then m Eigen values will be positive, say lambda1, lambda2, lambda m are positive and lambda m+1 to lambda p they are 0.

So I am assuming here actually the first term are positive and the remaining are 0. So in general for sigma this may not be true, because some in between it may be 0 and so on. But we can always arrange them in a sequence in such a way, because I can interchange the order of the vector. You are saying X has a sigma thing here, so I can interchange the components of sigma and arrange in such a way that lambda1 and lambda2, lambda m will be positive and the remaining will be 0.

Let us consider this gamma to be gamma1, gamma2, gamma p, so B0 is actually gamma D to the power 1/2, that is becoming something like B1, B2, Bm and the zeros, this is actually your p/p matrix. So we can say that B1, B2, Bm they are linearly independent vectors, and they are actually column vectors. So if I consider now T prime sigma T, then that will be = T prime B1, B2, Bm and then null vectors here * B1 transpose and so on, Bm transpose zeros and then T, this is also.

**(Refer Slide Time: 48:17)**

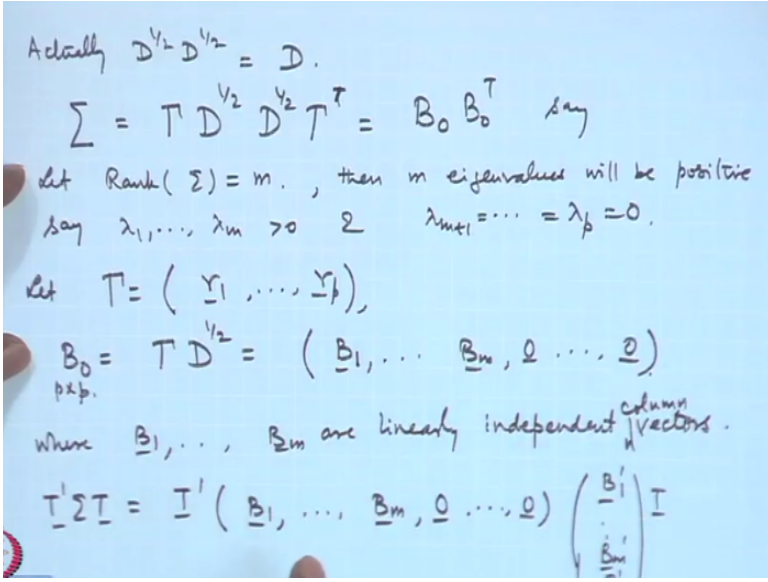



So we can consider it as simply sigma T prime Bj square, j = 1 to m. Now let us consider the characteristic function e to the power i T prime mu – 1/2 T prime sigma T. So that is = e to the power i T prime mu – 1/2 sigma T prime Bj square. So let us consider then Zi to be normal 0, 1 i = 1 to m and let Z1, Z2, Zm be independent. So we are considering a sequence of independent and identically distributed standard normal variables.

Let us further define say Yi = Bi Gi, so Yi is a p/1 vector. So characteristic function of Yi, that is = expectation of e to the power i T prime Yi, that is = expectation of e to the power T prime Bi, this is Zi, so let us change the notation here, like let us put it j, since Zj is normal 0, 1. So Bj, if you are putting this thing then this is becoming with the 0 here, and then you will get Bj Bj transpose there.

So it is becoming simply = expectation e to the power – 1/2 T prime Bi square. So now let us consider say, let us define X = mu + B1Z1 + B2Z2 + BmZm, we can write it as mu + B, this notation you can use for Z vector here. So here your Z is actually Z1, Z2, Zm and B is actually B1, B2, Bm.

**(Refer Slide Time: 51:47)**




Now let us consider the characteristic function of X, so that is = expectation e to the power i T prime X, that is = expectation of e to the power i T prime mu + i * T prime B1Z1 + so on BmZm. That is = e to the power i T prime mu expectation of e to the power i T prime B1Z1 + and so on BmZm, that is = e to the power i T prime mu – 1/2 sigma T prime Bj square. So what we have proved, we started with, let us summarise what we have done.

Suppose there is a vector mu and there is a real symmetric matrix sigma, which we are assuming positive semi-definite at least, we assume it to be positive semi-definite, then based on the decomposition of that I am defining, because if it is a positive semi-definite matrix I can decompose using a spectral decomposition where gamma is an orthogonal matrix and D is a diagonal matrix consisting of the Eigen values.

Then we put it in an order where we are considering actually the non-zeros first and then the 0, 1's and then the corresponding Eigen vectors are there, so I arrange them by multiplying this here. And this becomes 0 here, because we are multiplying by the 0 terms in the D 1/2 here. Now B1, B2, Bm are linearly independent column vectors, so if we consider T prime sigma T, then that is becoming T prime B1, B2, Bm then 0's.

And here we are getting the transposes of this into T here. Now using this, this is simply becoming the sum of the squares of scalars here, T prime Bj square. Now if I consider e to the power i T prime mu – 1/2 T prime sigma T we wanted to prove that there is a random variable for which this is the characteristic function. Now this characteristic function has a decomposition of this nature, e to the power i T prime mu – 1/2 summation T prime Bj.

Now from here what I consider, I realise that it is of the form 1/2 sigma square T square kind of thing, so I consider standard normal variable Zi, which are independent and I define Yi using this Bj, so I consider BiZi so this becomes multivariate here. I am multiplying by a T/1 vector into a standard normal variable here. So each of the components will become a different normal variable here but each of them will have mean 0, but variance will become Bi Bi transpose.

So if consider that thing, characteristic function of Yi will become e to the power i, because the mean term is 0 so it is simply becoming T prime Bi square. Now I define a vector X using these terms, so let us see mu was already given to me, B1, B2, Bm we found out, Z1, Z2, Zm are the standard normal variables. So using this I define a random vector X here, which is of course mu + BZ in the compact notation.

So if we use this then let us look at the characteristic function of X, then this is turning out to be simply, after using this decomposition I write it like this particular fashion and I use the characteristic function of the Yi here, these are Y1, Y2, Ym, so this is of this nature. So ultimately I get it as of the for e to the power i T prime mu – 1/2 T prime sigma T. So we are able to construct it random vector whose characteristic function is exactly the characteristic function of a multivariate normal distribution.

So basically it means that given a p dimensional random vector and a p/p positive semi-definite matrix we can define a random vector which will have a p dimensional multi normal distribution with that as the mean vector and that variance, covariance matrix. So this is a characterising property and the existence of the multivariate normal distribution. Here one thing is there we consider the rank to be = m, now m can be < p, m can be = p.

So if rank of sigma, that is = m is < p we call the multivariate normal distribution to be a singular distribution. So you can also associate it with the sigma matrix, sigma inversely exist if it is non singular if it is singular then sigma inverse will not exist. So later on when we consider the density then this point will be important.

Now in the next lecture I will consider some further properties of the multivariate normal distribution and we will look at the estimation of the parameters, etc., for the particular distribution.