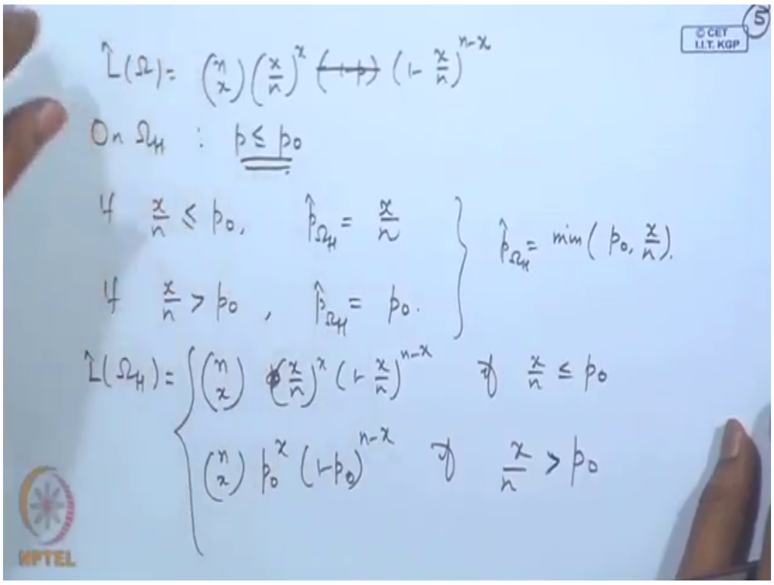
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 54
Likelihood Ratio Tests _ IV

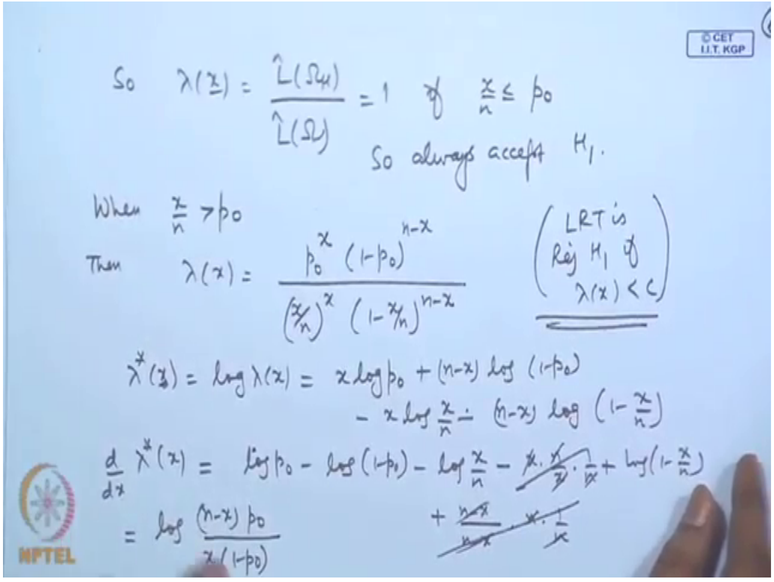
Suppose I ask for H 4 hypothesis that is if I consider p is equal to p naught, against p is not equal to p naught.

(Refer Slide Time: 00:27)



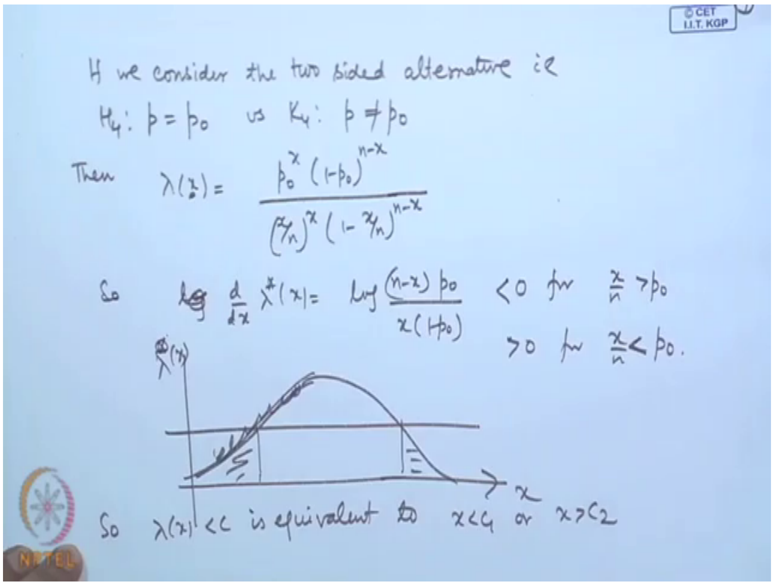
In place of p less than or equal to p naught if I modify that, then we can this region will not come only this portion will come and we will analyze the behavior of this function in a more appropriate fashion. Because there will be 2 regions; one will be corresponding to x by n greater than p naught. We have already seen that in this region we are getting the function to be a decreasing function.

(Refer Slide Time: 00:49)



However, if I consider x by n less than or equal to p naught this value will become positive. And therefore, lambda star x will become an increasing function. So, let me just give the complete analysis for the two sided testing problem.

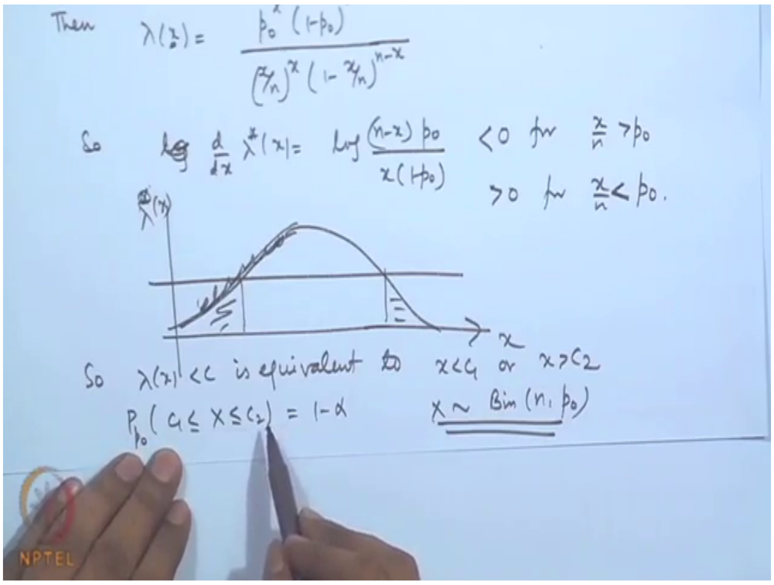
(Refer Slide Time: 01:25)



If we consider the two sided alternative that is H 4 p is equal to p naught against K 4 p is not equal p naught. In that case lambda x that is equal to simply p naught to the power x 1 minus p naught to the power n minus x divided by x by n to the power x 1 minus x by n to the power n minus x. So, lambda the behavior of lambda star we have seen that derivative of a d by dx of lambda star x that is equal to log of n minus x into p naught divided by x into 1 minus p naught. So, this is less than 0 for x by n greater than p naught and it is greater than 0 for x by n less than p naught.

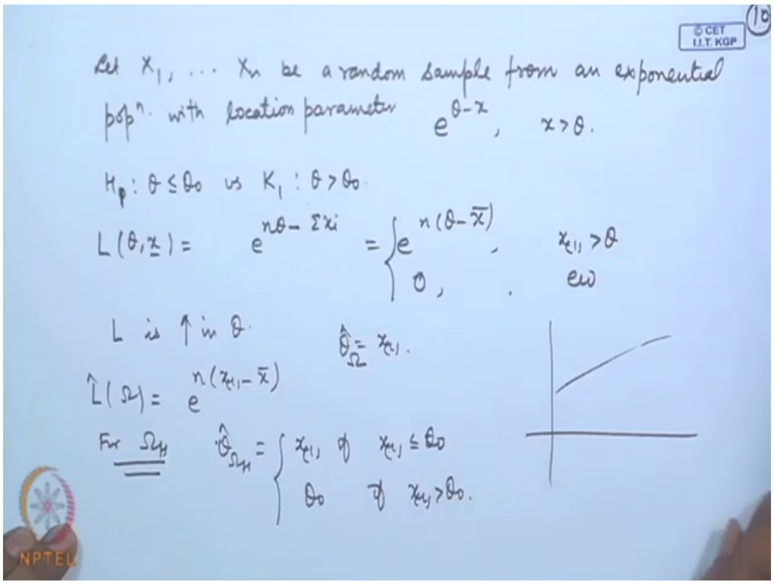
So, the nature of the function is then that it is increasing and then for x by n less than p naught it is increasing. So, if I am treating it as a function of x then it is firstly, for x by n yeah it is increasing and then it is decreasing on this side I have x on this side I have lambda x or lambda star x. So now, the region lambda x less than or equal to c, that is equal to this 2 regions. So, lambda x less than C is equivalent to x less than C 1 or x is greater than C 2.

(Refer Slide Time: 03:47)



And we should have probability of C 1 less than or equal to x less than or equal to C 2 is equal to 1 minus alpha where x follows binomial n p naught. So, this C 1 and C 2 can measurement from the tables of binomial distribution. I have shown you that, even in the case of discrete case the likelihood ratio test can give a nice and elegant solution. However, if the distribution is not in an exponential family let us consider one example.

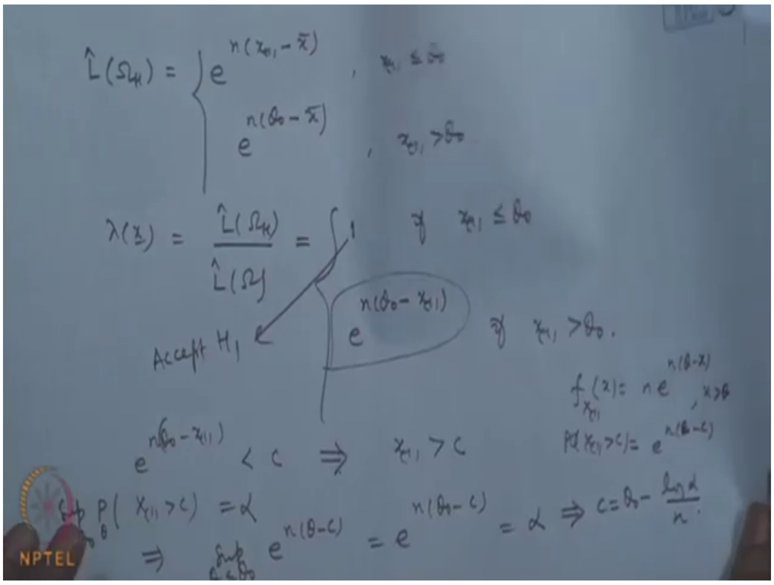
(Refer Slide Time: 04:21)



Let X 1, X 2, X n be a random sample from an exponential population with location parameter that is e to the power theta minus x is the density function. So, the likelihood ratio function is e to the power n theta minus sigma x i which I can write as e to the power n theta minus x bar. This is for x 1 greater than theta it is equal to 0, if x 1 is less than or equal to theta. So, naturally you can see that L is increasing in theta. So, theta hat is equal to x 1 this is over omega.

So, L hat omega is nothing but e to the power n x 1 minus x bar. Let us consider omega H. For omega H if I am considering, now this is an increasing function. In fact, you can see that theta is equal to x 1 this will be this value, but it is going like that. So, for omega H theta hat omega h that will be equal to x 1, if x 1 is less than or equal to theta naught and it is equal to theta naught if x 1 is greater than theta naught.

(Refer Slide Time: 06:35)

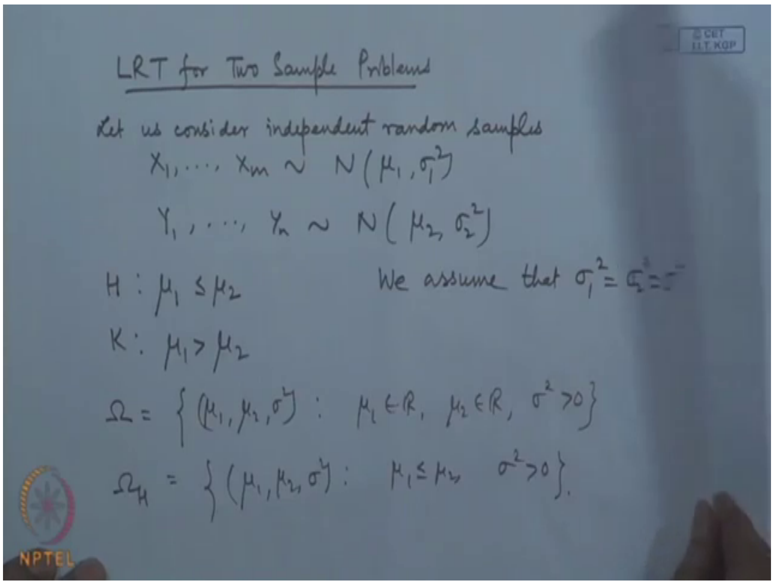


So, L hat omega H that is equal to e to the power n x 1 minus x bar, if x 1 is less than or equal to theta naught it is equal to e to the power n theta naught minus x bar if x 1 is greater than theta naught. So, the ratio that is L hat omega H by L hat omega that is equal to 1, if x 1 is less than or equal to theta naught and it is equal to e to the power n theta naught minus x 1.

If x 1 is greater than theta naught so, here in this case we will accept H 1 and in this particular case, if I am considering e to the power n theta naught minus x 1 less than C then this is equivalent to x 1 greater than C. So, this C will be determined from the size condition probability of x 1 greater than C. When theta we are taking maximum for theta less than or equal to theta naught this should be equal to alpha. Now, what is the distribution of x 1? That is equal to n e to the power n theta minus x for x greater than theta.

So, if I consider the probability of x 1 greater than C that is equal to e to the power n theta minus C. So, this value is actually e to the power n theta minus C. So, if I take supremum theta less than or equal to theta naught that is equal to e to the power n theta naught minus C, that is equal to alpha. That means C is equal to theta naught minus log of alpha by n.

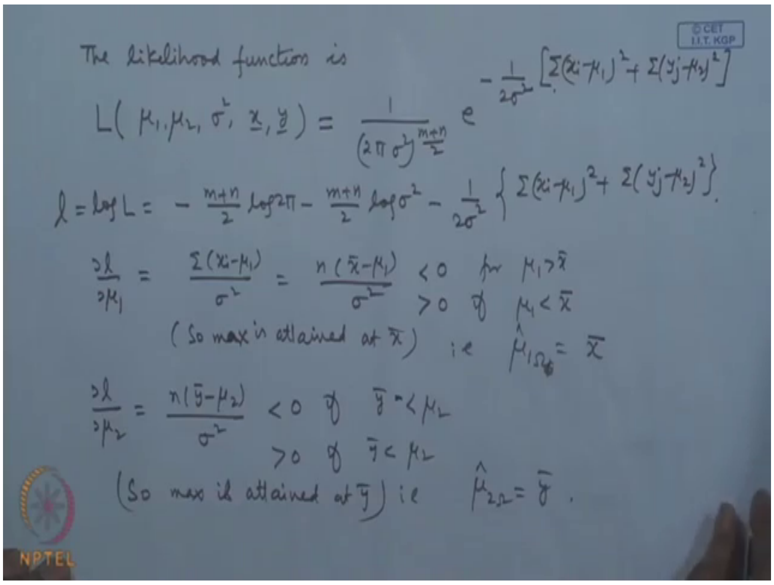
(Refer Slide Time: 08:41)



So, the likelihood ratio test is so, when x 1 is greater than theta naught reject H 1 if x 1 is greater than theta naught minus log of alpha by n. Now, we take up the likelihood ratio test for two sample problems. So, likelihood ratio test for two sample problems. To start with I consider random samples from 2 normal populations. Let us consider, let us consider independent random samples X 1, X 2, X m from normal mu 1, sigma 1 square, Y 1, Y 2, Y n from normal mu 2, sigma 2 square. Let us consider the hypothesis say, for testing of mu 1 less than or equal to mu 2 against say mu 1 greater than mu 2.

We assume here that sigma 1 square is equal to sigma 2 square. this case we have derived the UMP unbiased tests, when this is not equal that is sigma 1 square is not equal to sigma 2 square it was known as a Byron’s Fisher situation. So, our parameter space is mu 1, mu 2, sigma square: mu 1 is real, mu 2 is real and sigma square is positive and omega H that is mu 1, mu 2 sigma square: mu 1 is less than or equal to mu 2 and sigma square is greater than 0.

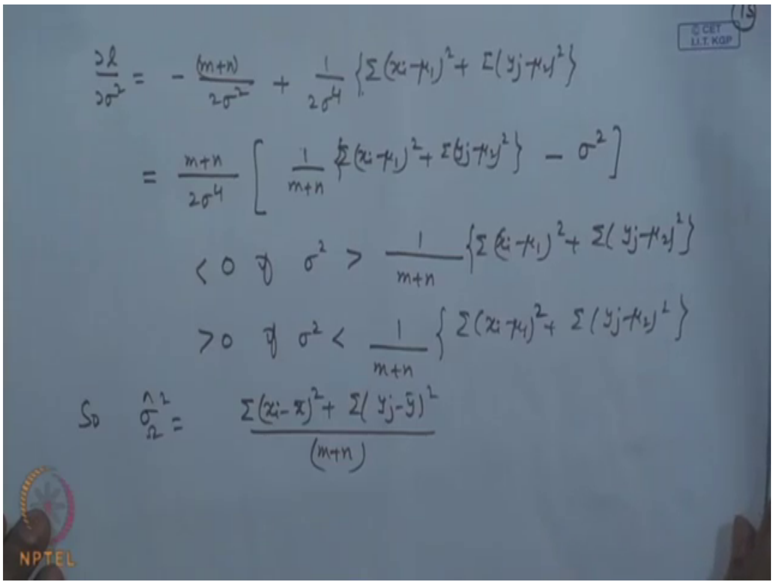
(Refer Slide Time: 11:33)



Let us write down the likelihood function. The likelihood function is given by L mu 1, mu 2, sigma square, x and y. So, that is equal to 1 by 2 pi sigma square to the power m plus n by 2 e to the power minus 1 by 2 sigma square sigma x I minus mu 1 square plus sigma y j minus mu 2 square. Log of l is equal to minus m plus n by 2 log of 2 pi minus m plus n by 2 log of sigma square minus 1 by 2 sigma square sigma x i minus mu 1 square plus sigma y j minus mu 2 square.

So, if I consider the maximization over the full parameter space. So, let me write this as small l. So, del l by del mu 1 that will give me sigma x i minus mu 1 by sigma square, that is n x bar minus mu 1 by sigma square. So, once again less than 0; if mu 1 is greater than x bar it is greater than 0. If mu 1 is less than x bar so, maximum is attained at x bar, that is mu 1 hat omega is equal to x bar. Similarly, if I take del l by del mu 2 that is equal to n y bar minus mu 2 by sigma square that is less than 0 if y bar is less than mu 2 it is greater than 0, if y bar is less than mu 2. So, maximum is attained at y bar. That is mu 2 omega hat that is equal to y bar.

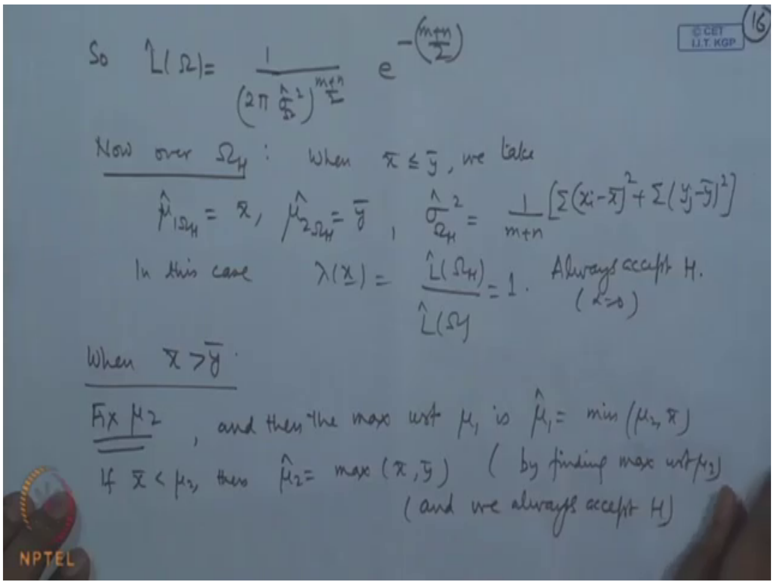
(Refer Slide Time: 14:25)



Let us also consider the maximization with respect to sigma square, so we get minus m plus n by 2 sigma square plus 1 by 2 sigma to the power 4 sigma x i minus mu 1 square plus sigma y j minus mu 2 square. So, that is equal to if I take m plus n by 2 sigma to the power 4 out then I get 1 by m plus n sigma x i minus mu 1 square plus sigma y j minus mu 2 square minus sigma square. Once, again it is less than 0 if sigma square is greater than 1 by m plus n sigma x i minus mu 1 square plus sigma y j minus mu 2 square, it is greater than 0 if sigma square is less than 1 by m plus n sigma x i minus mu 1 square plus sigma y j minus mu 2 square.

So, sigma omega hat square that is attend at this value where mu 1 and mu 2 are replaced by their respective maximum likelihood estimators. That means, it is attained at sigma x i minus x bar square plus sigma y j minus y bar square divided by m plus n. So, if we consider l hat omega.

(Refer Slide Time: 16:23)



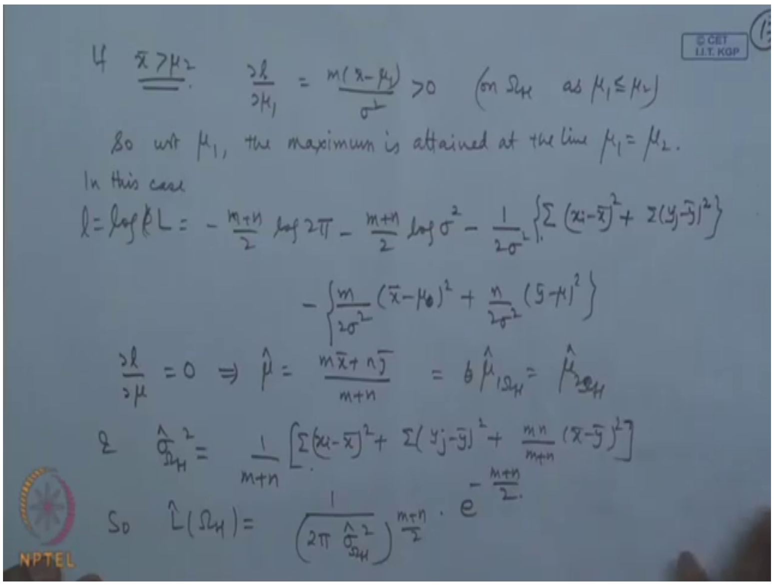
So, l hat omega is equal to 1 by 2 pi sigma omega hat square to the power m plus n by 2 e to the power minus m plus n by 2. Now let us consider over omega H, omega H. Over omega H what is happening? That let us look at the derivatives here the maximization were occurring at x bar and y bar for mu 1 and mu 2. Now, we are saying mu 1 is less than or equal to mu 2 over omega H. So, when x bar is less than or equal to y bar we can treat them as such.

So, when x bar is less than or equal to y bar we take, mu 1 omega H hat is equal to x bar, mu 2 omega h hat is equal to say y bar and sigma omega h hat square that is equal to 1 by m plus n sigma x i minus x bar square plus sigma y j minus y bar square. There is no change. In this case the ratio that is L hat omega h by L hat omega that is equal to 1.

So, we always accept H 1 the hypothesis that I mentioned here, mu 1 less than or equal to mu 2, against mu 1 greater than mu 2. So, we always accept H. So, that is alpha is equal to 0. Now the other possibility is when x bar is greater than y bar. Now when x bar is greater than y bar then we analyze. Firstly, you look at the behavior with respect to mu 1.

Now with respect to mu 1 you see it is initially increasing up to x bar and then decreasing. So, now if y bar is less than x bar then, we will have to modify this thing. What we do? Fix mu 2 first say ok. And then the maximum with respect to mu 1 is mu 1 hat is equal to minimum of mu 2 and x bar. Now if x bar is less than mu 2. Then mu 2 hat will be equal to maximum of x bar, y bar that is by finding the maximum with respect to mu 2. And we always accept H.

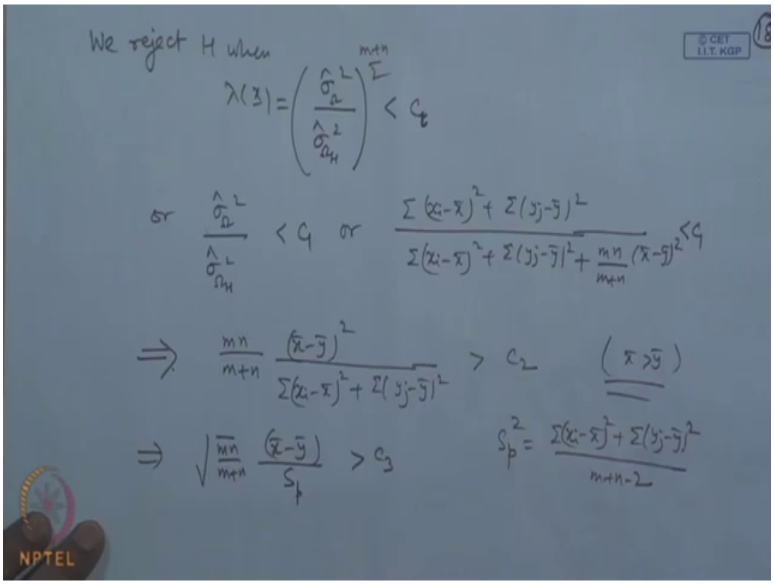
(Refer Slide Time: 20:03)



However, another case occurs when, if x bar is greater than mu 2. In this case the d l by del mu 1 that is equal to m x bar minus mu 1 by sigma square. I think I wrote a little bit wrongly in the first place this was m times not n times. So, this one will become greater than 0 on omega H because, we are having mu 1 less than or equal to mu 2 ok. So, with respect to mu 1 the maximum is attained at the line mu 2 is equal to mu 2. So, in this case the log likelihood that is L, I can write as minus m plus n by 2 log of 2 pi, minus m plus n by 2 log of sigma square minus 1 by 2 sigma square sigma x i minus x bar square plus sigma y j minus y bar square minus m by 2 sigma square x bar minus mu 1 square mu square plus n by 2 sigma square y bar minus mu square.

So, if I consider del l by del mu then I will get mu hat is equal to m x bar plus n y bar divided by m plus n. That is the grand mean; that is mu 1 omega H mu 2 omega H because, I have assumed them to be equal here. And sigma omega H hat square we will also get because there we were getting x i minus mu 1 and y j minus mu 2. So, this will become simply sigma x i minus x bar square plus sigma y j minus y bar square plus m n by m plus n x bar minus y bar square. So, L hat omega H that becomes equal to 1 by 2 pi sigma omega H square to the power m plus n by 2 e to the power minus m plus n by 2.

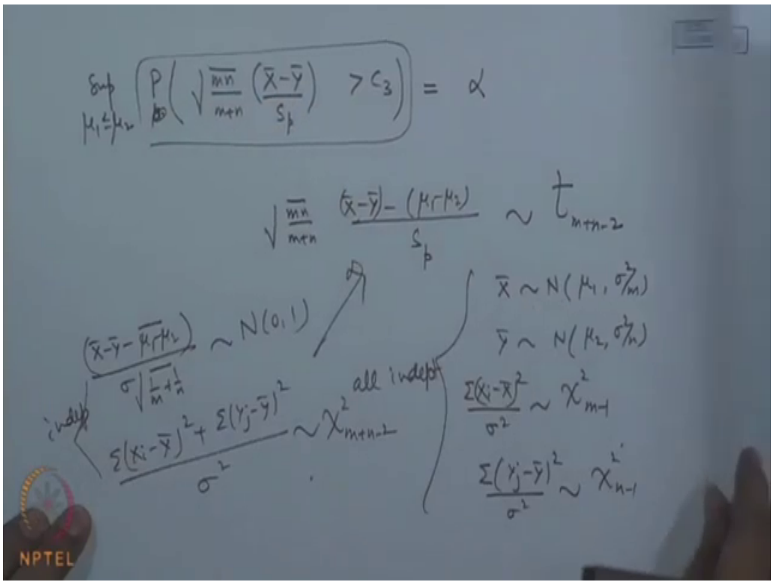
(Refer Slide Time: 22:57)



Now, the rejection region is, we reject H when, lambda x that is l hat omega h by l hat omega that will become sigma omega hat square by sigma omega H hat square less than say some to the power n by 2 will come, so I am to the power m plus n by 2 less than some C or we can say sigma omega hat square by sigma omega H hat square is less than C 1. Now, we substitute the values here, this is sigma x i minus x bar square plus sigma y j minus y bar square divided by m plus n and in the denominator then I have sigma x i minus x bar square plus sigma y j minus y bar square plus m n by m plus n x bar minus y bar square this is less than C 1.

You take the reciprocal and then subtract then this will give us actually m n by m plus n x bar minus y bar square divided by sigma x i minus x bar square plus sigma y j minus y bar square greater than say some C 2. Note here that x bar is greater than y bar. x bar is greater than y bar. So, this means I can take the square root here m n by m plus n x bar minus y bar divided by S p greater than some C 3, where s p square is nothing but sigma x i minus x bar square plus sigma y j minus y bar square divided by m plus n minus 2.

(Refer Slide Time: 25:15)

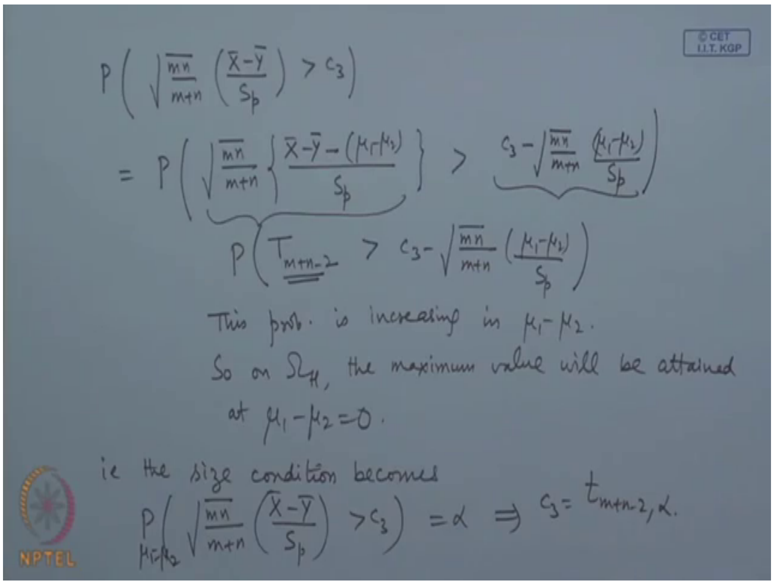


So, the condition that we will be getting here is probability of root m n by m plus n X bar minus Y bar by S p is greater than C 3 when mu 1 so, this is for general mu 1, mu 2 we have to take the supremum when mu 1 is less than or equal to mu 2. This should be equal to alpha. Now note here root m n by m plus n X bar minus Y bar minus mu 1 minus mu 2 divided by S p.

This will follow t distribution on m plus n minus 2 degrees of freedom. this part we have considered earlier, in the sampling distributions in my lectures on probability because you are having actually X bar following normal mu 1 sigma square by m, Y bar follows normal mu 2 sigma square by n. Then sigma X i minus X bar whole square by sigma square that follows chi square distribution on m minus 1 degrees of freedom. Sigma Y j minus Y bar square by sigma square follows chi square distribution on n minus 1 degrees of freedom and these are all independent.

So, X bar minus Y bar minus mu 1 minus mu 2; then if I consider this will follow if I divide by sigma root 1 by m plus 1 by n. This will follow normal 0, 1 and sigma X i minus X bar square plus sigma Y j minus Y bar square by sigma square this follows chi square distribution on m plus n minus 2 degrees of freedom. And we take the ratio that will give me this as t distribution.

(Refer Slide Time: 27:29)



If this is the t distribution then this probability if I consider, probability of root m n by m plus n X bar minus Y bar by S p greater than C 3. That is equal to probability of root m n by m plus n X bar minus Y bar minus mu 1 minus mu 2 divided by S p greater than C 3 minus root m n by m plus n mu 1 minus mu 2 by S p. Now this is a T distribution on m plus n minus 2 degrees of freedom. So, we are saying probability of a T variable on this is bigger than C 3 minus root m n by m plus n mu 1 minus mu 2 by S p.

Now, mu 1 minus mu 2 is less than or equal to 0. Actually if I increase mu 1 minus mu 2 then this term will decrease. If this term decreases this probability will increase; so, this probability is increasing in mu 1 minus mu 2. So, on omega H the maximum value will be attained at mu 1 minus mu 2 equal to 0. That is the size condition becomes probability of root m n by m plus n X bar minus Y bar by S p greater than C 3; when mu 1 is equal to mu 2. So, this means C 3 is nothing, but t m plus n minus 2 alpha. You can compare it with the ump unbiased test that we derived. This form is the same.

So, in the likelihood ratio test what is happening is that some part we are having slightly different in the 1 sided testing problems. But in the 2 sided testing problem this is exactly same as the ump unbiased test. In the next lecture I will derive this test likelihood ratio test for mu 1 equal to mu 2 against mu 1 not equal to mu 2. we will also consider the testing for the equality of variances. So, sigma 1 square less than or equal to sigma 2 square, sigma 1 square is not equal to sigma 2 square these hypothesis testing problems also I will consider in the next lecture.