Applied Multivariate Analysis

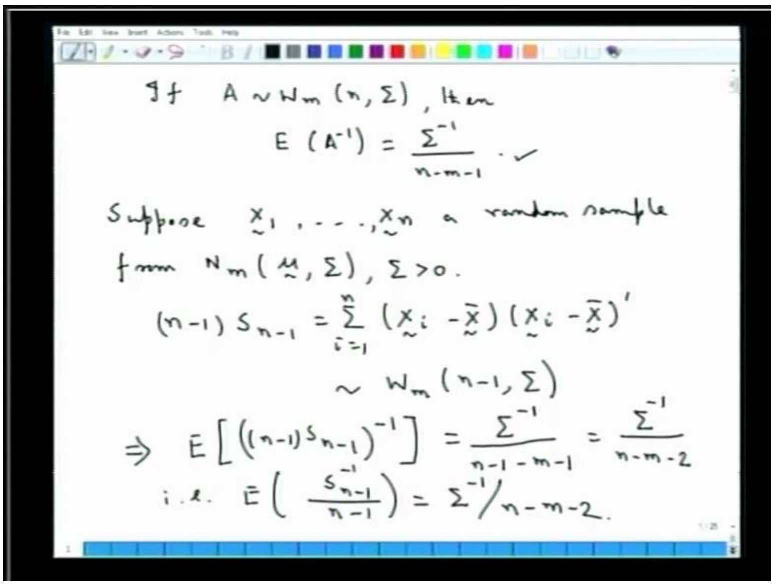
Prof. Amit Mitra
Prof. Shramishtha Mitra
Department of Mathematics and Statistics

Indian Institute of Technology, Kanpur

Lecture No. # 13

Hotelling's T2 Distribution and it's Applications

(Refer Slide Time: 00:26)



In the last lecture, we had started discussing about the inverted Wishart distribution, and we had proved in important result that if A has got a Wishart distribution, Wishart m n sigma; sigma is positive definite. Then expectation of A inverse A inverse was defined to be the inverted Wishart distribution.

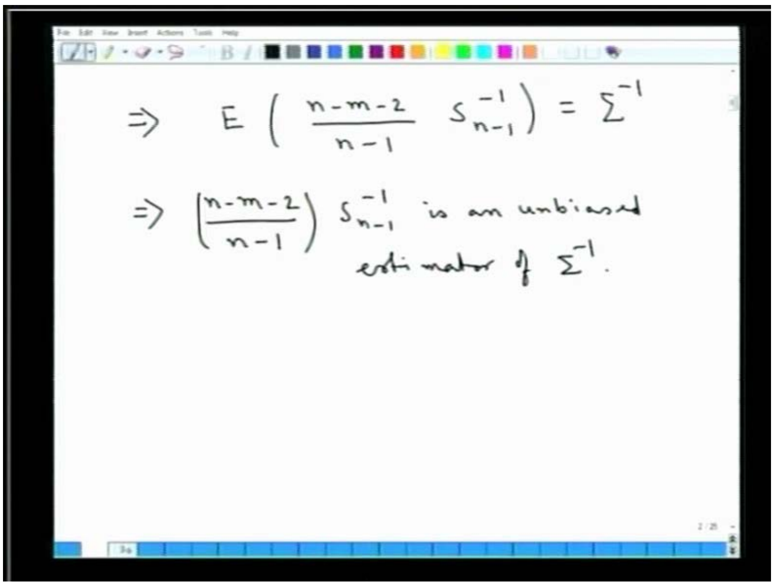
So, expectation of A inverse was equal to sigma inverse that divided by n minus m minus 1. Now, we will use this particular result that we had derived in the last lecture in order to get to an unbiased estimator of sigma inverse, when we have a random sampling from a multivariate normal distribution. That is suppose, we have got the following setup, suppose we have X 1, X 2, X n, a random sample from N multivariate normal distribution with mean vector mu, and the covariance matrix sigma. Then what is unbiased estimator of the sigma inverse matrix as one can feel that it is basically going to be based on this particular result that we had derived.

Now, what do we know? When we have random sampling X 1, X 2, X n, from a multivariate normal distribution which is given by multivariate normal m mu, sigma; sigma of course is a positive definite matrix. Then, we know that n minus 1 S, let me still right it as S n minus 1 to indicate that, this is sample variance, covariance matrix with a deviser n minus 1. This is given by i equal to 1 to up to n, x i minus x bar into x i minus x bar whole transpose, and this we had seen from the result proves in the last lecture that this has got a Wishart distribution on m dimension with n minus 1 degrees a freedom, and then associated variance, covariance matrix as sigma.

So, this would imply by this result out here, that expectation on of the inverse of this Wishart matrix. That is expectation of n minus one, S n minus 1 inverse of this, because we require expectation of the inverse of the Wishart matrix. That is invert Wishart this would be given by sigma inverse; whatever is the associated variance, covariance matrix here the associated variance, covariance matrix is sigma. So, this divided by we had here n minus m minus one. So, n is the degrees of freedom of the Wishart distribution here that is n minus one.

So, that we will have this as n minus 1 this minus m minus 1. So, that this is equal to sigma inverse this divided by n minus m minus two. So, this would imply further that this expectation of now we have here inverse of this quantity. So, we will have s n minus 1 inverse of it that divided by n minus 1 this is equal to this sigma inverse divided by n minus m minus two, and both thus that lead us to.

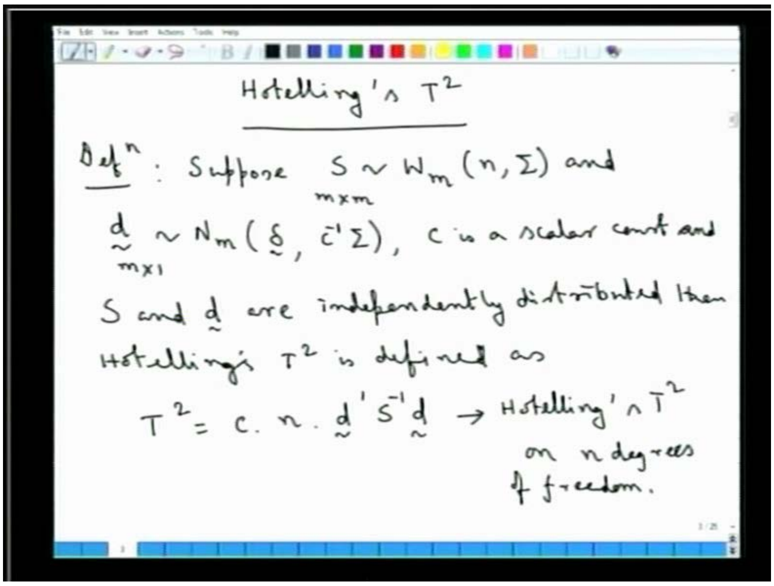
(Refer Slide Time: 03:56)



This would imply that expectation of this n minus m minus two; what we have on the right hand side that divided by n minus 1 times, S n minus 1 inverse, that is equal to sigma inverse.

So, this would imply further that this n minus m minus two; that divided by n minus 1 this constant multiplied times, S n minus 1 inverse matrix is an unbiased estimator of sigma inverse. So, we have the desire quantity that we have obtained a matrix which gives us which is unbiased estimator of the sigma inverse matrix. So, that is simple result.

(Refer Slide Time: 04:50)

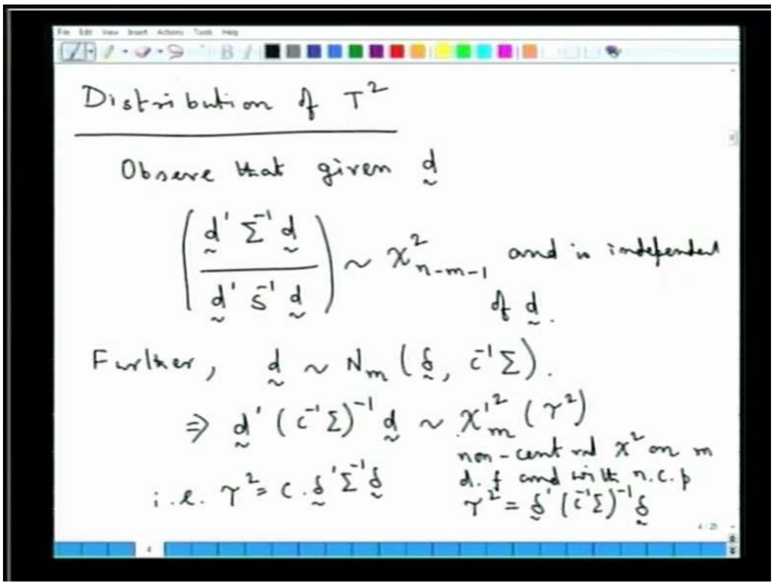


Now, remove on to hoteling's T square distribution; which is a multivariate generalization of the students T distribution for the invariate setup hotelling's T square distribution. Let me, first give the definition of what a hotelling's T square distribution is suppose, we have got a Wishart distribution; suppose S follows a Wishart distribution, Wishart m n sigma, and d be a random. Now, this S of course, S any random matrix which is of the order that it is m by m, and let d be a random vector m by one, which is got a multivariate normal distribution, multivariate normal m dimensional with a mean vector as say delta vector, and a covariance matrix given by c inverse sigma; where c is a scalar constant, c is a scalar constant and suppose we further have S, and d are independently distributed.

These two are independently distributed then hotelling's T square distribution is defined in the following way hotelling's T square is defined as, T square which is equal to c times, n times, d prime, s inverse d; now, this is what is the hotelling's T square hoteling's T square statistic hotelling's T square on n degrees of freedom. So, this has got a hotelling's T square distribution on n degrees of freedom.

The degrees a freedom what we have here n is associate with degrees of the freedom of the associated Wishart distribution which comes as a constant part of this hotelling'] T square distribution. Now, thing to be noted here is to that we are just looking at two sets of a random variable one is a random matrix another one is a random vector, and we have a Wishart distribution of the random matrix, and we have a multivariate normal distribution of the random vector, and critically, we would require independence of this random matrix S, and this random vector d. Now, the hotelling's T square distribution of course, is a very important distribution, and multivariate distribution theory as, I said that this hotelling's T square is the generalization of the usual students T distribution in case of invariate distribution. So, this is going to solve a rather the this is going to be used in problems similar to what we had use; say in student T distribution in case of invariate normal distribution theory.

(Refer Slide Time: 08:16)



But before, we can use that; let me, just try to look at what is going to be a distribution of this T square statistic. So, in order to get to the distribution of T what we do is? We go in two steps: in order to get to the distribution; now, first observe that given d; we will have this distribution d prime sigma inverse d. This divided by d prime S inverse d. Now, this is a quantity which we had introduced in the last lecture, and we know that this has got a central chi square distribution on n minus m minus one; decrease a freedom, and this independent of and this independent of this d vector.

We had derive this particular distribution taking at a first as conditional distribution, because once we say that observe that given d; that is condition on d be a given vector. We will have the distribution of this to be a central chi square on n minus m minus 1 degrees a freedom, and that is independent of d; it does not depend on the particular choice of fixing of d, and hence the unconditional distribution which is this d prime sigma inverse d divided by d prime, S inverse d, as got a chi square distribution on n minus m minus 1 a central chi square distribution.

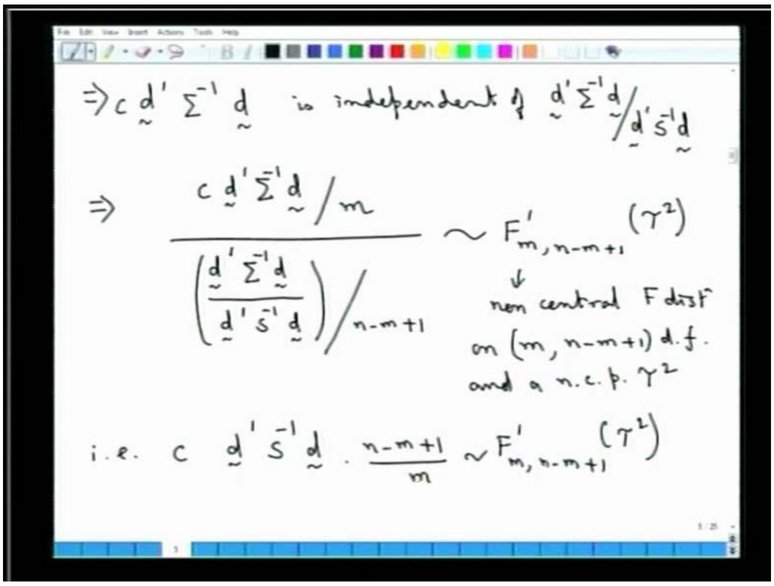
Now, further we have this d random vector to have a multivariate normal distribution, multivariate normal m delta. c is a scalar constant; remember c inverse sigma now from the results of the quadratic forms associated with a multivariate normal distribution; what we can say is the following: that this d prime, c inverse sigma; this is the covariance matrix inverse of that that multiplied d. This is going to have what distribution this is the quadratic form in d.

So, this will have a non-central chi square; a chi prime square on the degrees of freedom degrees a freedom would be the full of this particular matrix. So, it is the degree of freedom is m, and non-central parameter, say that is given by tau square. So, this is a non-central chi square on m on m degrees of freedom, and within non-centrality parameter with a non-centrality parameter tau square, and what is that tau square? The tau square term is going to be given by delta prime c prime sigma the inverse of it. That multiplied by this delta; that is in other words this non-centrality parameter here is just c times c a constant scalar constant delta prime sigma inverse delta.

Now, this quantity here this random variable is going to have a central chi square distribution. If we have this tau square to be equal to zero only under that particular condition this will have a central chi square distribution. Now, an important thing to note here and this particular point is that we have the distribution of this to have a chi square distribution which is independent of this d is the random vector; which is having a multivariate normal distribution.

Now, using that random vector d; this is having a multivariate normal distribution here. The quadratic form, what we have here as got a chi square distribution, and non-central chi square; now whatever be that the important thing is that since this quantity here is independent of d, and so, will be this independent of the quantity with which we have a obtained this non-central chi square.

(Refer Slide Time: 12:44)



So, what can we say this will imply that this d prime then we have that inverse of that quantity which is c sigma inverse d that was the term there is independent of the term which we had first define that is d prime, sigma inverse d that divided by d prime S inverse d why is that. So, because this is based on that d vector which is independent of this quantity; which is having a central chi square, and then, if we have that anything that is derived from that d vector would natural become in dependent of this quantity here.

Now, using that, and the fact, that this has got a non-central chi square on m degrees of freedom, and non-centrality parameter as tau square, and this has central chi square we can frame the following statistic; this would imply that this c d prime, sigma inverse d this is having a non-central chi square on m degrees of freedom.

So, we have this chi square non-central chi square divided by it is a degrees a freedom, that divided by the second chi square; which is independent of the first chi square critically d prime, sigma inverse d. This divided by d prime, s inverse d. So, this is the second chi square, and this divided by it is degrees a freedom what was it was n minus m. This is going to have a chi square on n minus m plus 1; actually, not minus 1 this is chi square on n minus m plus 1 degrees of freedom.

So, what we have here is this n minus m plus 1 degrees of freedom; this would follow, what distribution now this is a chi square random variate; a non-central chi square random variate on m degrees a freedom. So, we have meet that chi square divided by x degrees freedom chi square divided by x degrees freedom; the two are independent. So, this will have a non-central F distribution on the degrees a freedom which is m n minus m plus one, and the non-centrality parameter of this non central F distribution would be same as, the non-centrality parameter of this numerator chi square; which is a non-central chi square distribution. So, this is a non-central F distribution on m n minus m plus one; the two degrees freedom is associated with this, and a non-centrality parameter tau square. Now, tau square is what we had return out here; which is c times delta prime sigma inverse delta. right

So, in order to get to this statistics distribution, what we have used is the fact that this quantity here follows, a chi square on n minus m plus 1 degrees of freedom given d, and also since it is an independent of the particular choice of this the unconditional distribution of d prime, sigma inverse d divided by d prime, s inverse d follows, a central chi square on a n minus m plus 1 degrees of freedom, and that is going to be independent of this second chi square; which is c d prime, sigma inverse d, which is non-central chi square.

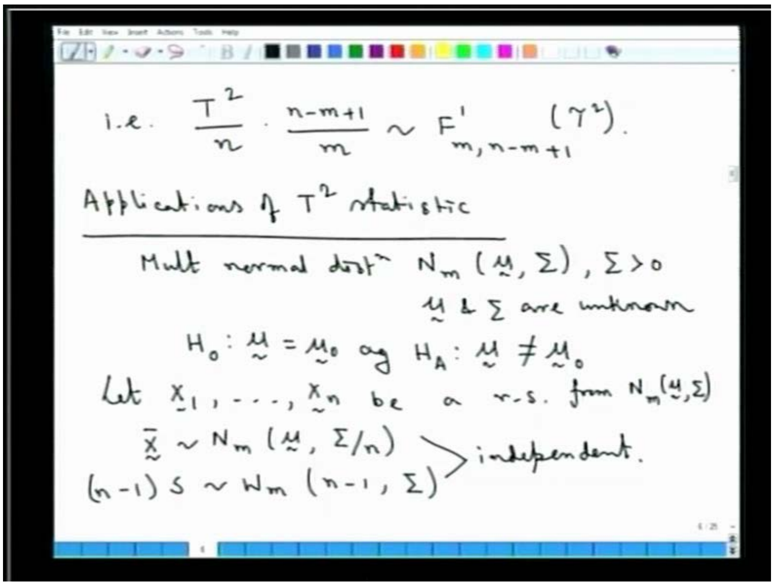
So, using this central, and the non-central chi square, we frame this particular ratio which is having a non-central F distribution on the degrees of freedom, and the non-centrality parameters as given above.

Now, if we simplify this particular term what will be getting is this term cancelling out d prime sigma inverse d. So, we will have this as c, d prime, s inverse d, and this multiplied by n minus m plus 1 that divided by m. This follows this, F prime the non-central F distribution on m n minus m plus 1 degrees of freedom, and the non-centrality parameter equal to tau square.

Now, it is easy to see that, what is a relationship between this statistic, and T square statistic; what was T square statistic we are define the T square statistic as c times n; n was degrees of freedom of the associated Wishart distribution.

So, c times n d prime S inverse d; let me, go back to that particular definition T square was defined as c n d prime S inverse d, and hence if we have a obtained the distribution of this quantity, we have obtain the distribution of d square.

(Refer Slide Time: 17:56)



That is what we will be having this T square by n T square by n would be just c times d prime S inverse d. So, it is T square by m in to n minus m plus 1 divided by m. This would follow that, non-central F distribution m n minus m plus one, and a non-centrality parameter as that equal to tau square.

So, this is the desired distribution of our hotelling's T square statistics; now let us, look at application of this hotelling's ; T square statistics applications of T square statistic, and how where actually this statistics is going to be used; now suppose, we have a multivariate normal distribution; say we have a multivariate normal distribution N m mu sigma; say sigma is positive definite both this mu, and sigma are unknown.

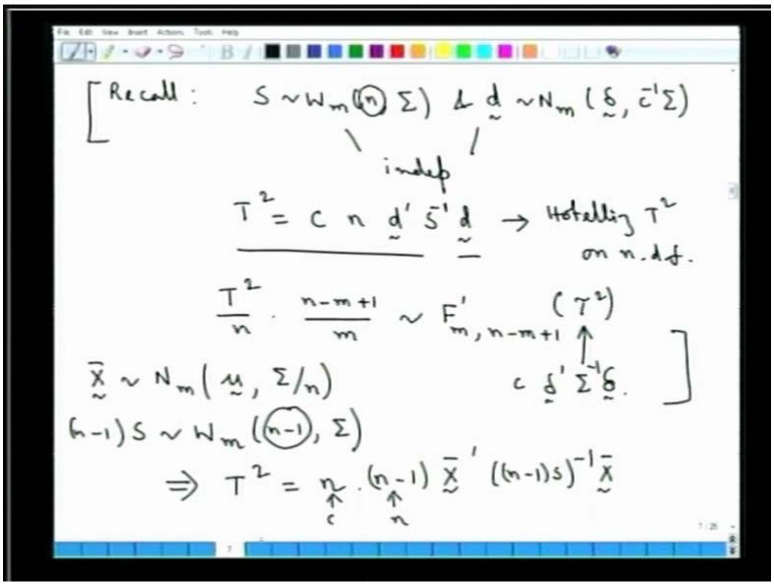
So, we have this particular setup; now under such a circumstances, suppose we have a null hypothesis H naught being framed as n mu equal to any specified vector mu naught. This is to be tested against, and alternate hypothesis H A which says that mu is not equal to mu naught. So, this is a type of testing; which we very frequently come across in multivariate theory where we are looking at this mu vector to be tested as taking this specified value mu naught; mu naught vector is of course specified. So, in order to tests this; what we do is we take a random sample as invariate theory. So, x 1, x 2, x n, be a random sample from this multivariate distribution multivariate normal m mu sigma, and then we will have to use this x 1, x 2, x n, in order to test this null hypothesis against this alternate hypothesis.

What do you know about, this random sampling we know the following fact that x bar the sample mean vector has got a multivariate normal distribution with a mean vector as mu, and a covariance matrix as sigma by n, and about the variance, covariance matrix say this S is S n minus one. I will drop this subscript n minus 1 we will say that this wherever I have this S is basically, denoting the sample variance covariance matrix with a deviser n minus one. So, n minus 1 S has got a Wishart distribution, Wishart m on n minus 1 degrees of freedom, and an associated variance, covariance matrix as sigma, and what is more about these two statistics is net x bar, and S are independent.

So, this is what we have already proved from concerning random sampling from a multivariate normal distribution.

Now, can we use this two information: in order to get to hotelling's T square distribution, and then frame the testing procedure for testing; this null hypothesis against this alternate hypothesis, yes because if we look back at the definition of hotelling's T square we would require a Wishart distribution. We would require a multivariate normal distribution, and we would require independence of the two.

(Refer Slide Time: 21:48)



So, what we have to use this that is the following: we may recall the theory, that we have just now learned is that, if S follows Wishart m n sigma. I will just write it once again, and this d is another random vector, which is having a multivariate normal m; let me, write it in the notation that had in introduced first.

So, this was a delta c inverse sigma, and if we have this two to be independent; then the hotelling's T square statistics was defined as c times n d prime s inverse d is. So, this was hotelling's T square on equal to n degrees of freedom, and for the more the quantity, which was T square by n, n minus m plus 1 divided by n. This was shown to have a non-central chi square on m n minus m plus 1 degrees of freedom, and non-centrality parameter equal to tau square; where this tau square is equal to c delta prime sigma inverse delta.

So, this is the fundamental result that, we have started today, and proved a part of it about the hotelling's T square distribution. So, we have here two constituent parts; which is A having a multivariate normal distribution, and this having a Wishart distribution, and they are independence this follows perfectly in line with this setup. So, further random sampling we will have the T square statistic to be given by c times n.

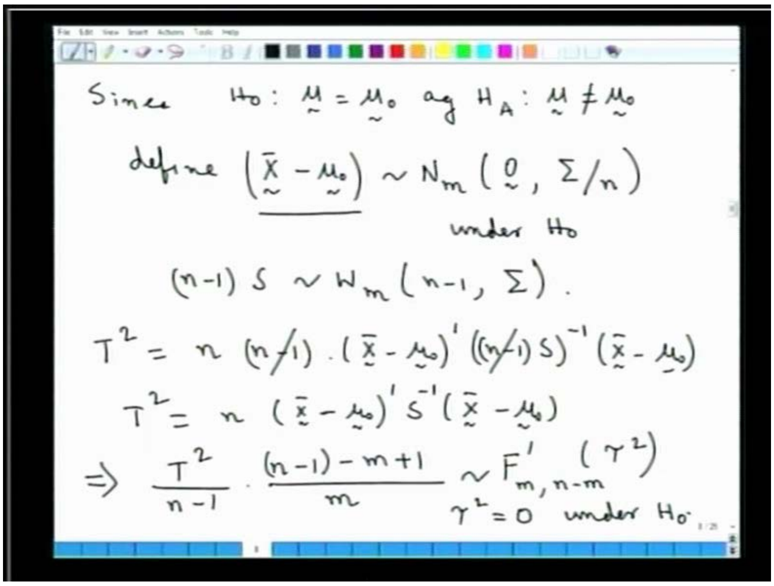
Now, let me right that statement once again here. So, that it would be easy for us to c y y the parameters are that this was having a multivariate normal distribution with a mean vector mu, and a covariance matrix as sigma by n, and our n minus 1 S, n minus 1 actually was having a Wishart distribution on n minus 1 degrees of freedom, and an associated variance, covariance matrix as sigma.

So, this falls in line with this particular definition. So, we will have this T square statistic being given by c times; now, what is c is c inverse is n inverse. So, c is our n, and what is our n. n In this definition of the hotelling's T square is the degrees of freedom is the degrees freedom associated with the Wishart distribution; the degrees freedom corresponding to our multivariate normal distribution is n minus one.

So, this basically is the c in this definition of hotelling's T square, and this is the degrees of freedom of the associated Wishart distribution; which was given by this n; now, what is d prime in our case it is x transpose prime x transpose prime, and this times S inverse, now what is our S inverse; this is our what I will do is I will just make this first: as this quantity, this is n minus 1 S whole inverse this is n minus 1 s whole inverse x bar right.

So, this is what is the counter part of this quantity further random sampling here, now this is going to have a non-central f distribution constant multiplier of that, what I will do; is that since our null hypothesis is mu equal to mu naught. since our null hypothesis is mu equal to mu naught Now, we know that if we have a null hypothesis. Then we will have to look at the distribution of the test statistics under the null hypothesis, and hence we have to bring in this a particular mu, mu equal to mu naught quantity; somewhere here, and that we will do here which is mu not equal to mu naught is the alternate hypothesis.

(Refer Slide Time: 25:57)



So, if we define this quantity mu naught remember is a known quantity. So, x bar minus mu naught; what is a distribution of this this is going to have a multivariate normal distribution multivariate normal m, and the mean vector would be zero under only the null hypothesis, and the sigma by n as a covariance matrix this under H naught.

Now, if that is not under H naught; then what would be the mean vector for this x bar minus mu naught statistics. It would be mu minus mu naught, because expectation of x bar would be equal to mu, and under the null hypothesis only that mu is equal to mu naught. So, under the null hypothesis we will have this to be a null vector, and sigma by n to be the associated covariance matrix nothing changes as for as the sample variance, covariance matrix is concerned; we will have that two have a Wishart distribution with n minus 1 degrees freedom, and associated with variance, covariance matrix sigma.

So, we will have this hoteling’s T square distribution hoteling’s T square statistics rather defined in terms of this new vector, which is centered in order to take care of this under the null hypothesis condition. So, what we will be having here is once again c, c remains n, and n is the degrees of freedoms. So, it is n minus one. So, we will have this n into n minus one; then d prime would now, the role of d prime; d prime will now be played by this x bar minus mu naught quantity.

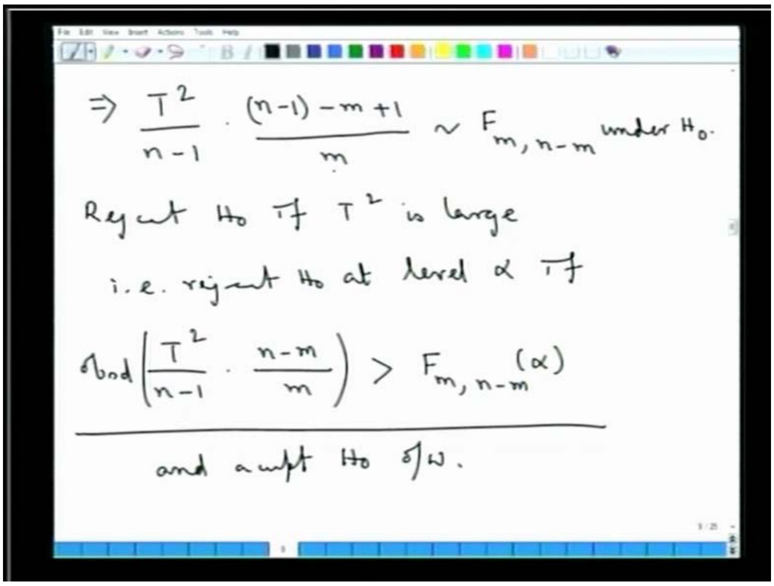
So, it is x bar minus mu naught transpose; then, the Wishart distributions inverse n minus 1 S whole inverse times x bar minus mu naught. So, what we see is that this n minus 1 term cancels with this one, and what we have is this n times x bar minus mu naught prime S inverse x bar minus mu naught in a need form.

Now, this would further imply from the distribution of this T square statistic; that this T square divided by n in the previous setup was degrees a freedom. So, we will have that as n minus one; this into n now, the rule of n is played by n minus one.

So, will have that n minus 1 minus m plus 1 that divided by m the dimension; this would follow, and F distribution F prime distribution with degrees of freedom as m; now, what is this equal to this is just equal to n equal to n minus m, and a non-centrality parameter equal to tau square. Now, this non-centrality parameter tau square remember was associated with the mean of the associated multivariate normal distribution. So, this is a null vector here. So, delta prime inverse of this particular matrix into delta would just be equal to zero, because delta the mean vector under the null hypothesis.

So, tau square will be equal to 0 under the null hypothesis H naught. So, we will have actually a central chi square central f distribution as the null distribution. So, this would imply that, the T square that we had defined; that divided by n minus 1 times. We have that as n minus 1 minus m plus 1 that divided by m. This follows, a central F distribution on m n minus m degrees of freedom under the null hypothesis H naught.

(Refer Slide Time: 30:18)

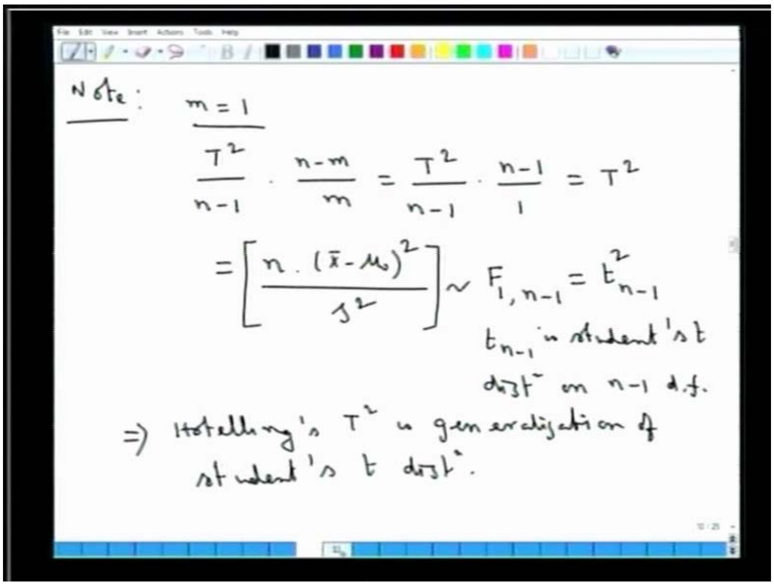


So, if we have that to be n central F distribution, we will use this distribution in order to reject or except the null hypothesis. So, the test for H naught against H A is 2 reject H naught. If T square is large, that is we will reject H naught at level alpha; if our T square by n minus 1 into n minus m by m. If observed value of this if observe value of this tested statistics exceeds F m n minus m; at the tabulated value, at alpha percentage point. So, this becomes finally, the testing procedure for testing the null hypothesis that, mu equal to mu naught against mu naught equal to be naught.

So, we look at the deviation x bar minus mu naught term here, and then this basically, so as the test statistic which is having a and F distribution on m n minus m degrees of freedom under the null hypothesis, and hence the test is that, if the observed value of this tested statistics exceeds, the upper alpha percent tabulated value of an F distribution on m n minus 1 degrees of freedom. We reject the null hypothesis, and except H naught otherwise.

So, that is the testing which we have an obtained; using the hoteling's T square distribution now this hoteling's T square distribution. I said to start with is the multivariate generalization of in variate students T distribution, and hence we will see shortly, how this is generalizing our a under now note that, if m is equal to 1 then we will basically, b having an invariate distribution, because the dimensionality of the multivariate underline multivariate distribution was m. So, if you take m equal to one; then this statistic reduces actually to students T statistic. Now, suppose we have got m equal to one, that is real looking at invariate distribution theory; suppose, we have got m equal to one; then, what happens m to this particular statistic out here, it is T square.

(Refer Slide Time: 33:16)



Now, what was T square? T square was given by this. So, this is equal to in such a situation given by T square by n minus 1 in to n minus m divided by m. So, for the value of m equal to one; this is equal to T square by n minus one. Then, this is also n minus 1 divided by one, and hence this is just equal to T square, and what is T square equal to T square for m equal to 1 would be let us, look that once again a to the form of T square it is n times this particular quantity here, now we are m equal to one. So, this x bar minus mu naught is a scalar quantity, and what is S inverse; if we have m equal to one; S is the sample variance, covariance matrix. If you have invariate distribution then this is just sample variance. So, what will be having as T square is n times x bar minus mu naught whole square that divided by S square; where S square is a sample variance.

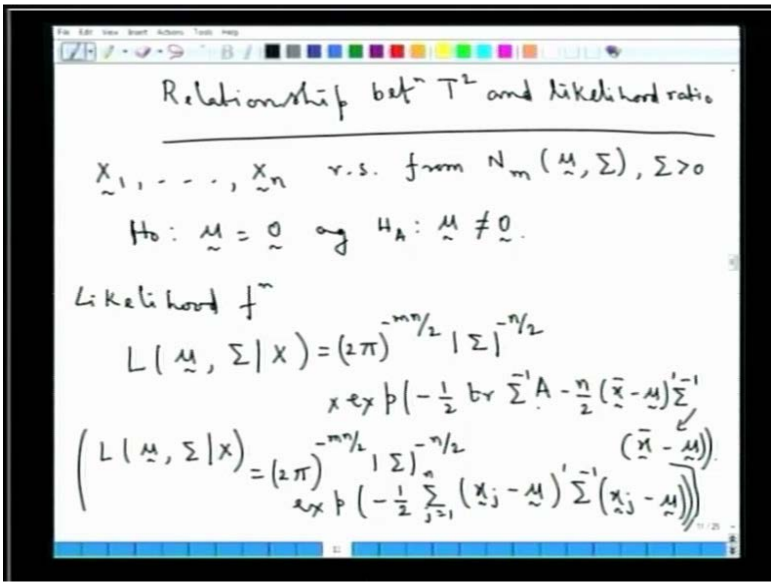
So, this would thus be given by n times x bar, minus mu naught, whole square that divided by S square; this will follow, and F distribution on one that is m n minus m is n minus 1 here, and this is equal to what this is a students T distribution square on n minus 1 degrees of freedom, right because and F distribution on 1 n degrees of freedom is equivalent to a T square on the same degrees of freedom.

So, we have (( )) this note that, what we have here which which is the random of that is square by n minus 1 into n minus equal to m by n for the special case; that, we have invariate theory. So, this is what we have? When this the square roots of this is precisely the students T statistics; that is, what is use? If you have x 1, x 2, x n, and invariate a random sample from invariate normal distribution with a mean equal to mu, and then unknown variance equal to sigma square; then the square root of this basically, or the square root that if you have n F statistics is what is precisely used in case of n invariate distribution theory. So, this t n minus 1 is students T distribution on n minus 1 degrees of freedom.

So, this implies does this hoteling's T square is generalization of this students T distribution in the multivariate setup. Now, the distinct procedure that, we had frame was for a general mu equal to mu naught against mu equal to mu naught; as a special case one can take this mu naught to be a null vector, and in such a situation that is T square testing would just n times x bar transpose s inverse x bar; an all other things remains exactly the same. So, for any specific choice of mu naught vector 1 can obtain this particular testing procedure.

We will extend this hoteling's T square distribution or other the testing procedure when we have a two sample, two multivariate normal population, but before that let me just look at the following important thing: Which is the relationship between this hoteling's T square, and a testing procedure which is based on the likely would ratio principle.

(Refer Slide Time: 37:16)



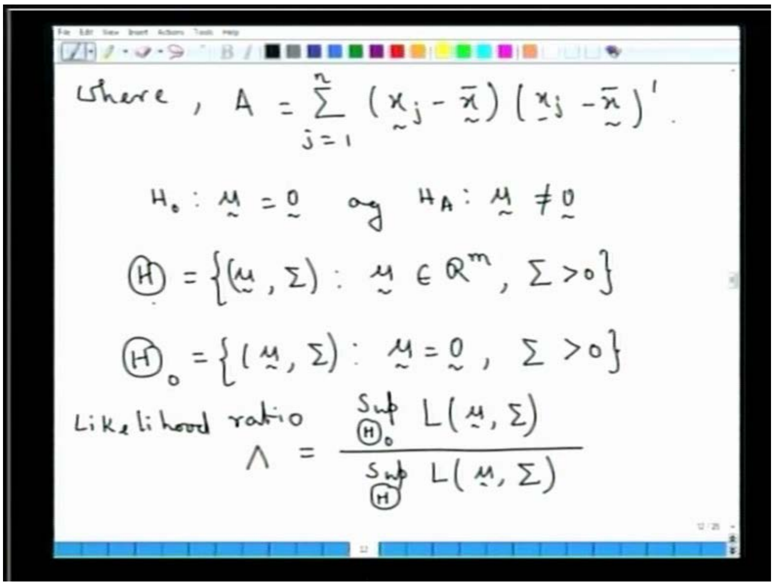
So, the relationship between T square the hoteling's T square, and likely hood ratio is: what is now the point of interest to us; now suppose, we have the same setup as that of the multivariate normal distribution, a random sampling. So, this is random sample from a multivariate normal m mu sigma; sigma is positive definite is unknown, and mu is also unknown. So, suppose we have got this null hypothesis mu equal to null vector. This is to be tested against the alternate hypothesis that mu is not equal to that null vector, and we can do just; now, we have seen that this type of testing we can carry forward using the hoteling's T square distribution.

Now, what we are now going to look at is what is the relationship of the T square statistic; that we had introduced with that of the general principle of the likely hood ratio; now, in order to implement the likely hood ratio what we would require is first the likely hood function. So, let us look at this likely hood function the likely hood function L mu sigma; this given I just write right x; x is the matrix which contains x 1, x 2, x n.

So, that is going to be given by as we have seen earlier 2 pi to the power minus m n by 2 determinant of sigma to the power minus n by two; Then E to the power minus half trace of sigma inverse A minus n by 2 x bar minus mu transpose sigma inverse; I do not have any please two write this here.

So, this is this x bar minus mu. So, that is what was the form of or the simplified form of this likely hood function. So, this is the constant part 2 pi to the power minus m n by two. Then determine sigma to the power minus n by two, and then this as we had seen perhaps would be better if, I right write the first form; then right write this simplified form as well this basically is coming from here. This exponent in the first form would be 1 mu sigma given x. This is going to be equal to the constant is as it is minus m n by 2 determinant of sigma to the power minus n by two. So, this is m n by 2 determinant of sigma to the power minus n by two, and then in the exponent what we have to start with is minus of summation T, equal to 1 to n x j minus mu transpose sigma inverse x j minus mu, and as we had seen when we are talking about sufficiency concept, and the related maximum likely hood estimation. We had seen that this particular likely hood function can be conveniently return in this particular form where A of course, s this matrix Where A is the summation of squares in the cross product matrix. That is A is summation j equal to 1 to up to n x j minus; this x bar into x j minus this x bar transpose this matrix right. So, if this is the likely hood function; then, the likely hood ratio A, before even introducing the likely hood ratio. Let me look at the spaces, this mu is the null hypothesis; mu equal to mu naught, this is to be tested against the alternate hypothesis; that this mu is not equal to mu naught..

(Refer Slide Time: 41:08)



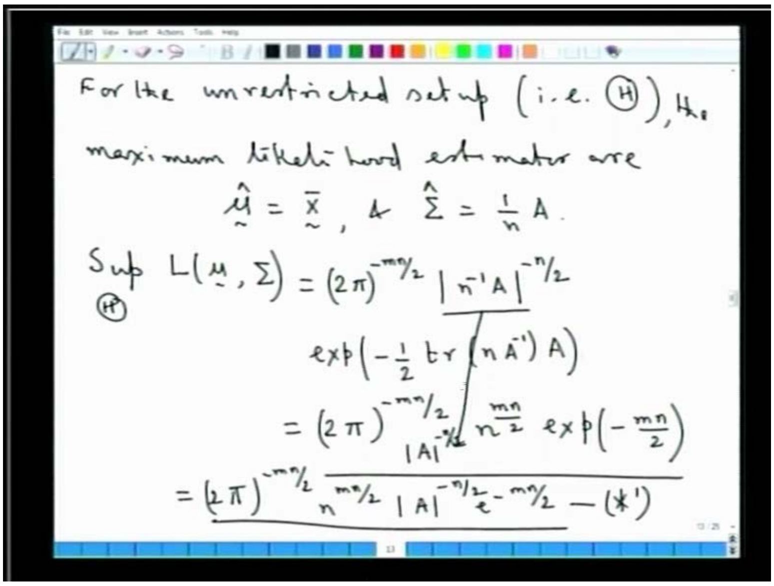
Now, let me denote by script theta the parameter space without any restriction. So, it is basically the setup all mu, and sigmas, such that mu belongs to R to the power m, because we have got that m dimension, and then the sigma matrix is all possible positive definite matrixes, and then we also look at this script theta naught; which is the parameter space undo the null hypothesis.

So, the parameter space script theta naught under the null hypothesis mu, H naught is going to be given by the setup all mu, and sigmas, such that this mu is fixed at the null vector point, and sigma is allowed to b any positive definite matrix. Now, under this two parameter spaces: the likely hood ratio is defined in the following way as usual likely hood ratio, this lambda is going to be given by the ratio of the 2 S supremums.

So, supremum under script theta naught of this likely hood function 1 mu sigma, I will just drop this condition drown x just keep it as 1 mu sigma. That divided by supremum over the entire parameters space script theta of this likely hood function. So, this is the likely hood ratio, that is what is going to be defined, and the likely hood ratio test would reject the null hypothesis H naught; in favor of the alternate hypothesis H A, if this likely hood ratio is small, because we are looking at the supremum of the theta naught by supremum of theta.

So, we will have to find out what is supremum under theta naught? What is the supremum under the unrestricted supremum? Under the parameter space script theta now, in order to do that we would require actually a what is the maximum likely hood estimates of the unknown parameters under the two setup.

(Refer Slide Time: 44:03)



For the unrestricted setup for the unrestricted setup that is, unrestricted is for script theta we know that, what the maximum likely hood estimators are the maximum likely hood estimators are in the following: maximum likely hood estimators are this mu hat is equal to x bar, and this sigma hat maximum likely hood estimator is 1 upon n times A; we had derived is 1 upon n times A. We had derived this particular result earlier; that these are the maximum likely wood estimators in case of a multivariate normal distribution.

Now, if we all looking at the supremum over script theta of this 1 mu sigma; then backed would be obtained, if we plug in this values as the corresponding maximum likely hood estimator, and then see what is the form of this quantity now. So, this would be the supremum of the likely hood function; this script theta would be this constant will remain as quantities, and then we will have determinant of sigma hat to the power minus n by two. So, what would that be equal to sigma hat is going to be replace by n inverse A to the power minus n by two.

So, that is what is doing that a task of this sigma hat; then E to the power minus half trace of sigma inverse A was the term, which we had in the likely hood function. It was this as trace of sigma inverse A. So, what would that be equal to sigma is n inverse A.

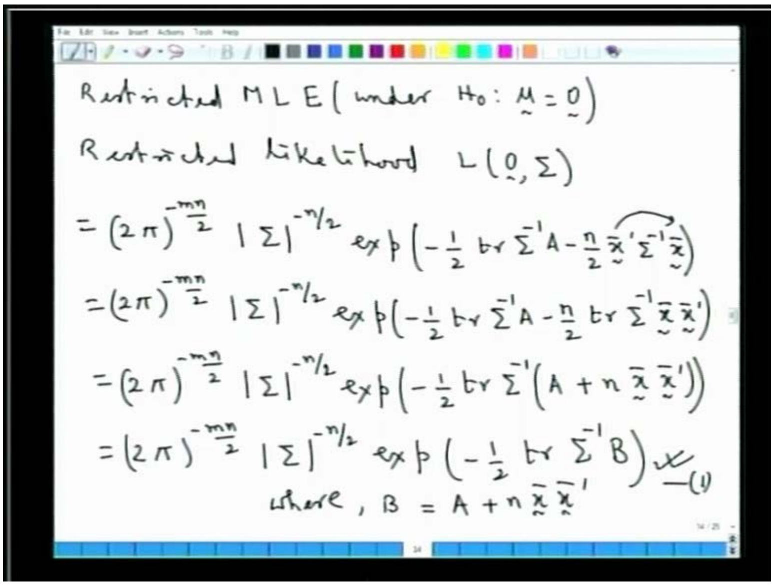
So, sigma hat inverse by invariance of the likely hood principle would be n times A inverse. So, this sigma n times A inverse; this is for sigma, and then we will have that A matrix coming as it is, and what happens to the second term? The second term is minus n by 2 x bar minus mu transpose sigma inverse x bar minus mu.

Now, at the point that mu is given the maximum likely hood point. So, that would be x bar. So, this term would be vanish simply. So, what we will be having is this term just for the supremize or other the supremum over script theta of this likely hood function; what is this going to be equal to A inverse; A is going to be identity matrix of order m. So, the trace of that would just be n m. So, that is this we can write as 2 pi to the power minus m n by two, and we can take this n inverse out from this particular determinant of n inverse A matrix, and what will happen to the power of n will get raise to the power m n by two, because we have negative sign here, we have a negative sign here and the order of this a matrixes m by n.

So, we will have this n raise to the power m n by two, and then we will have to the power minus of trace of n times i m. So, that would be n times m. So, that this is m n by two. So, this simple form is the supremum of the likely hood function at this script theta point. right.

Now, we have to obtain also the supremum of this likely hood function under the restriction that H naught mu equal to mu naught.

(Refer Slide Time: 47:49)



So, we will have to work out also, that restricted a M L E restricted under the setup that H naught mu equal to a null vector this of course, will be a the same a procedure would be followed; we have instead of this null vector any other A specify vector mu naught to be the null hypothesis point. Now, in order to find out the restricted M L E will have to look at the restricted likely hood the restricted likely hood at the restricted likely hood would be actually L naught sigma given x.

So, this is going to be equal to the constant to pi to the power minus m n by 2 remains as. It is then, we will have determinant of sigma right to the power minus n by 2; we will have E to the power minus half trace of that sigma inverse a trace of sigma inverse A; this A what I will do is? I will take this minus half a this trace of sigma inverse A.

Let me, first write it this trace of sigma inverse A; this minus n by two, and then x bar minus mu transpose sigma inverse. Now, under the restriction that H naught mu equal to mu naught, that mu is a null vector, and hence what will be having here is just x bar transpose sigma inverse x bar.

So, we can write this in this form that 2 pi to the power n minus m n by 2 determinant of sigma to the power minus n by two; then, we will have E to the power, if you take this minus half outside.

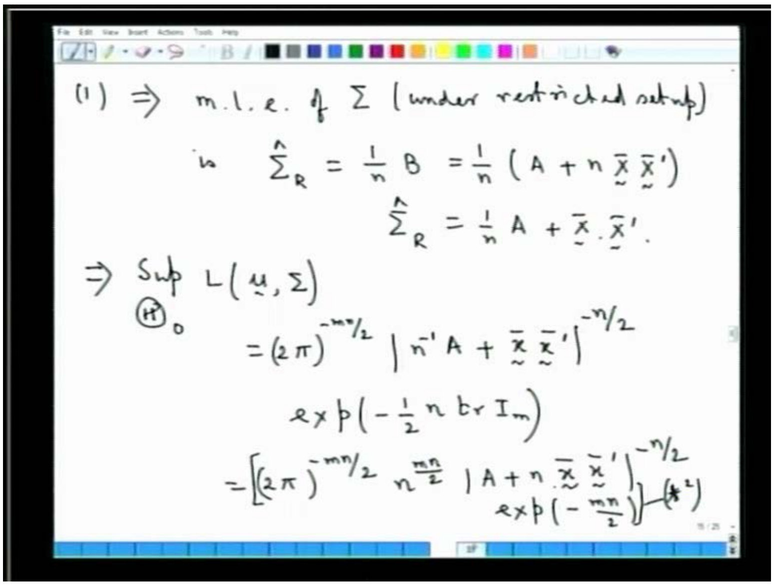
Now, note that we can put it trace out here. I will so, this calculation once more this trace of sigma inverse A minus n by two. Now, this is a scalar quantity, and hence this quantities itself; I can write trace of this term, and once I write trace of this x transpose sigma inverse x; then we can use the fact that trace A equal to trace of trace of a b equal to trace of b a, and then take this x bar on this side using the trace result, and then we will have this as trace of sigma inverse trace of sigma inverse x bar, x bar transpose.

So, we will have this term in a compact form as following: which would help us in order to get to the maximum likely hood estimator by using the same logic as what we had use in order to the maximum likely hood estimator for the unrestricted setup; this will be minus of trace of sigma inverse the entire term is common.

So, we will have that A plus; now this minus of trace sigma inverse is out now, then we will have n times x bar, x bar transpose as this particular matrix. We can write this as 2 pi; the power minus m by 2 determinant of sigma two; the power minus n by two, and E to the power minus half trace of sigma inverse of A matrix a which is A equal to B where B is nothing, but the matrix which is this one; where we have this B matrix to be given by A plus n times x bar, x bar transpose; now, note that this form. That we have here is similar in nature to the form which we had to encounter in order to find out the unrestricted maximum likely hood estimator in case of a multivariate normal distribution.

When we had unrestricted setup, then we had to find out the maximum of such a similar quantity with A replace by just this B A; I am sorry with this B just replace by this A. Now, the maximum likely hood estimator of sigma; in such a situation was 1 upon n A by using the same logic by analogy actually what we can say is that from this expression here, if I name this equation as one; then the maximum likely hood estimator of sigma in the restricted setup would just be given by 1 upon n B.

(Refer Slide Time: 52:47)



So, we will have this one implies using the same logic as that used in order to find out the maximum likely hood estimators estimator of sigma in the restricted setup; what we can says that the M L E of sigma under the restriction under restricted setup is going to be given by sigma hat are say restricted; that is equal to 1 upon n times B, and what is that equal to let us plug back what B was equal to.

So, it was equal to A plus n times x in the random variable from x bar, x bar transpose. So, that this is equal to just 1 upon n A plus, x bar, x bar transpose. So, this is the restricted maximum likely hood estimator, and this is the only maximum likely hood estimator; which we are going to optimizer? We are going to maximize other, because mu is given as mu naught. So, this would imply that is the other part which is require which is a supremum under script theta naught; under script theta naught mu is given as mu naught, and we have to find out the supremum of the likely hood L mu sigma at the point mu equal to mu naught.

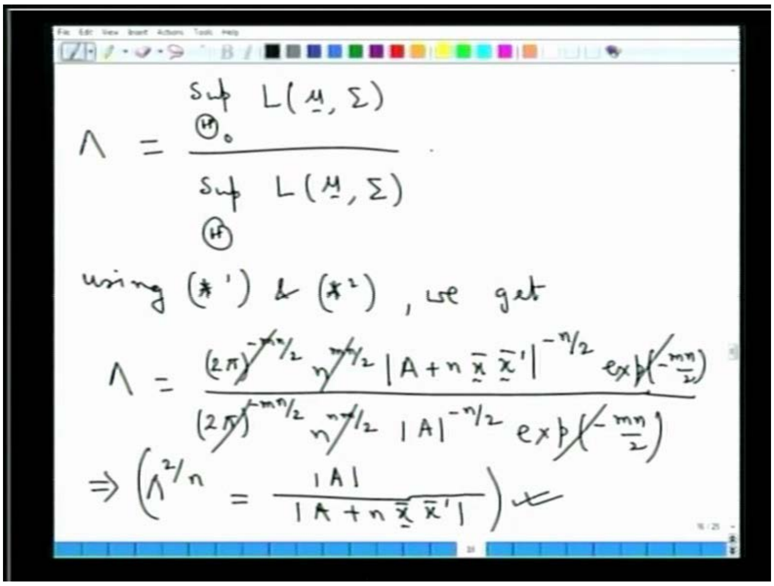
So, that we will just be using this in place of the previous matrix there. So, that this is thus equal to the constant remains as, it is p by to the power minus m n by 2 determinant of sigma, would be determinant of this quantity. Now, n inverse A, plus x bar, x bar transpose, whole raise to the power minus n by two, and then we will have E to the power minus half; let us, see what that was equal to.

So, this was the restricted likely hood. So, if we plug in here in place of sigma this quantity what will be having is this as 1 upon this sigma inverse as n times B inverse, and thus this would just, once again B equal to the type of term that, what we have this is going to n times trace of I n, and this thus would be equal to 2 pi to the power minus m n by 2 we can even take this n inverse outside.

So, we will have a term m here; n is basically, taken out in order to have a similar expression as to that what we had for the restrict a unrestricted maximized likely hood. So, this will be n raise to the power m n by 2, and that lives us with this A plus, n times x bar; now this x bar is going to be small x bar, because we have got to the data out here. We will have this return as small x bar, x bar transpose, and similarly, this would be the small x bar, because this is in the likely hood a function x bar, x bar transpose whole raise to the power minus n by two, and then we will have E to the power. This is going to give us m; m times m. So, this is minus m n by 2 right.

Now, we have both this parts which are required this is say star one; this sits in the denominator of the likely hood ratio. If this is given as star 1 number, then this is other expression; this entire expression is what we would require in the numerator of the likely hood ratio.

(Refer Slide Time: 56:59)



So, this would imply that this lambda; which was given by supremum under script theta naught of L mu sigma; this divided by the supremum, I think, I just write this in the center. So, this lambda, the likely hood ratio is given by this term here which is supremum over this.

So, this is the supremum under the unrestricted setup, and this sigma the supremum under the restricted setup. So, this thus using star one, and star two, we can straight away write what is this using star one; which was this term here, and star 2 equations, which was the restricted supremum? What we get is this form of lambda? Which is under the restriction its 2 pi to power minus m n by 2, n to the power m n by two. Then we have this determinant of what was it determinant of A plus n times x bar, x bar transpose whole raise to the power minus n by two.

So, A plus n times x bar, x bar transpose, whole raise to the power minus n by two, and then we had E to the power minus m n by two, and in the denominator; we have that star 1 expression. The star 1 expression remember had a similar constant; this constants are exactly the same, and if we look back at; what was star 1 in the star 1, I just missed out this term here, this term was sitting here. So, this term would come out here.

So, we will have this as determinant of A to the power minus n by 2 as well. So, this term is this this term comes in because this E to the power, this term is here. So, will have this as to pi to the power minus m n by two; then we have n to the power m n by 2 determinant of A to the power minus n by two, and E to the power minus m n by 2. So, that that is the correct expression of this star one, this term comes down here.

So, what will be having here is a determinant of A to the power minus n by 2, and then we will have p to the power minus m n by 2 terms here. So, lot of terms actually cancels out, this term cancel out, this term cancels out with this one, this term cancels out with this one.

So, what we have is this simple form that lambda to the power 2 pi n is going to be equal to lets flip this two terms, this will be determinant of A this divided by determinant of A plus n times x bar, x bar transpose. The determinant of this terms here in test this form. So, we will use this form, this is the likely hood ratio lambda raise to the power 2 by n; that is equal to determinant of A divided by determinant of A plus n times x bar x bar transpose. So, we will look at in the last… In the next lecture will start with this particular form, and see what is the correspondence of this likely hood ratio lambda, and that of the hotelling's T square distribution, thank you.