Applied Multivariate Analysis

Prof. Amit Mitra
Prof. Sharmishtha Mitra
Department of Mathematics and Statistics

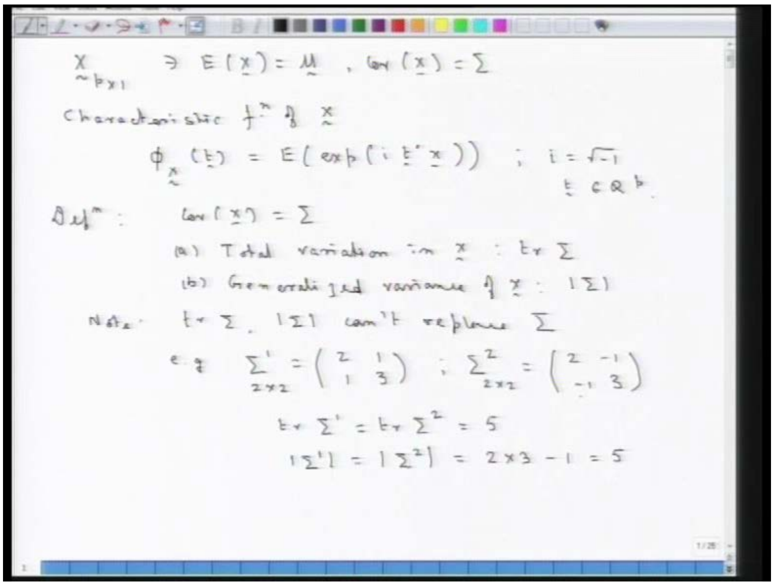
Indian Institute of Technology, Kanpur

Lecture No. # 02

Basic concepts on Multivariate Distribution – II

Let us start the day with some definitions. Suppose we have x, a p dimensional random vector such that, we have expectation of x as the mean vector mu and the covariance matrix of x to be denoted by sigma. We define the characteristic function.

(Refer Slide Time: 00:20)



Characteristic function of x is defined as phi x t to be expectation of e to the power i t prime x, where we have i to be square root of minus 1 and t, a vector belonging to R to the power p right. Now, this is the joint characteristic function of the components of x. So, it is the joint characteristic function for the random vector x. Now given the information about this particular joint characteristic function of the elements of x, one can actually get to the marginal characteristic functions of the respective components, that make up this particular random vector x; similar to what we had seen in the last lecture for the movement generating function.

Now, this characteristic function ofcourse completely determines a multivariate distribution. And the knowledge of this characteristic function about the multivariate random vector completely determines also the characteristic function and hence, the distribution of the respective components of x. Sometimes a quantity which is used to condense the information about, which is present in this covariance matrix sigma; other two following quantities just put it as definitions. So, suppose the covariance matrix of x is sigma; it is a p by p matrix. So, there are unknown elements which are present in this p by p matrix; it is a symmetric matrix.

So, the total number of unknown quantities in this particular matrix is p in to p plus 1 by 2. Now, in order to summarize this the information, that is present in the covariance matrix. The two following quantities are defined. The first is the total variance or total variation in x. So, the total variation in x is given by trace of the sigma matrix. So, it condenses the information that we get through the covariance matrix in terms of single quantity, which is given through the trace of this particular matrix. The second quantity which is of interest, what is referred to as a generalized variance?

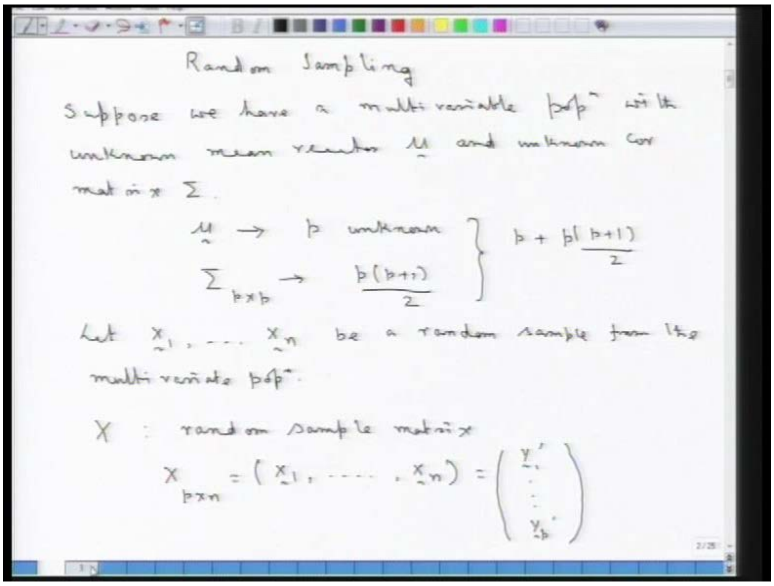
Generalized variance of this random vector x and that is given by determinant of sigma. So, once again it summarizes or rather condenses the information, that is present in this covariance matrix sigma in terms of two single quantities either trace of sigma or determinant of sigma. The first one is called the total variation in x. As we will see later on, then this total variation in x is the quantity of interest and which actually is looked upon as preserving type of approach, when we look at principle components analysis. Now, it should be noted that, the two quantities that we have defined just now.

The trace of sigma, which is the total variation in x and the generalized variance of x are not sufficient to replace this p by p matrix sigma. So, just put it as a note that, this trace of sigma and determinant of sigma cannot actually replace this sigma. It is obvious actually, if you look at this determinant and then, the corresponding condensation in terms of trace of sigma and determinant of sigma. Say for example, if you have a sigma matrix, suppose sigma is a 2 by 2 matrix which has elements say 2 3 on the diagonal; 1 1 on the off diagonal.

Suppose this is the first sigma matrix covariance matrix; we can have another sigma matrix, which is of 2 by 2 matrix; which has entries as 2 3, the same entries on the diagonal; minus 1 and plus 1 on the two off diagonal. So, these are two different covariance matrices. However, if we look at the total variation that is, it is given through trace of sigma. Trace of sigma 1 will be equal to trace of sigma 2 and that is equal to the some of the two diagonal elements that is 5. We can look at the determinant of the two covariance matrixes sigma 1 and sigma 2. So, these two quantities would be 2 into 3 minus 1. So, that is equal to 5 as well right.

So, the total variation that is there in this sigma 1 and sigma 2 are both equal to 5. The generalized variance of the underlined random vector x, which is given through this sigma 1 and sigma 2; both of them are also equal to 5. So, they have both of these two covariance matrixes have the same measures as far as total variation and the generalized variance of the associated random vector is concerned. However, the two covariance matrixes themselves are quite or clearly different right. So, these two quantities are not enough actually to replace the covariance matrix as such. However, the condensation that one gets is of interest at times.

(Refer Slide Time: 06:23)



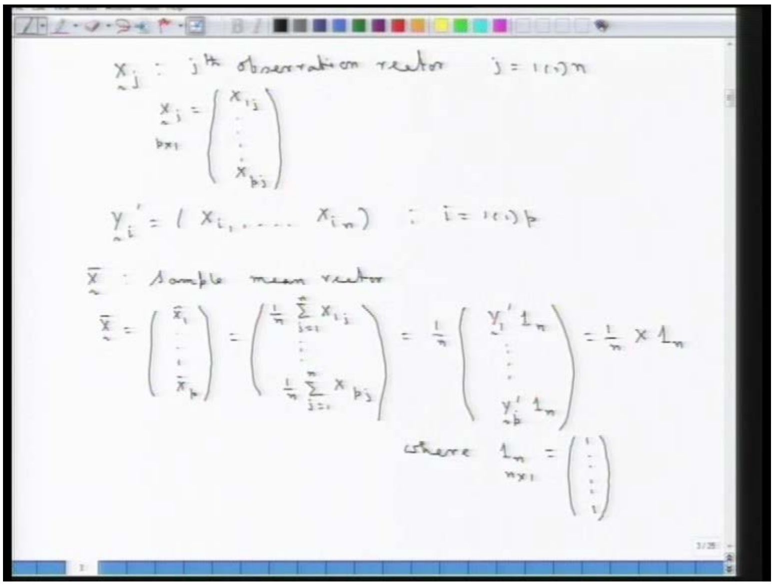
Now, let us move in to an important concept, which is random sampling from multivariate distributions or rather random sampling from multivariate distributions. Now, random sampling from multivariate distribution; what is what is basically the need of this random sampling? Suppose we have a multivariate multivariable population Suppose we have a multivariable population with unknown mean vector as mu, which is p by 1 and unknown covariance matrix as sigma. Since for all practical purposes, when we considering a multivariate population these two quantities; mu, the mean vector and the covariance matrix sigma are unknown.

So, the numbers of unknown quantities, the parameters which are actually present in that particular population are the following. So, this mean vector will lead us to p unknown quantities, because it is a p dimensional vector. So, p unknown quantities and when we look at the sigma matrix, which is p by p matrix. This is a symmetric matrix; it is a covariance matrix. Hence, it is a symmetric matrix and hence, the total number of unknown quantities which are present in this sigma matrix are p into p plus 1 by 2. So, the total number of unknown quantities which are present in this mu vector and the covariance matrix are p plus p into p plus 1 by 2 right.

So, in order to have inference about these unknown quantities and to have some estimators corresponding to the mean vector mu and the covariance matrix sigma, what is done is to go for random sampling from that particular population. Now, let us denote by the following quantities. Let x 1 vector, x 2 vector, x n vector be a random sample from the multivariate population. So, if we have that, then on the basis of this n random samples, now each of these x i's; x i vectors are p by 1; because that is basically random sample drawn from the p variate, a multivariate population. And hence in order to have the inference about the mean vector mu and the covariance matrix sigma, we will be using this particular set of random sample.

Now, let us denote by capital x, the random sample matrix. The random sample matrix, which is a p by n matrix basically comprises of the following vectors. So, where this x 1 vector, which is the first random sample to up to x n vector, which is the nth random sample. So, we have this p by n random matrix, which is actually having this entire random sample of dimension n. Now, this can alternatively be written in terms of the following. So, let us write this as y 1 prime, y 2 prime and y p prime, where these quantities are having the following interpretation.

(Refer Slide Time: 10:26)



So, we have this x j vector is basically the j th observation vector. We have n such observations. So, j equal to 1 to up to n, where this x j vector is given by the following, that this is say x 1 j, x p j. So, each of these are p dimensional, because that is j th observation vector corresponding to this p dimensional vector random vector. Similarly, this y i vector that we had defined here, that this is y 1 prime, y 2 prime, y n prime. This y i prime is what is holding the observations corresponding to the i th variable. So, these are basically y 1 prime. So, corresponding to the i th row of that matrix.

So, this would be x i 1, x i 2, x i n; where this i now is from 1 to up to p. So, this basically is the structure of data matrix; where this y i prime is basically the i th row of the x matrix and x j is the j th column of the x matrix right. In terms of these vectors, that we have defined through the random sampling; we will introduce the two important quantities, which is the random mean vector or sample mean vector and the sample variance covariance matrix, which would be used actually in order to have inference or rather the inference about the unknown mean vector mu and the unknown covariance matrix, that is sigma right.

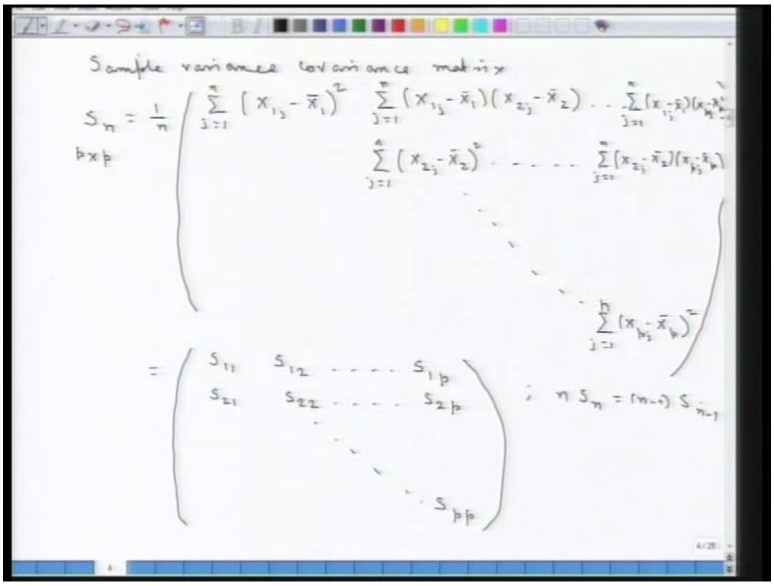
So, we will have this being defined, which is the x bar matrix or x bar vector rather. This is the sample mean vector. well In order to have the same notation as this one, what we will denote this by capital x i's. So, let us be consistent with this definition. So, you will have a capital x i's to denote the corresponding random variables and we will use small x i's to denote the observations. So, x capital x bar vector is the sample mean vector and it is given by the following. So, this is a p dimensional vector, which holds x 1 bar, x 2 bar and the p th variables mean x p bar.

So, this would be given by 1 upon n, then the summation of all the observations corresponding to the first variable. So, that is x 1 j say, this j is equal to 1 to up to n and the last element is 1 upon n summation j equal to 1 to n summation x p j right. So, these are the corresponding sample means of the respective variables in the x vector component. Now, in terms of this observation vector, this can be written in the following way. So, you can take 1 upon n outside and then, what we have here is? This is basically y 1 prime multiplied by I which is of dimension n; I n.

So, it is a column vector of dimension n with 1 as all the entries and the last one is, what is corresponding to the all the variables for the p th variable; all the observation corresponding to the p th variable, where this will just complete this particular stuff here. So that, this in terms of the random matrix, what we have is 1 upon n x times I n; where this I n is an n by 1 column vector, which has 1 as all its entries. So, this x bar which is the sample mean vector; it is a random vector once again; because it is comprising of the random variables that we have taken. So, the entries of that particular x vector are given by these, which in terms of the observation vectors that we had defined earlier, the p observation vectors.

So, these basically are these observation vector, y i primes and that random sample mean vector in terms of the random matrix, that the data matrix that we have defined. It is 1 upon n x times I n. Now, let us move on to looking at now this x bar vector, which is the sample mean vector is what is going to be used for inference that is based; that is inference about the population mean vector, which is unknown; which is the mu quantity. Now, in order to look at the other unknown quantity which is present; which is sigma the variance covariance matrix.

(Refer Slide Time: 15:50)



We introduce the sample variance covariance matrix. So, the sample variance covariance matrix can have divisor either n similar to the univariate set up can have a divisor as n can have a divisor n minus 1. Suppose we define that in terms of a divisor n, then this is how the sample variance covariance matrix is going to look like; so, this sample variance covariance matrix which is going to be based once again on the random samples that we have drawn. It is a p by p random matrix, which has the following entries. So, the 1 1th element of that is going to be given by this x 1 j minus x 1 bar.

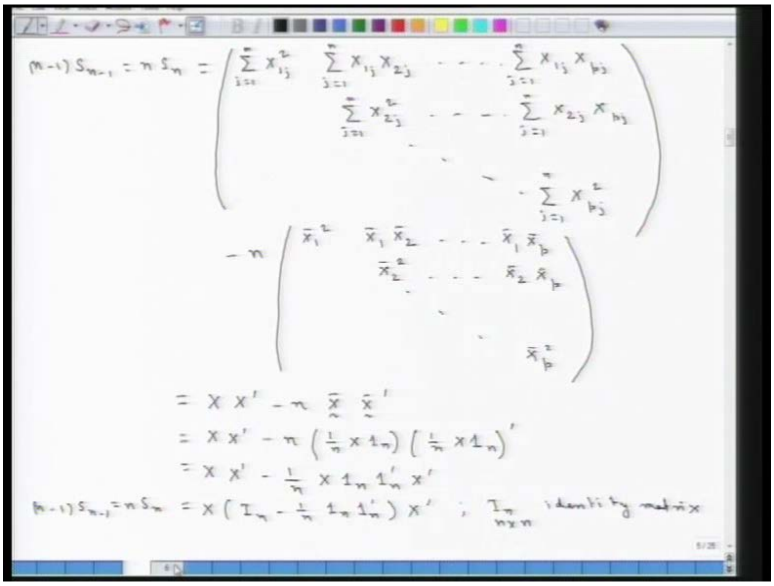
Now, x 1 bar is what we have already defined through this random vector here, the sample mean vector. So, that is what is going to be used here. Then, the 1 2 th element is going to be given by this x 1 j minus x 1 bar; this into x 2 j minus x 2 bar and the last entry here do not have enough space actually to write it. So, this would be given by this x 1 j minus x 1 bar that multiplied by x p j minus x p bar right. So, this is the matrix here. Then, this is going to be a symmetric matrix. So, this we only need to write the upper try upper triangle of this particular matrix.

The 2 2 th element would be the one, which is corresponding to the second variable; this is x 2 j minus x 2 bar whole square. The last entry in this row would be x 2 j minus x 2 bar this into x p j minus x p bar j equal to 1 to n and the last diagonal entry the p p th element of this matrix would be j equal to 1 to up to n x p j minus x p bar whole square. Now, if you look carefully at this sample variance covariance matrix that we have listed out here, it basically is holding say for example, this particular element; this you can associate with the variance covariance variance for the first component.

So, we can denote that by s 1 1 say, then 1 upon n of this quantity is what is sample covariance between the first and the second variable. The last entry in the first row here is the sample covariance between the first and the p th variable. So, this would be s 2 1; s 2 1 is just this element s 1 2 itself; because this is the variance covariance matrix. This is s 2 2, s 2 p and this would be s p p element, where the s i j is the i j th entry of the particular sample variance covariance matrix right. Now, if we define the sample variance covariance matrix through divisor n minus 1, we will see later on that; one of them would be unbiased estimator of the population covariance matrix sigma.

And the other is going to be associated with the maximum likelihood estimator, when we talk about random sampling from a multivariate normal distribution. So, one could have also define this, in terms of this your yes n minus 1 quantity. So, we can say that n times s n would be in such a situation will be given by n minus 1 s n minus 1 right; where the s n minus 1 matrix would be this matrix what we have with a divisor here as n minus 1 right. Now, let us try to write this particular sample variance covariance matrix, what we have defined in terms of the data matrix x that we had introduced.

(Refer Slide Time: 20:43)



So, we will have the n minus 1 s n minus 1 is equal to n times s n. Now, that if we look at this particular matrix here, now this element here can be written in terms of summation following. I will just write it out here that this entry here; we can write it as x 1 j square this minus n times x 1 bar square. So, all these entries similarly can be written like this. Say for example, the 1 2 th element here can be written as summation x 1 j x 2 j minus n times x 1 bar in to x 2 bar. So, keeping that in mind, we can write this particular matrix in two parts. The first part will hold the sum of squares and the cross products.

So, we can right that as x 1 j square; this j equal to 1 to n. The second entry, we will just have the first quantity which I said x 1 j x 2 j; the last entry here is summation from j equal to 1 to n x 1 j into x p j. This is going to be x 2 j square j equal to 1 to n and the last entry here similarly would be j equal to 1 to n x 2 j x p j and the last entry p p th element of this matrix is just the sum of squares corresponding to the p th component. So, this is the first entry, the first block actually and then, this minus n times the sum of squares entries which will be getting like here; n times x 1 bar will come here; n times x 1 bar into x 2 bar will come here and similarly, the other entries will follow.

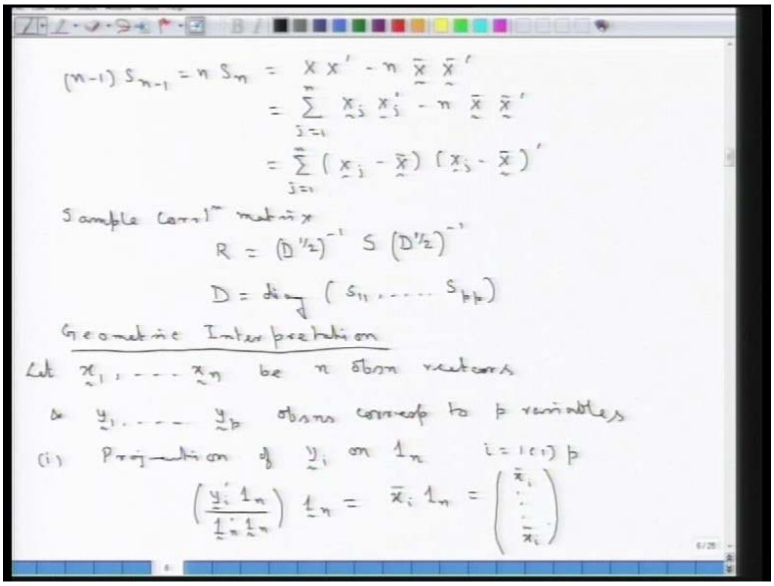
So, this can be written in terms of x 1 bar square. The second entry will be x 1 bar x 2 bar. The last entry in this first row would be x 1 bar x p bar x 2 bar square x 2 bar into x p bar. Then, the last diagonal entry would be x p bar square right. So, once we have written it in this particular form, it is easy to realize that the first matrix that we have written here in terms of the data matrix; that we had introduced that x. It is just x x transpose right. The x matrix we had defined here which was well we can actually we had written this as this matrix here. Let me see yeah this particular matrix here, this basically is the x matrix.

So, if we write the corresponding entries here x 1 vector, the first component; so, it will have the p components out there. Then, this x n will have once again the p components of that n th observation vector. And then, this matrix p by p would actually lead us to the product x x transpose, which would now be having the diagonal entries as the sum of the squares of the respective components and the off diagonal entries will hold the products. Now, this particular matrix block here can similarly be written in terms of the x bar vector. So, this is x bar vector that multiplied by this x bar vectors transpose.

What is x bar vector? x bar vector is what we have defined here as the respective x bar components for the p entries (( )) p elements. Now, we from this particular form, it is sometimes useful to reduce it to this particular form. Now, this is not yet in terms of entirely the data matrix. Now, as we had seen that, this x bar vector can be written in terms of the data matrix as x bar vector equal to 1 upon n x I n. So, we can use that same thing out here. So, it is 1 upon n x times I n and then, the transpose of that particular vector. So, this can be written as x x transpose minus 1 upon n; this one will get canceled out.

So, will have one one upon n remaining and then, this is x I n, then the transpose of this quantity. So, what will be having is I n prime and x prime. So, we can rewrite this in a more compact form as x I n minus 1 upon n I n I n prime; this multiplied by x transpose. So, this particular form what we have is what is now expressing this the sample variance covariance matrix with either the divisor n or the divisor n minus 1 in terms of the observed data matrix. Now, here I n of course is an n by n identity matrix. So, this is what we get. Now, an alternate form of this sample variance covariance matrix is also sometimes useful.

(Refer Slide Time: 26:36)



I will just write that, say n minus 1 s n minus 1 n times s n that what we have derived as x x transpose into x bar transpose. So, since this x vector, x matrix rather; x matrix is of this particular form x x transpose can be given by the following that, this is summation j equal to 1 to n x j x j transpose minus n times x bar x bar transpose. So, this can be written as following x j minus x bar vector into x j minus x bar transpose j equal to 1 to n right. So, this form also we will today itself use this form, in order to derive an unbiased estimator for the population covariance matrix, that is sigma. Now, similar to the population correlation matrix, one can also define the sample correlation matrix.

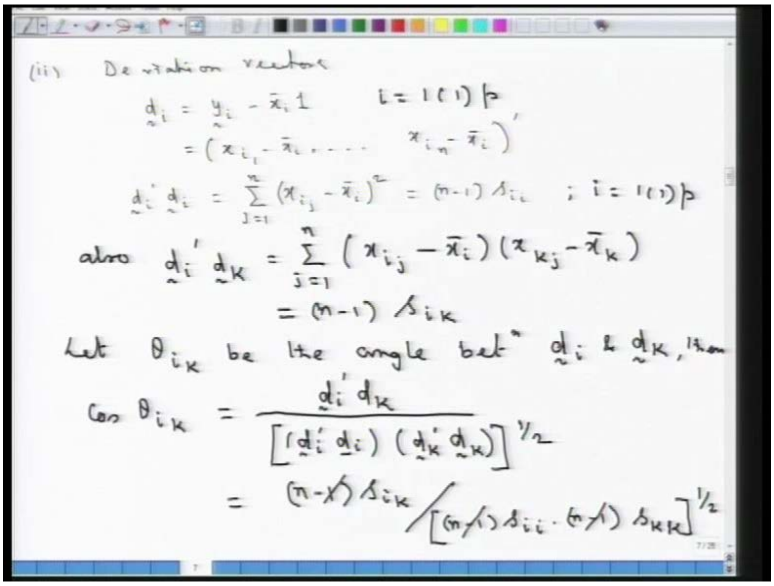
Say, the sample correlation matrix say denoted by R would be say D half inverse times s, where s is either with the divisor s n or n minus 1; this into D half to the power minus 1; where this D matrix is the diagonal matrix which is holding the sample variances of the respective components p in number right. So, through this diagonal matrix, if we look at pre and post multiplication using D half inverse; So, what will be getting is the sample correlation matrix; where the i j th element will actually be the sample correlation random variable for the x i and x j component of this right. Now, let us look at a bit of geometric interpretation for this random sampling.

So, when we talk about this geometric interpretation, we will look at too simple interpretations. Suppose we look at these as observation vectors, let x 1 x 2 x n be n observation vectors. Now, from these observation vectors we had seen that, one can also write this in terms of y i's. So, this y 1 vector, y 2 vector, y p vector; this would be observations corresponding to p variables. Now, if we look at the projection of any y i on one projection of say any y i vector on I n is going to be given by y i prime 1 divided by 1 prime 1; this multiplied by this 1 vector and what is this equal to?

This is the projection this is the projection vector, which is the projection of y i on I n. So, there are p such vectors, i equal to 1 to up to p. Corresponding to each of these p vectors, which now are holding all the observations corresponding to these variables. So, this is what is going to give us the sum of all the entries, which we have corresponding to the i th variable and 1 prime 1; this is I n I just drop this I n from all these places. So, later on actually without loss of any generality will just drop this particular I n index will say that, it is a vector which is holding 1's and its (( )) particular dimensions conforming to the other vector.

So, this is just the sum of all these observations and this is going to give us n. So, this is what we will be having is x i bar times I n. So, this basically is the vector, which is of dimension n and has entries x i bar at all the places. So, that is basically the projection. So, this is the mean mean of the i th variable and this is the n dimensional vector and that is what has got the interpretation that, it is just a projection of the y i vector. The i th vector correspond the vector corresponding to the i th variable are holding all the n observations on this I n vector and that is what is this.

(Refer Slide Time: 32:11)



Now, let us also look at the following deviation vectors. Let us denote by say d i, the deviation vector. Now, the deviation vector is the deviation, which we are going to define as y i; this minus x i bar times I right. So, this is what is going to hold the following quantities that, we have x i 1 minus x i bar; this is x i n, the nth observation minus x i bar. So, this deviation vector is what is going to give us in each of the components, the deviation of the respective observations corresponding to that i th variable.

This is the first observation in the i th variable that minus the mean corresponding to all the n observation of that i th variable. So, these are this is the deviation of the first observation from its mean from the mean corresponding to that variable and likewise, we have all these n entries like this. If we look at the square of the norm of this deviation vector; if we just look at this d i prime d i, what is that going to give us? That is going to give us the sum of squares of these deviation quantities. So, this is going to be given by x i j minus x i bar square this j equal to 1 to n. So, what is this quantity? This quantity is n minus 1 times s i i.

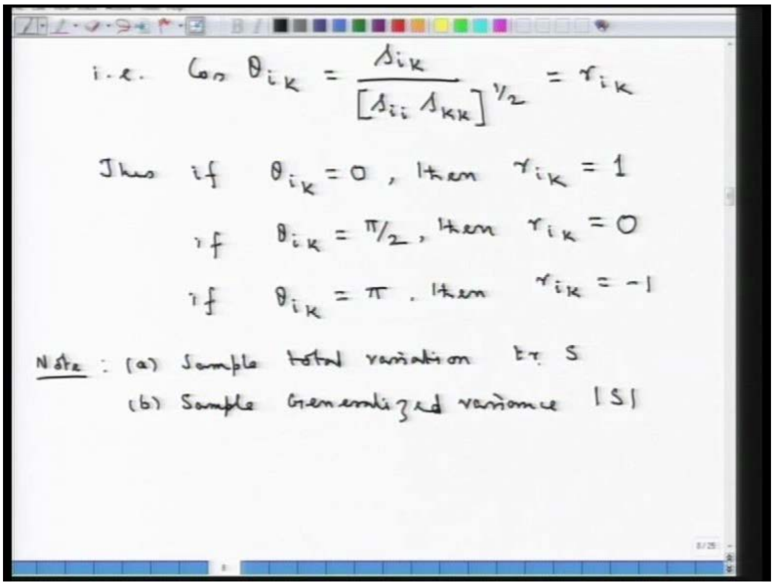
So, this is actually if we have the deviation vectors, d i is being denoted by this deviation as we have discussed that, these are the deviations basically from the respective mean components. Then, d i prime d i is nothing but the sum of squares of this l these entries in the deviation vector, which is associated with n minus 1 times s i i; where s i i is the sample observed variance components corresponding to the i th components. So, we have this deviation vectors p in numbers and hence, this also will be p in numbers. So, it is basically the norms square of this deviation vector are associated with the sample variances of the respective components with the multiplier n minus 1 as the divisor.

Now, similarly also what we can actually see that this d i prime d k, which is the dot product of the deviation for the i th variable and the dot product for the k th variable. This would be the cross product j equal to 1 to n x i j minus x i bar. So, this is x i j minus x i bar into x k j minus x k bar right. And this is what this quantity, which is the cross product between the deviation vector for the i th variable and k th variable. It is nothing but the covariance n minus 1 times s i k. So, where s i k is 1 upon n minus 1 of this product, which is the covariance between the i th and the k th variable.

Let us also now look at the angle between these two deviation vectors. Let us denote by theta i k. Let theta i k be the angle between d i, the deviation vector for the i th variable and d k, the deviation vector corresponding to the k th variable. Then, what we have is this cosine of this theta i k is going to be given by this, which is d i prime d k; this divided by d i prime d i, this for the i th vector d i; this into d k prime d k whole raise to the power half right. So, if we have these two vectors d i and d k, then the cosine of the angle between the two vectors two deviation vectors is given by this and what is this equal to?

As we have seen that, this d i prime d k is nothing but our n minus 1 times s i k, the covariance that divided by now this d i prime d i, as we had seen out here is n minus 1 times s i i. So, this comes down to n minus 1; this as s i i. So, that is coming from this and this d k prime d k is n minus 1 times s k k. So, this is this s k k whole raise to the power half. So, what we sees that, this n minus 1 factor cancels out and what we have as the cosine of the angle between these two deviation vectors is s i k divided by under root of s i i into s k k.

(Refer Slide Time: 38:06)



That is, the cosine of this angle theta i k is s i k that divided by s i i in to s k k whole raise to the power half and what is that equal to? That is, just the correlation the sample correlation between the i th and the k th variable. So, this gives us a nice geometric interpretation about these deviation vectors, that are associated with this random sampling; that the cosine of the angle between the two is nothing but the correlation between two random variables i and k. Now, given this particular expression here we can say that, if this theta i k the angle between the two deviation vector is 0. So, they are basically in the same direction.

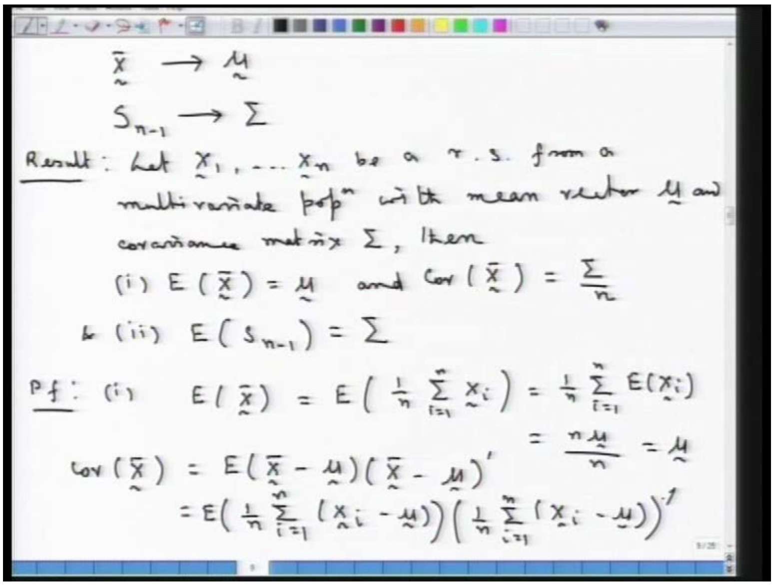
So, if this theta i k the angle between the two deviation vector is 0, then what we have this r i k is cosine of that angle 0 and hence, this is equal to 1. That justifies actually our intuition that, if two deviation vectors are in the same direction, then the correlation between the two random variables would be perfect linear correlation. So, that the correlation is equal to 1. Now if on the other hand, the two are orthogonal if theta i k equal to pi by two. So, we have the two deviation vectors orthogonal to one another. Then, what is the value of the correlation coefficient between the two as we would expect the two are orthogonal?

So, it is moving in orthogonal directions and hence, the correlation would just be equal to 0. In the other extreme, if they move in opposite direction not orthogonal, move perfectly in opposite direction; that is, if the angle between the two deviation vectors is pi. So, the two vectors are moving in completely different, but perfectly opposite direction. Then, what we have this r i k is what we expect that, they are exactly in the oppose perfect negative correlation. So, this gives us a nice feeling about verifying the intuition that, what happens to the deviation vectors and the angle that they are making and the corresponding values of the measure of association between the two any two variables.

So, this cosine theta i k ofcourse, it is for every pair i and k taken from the set of possible p variables right. Now you remember that, we had defined at the start that, two quantities which are associated with the covariance matrix sigma, the total variation in x given through the trace of sigma matrix and then, the generalized variance of x given by the determinant of this sigma matrix. Similar to that, one can also define the sample quantities. So, the sample quantities would be the sample total variance. One can define the sample total variation as trace of s matrix s either with a divisor n or with a n minus 1.

And the sample generalized variance as the determinant of s matrix and as we had argued that these ofcourse, gives us compression of the sample variance covariance matrix. However, these are not sufficient enough to replace s. Because we can construct in a similar way to what we had seen for the sigma matrix that, it is possible to have two different sample covariance matrix giving us the same total sample variation and the sample generalized radiance right. Now, we move on to one important result regarding this estimation procedure.

(Refer Slide Time: 42:35)



Now, as we had seen that, we are we have the two following quantities, which is the x bar vector; which we are going to associate with the population mean vector, which is mu and we have the sample variance covariance matrix say with a divisor n minus 1. Now, that is what is going to be used in order to have inference about the population variance covariance matrix that is sigma. The way of the following result, let me first state the result. So, let our x 1, x 2, x n be a random sample be a random sample from a from a multivariate population with mean vector unknown as mu and covariance matrix as sigma.

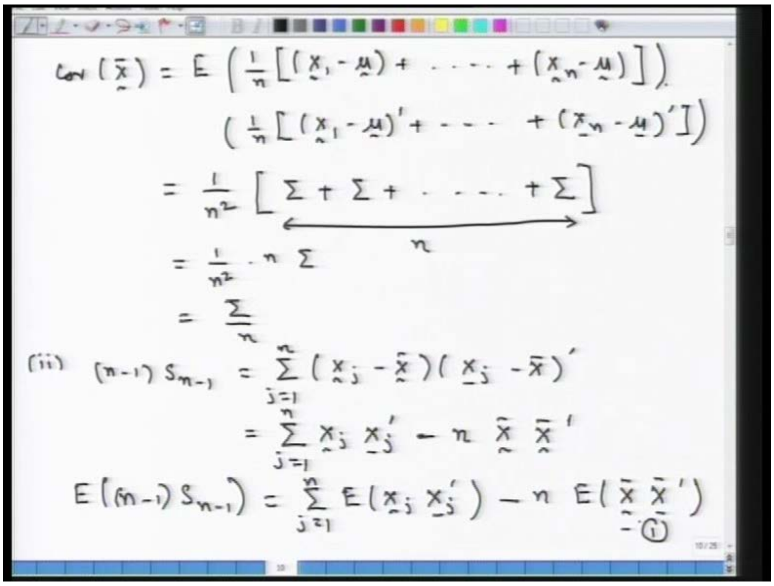
Then, the two quantities that we have defined as x bar and s n minus 1 has the following properties. Number one: expectation of this x bar vector is going to be equal to the mean vector mu and covariance matrix of this x bar vector is going to be sigma by n and number two: for the covariance matrix sample variance covariance matrix s n minus 1, expectation of s n minus 1 only is equal to sigma. In other words, we are trying to say that this x bar vector, the sample mean vector is an unbiased estimator of the population mean vector, which is mu.

The sample variance covariance matrix with a divisor n minus 1 is an unbiased estimator of the population variance covariance matrix, that is sigma right. And the covariance matrix of this x bar vector is going to be given by sigma by n. This gives us the feeling of generalization of the univariate result what we have to the multivariate set up. We had ofcourse; similar result when we had univariate set up. Let us quickly look into the proof of this particular result. Now, in order to prove this one, one is quite straight forward that, expectation of this x bar is expectation of the respective components.

So, its expectation of 1 upon n x bar; now x bar is going to be given by this x i bar. So, its all the all these observations i equal to 1 to n. Now, when we look at expectation of this; its expectation operators comes inside and what we have is i equal to 1 to n. Expectation of these x i components and expect x 1, x 2, x n are random sample from the same multivariate population. Each having identical mean as mu and hence, expectation of each of these exercise will be equal to mu. So, what will have this as n times mu; this divided by n and hence, that is equal to mu. So, we have this first component of this result.

We can also now look at the second result, which gives us the covariance of this x bar vector. Which by definition of the covariance matrix of any random vector would be given by x bar minus its expectation vector, which is mu as what we have derived; that multiplied by the transpose of that same quantity. So, what we have this is the following that, one can write this as 1 upon n summation i equal to 1 to n. Then, we will have this x i minus mu and then, the transpose of the same quantity, which is this i equal to 1 to n x i minus mu transpose. Now, what is this quantity equal to? This quantity would be equal to the following.

(Refer Slide Time: 47:10)



Let us just split this or rather write it term by term. So, the covariance matrix of this x bar vector is going to be given by expectation. Now, what is the first element? The first element is 1 upon n, as we had seen it is summation of those components. So, that it will hold these quantities. So, this is x 1 minus mu plus the last term is nth random sample vector. Now, this multiplied by the transpose of this. Transpose of this would be given by 1 upon n and then, the transpose of the first entry, which is x 1 minus mu transpose plus x n minus mu transpose right.

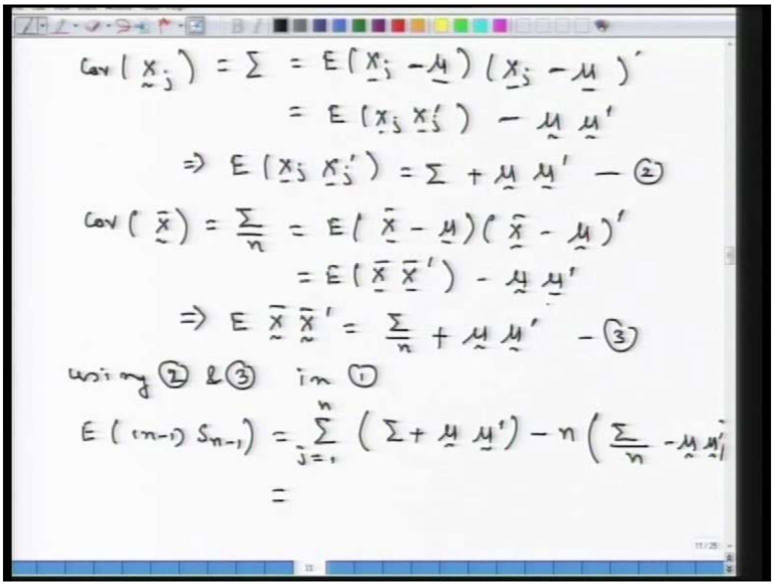
Now, when we look at taking expectation term by term of each of these entries with these; now, if we look at expectation of this term multiplied by this term, the first term here; what will be getting is the covariance matrix of this x, which is sigma apart from this multiplier. Now, when we look at the cross product say this with the next element here; now remember that this x 1 vector, x 2, x n; they are random sample and hence, they are independent. So that, we will have the expectation of the cross product of this with any of these here, except the first entry would be 0; because we have x i x any x i be independent of x j, if i is not equal to j.

And hence, the covariance matrix between the two would be equal to 0. So that, what will be having is the following; after we take expectation is 1 upon n square will be there and then, we will have this. When this is multiplied with each of these entries and then, expectation being taken only the first element would give us sigma and all the rest of the elements will be 0. When we look at the second entry here, once again the same entry here will give us the sigma matrix and the rest of the entries will be zeros. So, we will have these n sigma components. So, these are n in numbers.

Corresponding to this type of product 1 being 1 with 1, 2 with 2 and n with n, all the cross product entries will be zeros; because of independence of these components x 1 bar, x 1, x 2 and x n. So, this thus gets us down to 1 upon n square n times sigma matrix and that is nothing but, this sigma by n matrix right. So, we have proved the first part of this particular result. Let us now move on to proving the second part of the result, which establishes actually the unbiasness elements of this covariance matrix. So, what we had seen? this is the second part of the result. So, what we now have is, this n minus 1 s n minus 1; that is what, we had seen earlier is x j minus x bar into x j minus x bar transpose, this j equal to 1 to up to n right.

Now, we had also seen that, this particular term is written equivalently in the form that, this is j equal to 1 to n x j x j prime minus n times x bar x bar prime right. Now, we look at proving the results. So, if we now look at expectation of this n minus 1 s n minus 1. This is going to be given by expectation of this particular entire quantity. So, we take expectation term by term. We will have this as expectation of x j x j prime, this minus n times expectation of x bar x bar transpose. Let us give an equation number 1 to this because will be requiring this latter on. So, we need to find out those two quantities. What those expectations are? Expectation of x j x j prime and expectation of x bar x bar prime.

(Refer Slide Time: 51:52)



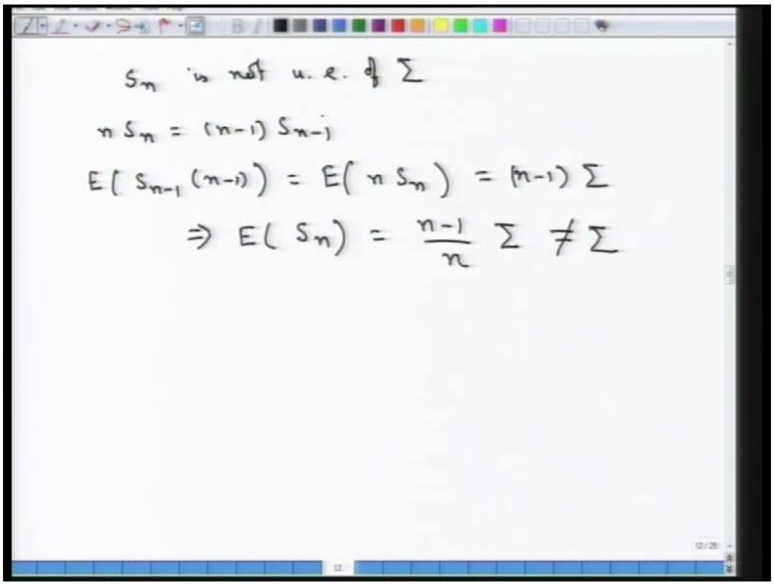
So, what we realize is that expectation this covariance matrix of x j. x j is the random sample drawn from that multivariate population. So, the covariance matrix of x j is nothing but sigma. So, that is equal to expectation of this x j minus mu into x j minus mu transpose. And hence, this is equal to as we had seen in the last lecture, this is nothing but x j x j prime; this minus mu mu prime. And hence, this would imply that expectation of x j x j prime, which is a quantity that we would be requiring in order to evaluate that expression 1 is given by this right.

Furthermore, if we recall the result that we had proved in a first part, covariance matrix of x bar is sigma by n and that is equal to expectation of x bar minus mu, its mean into x bar minus mu transpose right. So, by the same approach what one can show is that, this is expectation of x bar x bar prime; this minus mu mu prime right. So, what we will be having is this also, that expectation of x bar x bar prime is going to be equal to sigma by n; this minus mu mu prime. So, this is one that we have going to use and this is one we are going to use.

So, we will use this 1 and 2 or rather using 2 and 3 in 1. 1 is what; this particular expression. So, what will be having is this that expectation of n minus 1 s n minus 1; that is summation j equal to 1 to n, then expectation of this particular quantity. So, that what will be having here is sigma mu mu transpose; this minus we have this as n times expectation of this quantity. So, that sigma this is sigma by n; this minus mu mu transpose right. So, if one simplifies, this will cancel out.

So, will have one sigma from here; you will have n minus 1 sigma from here and then, this mu mu transpose term. Just a minute this is plus sign out here. Because if you take this expectation of x bar x bar prime to this side, then we will have sigma by n plus mu mu prime. So, this is the plus sign out here not a minus sign; then what will be having here is just n minus 1 times sigma. This would imply that, expectation of s n minus 1 is just going to be equal to sigma. So, this will imply that, s n minus 1 is an unbiased estimator of the population variance covariance matrix that is sigma.

(Refer Slide Time: 55:27)



Now, if on the other hand, we take s n is not going to be an unbiased estimator of sigma. Why? Simply, because this n times s n minus s n is n minus 1 times s n minus 1 and what we have proved is that, expectation of s n minus 1 that multiplied by n minus 1. So, that would also be equal to expectation of n times s n; that is equal to sigma. This would imply that, this was actually equal to n minus 1 times this sigma. This is n minus 1 times sigma. So, this would imply that, expectation of s n is going to be equal to n minus 1 by n times sigma and which is not equal to sigma. So, this proves that, though s n minus 1 with the sample covariance matrix with a divisor n minus 1 is an unbiased estimator of sigma. However, this s n is not an unbiased estimator of sigma. However, as n goes to infinity, this will be an unbiased system estimator and hence, this s n will be an unbiased estimator in the limit.