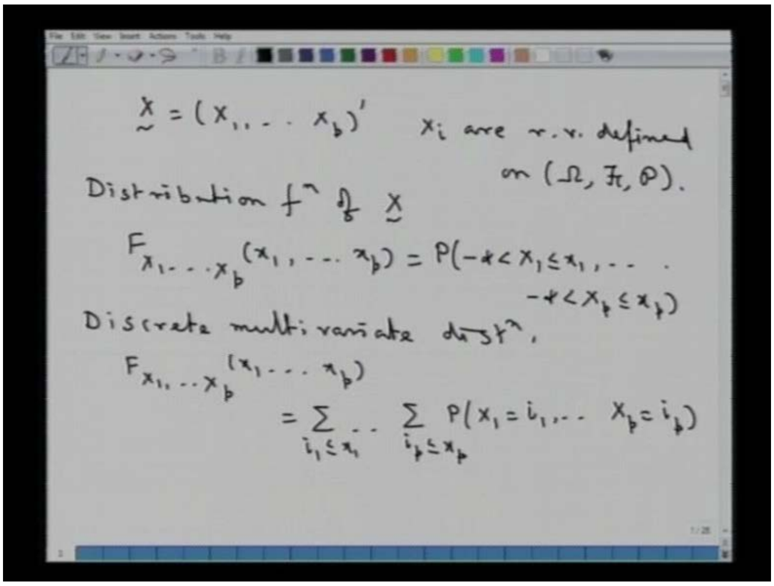
Applied Multivariate Analysis
Prof. Amit Mitra
Prof. Sharmishtha Mitra
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture No. # 01

Basic concepts on multivariate distribution – I

We will be starting with some basics concepts on multivariate analysis. So, let us define what we called by a multivariate random vector.

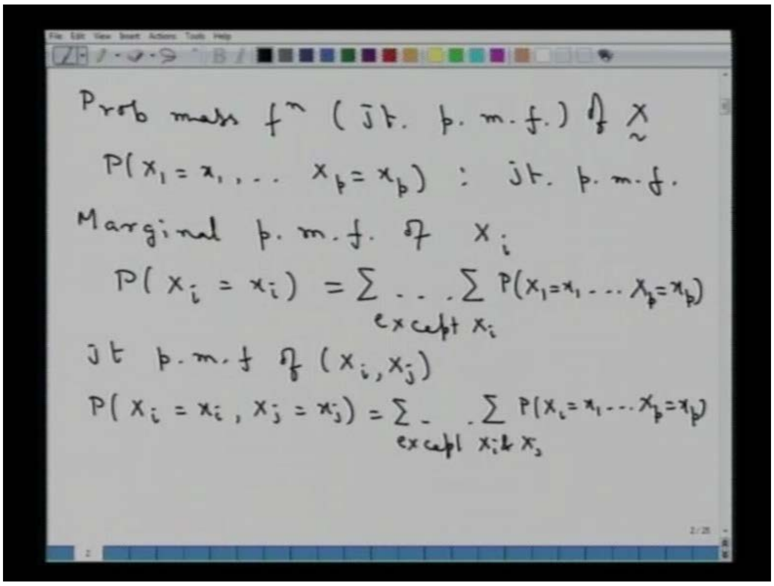
(Refer Slide Time: 00:27)



Let us denote by X; a set of p random variables x 1, x 2, x p. So, this is a random vector of random variables, where these X i's, actually X i's are random variables defined on a probability space omega, script f, p. Now, some basic concepts of multivariate analysis will be actually defining some simple concepts. We will make distinction between discrete multivariate random vector, and continuous multivariate random vectors. So, we define first, the distribution function. Distribution function of this multivariate random vector X is defined to be F X 1, X 2, X p at the points x 1, x 2, x p. This is defined as the following that, it is probability that minus infinity less than X 1 less than equal to x 1 and the last random variable minus infinity less X p less than equal to small x p. So, this is how, a distribution function of a multivariate random variable is defined.

Let us now make distinction between discrete multivariate random variable, and continuous multivariate random variable. Suppose we have discrete multivariate distribution; then this distribution function F X 1, X 2, X p at small x 1, x 2, x p points. This is defined to be summation (( )) p for rather. So, over all these points, where we have probability that X 1 equal to say i 1; X p equal to i p; where these now i 1 is less than equal to small x 1 and i p is less than equal to this x p point. So, this particular quantity for a discrete multivariate random variable is defined in this following way. Now, given the joint distribution function of this set of p random variables constituting this multivariate random vector, we can define marginal probability mass functions, joint mass functions.

(Refer Slide Time: 03:07)



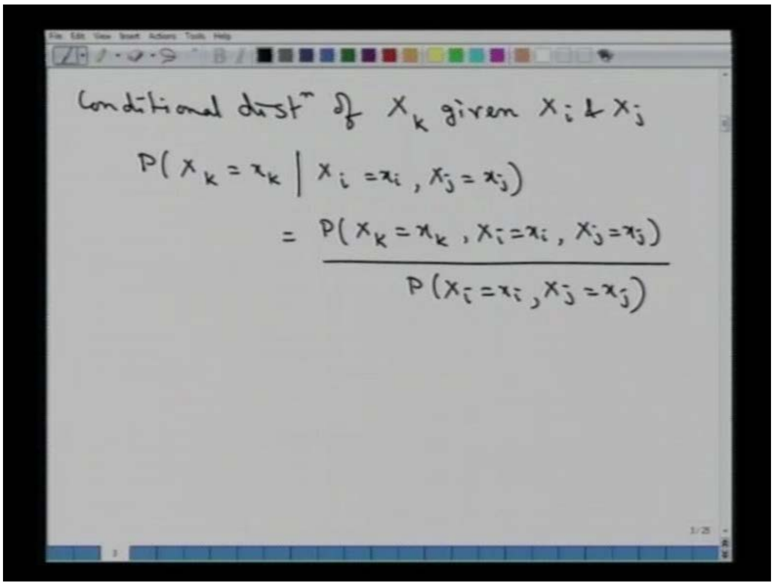
So, the probability mass function or rather the joint probability mass function of this multivariate random vector X is defined in the following way: that it is probability that X 1 equal to small x 1; X p equal to small x p. So, we call this particular quantity, which is the joint probability mass function. Now given this, the joint probability mass function of this random vector X, as we had defined. We can find out, what is called the marginal distributions. The marginal probability mass function of say any random variable in this particular set of p random variables x 1, x 2, x p say x i. This is defined to be probability that, X i equal to small x i.

So, this is given by p minus 1 fold summation, which is except this X i random variable. So, this is probability over the sum over the joint probability mass function. Like here we have, probability X 1 equal to x 1 extending up to… I am sorry this probability that, this we have the joint probability mass function right. Now, this is the marginal probability mass function of one random variable X i in the set of the multidimensional random vector, which is comprising of x 1, x 2, x p.

Now, in a similar way one can actually define the joint probability mass function of any set of variables, taken from this multidimensional random vector x 1, x 2, x p; say the joint probability mass function of two random variables x X i and X j taken from this set of this multidimensional random vector X i say X. The joint probability mass function of X i, X j is defined to be in a similar way to what we had defined for the marginal probability mass function of one random variable. So, this can be defined as probability X i equal to x i, X j equal to x j; this is sum.

Now, the sum is p minus one fold; it is over all X i's except X i and X j and then we have this joint probability mass function of all the random variables. When we have the joint probability mass function of any set of variables taken from the multidimensional random vector X comprising of those p components; one can define distributions, which are referred to as conditional distributions. say Suppose, we are interested in knowing what is the conditional distribution of any set of variables taken from this p dimensional random vector given another set of random variables.

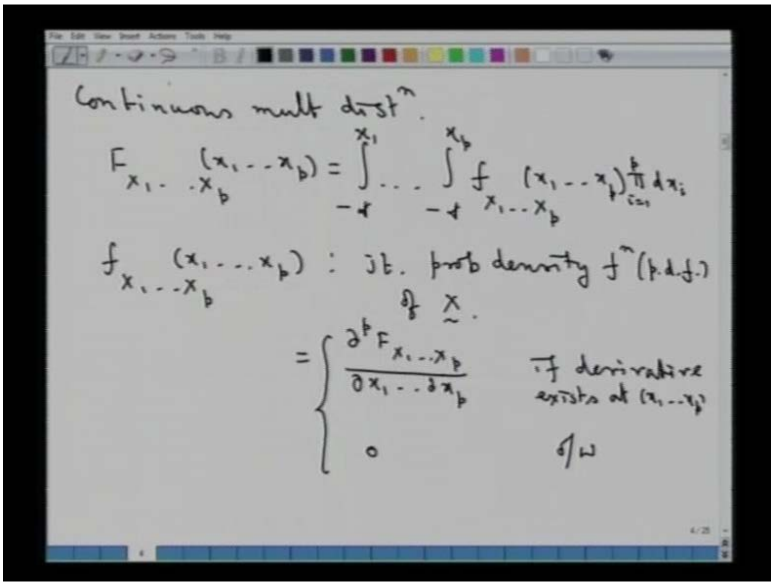
(Refer Slide Time: 06:37)



Say let us be simple and try to look at the following which we say that, say the conditional distribution of X, say any k given X i and X j. So, in order to find what is the conditional distribution of X k given X i and X j; we look at the following that probability that X k equal to x k; given that X i equal to small x i and X j, the random variable equal to small x j. So, this would be given by the following that this is the joint probability mass function of these three random variables X i, X j and X k; this divided by the marginal probability mass function of the two random variables X i and X j.

So, here what we have basically, in order to find out the conditional distribution of one random variable given two random variables X i; any two random variables X i and X j. Of course, X i and X j are not included in this X k here. It can also be. But when we look at this, it is basically the this numerator in this conditional distribution of X k given X i and X j is the joint probability mass function of X i, X j and X k and the denominator is X i and X k is joint probability mass function. right. Let us now move on to the case of conditional rather the marginal distributions, the conditional distributions in case of continuous multivariate random variables.

(Refer Slide Time: 08:22)

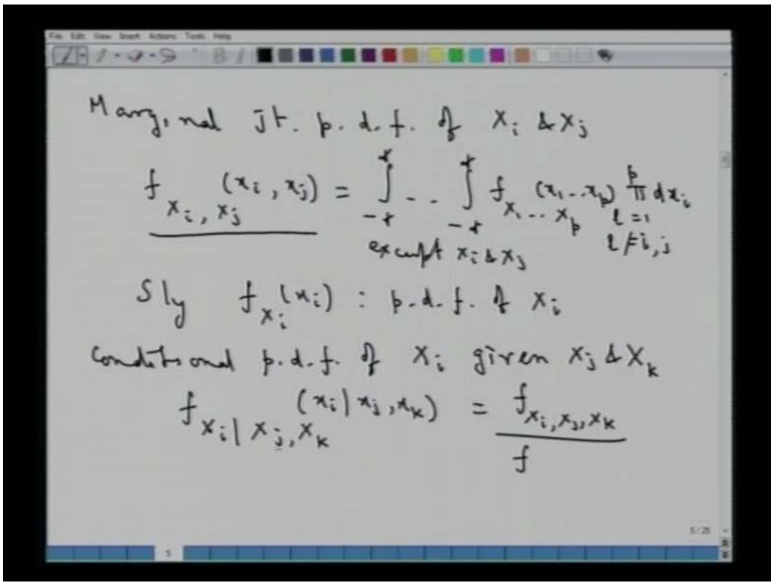


How it looks like? continuous multivariate distribution. So, we had first of all defined, what is the joint distribution function of any set of multivariate random components? So, in case of a continuous multivariate distribution comprising of these p elements, this for a continuous distribution would be defined through the integral. We assume that, the set of random variables form a set of absolutely continuous variables. So, we will have that been defined through the following functions, small f function which we are going to define shortly.

So, this this defined to be the following and then, this is product of x i's; i equal to 1 to p. Now, the function that we had introduced here is what is called the joint probability density function of these random variables. So, this quantity is what we referred to as the joint probability density function or in short, pdf of this random vector X. Now, this is given by; so, this quantity is given by the following. So, it is the p th partial differential, this with respect to the variables that we have; if the differentiation exists at the point x 1, x 2, x p if the derivative exists at x 1 x 2 x p and is equal to zero, if it is otherwise. Right.

So, for a continuous random vector X, this quantity is of interest and what we referred to that as, the joint probability density function. Now, given this joint probability density function in a way similar to what we had done for discrete random vector, we can define marginal probability density function of a set of random variables taken from that multivariate random vector. We can define conditional distributions exactly in the same way as, what we had done for a discrete distribution.

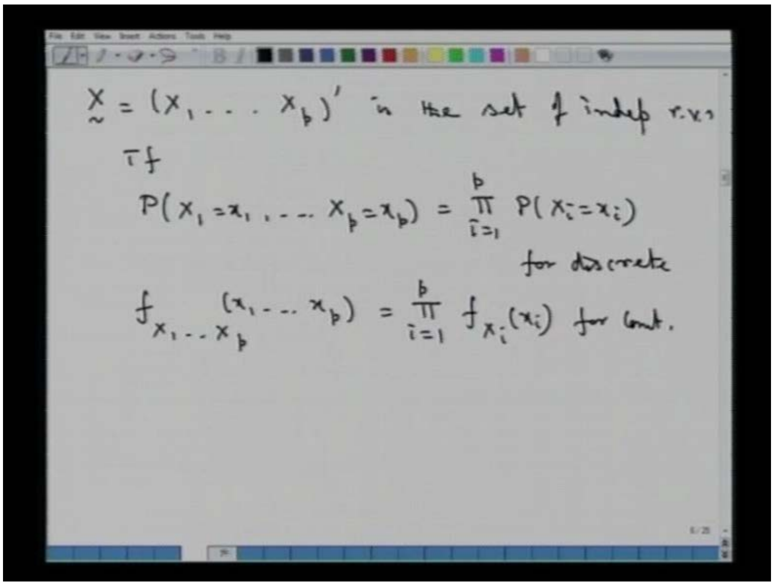
(Refer Slide Time: 11:08)



Say for example, we can look at the marginal say, joint pdf of X i and X j. So, there are two variables taken from this set of p random variables. So, this would be defined in the way that, this is f X i, X j at the point say x i and x j and we will have p minus one p minus two fold integral here over the entire range of those variables. These integrals are except the two variables X i and X j; leaving out these two variables, will have p minus two variables integrating over there range and then looking at this as the joint probability density function of x 1, x 2, x p. What will be getting here? This coefficient i equal to 1 to p then, i is not equal to… it is better to have a different notation here.

Say l equal to 1 to p with l not equal to i; this is l equal to one to p then l not equal to i it is not equal to j, dx i. So, we will have this particular quantity to give us the joint probability density function of X i and X j, two random variables taken from this set. Similarly, what we can look at is this. So, this would be a probability density function not the joint the probability density function of a particular random variable, X i taken from this particular set. So, we can also have a conditional distribution, conditional density function of say X i given X j and X k say; X i and X k, that would be given by the following that this is X i given X j and X k. So, this would be given by the joint probability density function of X i, X j, X k; this divided by the joint probability density function of the conditioning variables, that is X j and X k. right

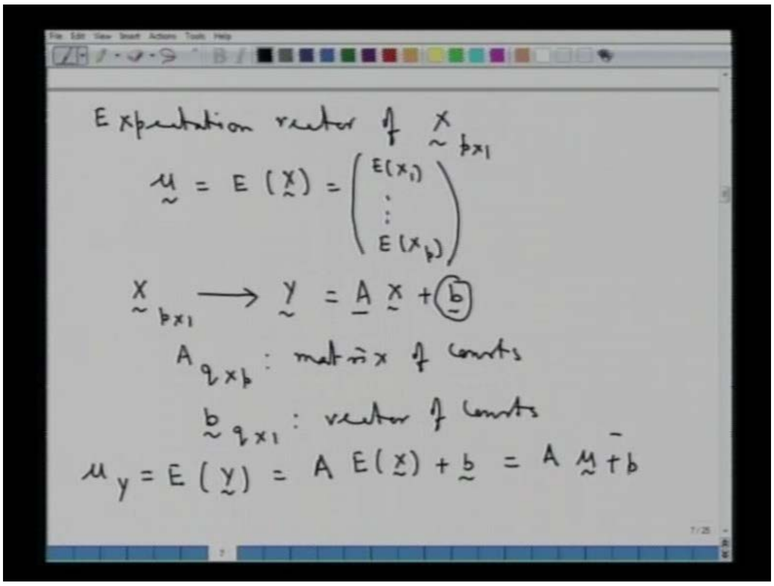
(Refer Slide Time: 13:52)



This random vector x, that we had defined comprising of these p elements X 1, X 2, X p is said to be a set of independent random variables; is the set of independent random variables independents defined by statistical independents. is the set of independent random variables. If we have the joint probability mass function or the joint probability density function given as the product of the respective marginal probability mass function or the probability joint probability mass density functions.

Say for example, in the case of discrete distributions, what we will be having for independent random variables is the following that, the joint probability mass function of X 1, X 2, X p would be given by the product of the respective probability mass functions given as the following. So, this is for the discrete distribution. for discrete distribution and If we have continuous distribution, then the joint probability density function of X 1, X 2, X p would be given by the product of the marginal probability density functions. This would be for the continuous random variable. Now, let us now move on to some concepts, which are going to be important for this particular course.

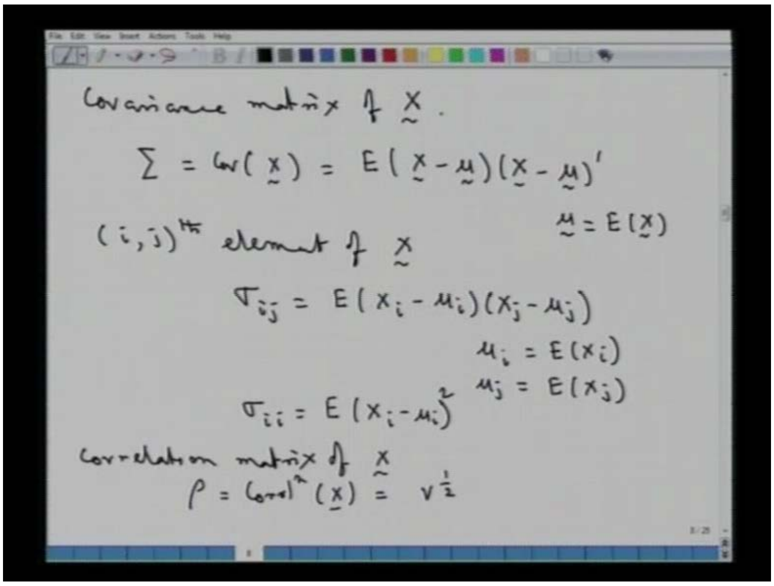
(Refer Slide Time: 15:33)



So, we define the expectation vector of this random vector X to be the following that let us denote that by the vector mu, which is also a p dimensional vector, which is expectation of this random vector. So, that is defined as the following. So, X component Y is expectation of the corresponding random variables X 1, X 2 and X p. Now, these expectations expectation x i in appearing in this particular vector; they basically are computed from the probability density function of the respective random variable X i, that is coming in this particular direction here. Now, given the information that X, the random vector which is a p dimensional random vector with an expectation vector as mu; if you make a transformation say suppose we make a transformation which is say p dimensional random vector to a random vector Y, which is given by say A X plus b.

Now, here this A matrix which may be is q by 1 which is matrix of constant. So, this is matrix of constants and b; so, this now becomes q by p and this is p by 1. So, this particular component out here is q by 1. So, b is say q by 1 vector. So, this is a vector of constants. So, we take here this A matrix and this b vector out here to be non-stochastic elements actually. And then, if we are now interested to find out, what is the expectation vector of this newly defined random vector, which is Y; say let us denote that by mu Y. So, this would be given by A matrix, which is a matrix of constants and that would then be given by expectation of X plus this b vector right. So, this would be given by A mu a mu plus this b. right Having defined this expectation vector, we move on to defining what we mean by covariance matrix of this random vector X?

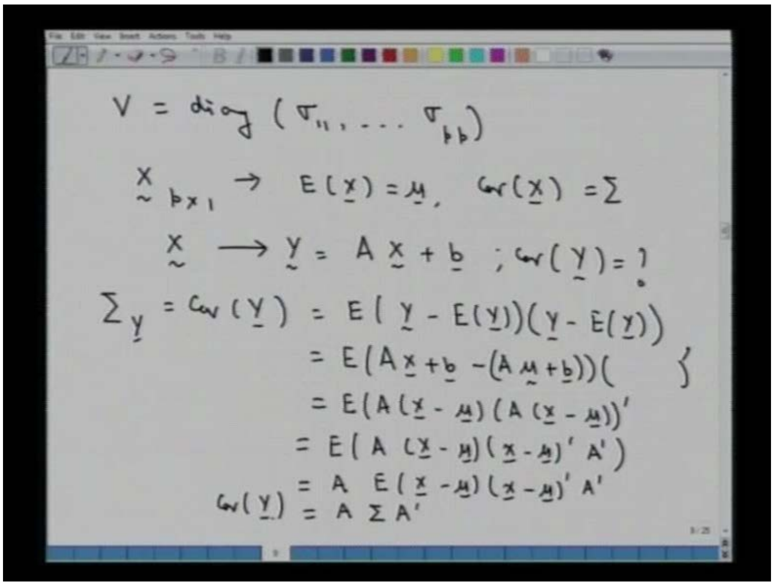
(Refer Slide Time: 18:06)



Let us define the covariance matrix of this random vector X. Now, let us denote that by sigma, which is covariance matrix of X; which is given by expectation of X minus expectation of X, which which we have already denoted by this mu vector that into X minus mu transpose. So, this mu is basically expectation vector of the random vector, which is X right. Now, the i j th element in this covariance matrix of X i j th element of this X is given by say sigma i j, which is simply the covariance between these two random variables X i and X j. So, let us denote by mu i to be expectation of the i th component of the mu vector. Then, this is X j minus mu j, where we have mu i is expectation of X i and similarly, mu j is expectation of the X j component right.

Now, this is the i j th element of this sigma matrix. So, the diagonal elements basically are giving us the variances of the respective components, which are p in number. So, this is just the variance of the i th component, which is expectation of X i minus mu i whole square right. Now, having defined this covariance matrix of this random vector X, we can define what is referred as the correlation matrix? Correlation matrix of X similarly can be defined. Say, this is a row matrix, which is holding the correlation components between the elements of this random vector. So, it is a correlation matrix of this.

(Refer Slide Time: 20:34)



Let us define that to be a matrix V half inverse times the sigma matrix into V half minus half. Let me also define, what we mean by this V matrix? So, this V matrix is a diagonal matrix holding the elements sigma 1 1, sigma 2 2 and the sigma p p. So, the V matrix is basically the matrix, which is having the diagonal entries only and comprising of the variances of the respective components. Now, we had seen what happens to when we have a random vector X? And then, we make a transformation from the random vector X to the random vector Y; which was from p dimension to a q dimension lower or higher dimension.

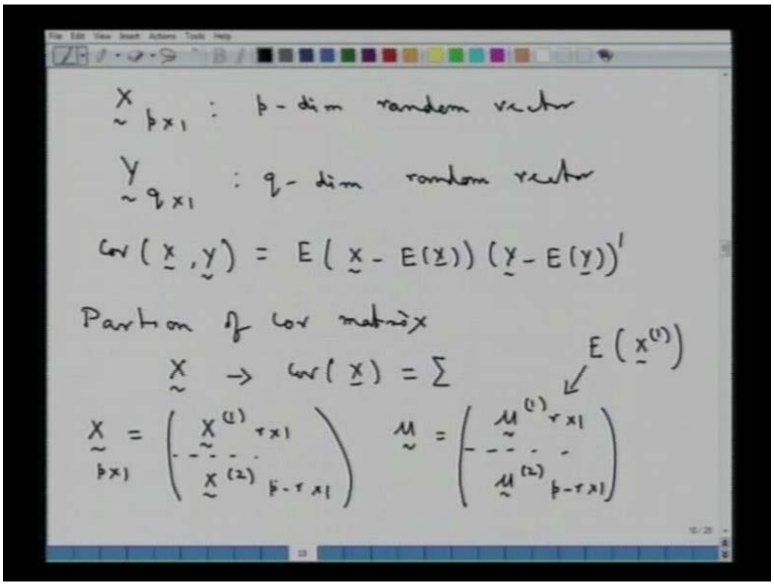
What we are, now what we can also look at is given the information that, X is a p dimensional random vector with expectation of X as a mu vector and the covariance matrix of X to be given by the sigma matrix. If we now make a similar transformation as to what we had done earlier; to a random vector Y, which is A X plus b. A and b are defined similarly as to what we had defined previously in this slide. So, this basically is defined exactly in the same way that, I am sorry. So, this is that A matrix. This A matrix is a matrix of constants; b is vector of constants.

So, given that information, we are now trying to look at, what is the covariance matrix of the new random vector that we have introduced, which is Y. Now, by definition that, the covariance matrix of Y would be given by… Let us denote that by sigma Y, for example. So, this sigma Y is expectation of Y minus expectation of Y into the transpose of this particular quantity. So, we have already computed, what is expectation of Y from this random vector A. So, this we just replace this by the component in terms of X. So, this would be expectation of A X plus b; this minus, now expectation Y as we had seen earlier is A times the mu vector; this plus this b vector.

So, we have this quantity here given by this b and then, the transpose of this particular quantity comes out here. So, what we will be having is, this b component cancelling out. So, we will have expectation of A X minus mu that multiplied by the transpose of this quantity. So, it is A X minus b. So, this would now be given by expectation of A into X minus mu X minus mu transpose and then we will have this A transpose. So, what we can see here is that this basically is the stochastic component here. This A matrix and the A transpose matrix basically are the two matrixes of constants.

And hence, we can take the expectation operator inside and what we will be having here is, A X minus mu to X minus mu transpose A transpose. And then this expectation of X minus mu into X minus mu transpose is nothing but the covariance matrix of X which is given earlier by sigma matrix. So, this is what we will be having as the covariance matrix of the newly defined random vector, which is Y. Now, suppose we have two random vectors X and Y; two different dimensions. We can also define; what is the covariance matrix between the two components here.

(Refer Slide Time: 24:29)

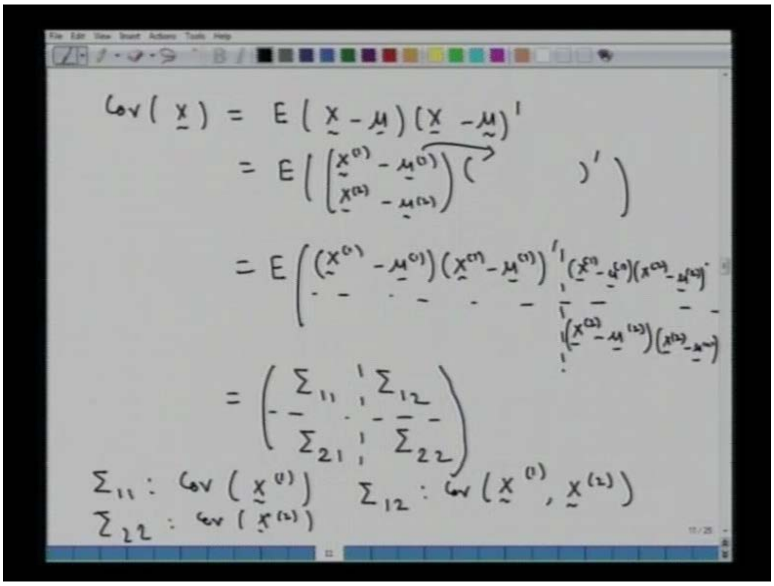


So, suppose we have this p dimensional random vector. So, suppose we have this as p dimensional random vector and we have Y, another q dimensional random vector. We can define, what is the covariance matrix between the elements of X and the elements of Y; similar to what we had defined by covariance of the components in that particular vector. This would be given now by expectation of X minus expectation of X vector; this multiplied by Y into expectation of this Y vector whole transpose. So, this is how, one defines the covariance matrix between two different sets of random vector.

It is important actually to look at the following, which we look at the partitions of the covariance matrix. What the partitions and what the elements actually can; what sort of interpretation actually can; the elements of those partitioned elements can be having actually? partition of covariance matrix. So, suppose we have this random vector X with the covariance matrix of X as sigma matrix. Let us make the following partition. Suppose, this X vector which was p dimensional vector; now is partitioned into two following sub vectors say X 1 and X 2.

Say suppose this is r by 1 and this is p minus r by 1; similar to this X vectors partition, we can look at the corresponding partition in the mu vector, which is the expectation vector. So, we can write the similar partition as mu 1 vector, which is r by 1 vector out here and then, we will have the second sub vector as p minus r dimensional. So, this quantity here is expectation of the corresponding sub vector, what we had as X 1 written out there. So, this would be expectation of X 1 and similarly, this would just be expectation of this particular component X 2 from here.

(Refer Slide Time: 27:12)



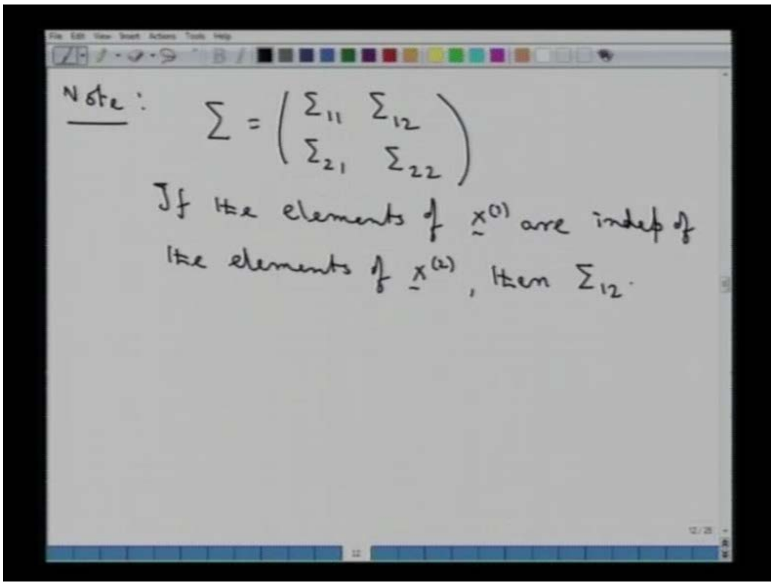
Now, if we look at the covariance, the partition that we have corresponding to the partitioning of the X vector that is what we had written. So, this by definition of the covariance matrix is given by the following that it is this particular quantity. Then we look at what happens to this particular element, when we look at this as the sub vectors. So, this would be X 1 minus mu 1. So, this vector would be the second component X 2 minus mu 2 and then, the transpose of this comes out here. Now, if one multiply the two and then take the expectation inside or (( )) to that one can just look at, what would be this particular matrix?

This would be given by this X 1 minus mu 1 X 1 minus mu 1 transpose and then this block here. So, this would comprise of four blocks. This would be the second component X 2 minus mu 2; then this is X 2 minus mu 2 transpose. And then this element would be X 1 minus mu 1 into X 2 minus mu 2 transpose and this would just be the transpose of this particular element there. So, this element once we take the expectation operator inside would be this would be a p by p matrix, which we denote by sigma 1 1 say; this would be a p by q matrix, which we denote by sigma 1 2. And then the expectation of the quantity which is here would be a q by p, which is denoted by say sigma 2 1.

This element, the last block here will have this partitioning actually carried forward. So, when we take the expectation of this block here, that is denoted by sigma 1 1. When we take expectation of this particular block here, that is going to be denoted by sigma 2 2 block, which is a q by q dimensional matrix there. When we take the expectation of this block here, it is basically denoted by sigma 1 2. Now, this sigma 1 1 element; if you look at carefully, this sigma 1 1 block here is expectation of this particular matrix here and what is that? That that is, basically is the covariance matrix of the sub vector, what we had denoted by X 1.

Similarly, if we are looking at sigma 2 2; sigma 2 2 is nothing but the covariance matrix of the second block actually that, this particular X 2 block of elements and this sigma 1 2 matrix, which is the off diagonal blocks in this sigma matrix. So, this is the covariance matrix of the two sub vectors X 1 and X 2 that is what we had defined. So, this is the covariance matrix of the two blocks X 1 and X 2 right. This is just the transpose of this particular matrix and that one can say that, it is the covariance matrix of X 2 and X 1. So, it just differing by that particular transpose.

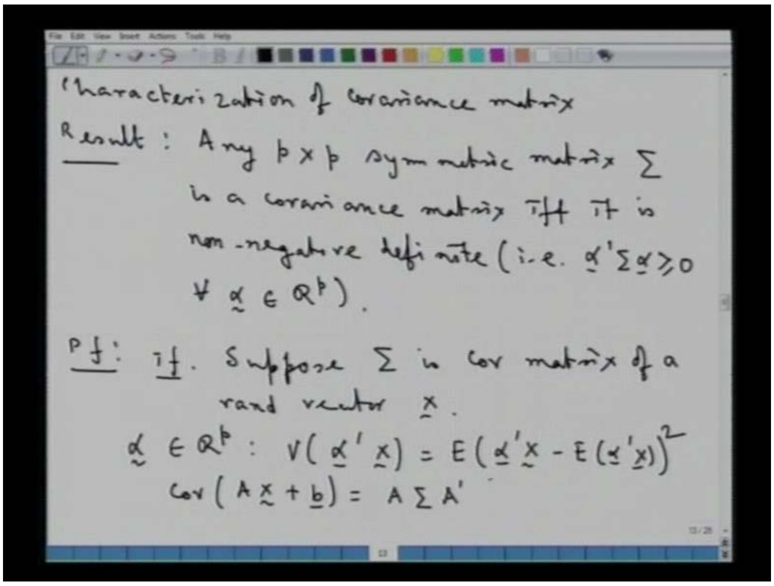
(Refer Slide Time: 31:00)



Now, we make a small note here; that, we had this sigma matrix in the partition form written as sigma 1 2 sigma 1 1 sigma 1 2 sigma 2 1 sigma 2 2. Now, if we have X 1 and X 2, independent set of random variables then, what will be having is this sigma 1 2 will be given by a null matrix. Now, if the elements of X 1 are independent independently distributed of the elements of the X 2 sub vector; then we will be having this off diagonal block here sigma 1 2 to be given by a null matrix. However, the converse of this particular result is not true.

What I mean by saying that is that, if we have sigma 1 2 to be given by null matrix then, it does not imply that, the elements of X 1 and X 2 are independent. It does not imply. It would only be implied, if we have some other condition like, what we had for bivariate distribution that; if we have the joint distribution of X 1 and X 2 to be given by a multivariate normal. Then only, we will have sigma 1 2 to be given by a null matrix; that would imply in such a situation that, the element of X 1 and X 2 are independent right. Now, let us move on to one characterization of covariance matrix, what we are defining. So, let us now look at one characterization.

(Refer Slide Time: 32:57)



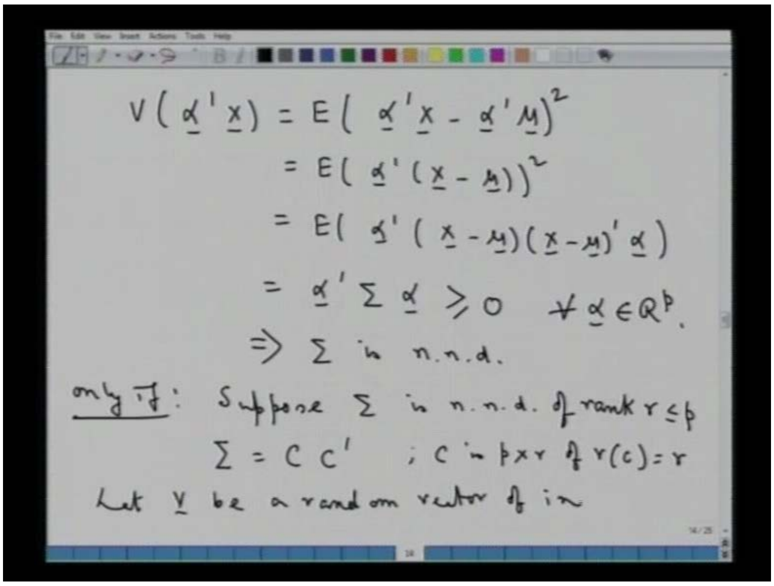
How we can characterize a covariance matrix? So, characterization of a covariance matrix, we have the following result. Let me first state the result that if we have any matrix which is positive semi definite, it can be associated with covariance matrix of some random variable. So, if we have a p by p symmetric matrix sigma, will be a covariance matrix if and only if, it is positive semi definite or non-negative definite. So, let me write the result first that, any p by p symmetric matrix sigma is a covariance matrix if and only if, it is non-negative definite. That is, alpha prime sigma alpha is greater than or equal to 0 for every alpha belonging to R p right.

So, that gives us a characterization of a covariance matrix. So, let us look at the proof of this particular simple fundamental result. So, what it says is that, if we have sigma to be the covariance matrix of a of some random vector X, then it is going to be positive semi definite, non-negative definite. Other way round, if we have any matrix which is symmetric and it is non-negative definite, then we can associate some random vector to that particular matrix. For which, the random vector is going to have that matrix as its covariance matrix. So, let us look at the proof of this result. Let us first proof a look at the proof of the if part.

Now, suppose sigma is is covariance matrix covariance matrix of some random variable x covariance matrix of a random vector X; then we take a alpha belonging to R to the power p. Then, what we can say is that the following that variance of alpha prime X. Now, note that this alpha prime X is a scalar random variable. So, it is a linear combination of the elements of that p dimensional random vector. By definition, this actually would be given by expectation of this random variable alpha prime X minus expectation of this alpha prime X whole square right.

Now, we actually had proved a result that variance of or rather the covariance matrix we had proved the result; so that, we can even use that particular result. We had proved the result that, covariance matrix of A X plus b; this matrix is given by A covariance matrix of X, which is denoted by sigma; A sigma A prime. So, we can use this particular result, in order to find what is the variance of X? Otherwise from definition, one can also find out what is the what is the expectation of this particular quantity? Let us look at that.

(Refer Slide Time: 36:51)



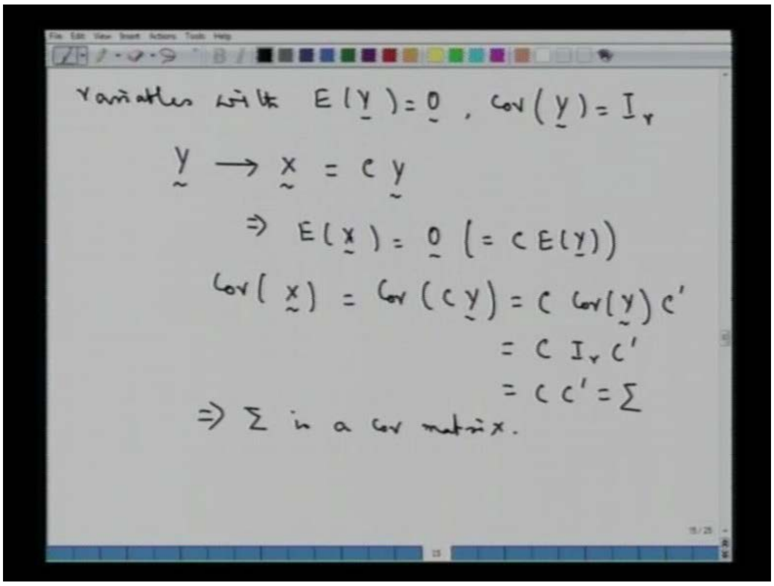
So, this is variance of alpha prime X; this is given by expectation of this alpha prime X minus alpha prime mu. mu is the expectation vector of that covariance that particular element, which is alpha prime X this whole square. This would be given by expectation;

we take alpha prime out. So, this is X minus mu whole square. So, this would be given by expectation of alpha prime X minus mu. Then this is the transpose of that; that multiplied this alpha vector. So, this would just be given by alpha prime; then the covariance matrix of X, which is the sigma alpha. This is the variance of X.

Now, variance of alpha prime X is going to be given by this alpha is belonging to R to the power p. So, since this is the variance of this random variable alpha prime X, which is going to be greater than or equal to 0. So, we have proved that, if sigma is actually the covariance matrix of any random vector X, then for every alpha belonging to R to the power p alpha prime sigma alpha is going to be greater than or equal to 0. So, that it is non-negative. It is going to be non-negative definite for every alpha belonging to R to the power p. So, this basically implies that, sigma matrix is non-negative definite. So, let me just write that as, n n d. So, that is non-negative definite.

So, the other way round, the only if part we look at, suppose sigma is non-negative definite we will prove that, sigma is a covariance matrix associated with some random vector x. Now, since we have assumed that sigma is non-negative definite, say suppose it is a non-negative definite of rank r which is less than or equal to the dimension of this random of this matrix sigma. Since sigma is non-negative definite of rank r, we can write this sigma matrix in the following form that, it is c c prime; where c is p by r of full column rank right. Now, since we have factorized sigma into this c c prime form, let us now define a random vector Y.

(Refer Slide Time: 39:59)



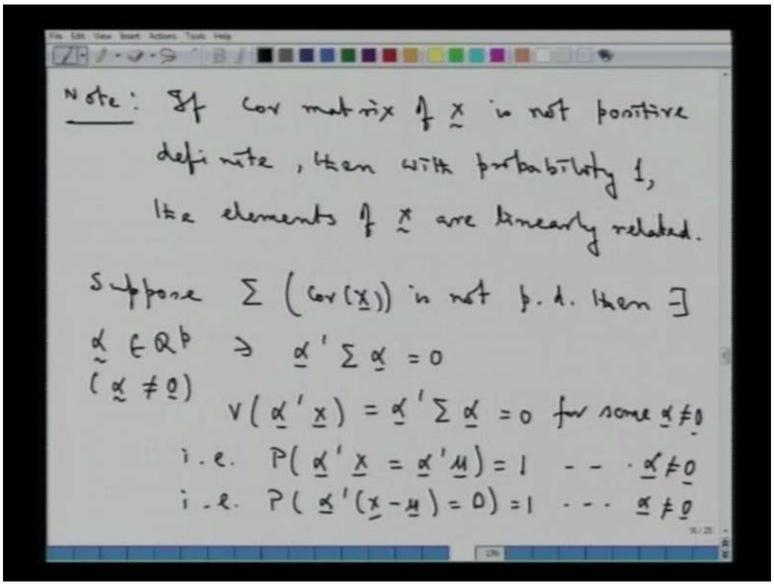
Let Y be a random vector of independent random variables of independent random variables with say expectation of Y given by a null vector. Now, since we assume that, the components Y 1, Y 2, Y r; what we have for the random vector Y? They are independent. Then, the covariance matrix actually would be given by a diagonal matrix. Using the result that we had stated earlier that, if we have the components of X 1, that sub vector to be independent of the components of the X 2 sub vector. Then, the block which is holding the covariance's of the components of X 1 and X 2 are going to be 0. So, this would imply that, the covariance matrixes diagonal without loss of generality, we take that to be an identity matrix.

So, the covariance matrix of Y is given by say this identity matrix of dimension r right. Now, Y is that set of independent random variables with the expectation of Y vector to be a null vector and the covariance matrix of this Y vector to be an identity matrix of dimension r. Now, let us make a transformation from Y to X, where X is given by c times this Y. Now, the matrix c what we had defined earlier, which is a p by r matrix; which is a matrix of constants of full column rank. So, that is of rank r. So, with that c matrix, if we now look at what is this X matrix characteristic?

As far as its expectation vector and its covariance matrix are concerned, we will have; this implies that, expectation of this X vector would be given by a null vector; because this is equal to C times expectation of this y vector and expectation of a Y vector being a null vector. This would imply that, it is just a null vector. And then, the covariance matrix of this X vector is covariance matrix of this C times Y vector. That would be given by C times covariance matrix of Y into this C transpose matrix. Covariance matrix of Y, what we have assumed is an identity matrix. It is a the covariance matrix associated with i i d random variables actually with variance is equal to 1.

So, this would be given by C times this I r matrix C transpose. So, this is equal to C C prime and which is nothing but that sigma matrix. This would imply that, this sigma matrix what we had just taken as a non-negative definite matrix. It is now the covariance matrix of this random vector X right. So, this would actually prove that, sigma is a covariance matrix. So, we have proved the result that any p by p symmetric matrix which is non-negative definite is if a p by p symmetric matrix is non-negative definite is a covariance matrix if and only if, it is non-negative definite.

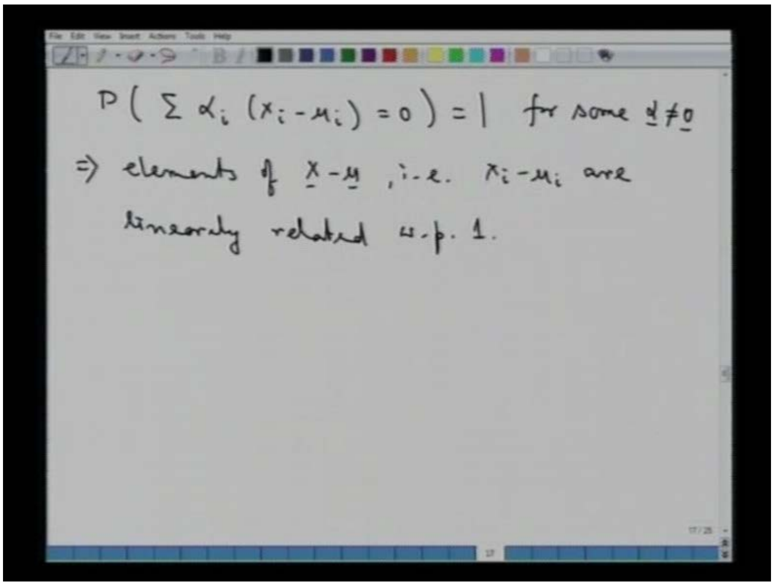
(Refer Slide Time: 43:28)



We just put it as a note the following result that, if covariance matrix of X is not positive definite, then with probability 1, the elements of X are linearly related with probability 1 the elements of x are linearly related. What do we mean by this is the following that, suppose sigma, the covariance matrix of this X random vector is not positive definite, then there exists an alpha belonging to R to the power p such that, we will have alpha prime. Now, these alpha vectors are belonging to R to the power p ofcourse, alpha is not equal to 0. So, this goes without saying that, this alpha is not equal to 0. Then, this alpha prime sigma alpha that is equal to 0. So, there exists some alpha such that, this alpha prime sigma alpha is equal to 0. Now, what does this mean?

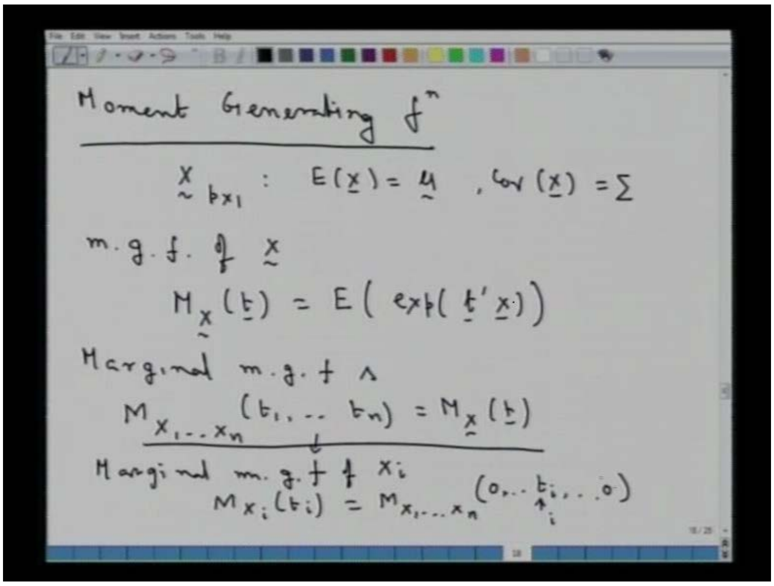
This means basically that, what we had seen earlier is that, this alpha prime sigma alpha is nothing but variance of this alpha prime X. So, we have this variance of alpha prime X given by alpha prime sigma alpha; this is equal to 0 for some alpha which is not equal to a null vector. What it means is the following that, probability that alpha prime X is equal to its expectation alpha prime mu; this is equal to 1 for some alpha which is not equal to 0 right. That is what we have is alpha prime X minus mu; this is equal to 0; this probability is 1 for some alpha, which is not equal to 0 right. So, this basically would imply from the statement that we have made here that, probability that this particular linear combination is equal to 0; this is equal to 1 for alpha not equal to 0.

(Refer Slide Time: 46:25)



So, this would imply that, probability that summation alpha i; the i th component X i minus mu i; this is equal to 0 is equal to 1 for some alpha, which is not equal to a null vector. So, this would imply that, the elements of X minus mu; that is, these X i minus mu i's are linearly related with probability 1 right. So, we will have the elements of the X i components; because these mu i components, they are basically constant components. So, we will have the elements X i is actually being linearly related, if we have the matrix sigma, the covariance matrix to be a matrix such that, it is not positive definite right. Before we move on to random sampling of from a multivariate distribution, we look at to concept, which is moment generating function. From where, we can get actually joint moments of the components, which are there in that multidimensional random vector.

(Refer Slide Time: 48:00)



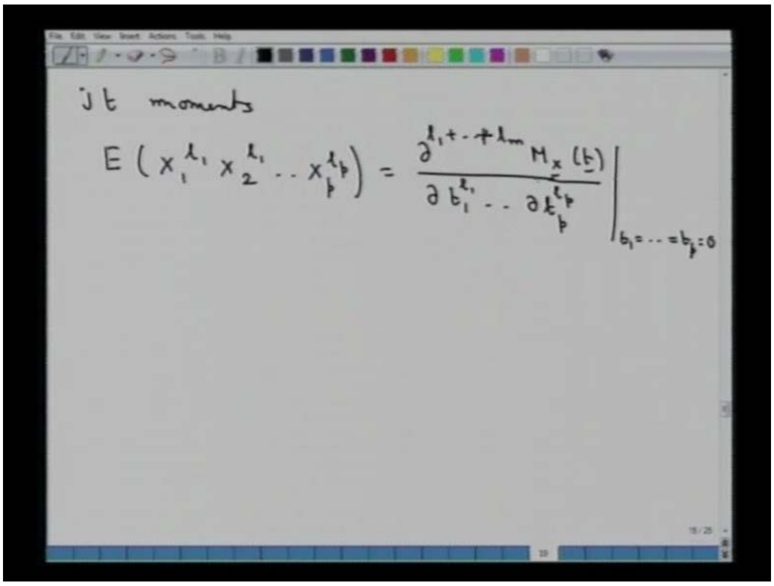
So, let us look at that moment generating function. So, the concept of this moment generating function of the random vector X is, basically trying to generalize what we actually have for univariate random variable. So, we have say this X, a p dimensional random vector with expectation of X given by this mu vector. Covariance matrix of X is being given by the sigma matrix. Then the moment generating function mgf of this random vector X is defined to be the following that, it is expectation of E to the power t prime X right.

So, this is how, one defines the moment generating function of the random vector, the p dimensional random vector X. Now, given the information; now, this provided the expectation exists, now the moment generating function of this random vector X actually would also lead us to the marginal moment generating function of the components of X in the following way. Now, if you look at, how to get the marginal moment generating functions marginal moment generating functions. Say, we are looking at this the moment generating function X 1, X 2, X n at these points x 1, x 2, x n; that is what is our M x(t), the moment generating function.

So from here, if we want to find the marginal mgf of X i, that would be derived from this; which is the joint moment generating function of the components of X in the following way. That, the marginal moment generating function of X i at the point t i; that would be given from the joint moment generating function of X 1, X 2, X n with 0 at all other points, except the i th position which is say t i. So, this is the i th point. So, this is what is going to give us. The marginal moment generating function of the component, which is X i.

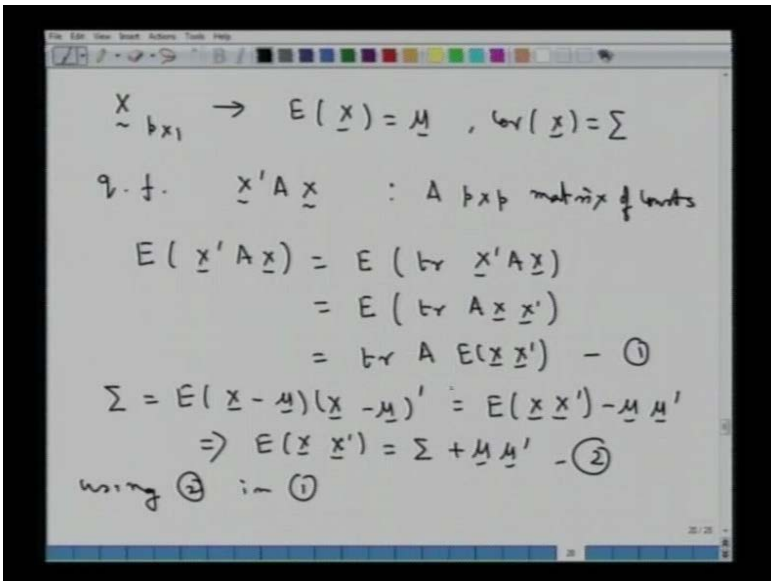
To see how that is true, it is trivial to look at the expression of this particular joint moment generating function of the random vector X. So, here if we take all the t i's except one t i component here as 0, then this particular term here; which is this term is the summation t i X i. So, here if all the t i's except one particular t i is 0. Then this sum is nothing but just t i X i. So, expectation of e to the power t i X i is just the moment generating function of this particular component which is X i. Now, given given this particular moment generating function, one can actually look at the joints moments.

(Refer Slide Time: 51:31)



Joint moments, actually of any order provided the moment the corresponding moments exists can be obtained from the following. Say, suppose we are looking at this particular joint moment, we can obtain the this joint moment for this set of random variables X 1, X 2, X p or any set of random variables derived from it. Using the moment generating function in the following way that, it is the partial derivative of the joint moment generating function; this, the derivative with respect to all these t i's. This evaluated at t 1 equal to t 2 equal to t p; this equal to 0. So, this is how, from the joint moment generating function M x(t), one can get to the joint moments of any order provided that order moment exists; one can get to the joint moments derived from that right.

(Refer Slide Time: 53:14)

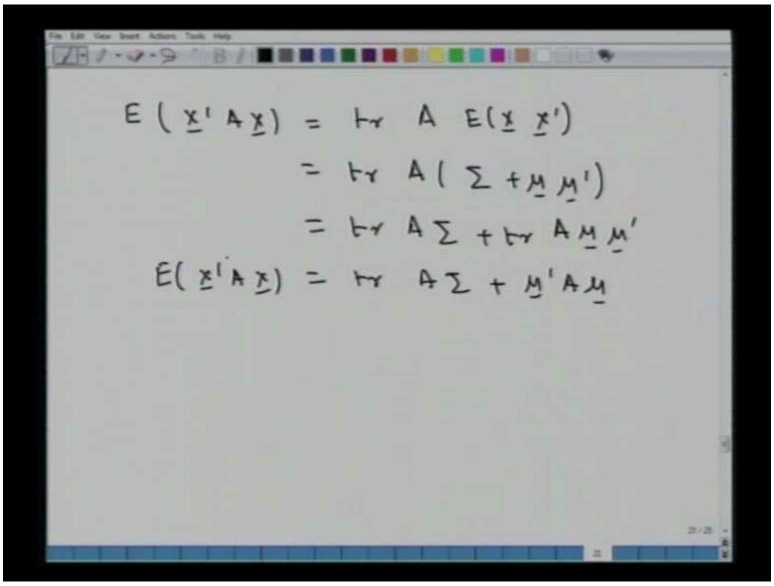


Now, a result which is particularly useful in some applications; as we move, when this particular multivariate, the normal course; we will see that, we are very frequently concerned with the a quadratic form of the following nature that, suppose we have this X; this multivariate vector with expectation of X to be this mu vector and the covariance matrix of X to be this sigma matrix. Then, if we define the quadratic form as, X prime A X; this is a quadratic form in X. So, where this A has ofcourse, the interpretation; that it is a matrix of constants. So, it is a non-stochastic matrix.

So, this is say A p by p matrix of constants, then expectation of X prime A X is the expectation of this particular quadratic form. So, this actually can be obtained in the following way that, now note that this is scalar quantity. This, we can write as expectation of the trace of this X transpose A X. We can take, we can use the result that trace of A b equal to trace of b A. So, what we can write is the following that, this is trace of A X X transpose. We can now take the expectation operator inside the trace. So, this would be the trace of A expectation of X X transpose. Now, let me give this equation number 1.

Now, in order to find out what is the expectation of X X transpose? We see that, this sigma is expectation of X minus mu into X minus mu transpose. If you look at, what the expectation of this quantity is. So, it would turn out that, this is expectation of X X transpose minus mu mu transpose; just open it up, then take expectation inside. So, this would tell us that, expectation of this X X transpose is nothing but sigma plus mu mu transpose. So, if you use this equation number 2 in equation 1. So, using this equation number 2 in 1, what we get is the following that what we had there was trace of this particular quantity.

(Refer Slide Time: 55:39)



So, we have expectation of X transpose A X, that was given by trace of A expectation of X X transpose. Now, we have obtained that in the form of this two here; that expectation of X X transpose is sigma plus mu mu transpose. So, this would be now be given by sigma plus mu mu transpose. So, this expression now would just be given by trace of A sigma; this plus trace of A mu mu transpose. And then, that can be written actually as the following which is mu transpose A mu; using once again the result that trace of A b equal to trace of b A. So, what we have is the following result that if we have that multi-dimensional random vector X with a mean vector equal to mu and a covariance matrix equal to sigma, then expectation of this quadratic form is just given by this trace of A sigma plus mu transpose A mu; similar result for the variance of this quadratic form can also be derived.