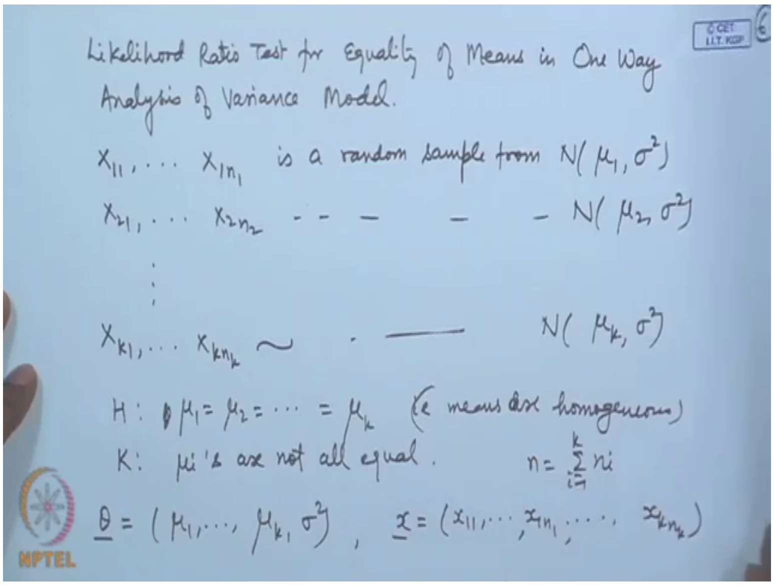
**Statistical Inference**
**Prof. Somesh Kumar**
**Department of Mathematics**
**Indian Institute of Technology, Kharagpur**

**Lecture – 58**
**Likelihood Ratio Tests – VIII**

In practice there can be even more complex situations. For example, I may have k different populations and I may like to check whether their means or whether their variances are equal. Now in this case suddenly the UMP test or UMP unbiased tests are very difficult to derive. In fact, we cannot write down the form of the joint density function in a in the form of multi parameter exponential family so that this parameter which is to be considered occurs there.

And therefore, the likelihood ratio test seems to be a good option, only thing is we should be able to derive the maximum likelihood estimators under omega as well as under omega h. So, I will give a couple of examples for applications when we are dealing with multiparameter situations and the number of populations may be more than two also.

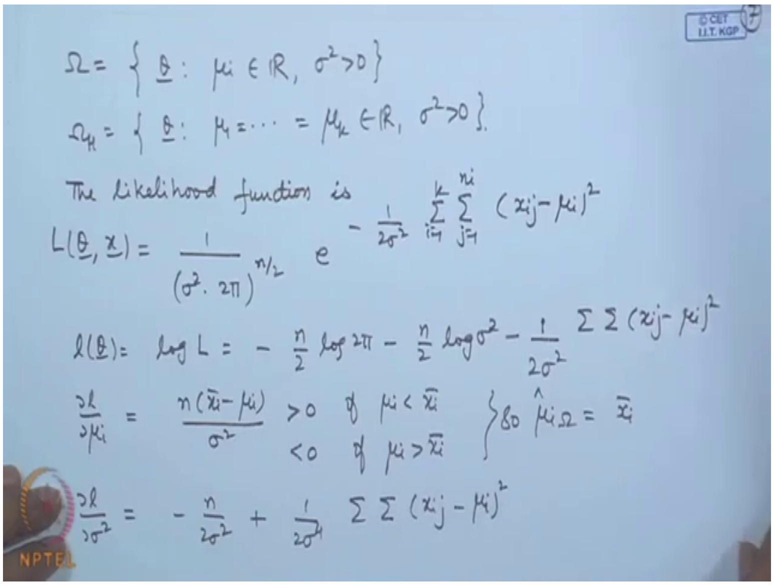
(Refer Slide Time: 01:19)




So, the first one I will do likelihood ratio test for equality of means in one way analysis of variance model. So, the set up is like this we are having X 11 and so on X 1 n 1 this is a random sample from say normal mu 1 sigma square. X 2 1 and so on X 2 n 2 this is a random sample from normal mu 2 sigma square and so on. X k 1 and so on, X k n k that is a random sample from normal mu k sigma square. Note here I have taken the variances to be common. So, this is actually the situation of a one way analysis of variance model and we are considering the variances to be the same. Our testing problem is to test that whether the means are the same or not this is called homogeneity of that is means of are homogeneous; that means, the populations are homogeneous basically.

If mu is are the same then basically it means that we have same population against that mu i's are not all equal. Now let me introduce some notation theta is now mu 1 mu 2 mu k sigma square. So, this k plus 1 dimensional I will use the notation x for x 1 1 and so on x 1 n 1 and so on, x k n k all the n 1 plus n 2 plus n k observations that is sigma n i am calling a n these observations I call as x.

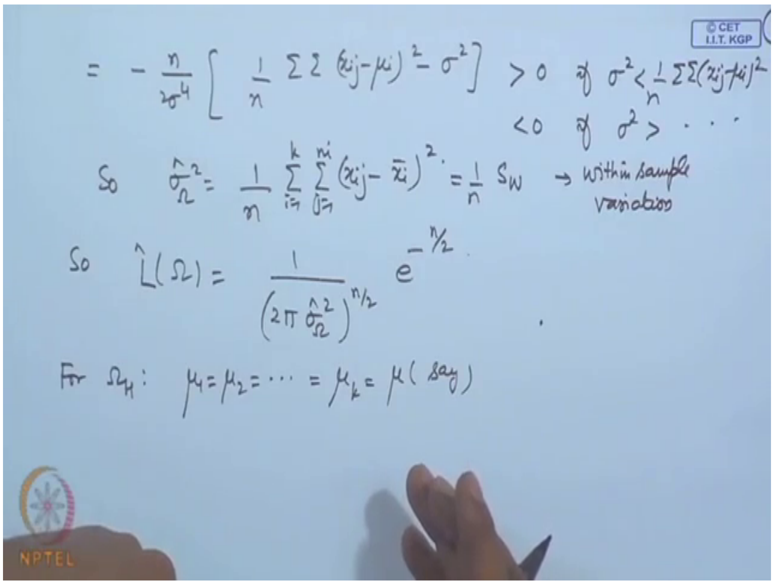
(Refer Slide Time: 03:52)




The parameter is phase, the full parameter is phase if we look at then the mu i’s belong to R that is k dimension Euclidean space in to positive half of the real line and under omega H you are dealing with 2-dimensional. Now the likelihood function is L theta x that is equal to 1 by sigma square 2 pi to the power n by 2 e to the power minus 1 by 2 sigma square sigma x ij minus mu i square j is equal to 1 to n i is equal to 1 to k.

So, we take the log likelihood here that is minus n by 2 log 2 pi minus n by 2 log of sigma square minus 1 by 2 sigma square sigma x ij minus mu i square. Now easily you can see that if I consider the derivative with respect to mu i then I will get n x i bar minus mu i divided by sigma square. So, this is greater than 0 for mu i less than x i bar and it is less than 0 if mu i is greater than x i bar. So, you get mu i omega hat is equal to x i bar because the maximum will occur at x i bar. Now we can see that the derivative with respect to sigma square and I get minus n by 2 sigma square plus 1 by 2 sigma to the power 4 double summation x ij minus mu i square which I can write as minus n by 2 sigma to the power 4 1 by n double summation x ij minus mu i square minus sigma square.

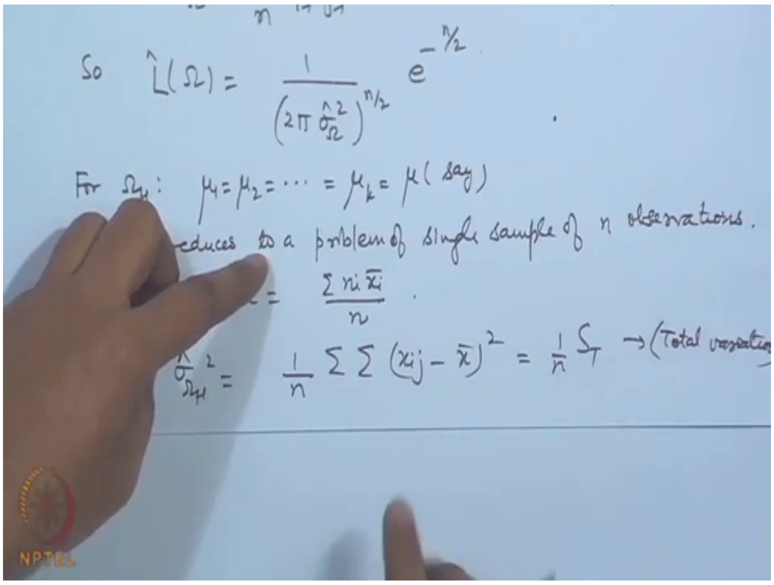
(Refer Slide Time: 06:20)




So, once again you can see it is greater than 0 if sigma square is less than 1 by n double summation x ij minus mu i square and it is less than 0 if sigma square is greater than this. So; obviously, the maximization is occurring at this point and since mu is estimated to be x i bar. So, we get sigma omega hat square that is equal to 1 by n double summation x ij minus x i bar square. I give a notation to this, this is equal to S W that is 1 by n within sample variation because sigma x ij minus x i bar whole square denotes the variation within the i x sample and then I am taking sum over all such.

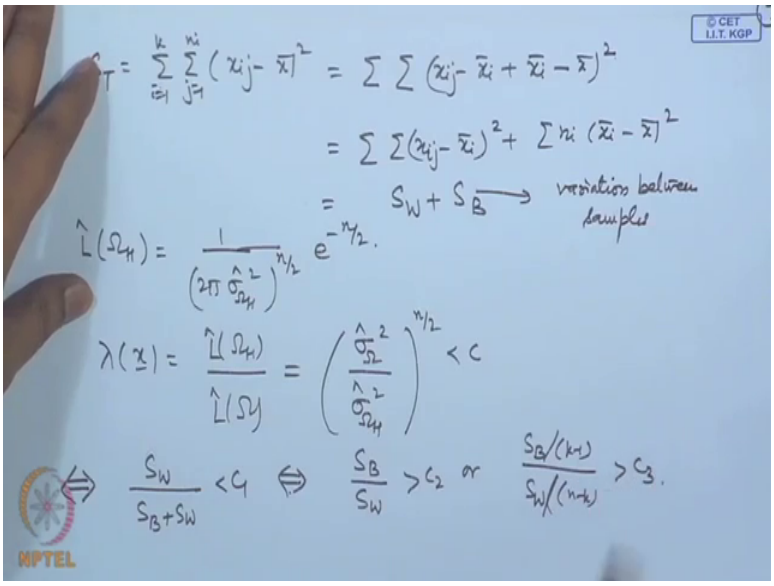
So, this is the total variation within each sample I have considered here. So, if I consider L hat omega that is by substituting the values of estimated values of sigma square and mu i I get 1 by 2 pi sigma omega hat square to the power n by 2 e to the power minus n by 2. Now for omega H you are having mu 1 is equal to mu 2 is equal to mu k is equal to say mu.

(Refer Slide Time: 08:27)




So, this actually becomes a problem of single sample, this reduces to a problem of single sample of n observations. So, mu hat omega H will become simply x bar that is sigma n i x i bar divided by n and sigma omega H hat square that will become equal to 1 by n sigma x ij minus x bar square, which I call as 1 by n S T that is the total variation ok. Because I am considering the difference of each unit from the grand mean and then I am taking the sum of the squares here this is the total variation.

(Refer Slide Time: 09:37)



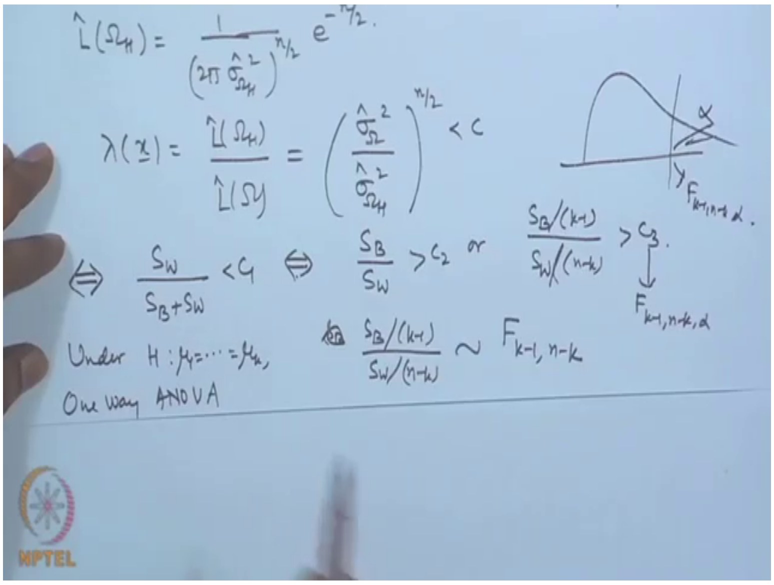
So, this purpose of writing this S W and S T is to explain the type of terms that we are getting and this S T you can actually write as double summation x ij minus x bar square, here you add and subtract x i bar. So, this gives me x ij minus x i bar plus x i bar minus x bar whole square. So, this becomes sigma x ij minus x i bar square plus sigma n i x i bar minus x bar square.

So, this is equal to S W plus S B, now this S B this term is actually variation between samples because we are considering x i bar that is i x sample mean and x bar is the grand mean. So, this is nothing, but the and then I am taking square and then taking all such cases. So, this is actually how much variation is there between the different samples.

So, now, we can utilize this and write down the form of L hat omega H. So, that turns out to be simply 1 by 2 pi to the power n by 2 e to the power minus n by 2. So, let us look at these two terms here that we are getting now, L hat omega here e to the power minus n by 2 and here sigma omega hat square to the power n by 2, where sigma omega hat is the S W by n term.

And in L hat omega H I get the same thing, here only omega H is replaced and the value of omega H that we calculated as S T by n and this S T I wrote again as S W plus S B. Therefore, the form of the likelihood ratio test that is lambda x that is L hat omega H by L hat omega that becomes equal to sigma omega hat square by sigma omega H hat square to the power n by 2 less than C which is equivalent to now by because of writing down this equation sigma S W divided by S B plus S W less than C 1. Then you take the reciprocal and subtract 1, so we get greater than C 2 or we can write S B divided by k minus 1 divided by S W divided by n minus k greater than say some C 3.

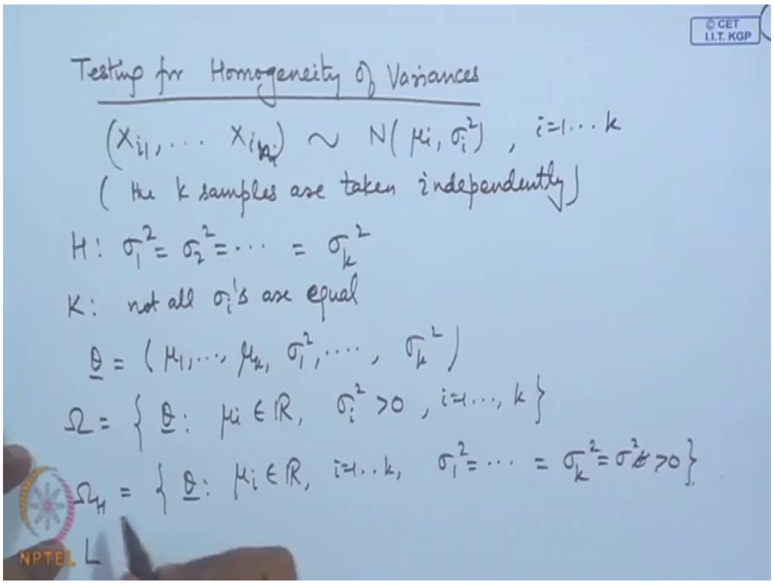
(Refer Slide Time: 12:43)




Now, under H that is mu 1 is equal to mu 2 is equal to mu k this S B that is S B by k minus 1 divided by S W by n minus k, this follows f distribution on k minus 1 n minus k degrees of freedom. So, this point C then is nothing, but a point on the curve of the density of F k minus 1 n minus k distribution that is the upper handed alpha percent point this probability is alpha. So, then this is F of k minus 1 n minus k alpha.

So, this is a usual test which is there in the one way analysis of variance, for testing the homogeneity of the means. I will repeat this problem, I made the assumption here that the variances are common. But when we are discussing general sampling from various populations then many times this assumption needs to be checked; that means, we are not sure whether variances are of same, if the variances are to be checked then we can derive a likelihood ratio test for this also. Let me just give a brief sketch of this likelihood ratio test here.

(Refer Slide Time: 14:26)

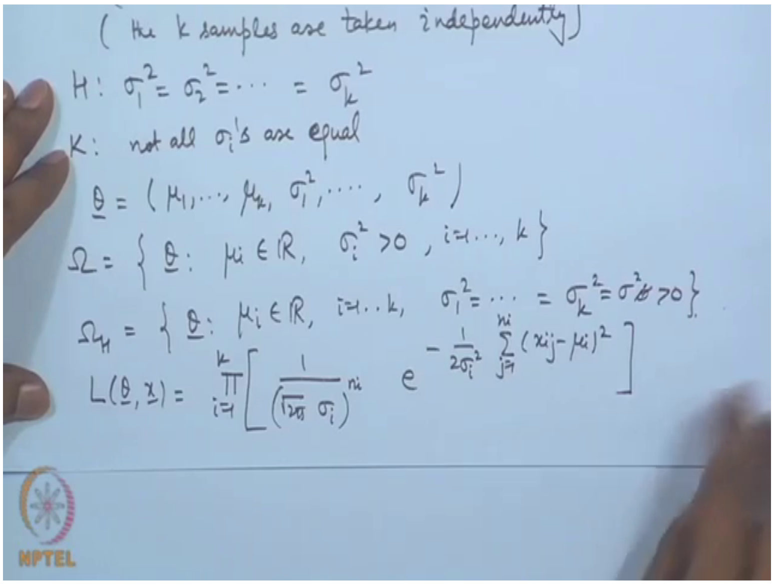



Let me consider testing for homogeneity of variances. So, you have seen actually this likelihood ratio test is applicable to very very general situations, I considered a k population model and I could actually derive an exact test. Now here let us consider the sampling X i 1 X i 2 X i k X i n i, this is from normal mu i sigma i square for i is equal to 1 to k and the k samples are taken independently.

So, our hypothesis is sigma 1 square is equal to sigma 2 square is equal to sigma k square against not all sigma i squares are equal our parameter phase has become little bit larger, it is actually 2 k dimensional now. The parameter theta is actually mu 1 mu 2 mu k sigma 1 square sigma 2 square sigma k square.

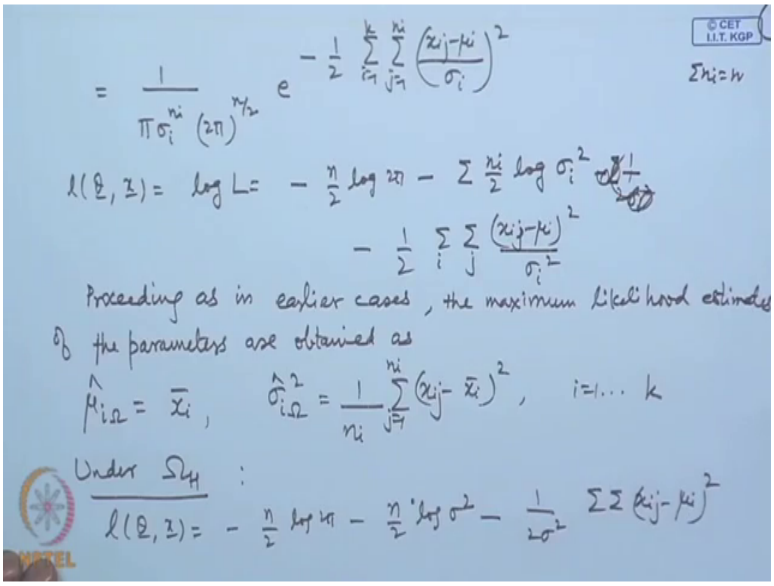
The full parameter space is that this mu i s are real numbers and sigma i squares are the positive real numbers and under omega H it becomes k plus 1 dimensional because mu i is remain as such, but sigma 1 squares become equal. So, the dimension of this part has reduced to 1.

(Refer Slide Time: 16:50)




So, this is k; so k plus 1 dimensional. Now as before I will not do the detailed calculations here. The likelihood function that is equal to since I have to write for the k different populations I am writing it in this form product i is equal to 1 to k 1 by root 2 pi sigma i to the power n i e to the power minus 1 by 2 sigma i square sigma x ij minus mu i square j is equal to 1 to n i.

(Refer Slide Time: 17:28)

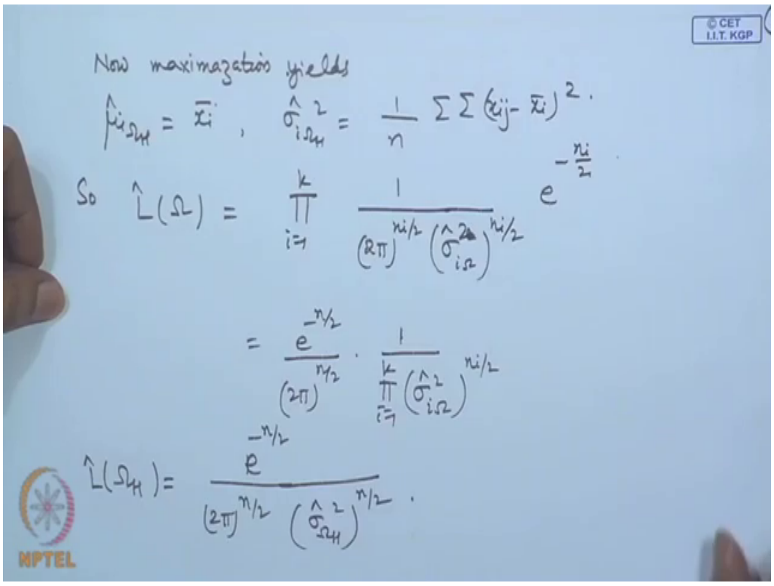


We can simplify this that is equal to 1 by product sigma i to the power n i 2 pi to the power n by 2 where sigma n i is equal to n then e to the power minus half sigma x ij minus mu i by sigma i square I is equal to 1 to k and j is equal to 1 to n i.

So, we take the log likelihood function that is equal to minus n by 2 log of 2 pi minus sigma n i by 2 log of sigma i square minus 1 by 2 sigma i square sigma ok. Let me write it as minus half double summation x ij minus mu i square by sigma i square. So, we can easily see that if I consider the usual maximization with respect to mu i s it will occur at x i bars and if I do with respect to sigma i square I will get it at 1 by n i sigma x ij minus x bar square.

So, proceeding as in earlier cases the maximum likelihood estimates of the parameters mu i omega hat is equal to x i bar and sigma i omega hat square that is equal to 1 by n i sigma x ij minus x i bar square j is equal to 1 to n i i is equal to 1 to k. Under omega H that is when we are taking sigma 1 square is equal to sigma 2 square is equal to sigma square then this l gets modified l will become minus n by 2 log of 2 pi minus n by 2 log of sigma square minus 1 by 2 sigma, square double summation x ij minus mu i square.

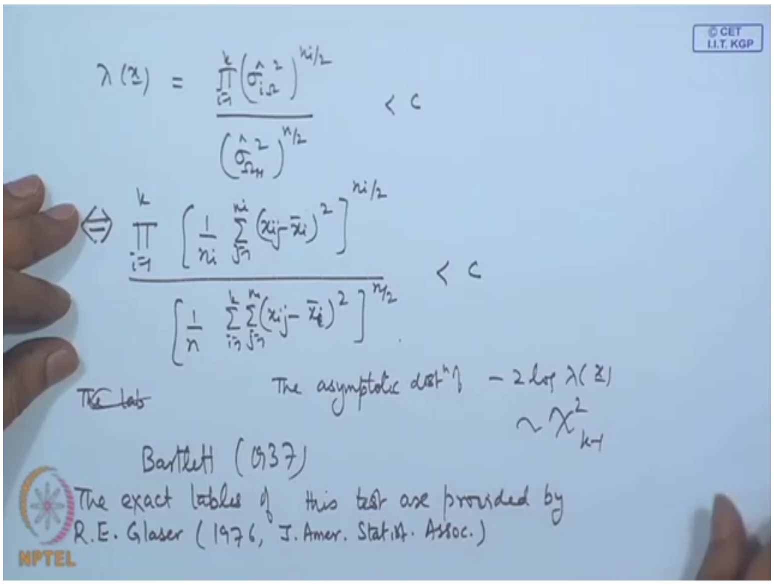
(Refer Slide Time: 20:39)




So, now if we consider the maximization, maximization yields mu i omega H hat as before that is x i bar, but the value of sigma i omega H hat square that becomes 1 by n double summation x ij minus x i bar square. So, L hat omega and L hat omega H can be calculated, I will write down the simplified expressions here that is product I is equal to 1 to k 1 by 2 pi to the power n i by 2 sigma i omega hat square ok, square part you can remove to the power n i.

And then we get if I write the square here then n i by 1 I can write e to the power minus n i by 1 that is equal to e to the power minus n by 2 divided by 2 pi to the power n by 2, 1 by product sigma i omega hat square to the power n i by 2 i is equal to 1 to k. And L hat omega H that is equal to 1 by 2 pi to the power n by 2 e to the power minus n by 2 and sigma omega H hat square to the power n by 2.

(Refer Slide Time: 22:39)

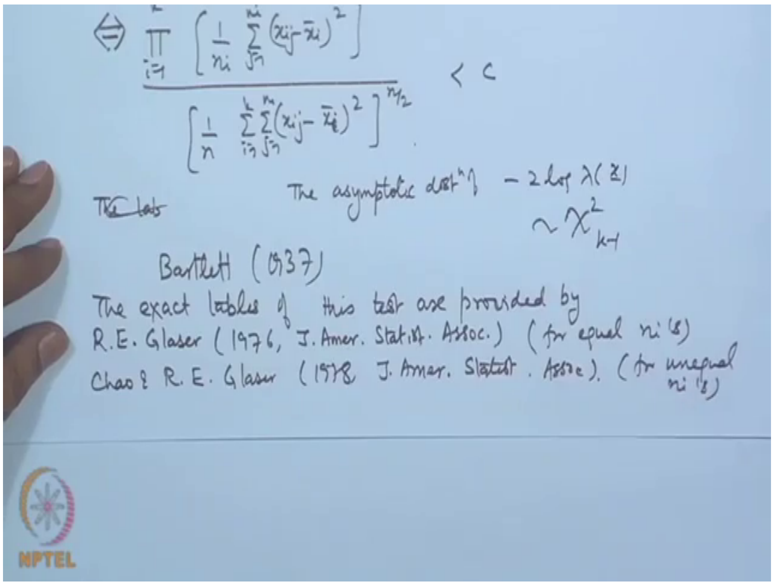



So, lambda x that is equal to sigma i omega hat square product i is equal to 1 to k to the power n i by 2 divided by sigma omega H hat square to the power n by 2. So, this less than C, this is equivalent to saying that now this sigma i omega hat that we have already written the expression that is 1 by n i sigma x ij minus x i bar square to the power n i by 2 product i is equal to 1 to k divided by 1 by n double summation x ij minus x bar x i bar square. This is j is equal to 1 to n i j is equal to 1 to n i i is equal to 1 to k whole to the power n by 2 less than C.

So, basically it is in the terms of the sums of the squares and this is as we have already defined it is the within sample variance for each sample and this is the total variance sample thing. So, this is coming in terms of the sum of that earlier we got 2-dimensional sum that is S B plus S W thing here I am getting k dimensional terms here, the tables actually the exact distribution of this is not so simple; however, it has been derived. But if we use the asymptotic distribution that will become chi square on how many degrees of freedom 2 k minus k plus 1.

So, if I consider the asymptotic distribution here of minus 2 log of lambda x that is chi square on k minus 1 degrees of freedom. So, this was provided by Bartlett in 1937 and the exact tables of this test are provided by R E Glaser in 1976 in Journal of American Statistical Association.

(Refer Slide Time: 25:45)



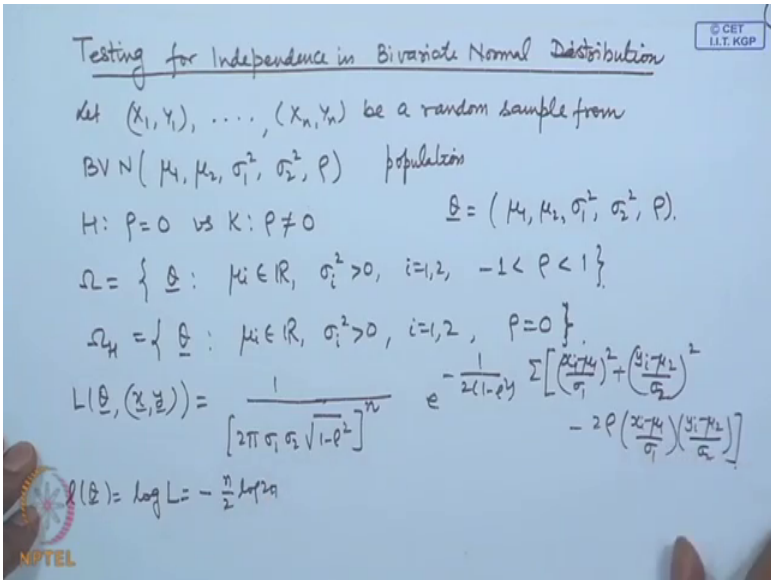

This part he considered for equal sample sizes and for unequal sample sizes Chao and R E Glaser in 1978 Journal of American Statistical Association, this is for unequal n is, this test is shown to be unbiased test. So, there are desirable properties of this test that are known. So, if we use this test to check the equality of the variances and if the equality is there then we can apply the one way analysis of variance test for mu 1 is equal to mu 2 is equal to mu k. And so what I have demonstrated here that the likelihood ratio test gives the solutions here.

Whereas, we cannot directly apply any result for the UMP or UMP unbiased test theory here, those things are not applicable here. Let me take up another problem here which is related. Earlier I have explained one problem testing for independence in a contingency table. Now in a contingency table you are having two attributes, but when we have quantitative data then how to test the independence. So, the simplest model that we can think of is a bivariate normal population.

Now we know that in a bivariate normal population you have five parameters; that means, the two means the two variances and then there is a correlation coefficient rho. And it has been it is known that the independence is equivalent to rho being equal to 0 in the bivariate normal population independence condition and the uncoded conditions are the same.

So, if we test rho is equal to 0 in a bivariate normal setting then it becomes the test for independence. Let me show that a likelihood ratio test can be derived here.

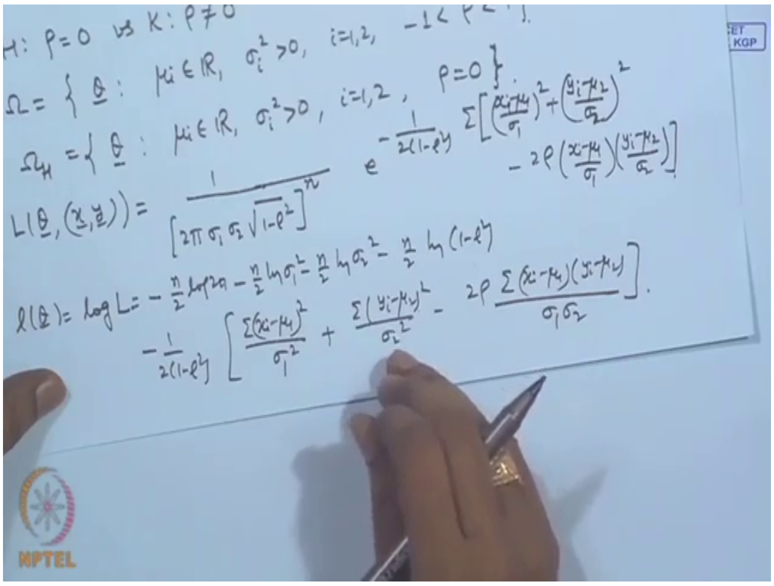
(Refer Slide Time: 27:57)




So, this will be my last example in the likelihood ratio test and then we will proceed to another theory here. So, testing for independence in bivariate normal distributions. So, let X 1 Y 1, X 2 Y 2, X n Y n be a random sample from bivariate normal mu 1, mu 2, sigma 1 square, sigma 2 square, rho population. And we want to test the hypothesis whether rho is equal to 0 against rho is not equal to 0, this is the condition for the independence our parameter here is the 5-dimensional mu 1, mu 2, sigma 1 square, sigma 2 square, rho. The full parameter space here is mu is are real sigma i squares are positive and rho lies between minus 1 and 1.

Under the null hypothesis the dimension becomes one less, mu i's and sigma i squares remain the same; however, rho becomes 0. So, the likelihood function here that is equal to 1 by 2 pi sigma 1 sigma 2 root 1 minus rho square to the power n e to the power minus 1 by 2 1 minus rho square sigma x i minus mu 1 by sigma 1 square plus y I minus mu 2 by sigma 2 square minus 2 rho x i minus mu 1 by sigma 1 y i minus mu 2 by sigma 2. Now you take the log likelihood here.

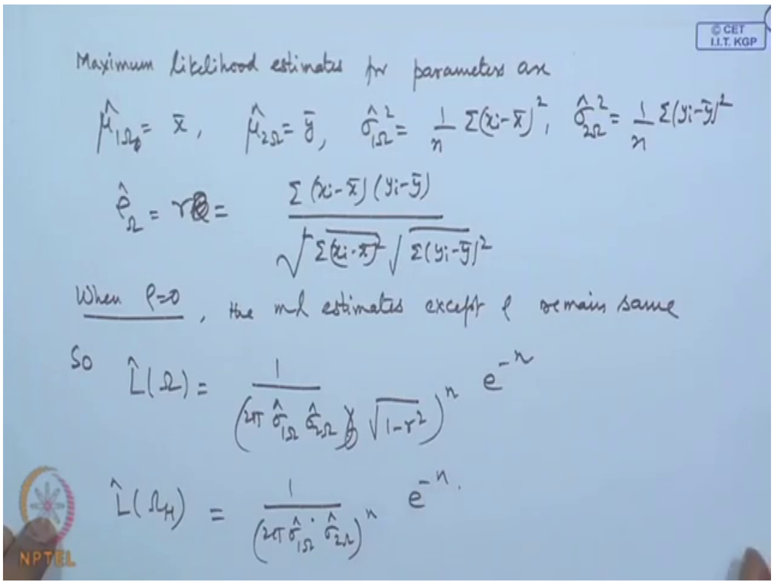
(Refer Slide Time: 30:50)




And that is equal to minus n by 2 log of 2 pi minus n by 2 log of sigma 1 square minus n by 2 log of sigma 2 square minus n by 2 log of 1 minus rho square and then this expression that is minus 1 by twice 1 minus rho square sigma x i minus mu 1 square by sigma 1 square plus sigma y i minus mu 2 square by sigma 2 square minus twice rho sigma x i minus mu 1 in to y I minus mu 2 divided by sigma 1 sigma 2.

Now, if we proceed in the usual fashion for the maximization with respect to the parameters for mu i's I will get the sample means and for the sigma i squares we will get the sample variances and for rho I will get the sample correlation coefficient.

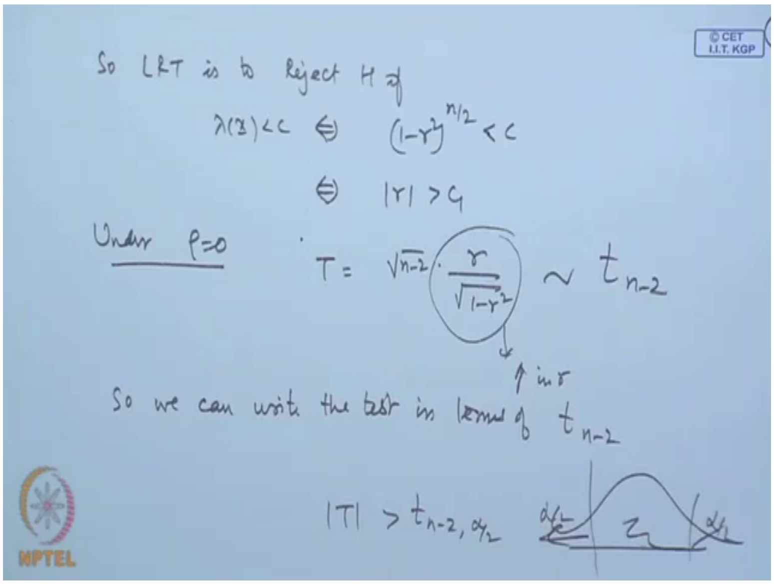
(Refer Slide Time: 31:57)




So, without getting in to too much of the derivations let me write the final answers here. So, the maximum likelihood estimates for parameters are. So, mu 1 omega hat that is equal to x bar, mu 2 omega hat is equal to y bar sigma 1 omega hat square is equal to 1 by n sigma x i minus x bar whole square, sigma 1 omega hat square is equal to 1 by n sigma y i minus y i bar square. And rho hat omega that is equal to R that is equal to sigma x i minus x bar, let me write a small r here y i minus y bar divided by square root sigma x i minus x bar square and sigma y i minus y bar square.

Now, when rho is equal to 0 then the m l estimates except rho remain same. So, we get then L hat omega that is equal to 1 by 1 by 2 pi sigma H hat omega sigma 2 hat omega to the power into square root 1 minus r square whole to the power n e to the power minus n and in L hat omega H except this term all other things will be same. So, it will become equal to 1 by 2 pi sigma 1 omega hat sigma 2 omega hat to the power n e to the power minus n.

(Refer Slide Time: 34:02)




So, the test is simply reducing to so the likelihood ratio test is to reject H if lambda x is less than c which is equivalent to saying one minus r square to the power n by 2 is less than C, which is also equivalent to saying modulus r is greater than C 1. Now under rho is equal to 0 that is under H, the distribution of t that is root n minus 2 r by root 1 minus r square this is t on n minus two degrees of freedom and this term is increasing in r. So, we can write the test in terms of t n minus 2.

So, basically what we can say we reject when modulus T is greater than t n minus 2 alpha by 2 because t distribution is symmetric. So, we can consider this alpha by 2 and this probability has alpha by 2 here this is the acceptance region. So, you can see here also we are able to nicely get the likelihood ratio test for testing for the independence in a bivariate normal population. I have discussed in detail the large sample test this likelihood ratio test which is having a large sample optimality property that is the asymptotic distribution is chi square.

And I have derived the exact distributions for testing problems in the normal populations and also some examples in binominal or exponential distributions also have been worked out. In the next lecture I will consider another concept that is of invariance in the testing problems.