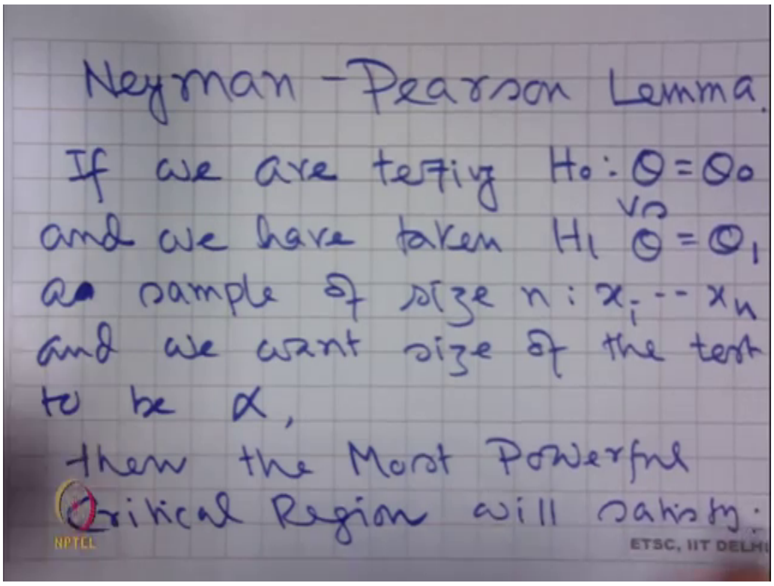
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 21
Statistical Inference

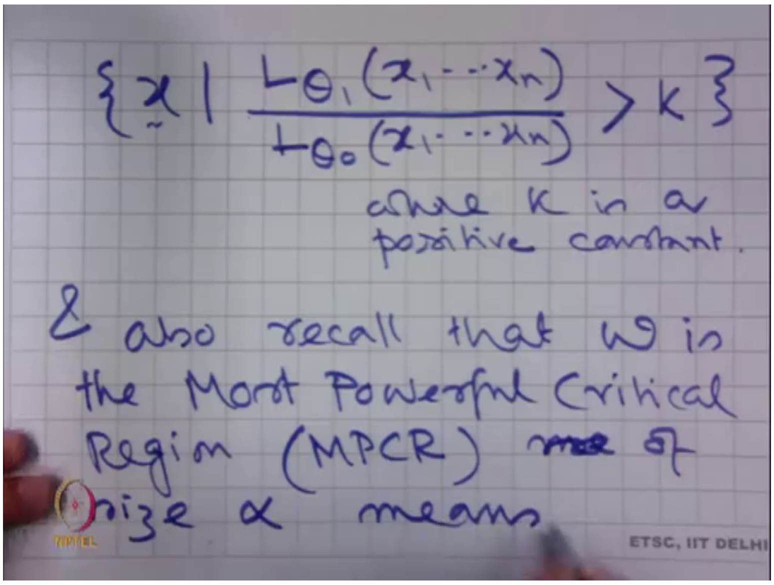
Welcome students to the MOOCs series of lectures on Statistical Inference. This is lecture number 21 and you remember that we have been discussing testing of hypothesis. In particular our focus has been on Neyman-Pearson Lemma.

(Refer Slide Time: 00:43)



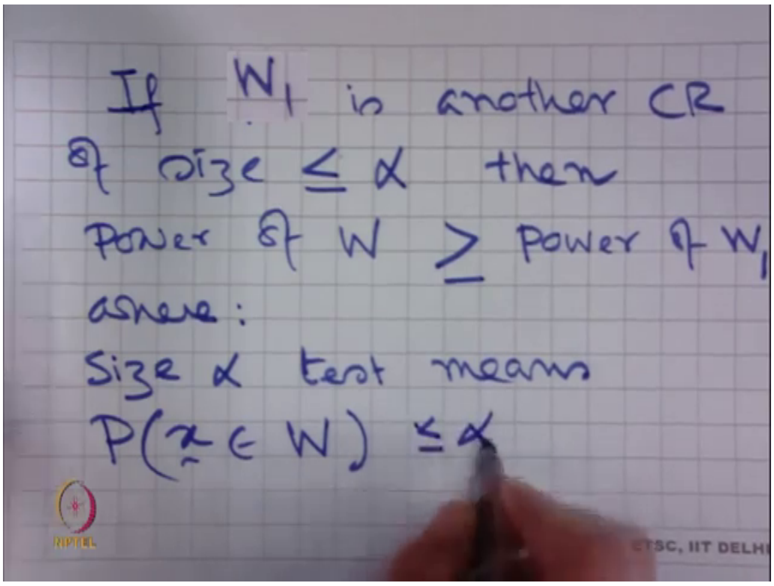
So, what it says? That if we are testing H naught, theta is equal to theta naught versus H 1 theta is equal to theta 1, and we have taken a sample of size n: namely x 1, x 2, x n and we want size of the test to be alpha then the most powerful critical region will satisfy the following.

(Refer Slide Time: 02:23)



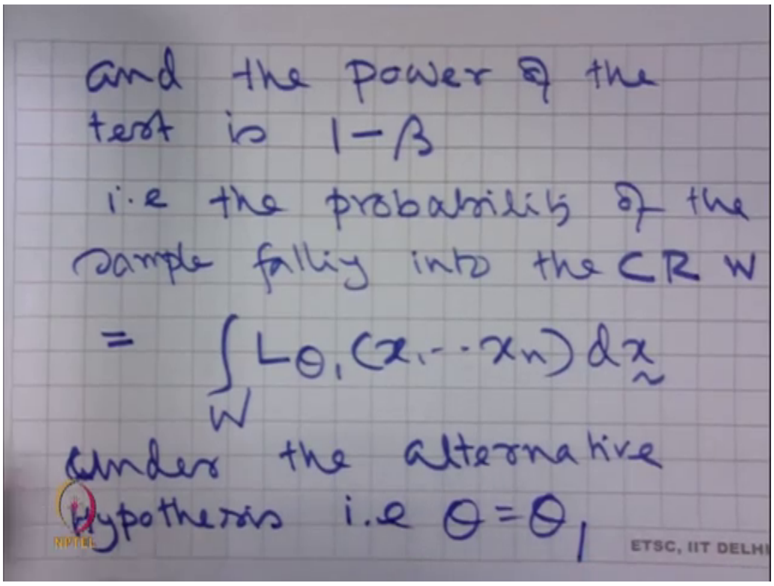
It is the set of observations x, such that L theta 1 of x 1, x 2, x n where this is the likelihood function and under the alternative hypothesis is H 1 upon L theta 0 of x 1, x 2, x n which is the likelihood function under H naught has to be greater than K, where K is a positive constant. And also recall that W is the most powerful critical region that is MPCR means of size alpha, means:

(Refer Slide Time: 04:05)



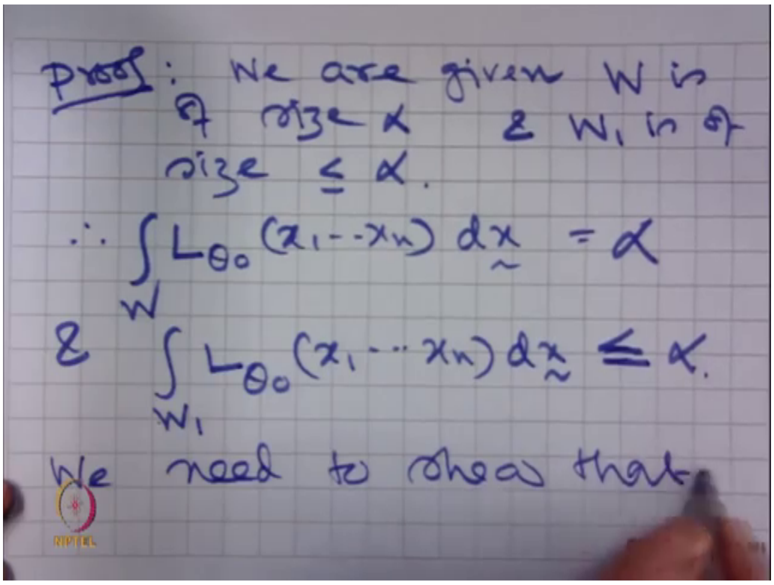
If W 1 is another critical region of size less than equal to alpha, then power of W is greater than equal to the power of W 1, where size alpha test means probability x belonging to the W is less than equal to alpha and the power of the test is 1 minus beta.

(Refer Slide Time: 05:05)



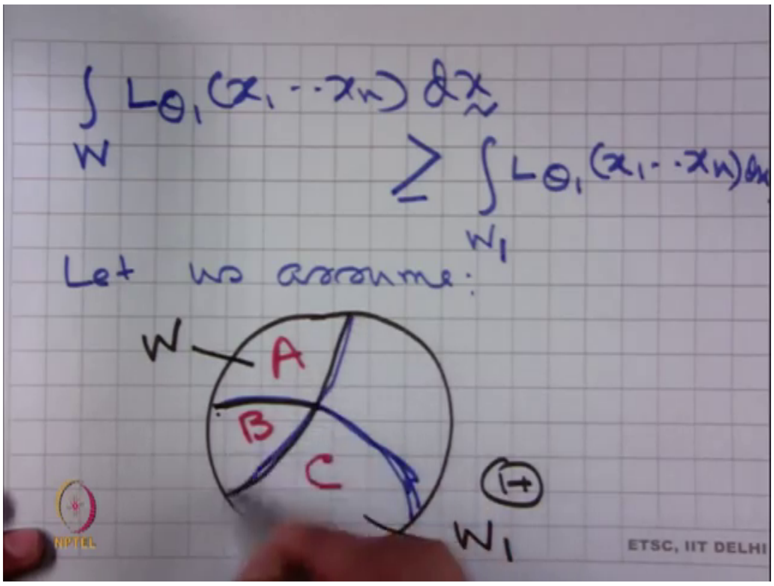
That is the probability of the samples falling into the critical region W and that is equal to integration over W L theta 1 x 1, x 2, x n dx under the alternative hypothesis that is theta is equal to theta 1.

(Refer Slide Time: 06:23)



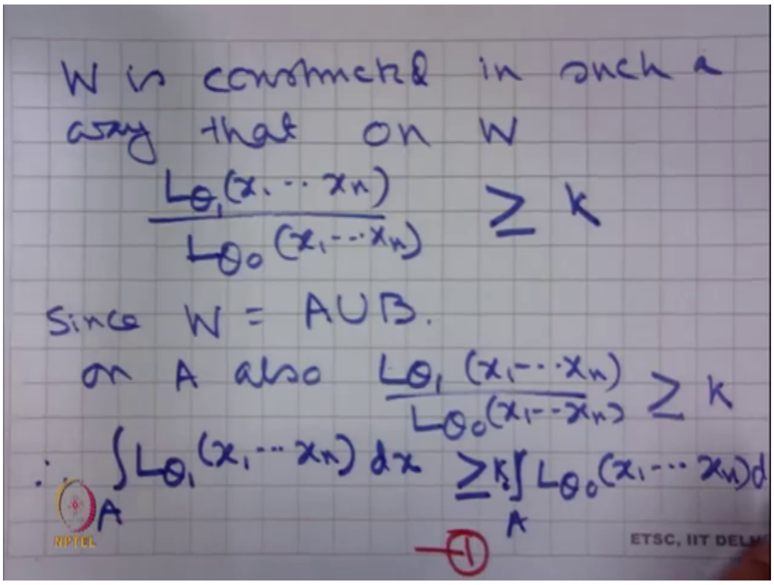
So, we want to prove this we are given W is of size alpha and W 1 is of size less than equal to alpha therefore, integration over W of L theta 0 of x 1, x 2, x n dx is equal to alpha and integration over W 1 L theta 0 x 1, x 2, x n dx is equal to or less than equal to alpha. We need to show that integration over W L theta 1 of x 1, x 2, x n dx is greater than equal to integration over W 1 of a theta 1 of x 1, x 2, x n dx.

(Refer Slide Time: 07:25)



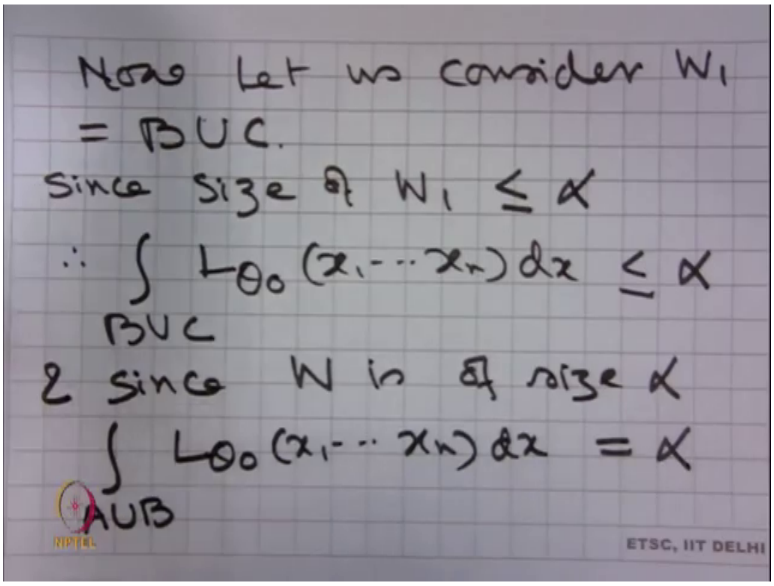
So, let us assume, this is the sample space theta. Let W be this is W which is A union B and suppose this is W 1 which is B union C. So, that is going to be that general diagram W is a subset of the parameter space theta and W 1 is another subset of the parameter space theta. In general they may have an intersection and according to the diagram the intersection is B.

(Refer Slide Time: 09:00)



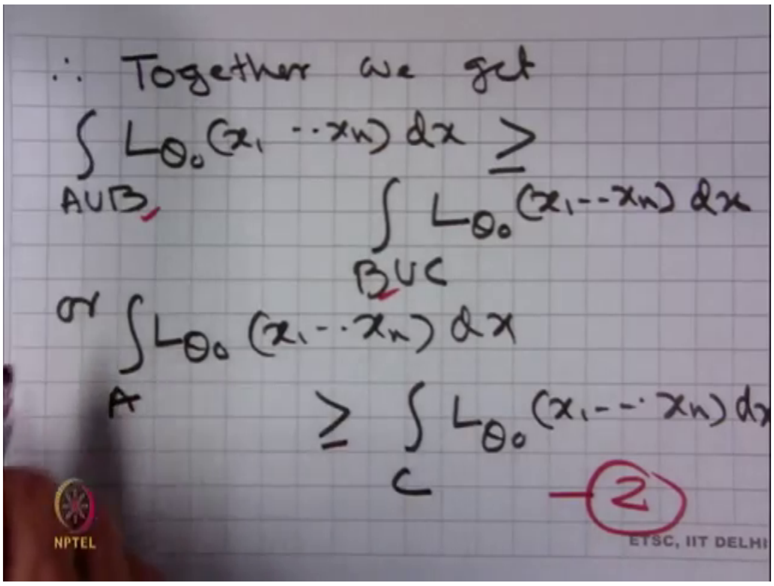
So, W is constructed in such a way that on W L theta 1 of x 1, x 2, x n upon L theta 0 of x 1, x 2, x n is greater than equal to K. Since W is equal to A union B, on A also L theta 1 of x 1, x 2, x n upon L theta 0 of x 1, x 2, x n is greater than equal to K. Therefore, integration over a of L theta 1 x 1, x 2, x n, dx is greater than equal to integration over a L theta 0 of x 1, x 2, x n dx multiplied by K. Let us call it inequality 1.

(Refer Slide Time: 10:35)



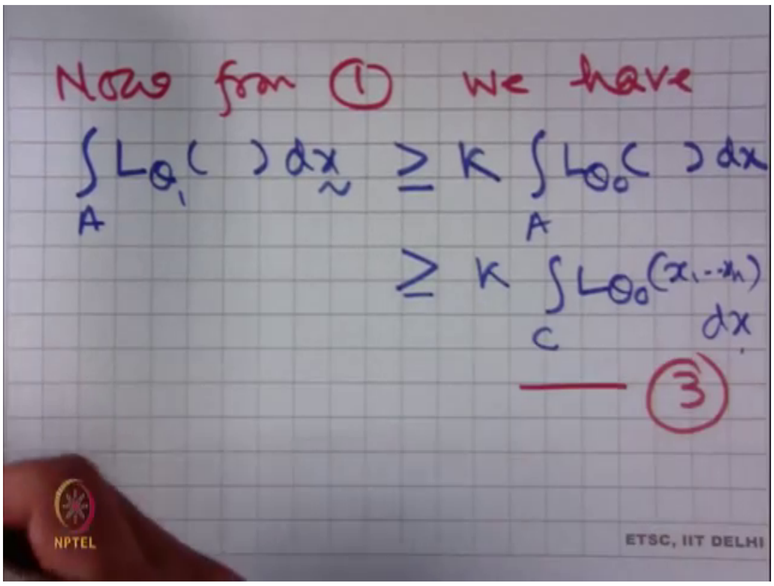
Now, let us consider W 1 is equal to B union C. Since size of W 1 is less than equal to alpha, therefore integration over B union C L theta 0 of x 1, x 2, x n dx is less than equal to alpha. And since W is of size alpha integration over A union B, L theta 0 of x 1, x 2, x n dx is equal to alpha. Therefore, together we get integration over A union B of L theta 0 of x 1, x 2, x n dx is greater than equal to the integration over B union C L theta 0 of x 1, x 2, x n dx or integration of over A, L theta 0 of x 1, x 2, x n dx is greater than equal to integration over C, L theta 0 x 1, x 2, x n dx this happens because B is common to both and therefore we can take away the B part and we get this inequality.

(Refer Slide Time: 11:49)



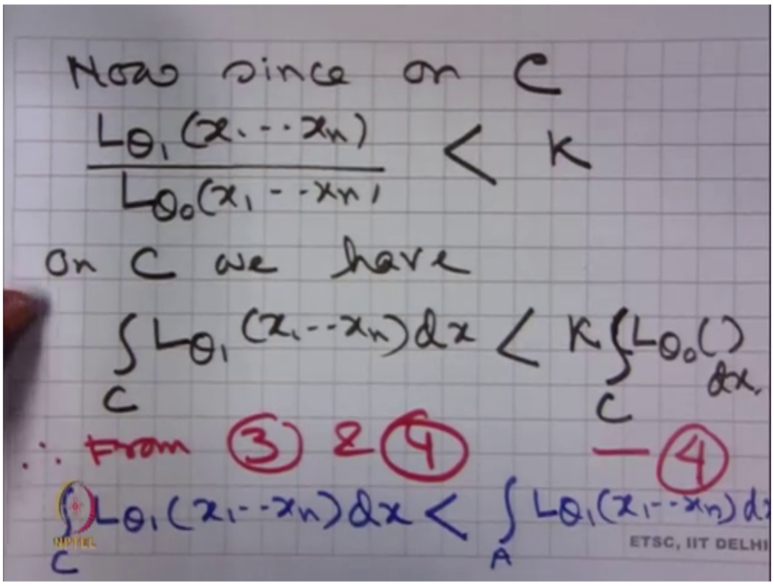
Let us call it 2.

(Refer Slide Time: 12:59)



Now, from 1, we have you can see from here that L theta 1 is greater than equal to K times L theta naught 1 A. So, if I use little bit of shortcut let me write it as integration over A, L theta 1 of dx is greater than equal to K times integration over a L theta naught of dx. And from 2 since we have integration over A, L theta naught is greater than equal to integration over C of L theta naught we can write this is greater than equal to K integration over C L theta naught of x 1, x 2, x n dx. So, let us call it 3.

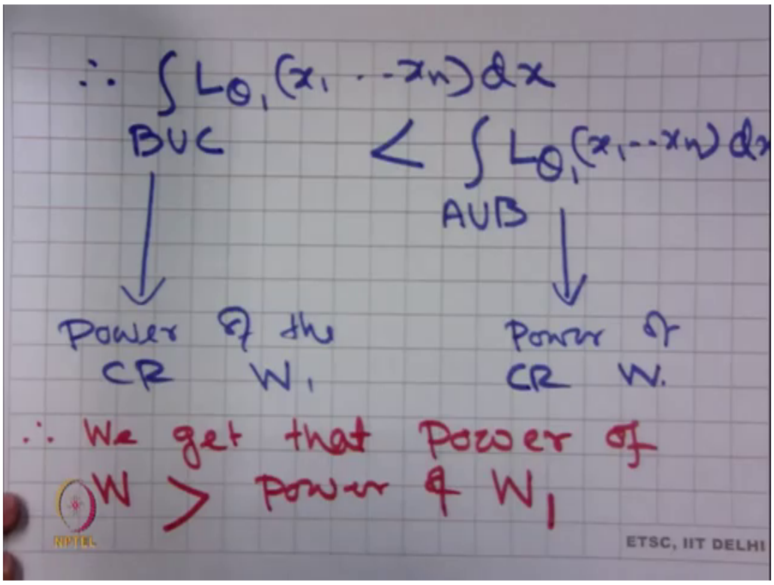
(Refer Slide Time: 14:25)



Now, since on C L theta 1 of x 1, x 2, x n upon L theta 0 x 1, x 2, x n is less than Km on C we have integration over C L theta 1 of x 1, x 2, x n dx is less than K times L theta 0 of x 1, x 2, x n dx. So, let us call it 4.

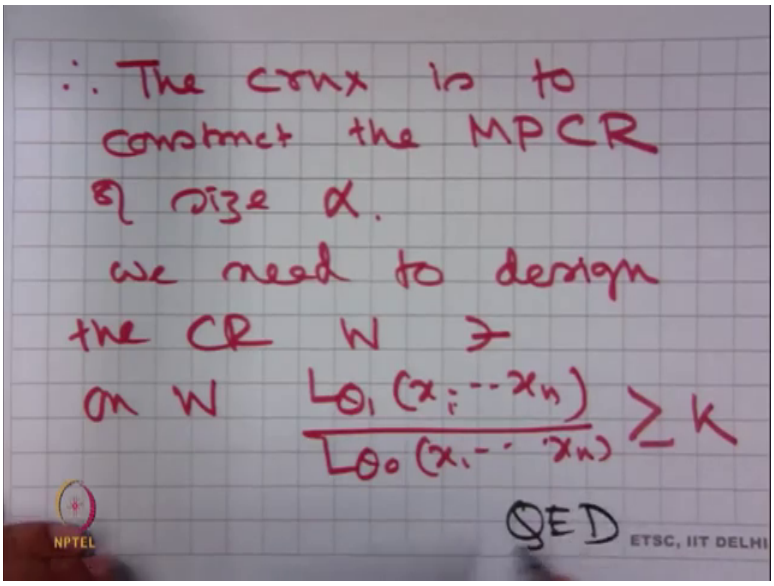
So, from 3 and 4 from here we have integration over C, L theta 1 of x 1, x 2, x n is less than this which in turn is less than this which in turn is less than this. Therefore, together we write the integration over C L theta 1 of x 1, x 2, x n dx is less than the integration over A, L theta 1 of x 1, x 2, x n dx.

(Refer Slide Time: 16:18)



Therefore, the integration over B union C of L theta 1 x 1, x 2, x n dx is less than integration over A union B of L theta 1 of x 1, x 2, x n dx. And by definition we know this is the power of the critical region W 1 and this is the power of critical region W therefore, we get that power of W is greater than power of W 1.

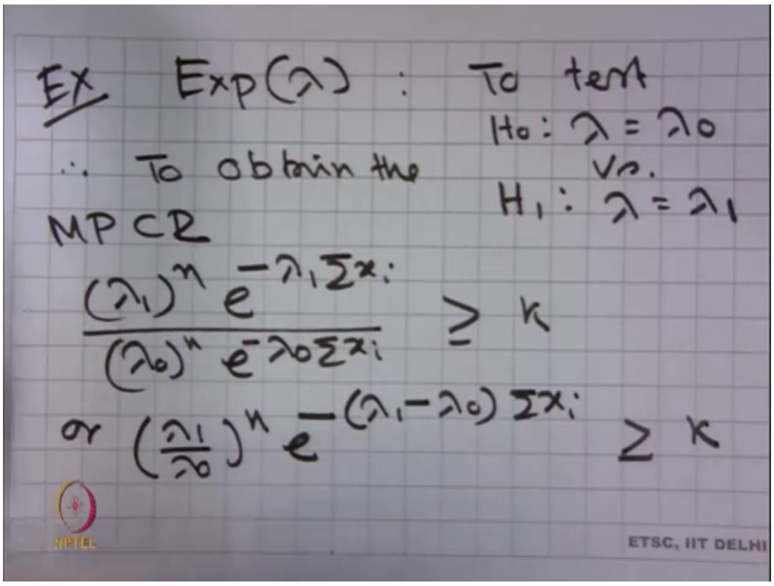
(Refer Slide Time: 17:30)



Therefore the crux is to construct the most powerful critical region of size alpha we need to design the critical region W such that on W L theta 1 of x 1, x 2, x n upon L theta 0 of x 1, x 2, x n is greater than equal to K.

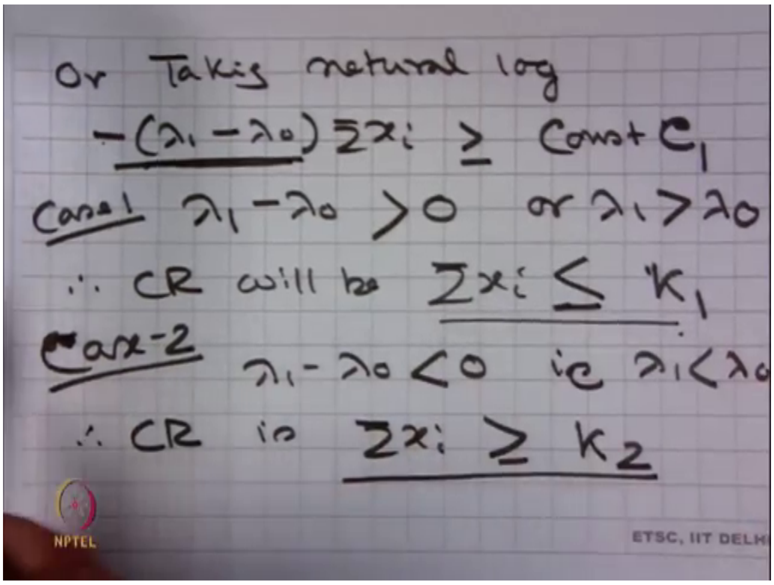
Now, in the last class I have given you two exercises to find the most powerful critical region, to test lambda is equal to lambda 0, against lambda is equal to lambda 1 for exponential distribution. So, let me relook at the problem.

(Refer Slide Time: 19:00)



So, example, the exponential lambda 2 test H naught lambda is equal to lambda 0 versus H 1 lambda is equal to lambda 1. Therefore, to obtain the most powerful critical region we have tried this that lambda 1 to the power n, e to the power minus lambda 1 sigma x i upon lambda 0 to the power n e to the power minus lambda 0 sigma x i is greater than equal to K or lambda 1 upon lambda 0 whole to the power n, e to the power minus lambda 1 minus lambda 0 sigma x i is greater than equal to K.

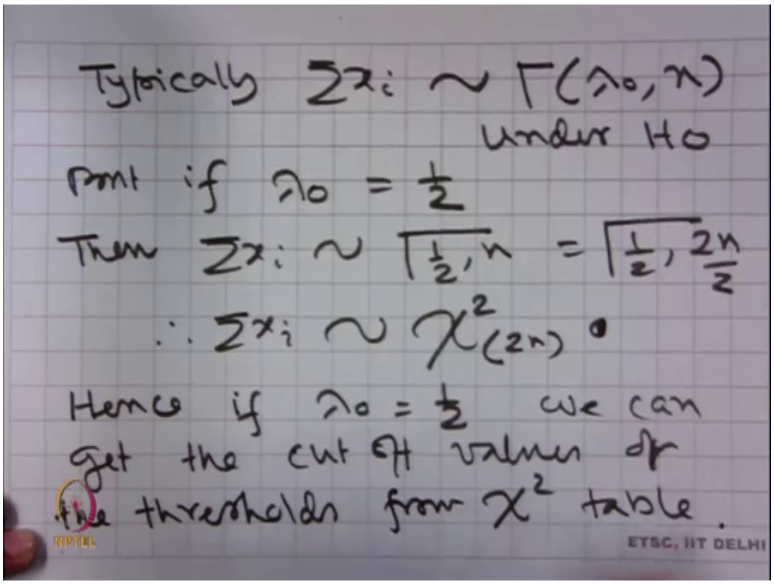
(Refer Slide Time: 20:09)



Or taking natural log minus lambda 1 minus lambda 0 sigma x i is greater than equal to some constant say C 1 which we get by manipulating the algebraic terms. But I am not interested in that I am interested in the relative values of lambda 1 and the lambda 0.

So, case 1, lambda 1 minus lambda 0 is greater than 0 or lambda 1 greater than lambda 0. In that case this term becomes negative with the minus sign therefore, the critical region will be sigma x i is less than equal to some constant say K 1. And case 2 lambda 1 minus lambda 0 is less than 0, that is lambda 1 is less than lambda 0 in that case this is going to be positive therefore, critical region is going to be sigma x i is greater than equal to some constant K 2. Up to this we have done in the last class.

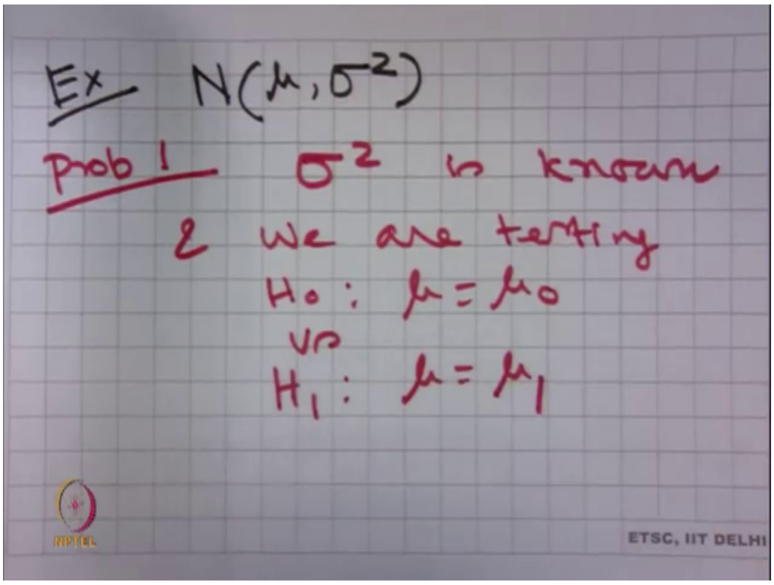
(Refer Slide Time: 22:11)



Typically, sigma x i is distributed as gamma with lambda naught and n under H naught, but if lambda naught is equal to half, then sigma x i is distributed as gamma half comma n which is equal to gamma half comma 2, n by 2 and therefore, sigma x i is distributed as chi square with 2 n degrees of freedom. Hence if lambda naught is equal to half we can get the cutoff values or the thresholds from chi square table.

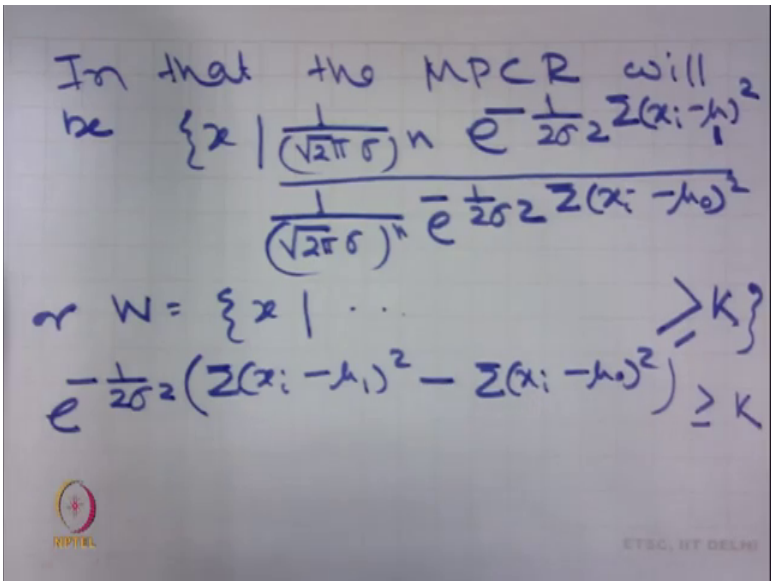
This is a very special case of exponential distribution. In general it will not work, but if we are testing with lambda naught is equal to half then we get from chi square table. Otherwise as I mentioned in the last class that we have to integrate or we have to do it numerically to find that constant, so that we can construct the most powerful critical region another example, that I did in the last class is normal mu coma sigma square.

(Refer Slide Time: 24:19)



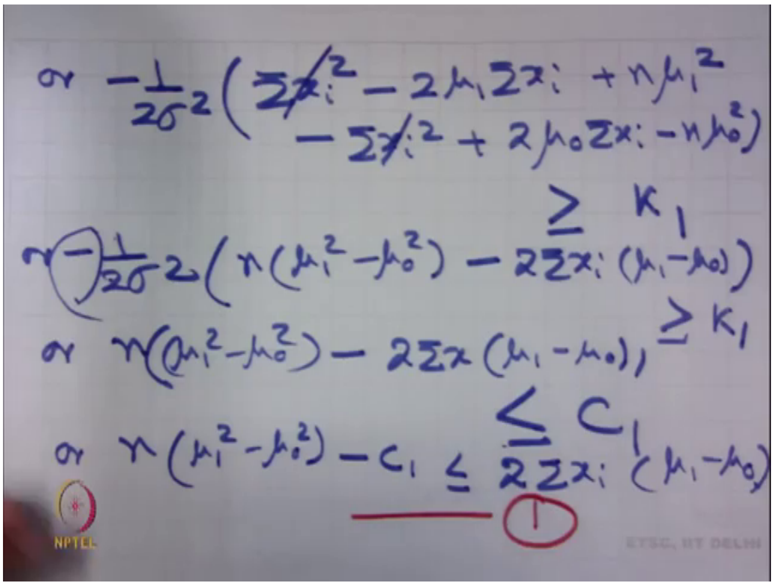
Let us look at problem 1, sigma square is known and we are testing H naught mu is equal to mu naught versus H 1 mu is equal to mu 1.

(Refer Slide Time: 25:05)



And we found that in that case, the most powerful critical region will be all those x such that 1 over root over 2 pi sigma whole to the power n, e to the power minus 1 upon 2 sigma square sigma x i minus mu whole square upon mu 1 whole square upon 1 over root over 2 pi sigma whole to the power n, e to the power minus 1 upon 2 sigma square. Sigma square is known for both of them and it is the same sigma x i minus mu naught whole square this ratio has to be greater than equal to K or W is equal to all those x, such that e to the power minus 1 upon 2 sigma square into sigma x i minus mu 1 whole square minus sigma x i minus mu naught whole square is greater than equal to K.

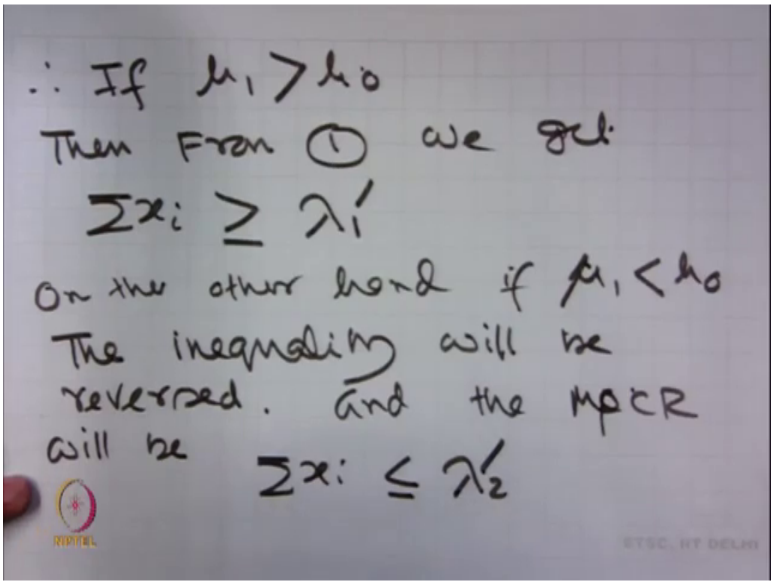
(Refer Slide Time: 26:43)



Or by taking natural log minus 1 upon 2 sigma square into sigma x i square minus 2 mu 1 sigma x i plus m mu 1 square minus sigma x i square plus 2 mu 0 sigma x i minus n mu naught square is greater than equal to some constant say K 1 or minus 1 upon 2 sigma square. Now, this gets cancelled and we have n times mu 1 square minus mu naught square minus 2 sigma x i into mu 1 minus mu naught is greater than equal to K 1 or n times mu 1 square minus mu naught square minus 2 sigma x i mu 1 minus mu naught is less than equal to some constant say C 1.

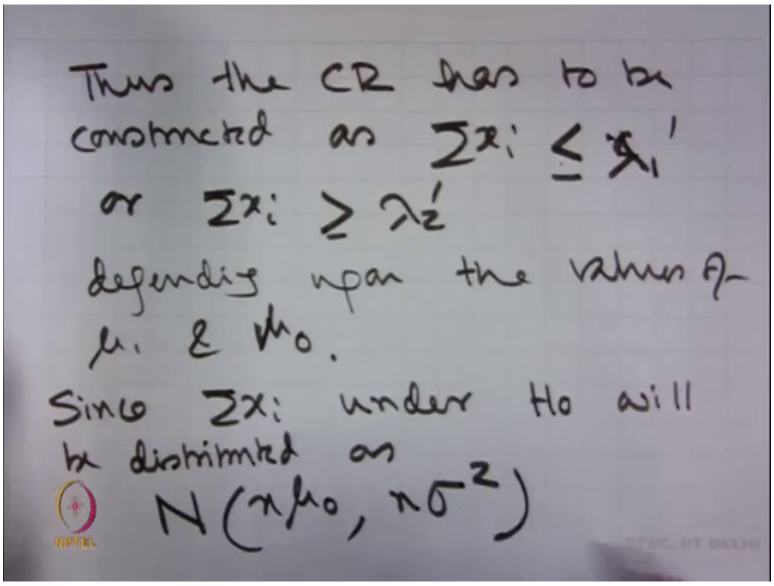
This comes because this minus sign is there therefore, the inequality will be reversed and when we multiply by 2 sigma square we get a different constant C 1 or in mu 1 square minus mu naught square minus C 1 is less than equal to 2 sigma x i into mu 1 minus mu naught. So, let us call it 1.

(Refer Slide Time: 29:00)



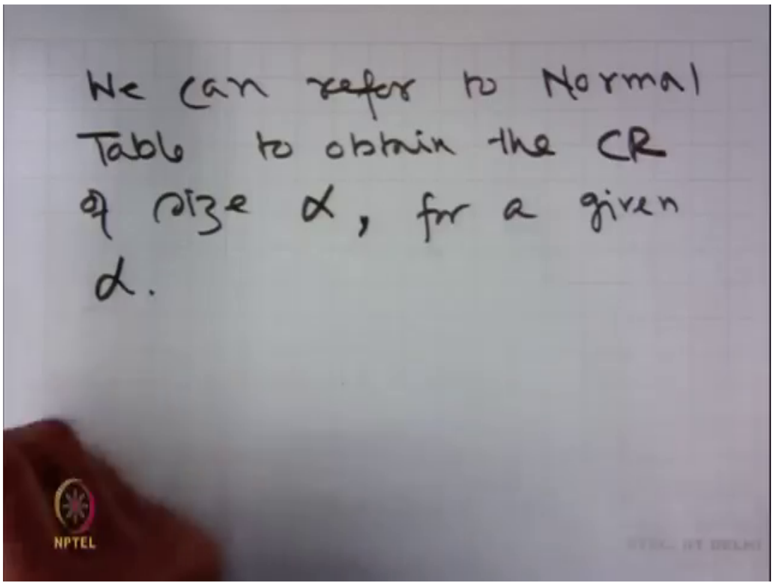
Therefore, if mu 1 is greater than mu naught, then from 1 we get if we divide both the sides by mu 1 minus mu naught which is positive we get that sigma x i has to be greater than equal to some constant lambda 1 prime. As you can see from here mu 1 minus mu naught is positive therefore, if I divide both the sides then we get sigma x i has to be greater than some constant. On the other hand if mu 1 is less than mu naught, then if we divide both sides by mu 1 minus mu naught. The inequality will change will be reversed and the MPCR will be sigma x i less than equal to some constant lambda 2 prime.

(Refer Slide Time: 30:59)



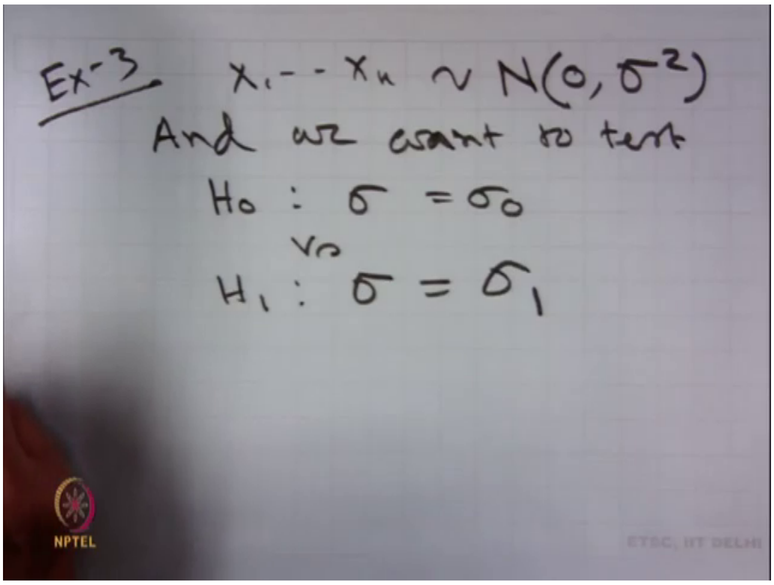
Thus, the critical region has to be constructed as sigma x i less than equal to some constant lambda 1 for sigma x i greater than equal to some constant lambda 2 prime, depending upon the values of mu 1 and mu naught. Since sigma x i under H naught will be distributed as normal with N mu naught and the variance is going to be n sigma square.

(Refer Slide Time: 32:17)



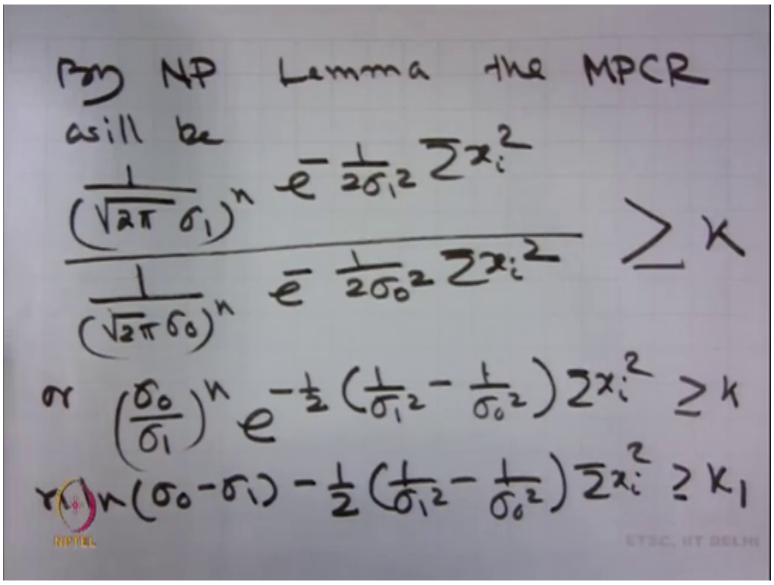
We can refer to normal table to obtain the critical region of size alpha, for a given alpha.

(Refer Slide Time: 32:55)



Now, suppose we want to test for sigma. Suppose we have the following problem x 1, x 2, x n are from normal with mean 0 and variance sigma square. And we want to test H naught sigma is equal to sigma naught versus H 1 sigma is equal to sigma 1. Note that we have kept the mean to be 0. This is only for ease of calculation if that normal distribution has the mean known which is mu then instead of x 1, x 2, x n we can always look at y 1, y 2, y n, where y is x i minus mu and therefore, we will come back to normal 0 sigma square.

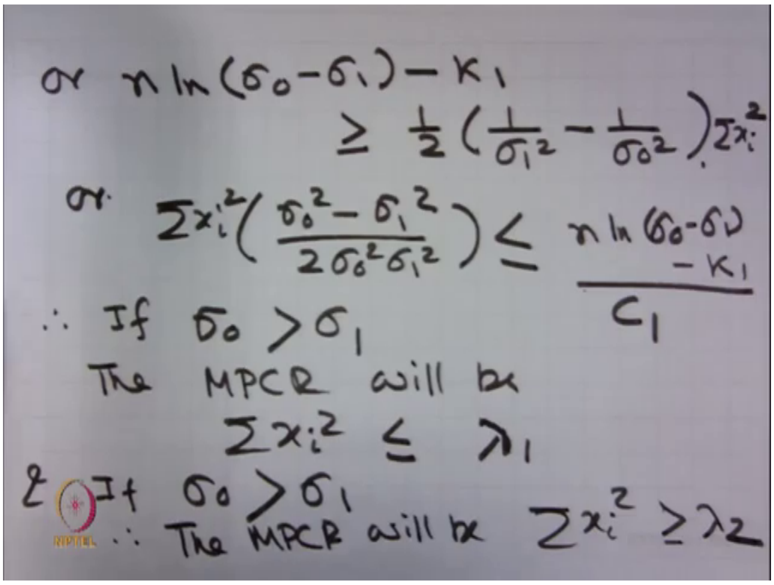
(Refer Slide Time: 34:07)



By Neyman-Pearson Lemma the most powerful critical region will be 1 upon root over 2 pi sigma 1 whole to the power n into e to the power minus 1 upon 2 sigma 1 square sigma x i square upon 1 over root over 2 pi sigma naught whole to the power n, e to the power minus 1 upon 2 sigma naught square sigma x i square.

This has to be greater than equal to K or sigma naught upon sigma 1 whole to the power n, e to the power minus half; 1 upon sigma 1 square minus 1 upon sigma naught square into sigma x i square is greater than equal to K. Or taking log in log sigma naught minus sigma 1 minus half 1 upon sigma 1 square minus 1 upon sigma naught square sigma x i square is greater than equal to some constant which will come through log operation on K. So, let me call it K 1.

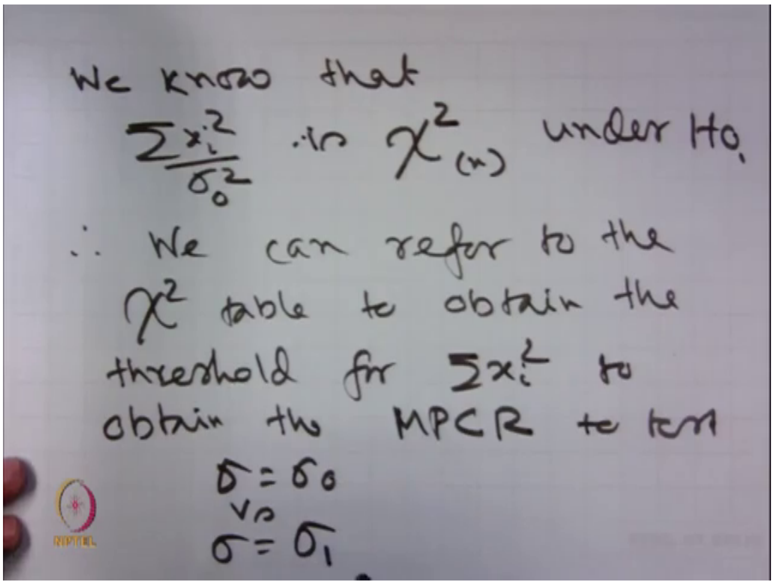
(Refer Slide Time: 35:51)



Or n log of sigma naught minus sigma 1 minus K 1 is greater than equal to half of 1 upon sigma 1 square minus sigma naught square or into sigma x i square. Or sigma x i square into sigma naught square minus sigma 1 square upon 2 sigma naught square sigma 1 square is less than equal to n log n sigma naught minus sigma 1 minus K 1. Let me call it C 1.

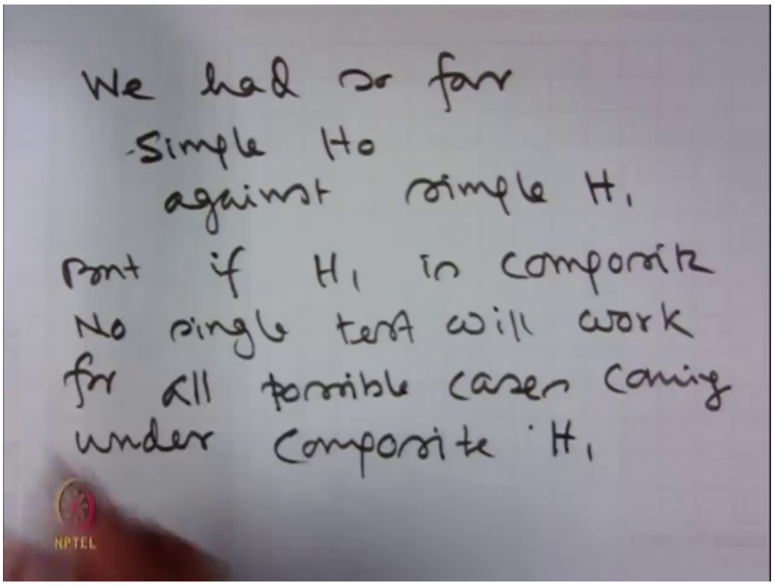
Therefore, if sigma naught is greater than sigma 1 then this term becomes positive the MPCR will be sigma x i square is less than equal to some constant say lambda 1. And if sigma naught is greater than sigma 1 then because of the negative sign the inequality will be reversed and therefore, the MPCR will be sigma x i square which greater than equal to lambda 2.

(Refer Slide Time: 38:01)



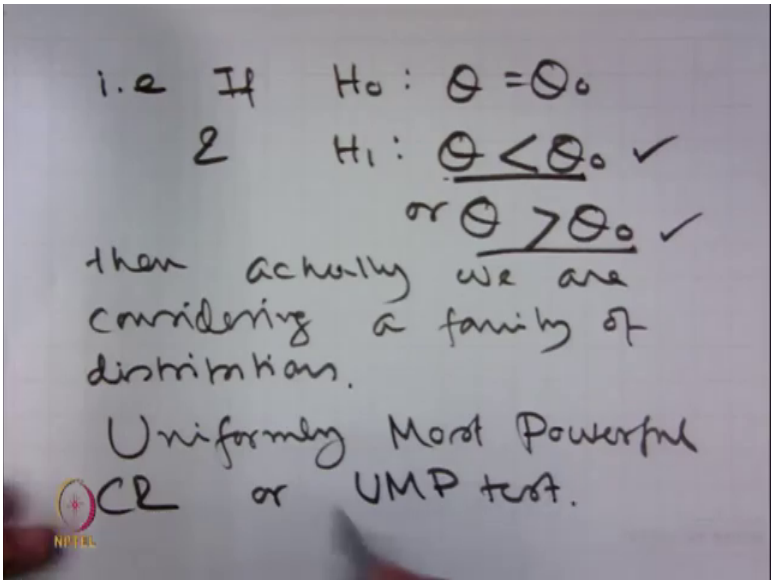
Now, we know that sigma x i square upon sigma square is chi square with n degrees of freedom under H naught. Therefore, we can refer to the chi square table to obtain the threshold for sigma x i square to obtain the most powerful critical region to test sigma is equal to sigma naught versus sigma is equal to sigma 1.

(Refer Slide Time: 39:19)



In all the examples that I have given so far we had so far simple H naught against simple H 1. But if H 1 is composite, no single test will work for all possible cases coming under composite H 1.

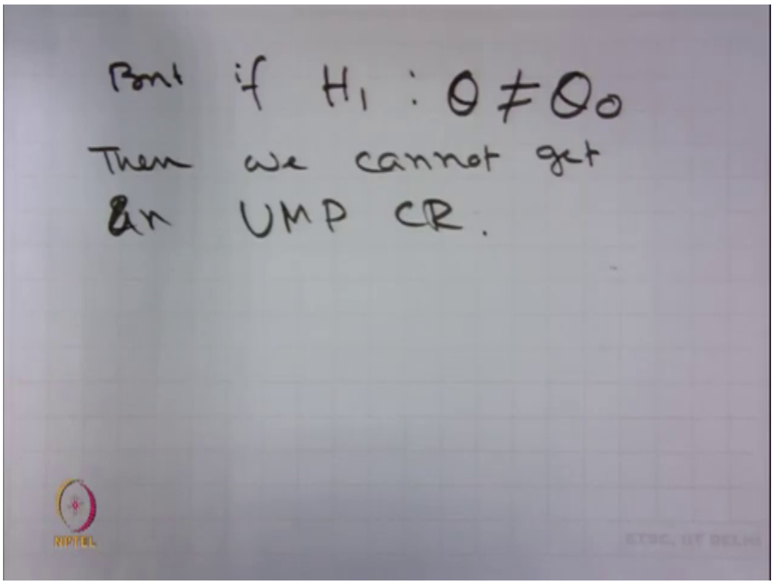
(Refer Slide Time: 40:24)



That is, if H naught is theta is equal to theta naught and H 1 is theta less than theta naught or theta greater than theta naught then actually we are considering a family of distributions.

But in some cases as we have seen in all the examples, when we are creating the most powerful critical region actually the value of theta 1 has not come into picture in constructing the MPCR. Therefore, if we are testing alternatives of this type that is only one sided inequality then the MPCR that we have obtained there that we will work for all the family of distribution satisfying this or satisfying this. And therefore, the most powerful critical region obtained in that way can be called uniformly most powerful critical region or corresponding test is called uniformly most powerful test.

(Refer Slide Time: 42:19)



But if H 1 is of the form theta naught equal to theta naught then we cannot get and uniformly most powerful critical region in general.

Ok students, with that I conclude my series of lectures on statistical inference. Over the last 21 lectures we started with some basic probability distributions and we have stressed upon on several topics including order statistics, theory of estimation and of course, testing of hypothesis. I hope you found that course useful and if you have any doubts or queries please do not hesitate to contact me or my TLs to clarify your doubts.

Thank you very much.