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

Lecture – 51
Likelihood Ratio Tests- I

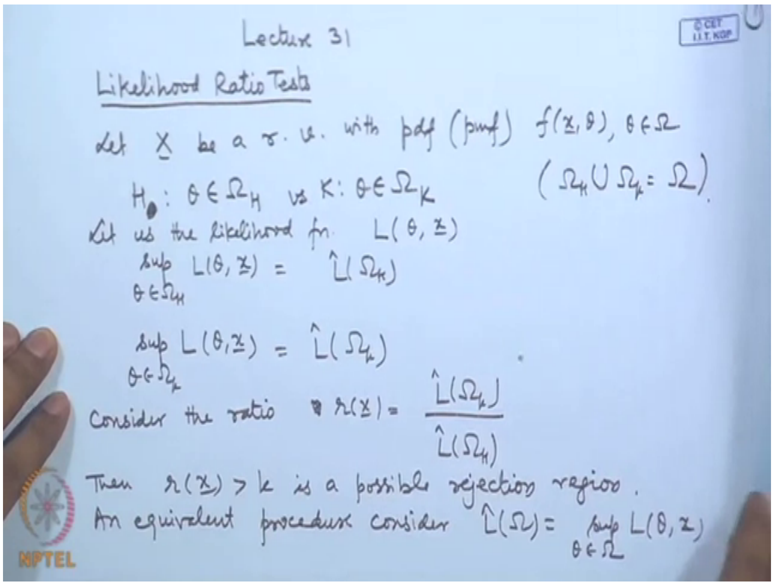
In the last few lectures, I have developed the theory of most powerful uniformly most powerful; uniformly most powerful unbiased tests. So, the basic building block of these tests was the Neyman and Pearson fundamental Lemma whose philosophy was that we fix level of significance, for a fixed level of significance you derive the most powerful uniformly most powerful or uniformly most powerful unbiased tests.

However, we have seen that in a wide variety of problems these type of tests are not applicable. For example, when we want to derive UMP tests then we are imposing a condition on the family of distributions that they should have monotone likelihood ratio property. Now there are large number of distributions which may not have monotone Likelihood Ratio property. Further when we develop that theory of unbiased UMP tests for hypothesis of that where the null hypothesis could be 1 sided or point null hypothesis or an interval, but the alternative hypothesis was 2 sided, in that case the UMP test was not available.

In fact, UMP unbiased test was derived, but it was for a 1 parameter exponential family. Later on when we have developed the general theory of UMP unbiased tests, we have considered multi parameter exponential families only and that too we should have the complete sufficient exact statistic. Now there can be many practical applications where these conditions will not be satisfied and therefore, we need certain other method for deriving the tests.

Now, as we have seen in the estimation problem one can restrict attention to the joint distribution which we term as the likelihood function. In the estimation problems we considered that value of the parameter to be the maximum likelihood estimator which maximize that likelihood function. Therefore, this maximization of the likelihood function is a heuristic thing; that means, on your own as a layman we think that we should consider the probable values are the most probable value for the likelihood function. Now using this philosophy likelihood ratio tests are derived and let me introduce the theory of likelihood ratio test.

(Refer Slide Time: 03:11)



So, let X be a random vector with probability density function or probability mass function say f x theta; theta belonging to omega. We want to test say H naught theta so let me call H K, we have introduced the notation H for the null hypothesis and K for the alternative hypothesis when we were developing the UMP test, we had considered 4 types of hypothesis H 1 K 1, H 2 K 2, H 3 K 3 and H 4 K 4.

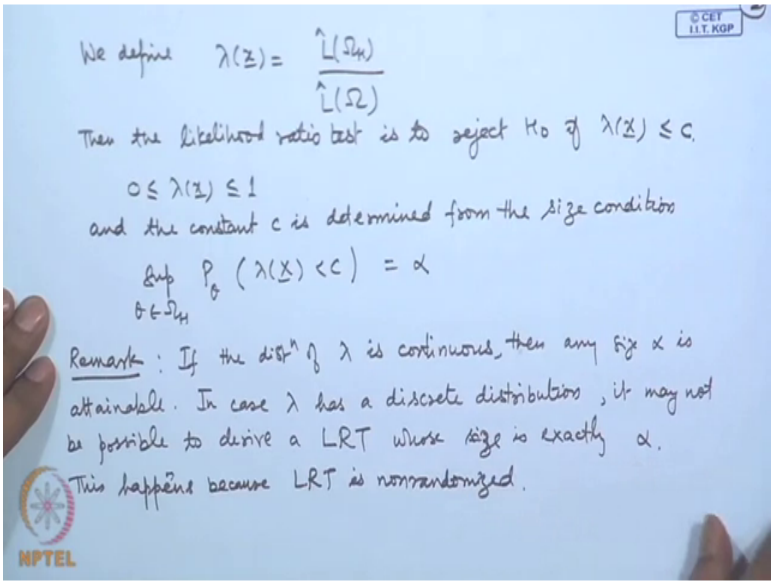
So, in general I will consider theta belonging to omega H versus theta belonging to omega k as the alternative hypothesis where omega H union omega k is equal to the parameter space. Now, we may consider say maximum of the let us consider the likelihood function. So, we call it L theta x that is nothing, but the joint distribution of the random variables under consideration, we consider the maximum value or the sup of L theta x, let me call it L hat omega H.

And we also consider the maximum of the likelihood function under omega k then a very natural procedure is to consider the ratio, let me call it r x that is equal to L hat omega k divided by L hat omega H. So, then r x greater than k is a possible rejection region, what is the criteria for this?

Because if the alternative hypothesis is more likely to happen, then L hat omega k will have a higher value and if null hypothesis is more likely to happen, then L hat omega H will be higher value or bigger value. Therefore, the rejection region should be for the larger values of this ratio, acceptance region for lower values.

Now, this requires 2 maximizations and at least one of the maximizations could be more complicated we have seen the problems like 1 sided or 2 sided hypothesis testing problems. So, overall consideration of L hat omega H and L hat omega k may be slightly difficult and equivalent procedure considers say L hat omega that is the maximum of the likelihood function over the full parameter space.

(Refer Slide Time: 06:51)

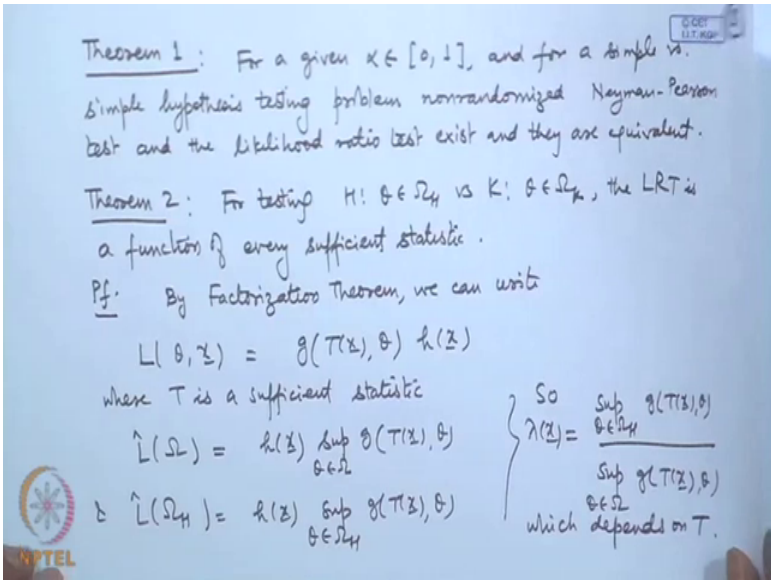


And we define the ratio lambda x is equal to L hat omega H by L hat omega. Then the likelihood ratio test is to reject H naught if lambda x is less than or equal to c and we will determine and of course, see this is maximization over the full parameter space. So, naturally you will have this thing between u and 1 and the constant c is determined from the size condition supremum of the probability of lambda x less than c theta belonging to omega H is equal to alpha.

Now as a remark let me say that, if the distribution of lambda is continuous, then any size alpha is attainable in case x lambda has a discrete distribution, it may not be possible to derive a likelihood ratio test whose size is exactly alpha.

However, this is happening because likelihood ratio test the way I am defining it is actually a non randomized test. This happens in the Nyman Pearson theory we were allowing the randomization, but in the likelihood ratio test randomization is not there. So, this happens because likelihood ratio test is non randomized.

(Refer Slide Time: 10:05)



We have the following equivalence result, let me call it theorem 1; for a given alpha belonging to 0 to 1 and for a simple versus, simple hypothesis distinct problem non randomized Nyman Pearson test and the likelihood ratio test exist and they are equivalent.

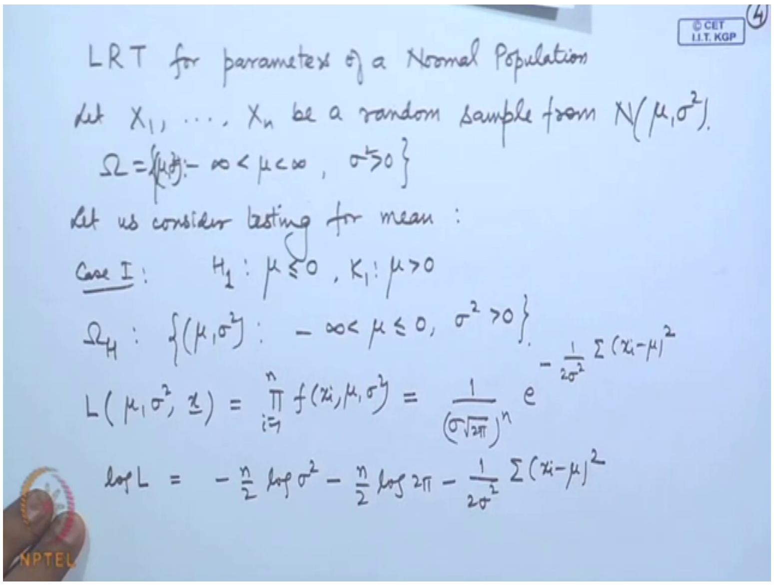
Because in the Nyman Pearson theory if you look at the simple verses simple you had f 1 by f naught which is nothing, but the likelihood ratio corresponding to the null and alternative hypotheses as f naught and f 1. We have another result as we have seen in the maximum likelihood estimation because of the factorization theorem the likelihood function can be written as a product of a function which is free from the parameter and another term which is involved in the parameter.

So, when we take the ratio the term which is not having the parameter becomes gets cancelled out and therefore, you are getting only the sufficient statistic. Therefore, for testing say H theta belonging to omega H versus K theta belonging to say omega k, the likelihood ratio test is a function of every sufficient statistic. Proof is very simple by factorization theorem, we can write the likelihood function as g T x theta into h x, where

T is a sufficient statistic. So, L hat omega that will be equal to h x into supremum of g T x theta where theta belongs to omega and L hat omega H that is equal to hx into supremum of g T x theta; theta belonging to omega H.

So, if I consider lambda x that will be simply supremum of theta belonging to omega H g of T x theta divided by supremum of theta belonging to omega g of T x theta, which depends on T. Therefore, the likelihood ratio test it will depend only on the sufficient statistic. Now let me derive the likelihood ratio test for various problems for which we have derived the Nyman Pearson tests that is the UMP unbiased test etcetera. For similar problems let me derive the likelihood ratio test. So, let me start with say normal distributions.

(Refer Slide Time: 14:34)



So, likelihood ratio tests for parameters of a normal population. So, we have let X 1, X 2, X n be a random sample, from say normal mu sigma square distribution. So, as before the full parameter space omega it is a set of all mu sigma, such that mu is from minus infinity to infinity and sigma is positive, if I say sigma square then I can write like that. Let me consider; let us consider testing for mean, note here that I am considering here full model; that means, both mu and sigma square are unknown.

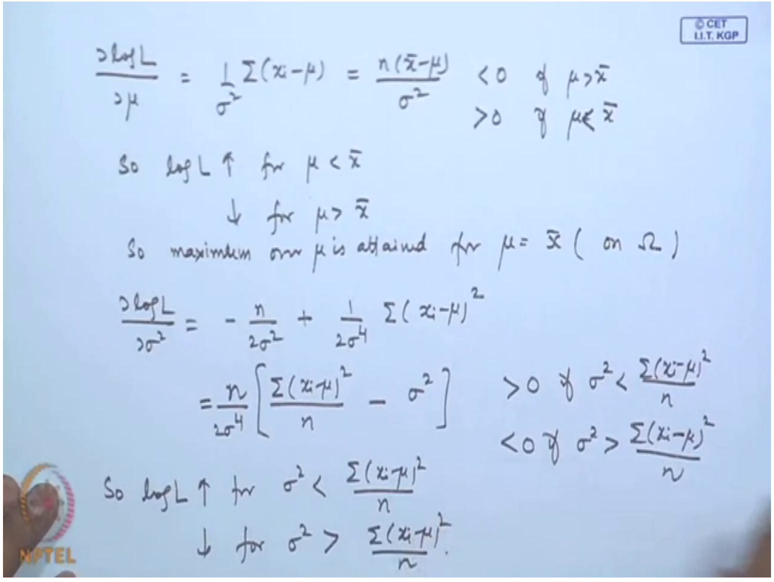
If you remember the Nyman Pearson theory I have considered various possibilities, initially when I tested for the mean I had assumed variance to be known and when I tested for variance I had assumed mean to be known. And then later on we have derived the test when both are unknown and in the final analysis we got the uniformly most powerful unbiased tests for those situations. Now in the likelihood ratio test I am considering the problem, where both the parameters mu and sigma square are unknown.

Now let us consider say the hypothesis of the nature that H 1, that is mu less than or equal to mu naught against mu greater than or equal to mu naught. Now, without loss of generality I can take mu naught to be 0. So, I can consider the hypothesis mu less than or equal to 0 against mu greater than 0. Now we need to consider if we want to apply the likelihood ratio test, then I need to consider L hat omega H by L hat omega and what is L hat omega H? L hat omega H is nothing, but the maximization of the likelihood function over the null hypothesis parameter space.

Similarly, L hat omega is the maximization of the likelihood function over the full parameter space. So, if we take care of these values then let us write down omega is written here, what is omega H then? Omega H is the set of all those mu sigma square for which mu is less than or equal to 0 and sigma square is positive. So, note here this has become a subset of omega let us write down the likelihood function, here we have 2 parameters. So, this is a joint distribution of X 1, X 2, X n. So, we have been writing this terms several times.

So, it is 1 by sigma 2 pi to the power n e to the power minus 1 by 2 sigma square sigma xi minus mu square. Now if we consider log of L, then that is equal to minus n by 2 log of sigma square minus n by 2 log of 2 pi minus 1 minus 2 sigma square sigma xi minus mu square.

(Refer Slide Time: 18:36)

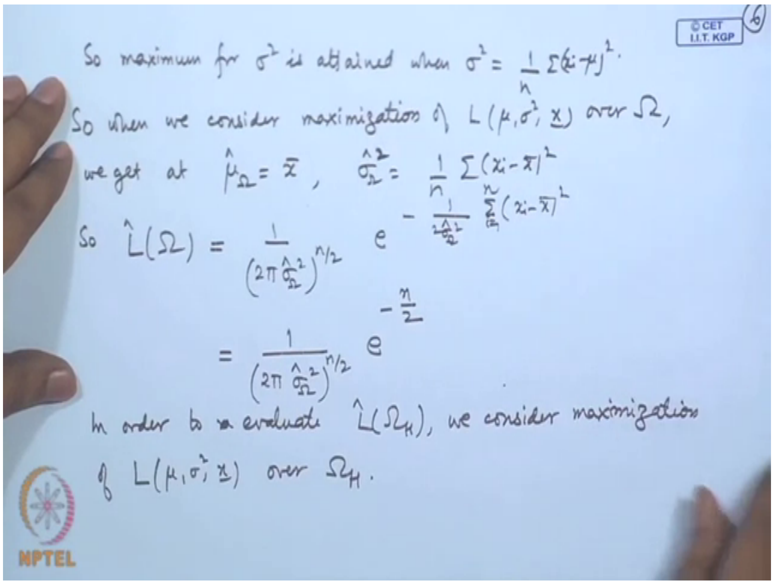


If you look at the derivative del mu that is equal to sigma xi minus mu by sigma square that is equal to n x bar minus mu by sigma square. So, this is less than 0, if mu is greater than x bar it is greater than 0, if mu is equal to x bar. So, log L is increasing sorry this is mu less than x bar. So, log l is increasing for mu less than x bar it is decreasing for mu greater than x bar.

So, maximum over mu is attained for mu is equal to x bar, this is on omega. See we are firstly, considering the maximization over omega and then over omega H. So, and then if I consider derivative with respect to sigma squared then I will get minus n by 2 sigma square plus 1 by 2 sigma to the power 4 sigma xi minus mu square.

Now, once again we combine the terms I can write it as sigma xi minus mu square, I can take out 1 by 2 sigma to the power 4 maybe I can take out n also, so this divided by n minus sigma square. Once again you note here this is greater than 0 if sigma square is less than sigma xi minus mu square by n and it is less than 0 if sigma square is greater than sigma xi minus mu square by n. So, log L is increasing for sigma square less than sigma xi minus mu square by n and decreasing for sigma square greater than sigma xi minus mu square by n.

(Refer Slide Time: 21:15)

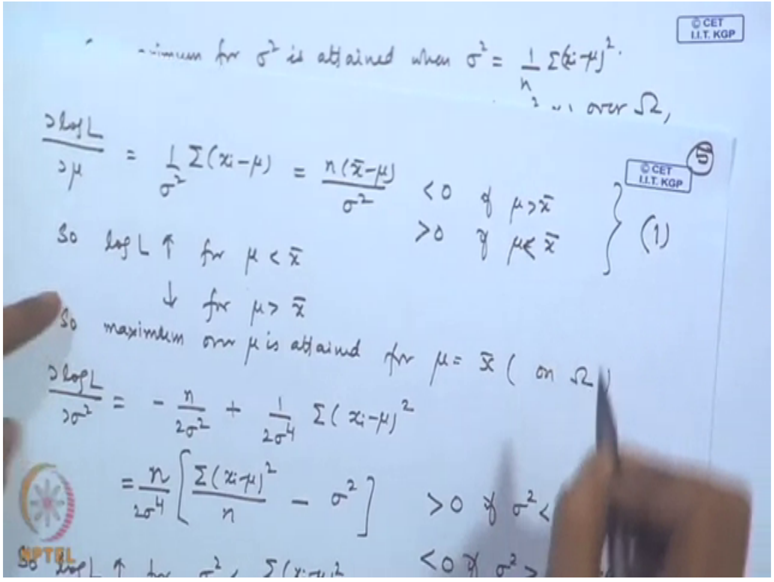


So, maximum for sigma square is attained when sigma square is equal to 1 by n sigma xi minus mu square. So, when we consider maximization of L over omega, we get at mu hat let me write mu hat omega is equal to x bar, sigma hat square omega is equal to 1 by n sigma xi minus x bar square.

So, L hat omega is nothing, but the value evaluated at this point, that is 1 by 2 pi sigma hat omega square to the power n by 2 e to the power minus 1 by 2 sigma omega hat square sigma xi minus x bar whole square i is equal to 1 to n by n sigma xi minus x bar whole square. So, if I substitute this term will get cancelled out and I will get it as simply 1 by 2 pi sigma omega hat square to the power n by 2 e to the power minus n by 2.

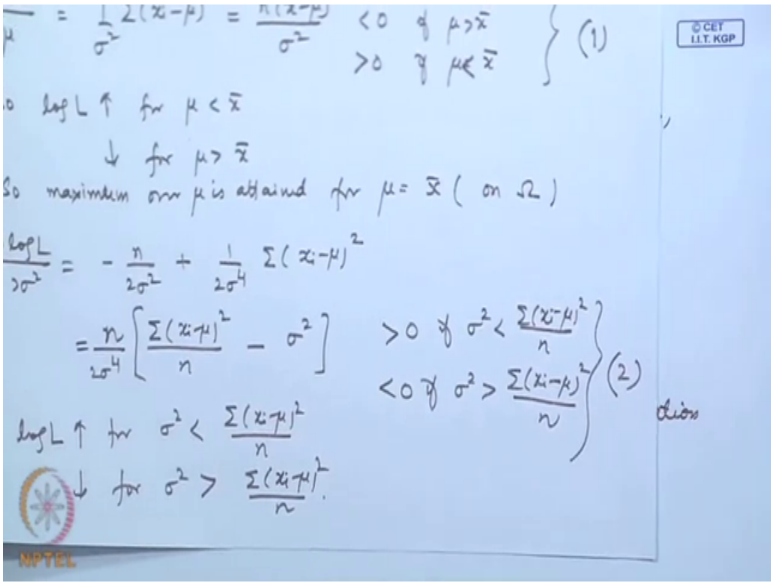
So, we have evaluated the maximization of the likelihood function over the full parameter space, in order to apply the likelihood ratio test I also need to consider in order to evaluate L hat omega H, we consider maximization of L mu sigma square x over omega H. Now, let us analyze over omega H.

(Refer Slide Time: 23:54)



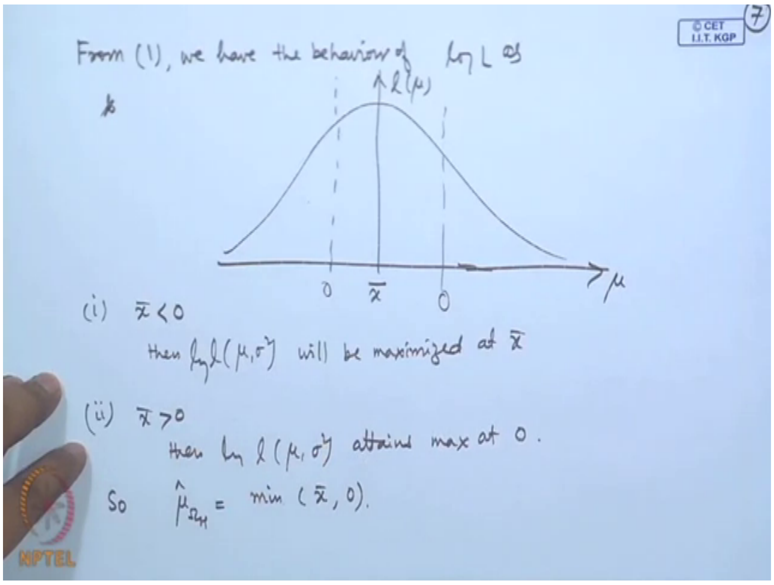
We have already looked at the derivatives here, so look at del log L by del mu, we have seen the behavior of it. So, let me give this some numbering here this is say 1 and this one is say 2.

(Refer Slide Time: 23:57)



Because we will be using these expressions of del log L by del mu and del log L by del sigma square.

(Refer Slide Time: 24:18)

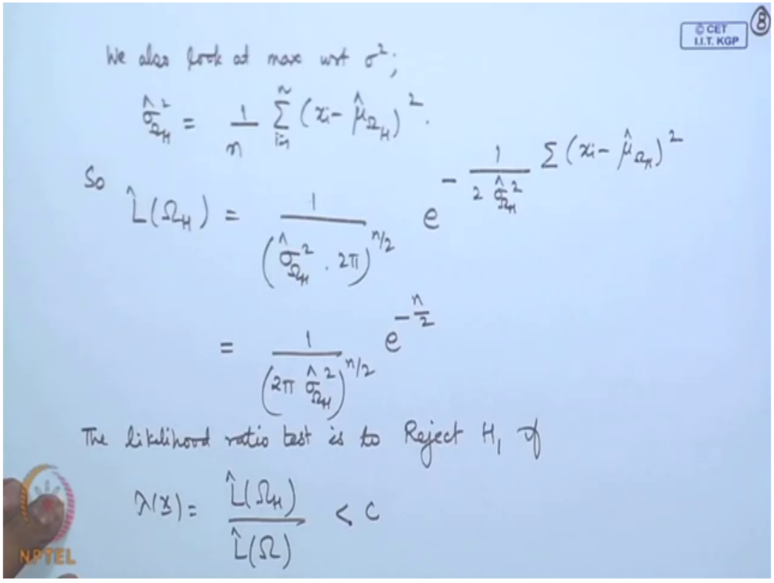


So, from 1, we have the behavior of log L as, so let me just draw. So, it is something like this is increasing up to x bar, so I am treating it as a function of mu and this is function of mu. I am treating it as a function of, so as a function of mu it increases up to x bar and thereafter it is decreasing. Now it will depend upon where what is the position of 0 because in the omega H we have to maximize over minus infinity less than mu less than or equal to 0 sigma square greater than 0; that means, mu is less than or equal to 0. So, when we look at this let us look at there are 2 cases.

One case x bar is less than 0, if x bar is less than 0; that means, 0 is say here in that case l mu sigma square log L will be maximized at x bar. Now second case is x bar is greater than 0, if x bar is greater than 0; that means, the position of 0 is here, if the position of 0 is here this is the likelihood function, it is increasing up to 0 and we are only looking at this region, so the maximum value is attained at 0.

So, then log L mu sigma square attains maximum at 0. So, what we are getting mu hat omega H it is equal to minimum of x bar and 0, because it is x bar if x bar is less than 0 and it is 0 if 0 is less than x bar. So, at 1 place we may put equality here.

(Refer Slide Time: 26:42)

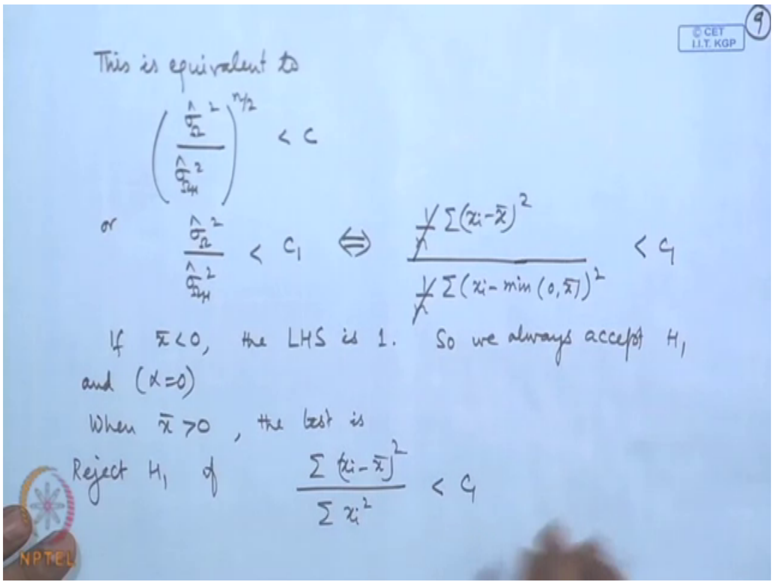


Let us look at the maximization of we also look at max with respect to sigma square. Now when we look at the max with respect to sigma square, we have shown here that the max is occurring at sigma square is equal to sigma xi minus mu whole square by n that is a statement we gave here. For sigma square the max is occurring at sigma square is equal to this quantity. Now this mu I had substituted as mu hat, now in this case mu hat is modified.

So, accordingly sigma square will get modified and we get sigma omega H hat square that will become 1 by n sigma xi minus mu hat omega H square i is equal to 1 to n. So, L hat omega H that will be equal to 1 by sigma omega H hat square into 2 pi e to the power n by 2 e to the power minus 1 by 2 sigma omega H hat square sigma xi minus mu hat omega H square.

Once again if I substitute the value of sigma omega H hat square then I will get n by 2 here. So, this is equal to 1 by 2 pi sigma omega H hat square to the power n by 2 e to the power minus n by 2. So, the likelihood ratio test is to reject H 1 if lambda x that is equal to L hat omega H by L hat omega less than say some constant c.

(Refer Slide Time: 29:01)

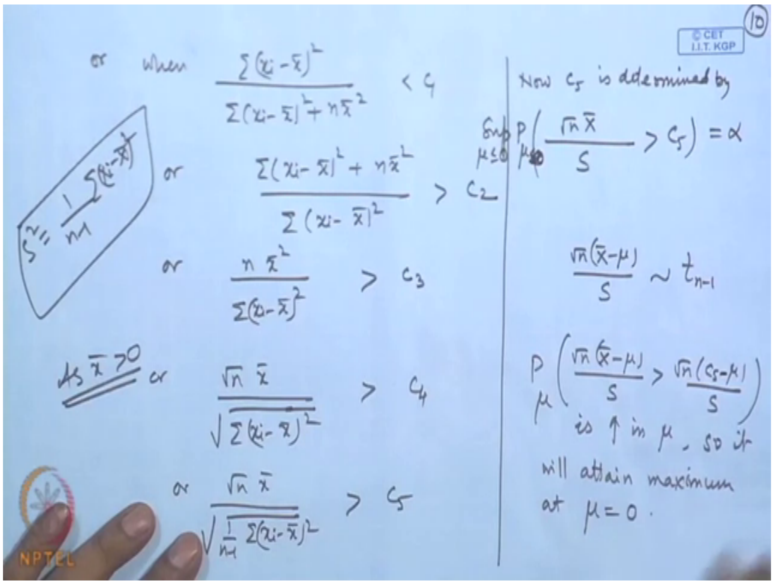


So, if I substitute the values of L hat omega H and L hat omega, then L hat omega is this quantity and L hat omega L hat omega H and L hat omega both are given here. So, e to the power minus n by 2 terms will get cancelled out and we will get the ratio as sigma omega hat square divided by sigma omega H hat square to the power n by 2 less than c.

Now, I can take this power 2 by n here, so this is equivalent to sigma omega hat square by sigma omega H square less than another constant say c 1. Now let us substitute the values here this is 1 by n sigma xi minus x bar square divided by 1 by n sigma xi minus minimum of 0 and x bar square, so this gets cancelled out less than c 1. Now, if x bar is less than 0 the left hand side is 1, if this is 1 then basically what we have said likelihood ratio is always between 0 to 1 because in the numerator I have a smaller quantity than the denominator.

So, this is the extreme case, if it is a extreme case then we should always accept H 1. So, we always accept H 1 and actually the probability alpha will become 0. Now, when x bar is greater than 0, then here minimum will become 0, so the test is reject H1 if sigma xi minus x bar whole square divided by sigma xi square is less than c 1.

(Refer Slide Time: 31:25)



Or when sigma xi minus x bar square divided by sigma xi minus x bar square plus nx bar square less than c 1 or if I take the reciprocal greater than say c 2. So, I can write it as in n x bar square by sigma xi minus x bar whole square greater than say some c 3.

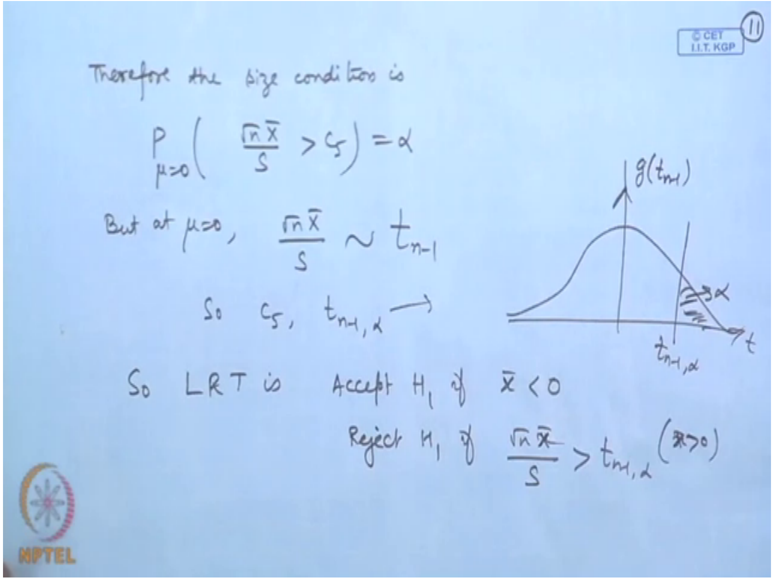
I can take the square root because x bar is positive as x bar is positive I can take the square root here this becomes the square root sigma xi minus x bar whole square greater than say c 4, I can adjust again and write. So, basically if you remember this is the test which we got in the UMP test also.

If I consider root n x bar divided by root 1 by n minus 1 sigma xi minus x bar whole square greater than say c 5. Now, this c 5 is determined by probability of root n x bar divided by S greater than c 5. If you remember the definition of S square that was 1 by n minus 1 sigma xi minus x bar whole square. So, n minus 1 S square by sigma square follows chi square distribution on n minus 1 degrees of freedom. So, when mu is equal to 0 this will follow T distribution on n minus 1 degrees of freedom.

So, here this is determined by supremum of this when mu is less than or equal to 0 equal to alpha, this supremum is over this. Now let us look at this thing root what is the distribution here? You are having root n x bar minus mu by S that follows T distribution on n minus 1 degrees of freedom ok. So, this probability then can be written as root n x bar minus mu by S greater than root n c 5 minus mu by S probability of this.

Now, if mu increases; if mu increases this limit will decrease the lower bound here because here it is coming as minus mu ok. So, if mu increases this will decrease, so this probability will increase, this probability is increasing in mu. So, it will attain maximum at mu is equal to 0.

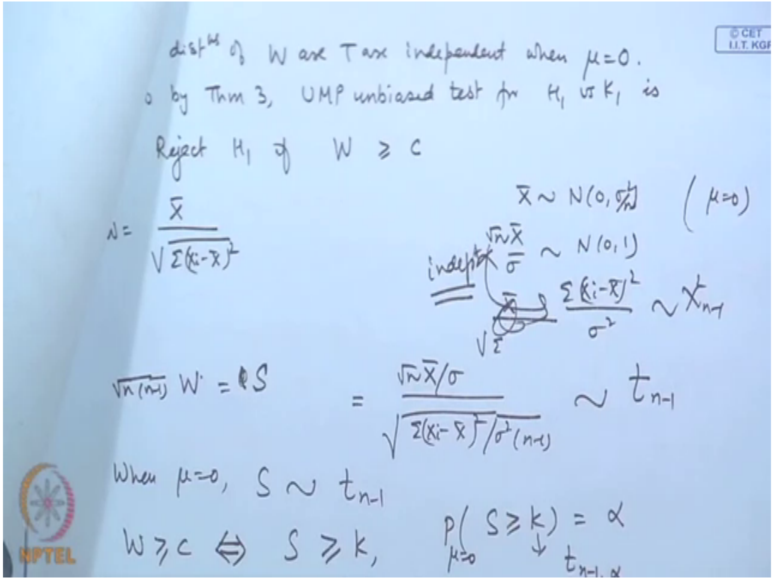
(Refer Slide Time: 35:07)



Therefore, the size condition is probability of root n x bar by S greater than c 5 at mu equal to 0 is equal to alpha, but at mu is equal to 0 root n x bar by S this will follow T distribution on n minus 1 degrees of freedom. So, c 5 is nothing, but the upper 100 alpha percent point of the T distribution on n minus 1 degrees of freedom. If this is the curve of T distribution on n minus 1 degrees of freedom then on the side you have T, then this probably should be alpha. So, likelihood ratio test is accept H if that is always accept H 1 if x bar is less than 0 and reject H naught H 1 if a root n x bar by S is greater than t n minus 1 alpha for x bar positive.

So, there is a slight modification from the UMP test and UMP unbiased test for the situation we have derived if you remember in the lecture which I gave earlier.

(Refer Slide Time: 36:56)



The test that we derived here root n x bar; root n x bar by sigma xi minus x bar whole square by n minus 1 square root. So, the test was in the terms of this. So, in many situations the likelihood ratio test yield the similar test as the in the Nyman Pearson theory; that means, they are also the UMP unbiased tests and in many situation of 1 parameter they may be UMP tests also.

However, we have seen that sometimes it may not happen like that and in that case we need certain asymptotic properties fortunately for the likelihood ratio test the asymptotic properties do hold; that means, asymptotic distribution of the test S statistic is nice it becomes actually the chi square.