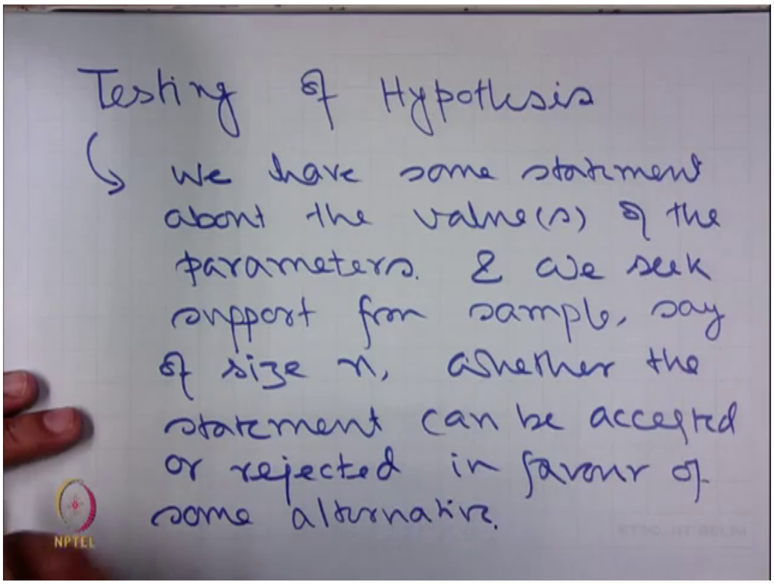
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 20
Statistical Inference

Welcome, students to the MOOCS lecture series on Statistical Inference, this is lecture number 20.

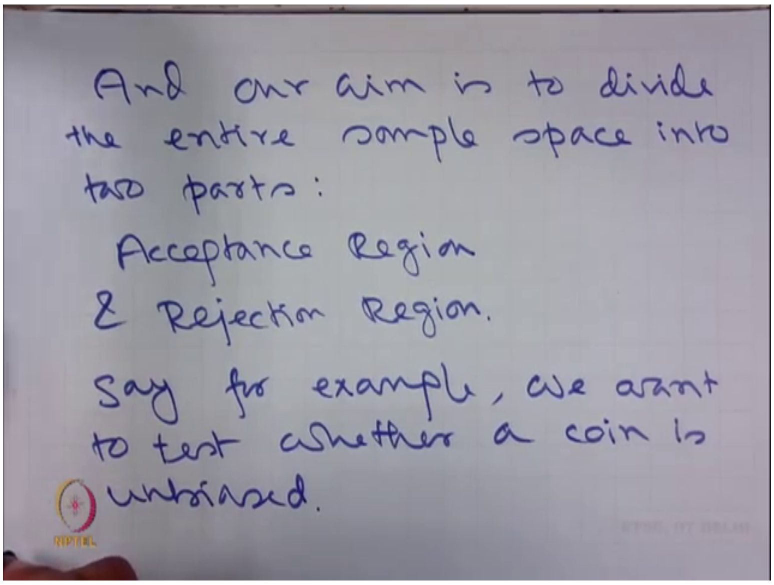
(Refer Slide Time: 00:37)



If you remember we are doing testing of hypothesis as the last topic of the series of lectures. So, testing of hypothesis means we have some statement about the value or values of the parameters. And we seek support from sample say of size n, whether the statement can be accepted or rejected in favor of some alternative.

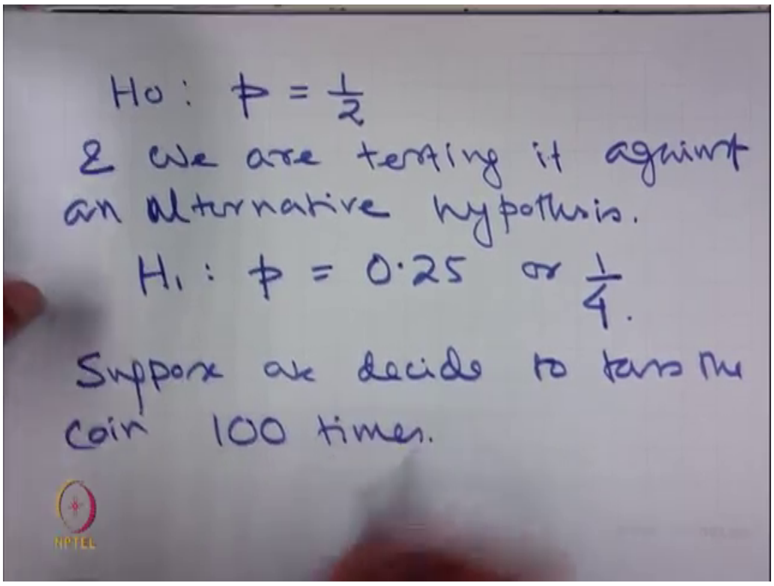
So, that is the whole purpose of testing of hypothesis.

(Refer Slide Time: 02:18)



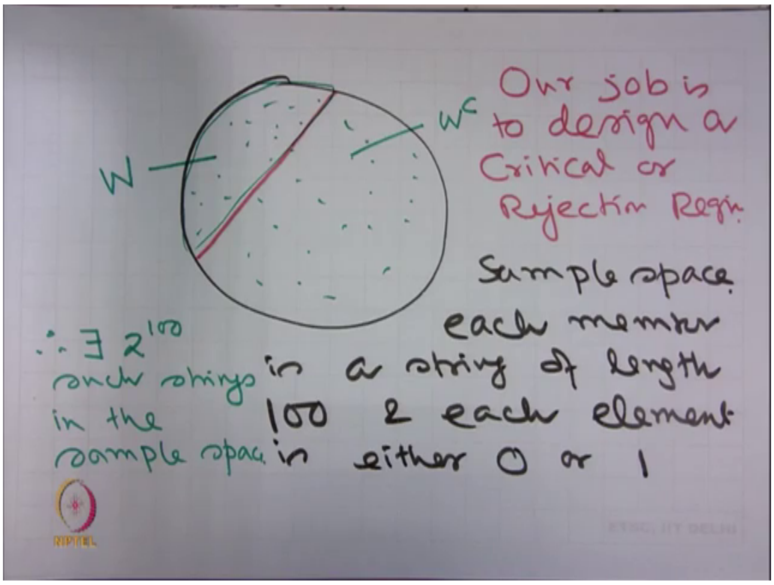
Our aim is to divide the entire sample space into two parts: Acceptance region and Rejection region Say for example, we want to test whether a coin is unbiased. So, if we look at a simple null hypothesis.

(Refer Slide Time: 03:33)



Then our h naught is p is equal to half and we are testing it against an alternative hypothesis H 1 p is equal to say 0.25 or 1 by 4. Suppose, we decide to toss the coin 100 times then what is going to be the sample space.

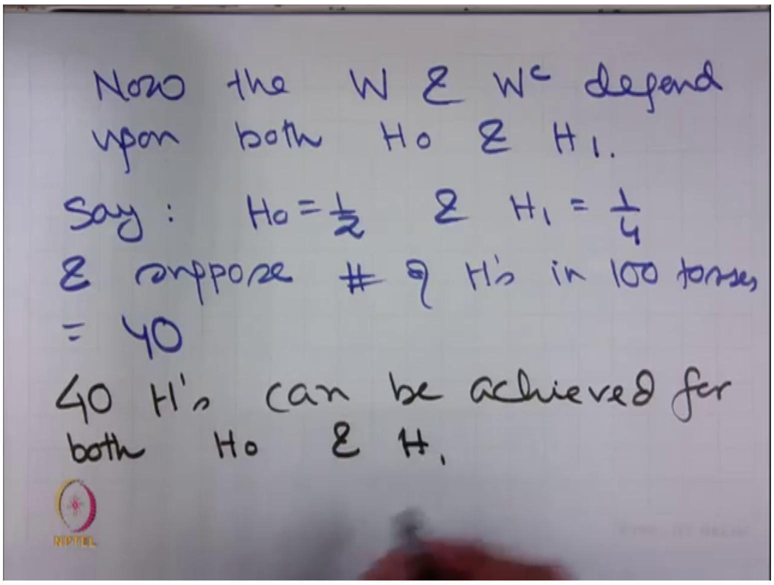
(Refer Slide Time: 04:51)



So, the sample space is and each member of the sample space is a string of length 100 and each element is either 0 or 1. Therefore, there are 2 to the power 100 such strings in the sample space. Suppose these are the points in the sample space. Now our job is to design a critical region or rejection region say we decide this is to be our critical region.

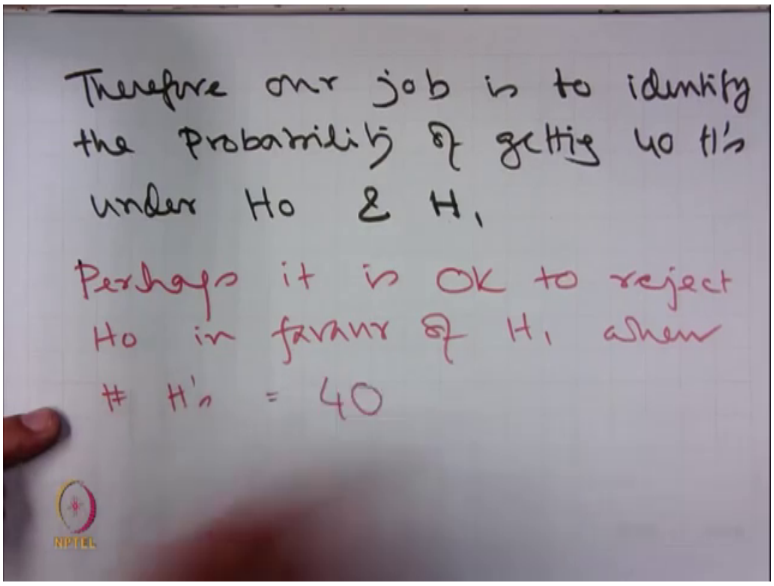
So, that if the outcome is falling in this part of the sample space then we are going to reject the null hypothesis. So, this is called w and this side is called w complement which gives us the acceptance region.

(Refer Slide Time: 07:03)



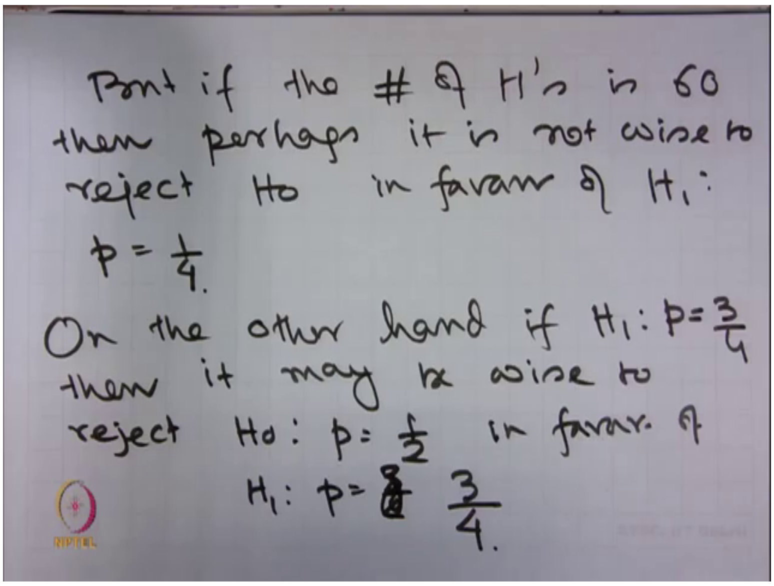
Now, the W and W complement depend upon both H naught and H 1 say H naught is half as I have already mentioned and H 1 is 1 by 4 and suppose number of heads in 100 tosses is equal to 40. Now, 40 heads can be achieved for both H naught and H 1.

(Refer Slide Time: 08:16)



Therefore our job is to identify the probability of getting 40 heads under H naught and H 1. Perhaps, it is to reject H naught in favor of H 1 when number of head is 40.

(Refer Slide Time: 09:26)

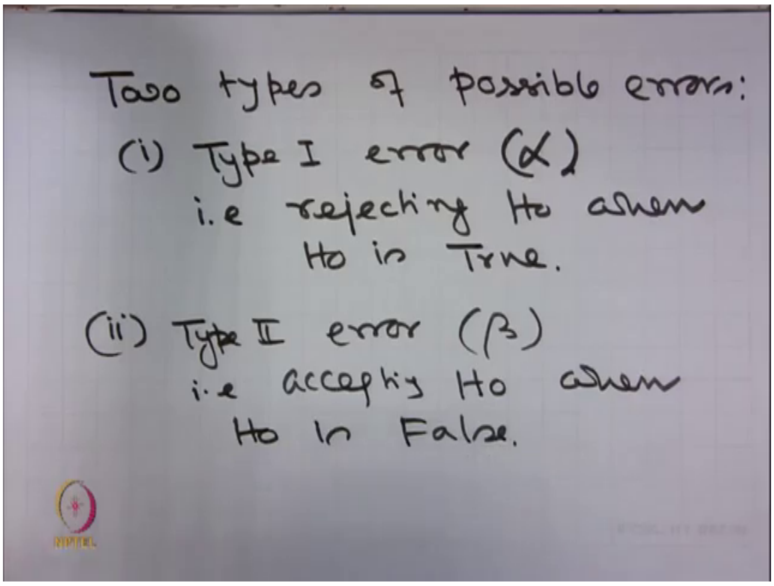


But if the number of heads is 60 then perhaps it is not wise to reject the null hypothesis H naught in favor of H 1 that p is equal to 1 by 4. On the other hand, if H 1 is p is equal to

3 by 4 then it may be wise to reject H naught p is equal to half in favor of H 1 that p is equal to 3 by 4.

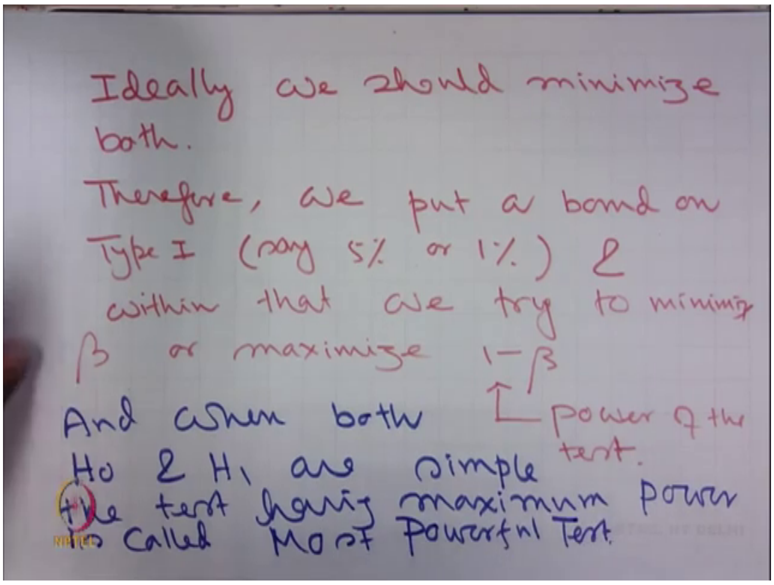
So, the job of a statistician for testing of hypothesis is that we have to decide that test. So, that we know the sample space and we need to decide the testing criteria. So, that we can accept or reject the null hypothesis against an alternative.

(Refer Slide Time: 11:25)



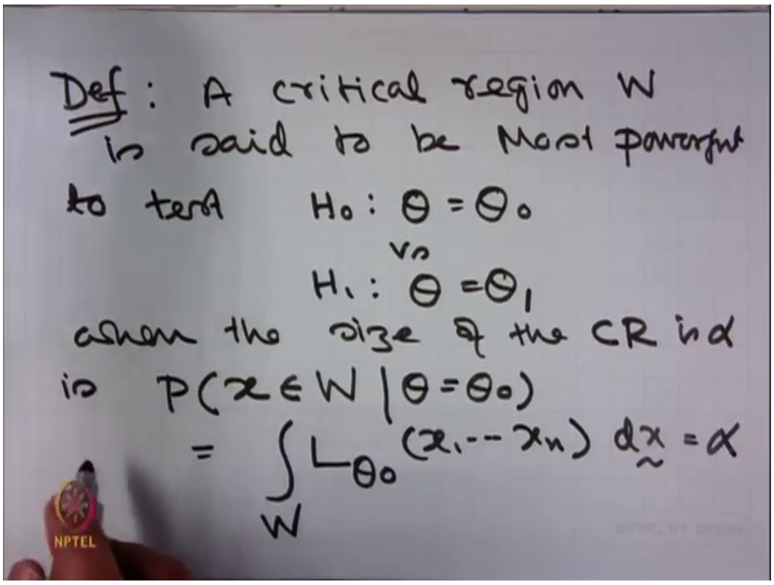
So, as I discussed two types of possible errors, one is type I error which you call alpha that is rejecting the H naught when h naught is true. That means, actually the coin is unbiased what we are rejecting that hypothesis based on the sample obtained. And, if I reject when the H naught is true then we are committing an error which you call type I error and type II error is which you call beta is accepting H naught when H 1 H naught is false.

(Refer Slide Time: 12:52)



Ideally, we should minimize both, but as I have explained that that cannot be done. Therefore, we put a bound on type I say 5 percent or 1 percent; that means, we can make a mistake only 5 out of 100 cases or 1 out of 100 cases. And within that we try to minimize beta or maximize 1 minus beta, this quantity is called the power of the test. And when both H naught and H 1 are simple, the test having maximum power is called the most powerful test.

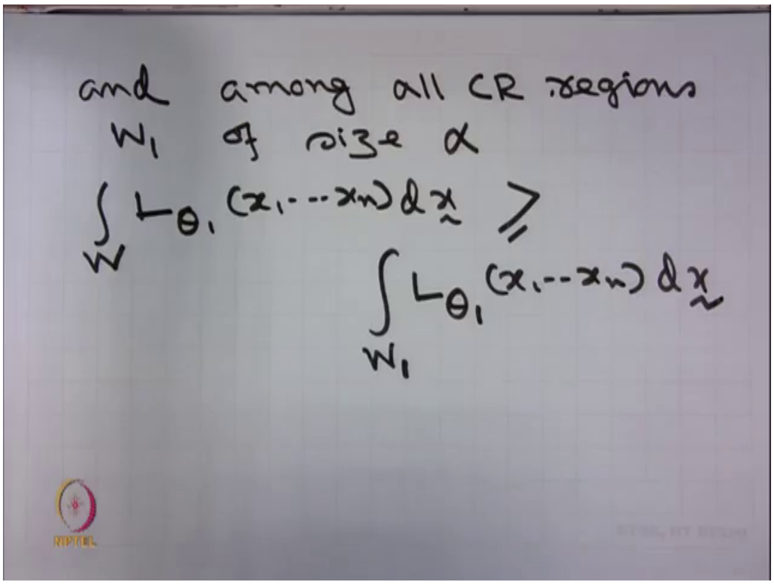
(Refer Slide Time: 15:14)



So, definition are critical region W, we said to be most powerful to test H naught theta is equal to theta naught versus H 1 theta is equal to theta 1. When the size of the critical region is alpha such that probability x belonging to W given theta is equal to theta naught is equal to integration over W L theta naught of x 1, x 2, x n dx is equal to alpha.

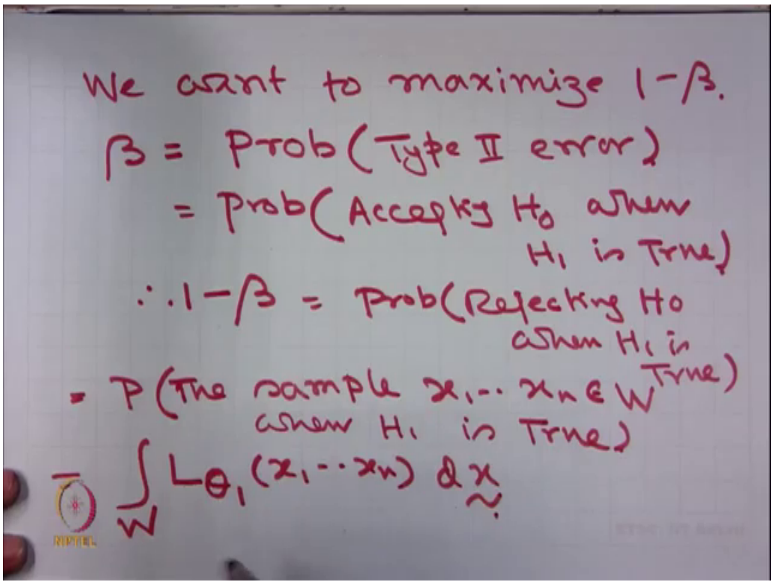
So, we are saying that the making a mistake of rejecting or correct null hypothesis is this probability that the obtained sample x will belong to the rejection region or critical region W, these probabilities integrating on the space W of L theta naught x 1, x 2, x n this is the joint pdf of obtaining the sample x 1, x 2, x n under the null hypothesis that is the parameter theta is equal to theta naught and that probability should be alpha and.

(Refer Slide Time: 17:37)



Among all critical regions W 1 of size alpha integration over W of L theta one of x 1, x 2, x n dx is greater than equal to integration of W 1 of L theta one of x 1, x 2, x n dx.

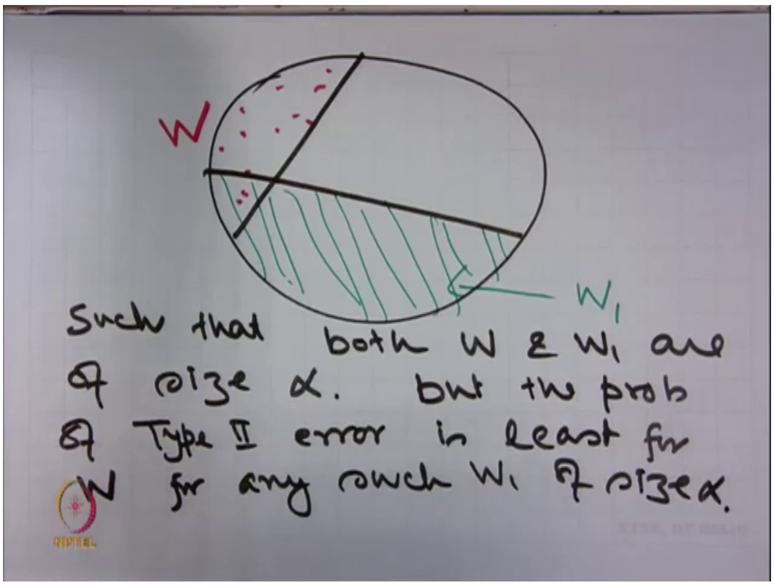
(Refer Slide Time: 18:29)



This is because we want to maximize 1 minus beta, what is beta? Beta is equal to probability of type II error is equal to probability of accepting H naught when H 1 is true. Therefore, 1 minus beta is equal to probability of rejecting H naught when H 1 is true is equal to probability that the sample x 1, x 2, x n belongs to the critical region when H 1 is true.

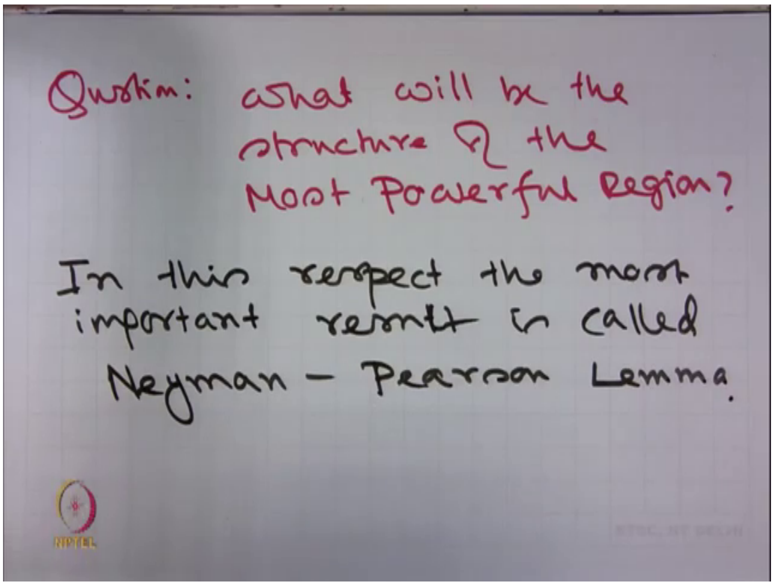
And therefore, this is, is equal to integration over W L theta 1, because H 1 is true the parameter is theta 1 and this is x 1, x 2, x n dx.

(Refer Slide Time: 20:19)



Therefore, if this is the sample space and suppose this is one critical region W. And this is another critical region, say W 1 such that both of them are of size alpha, but the probability of type II error is least for W for any such W 1 of size alpha.

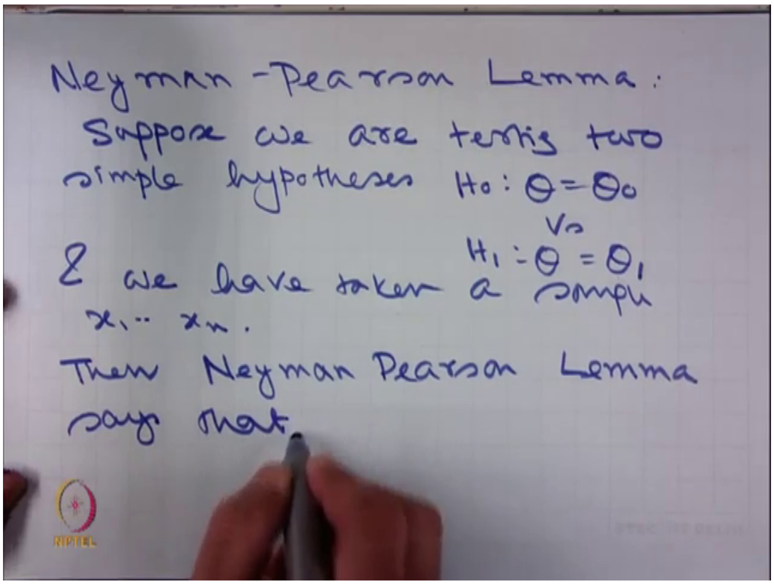
(Refer Slide Time: 21:42)



In this respect questions, what will be the structure of the most powerful region or in other words which function of the sample X 1, X 2, X n we should take?

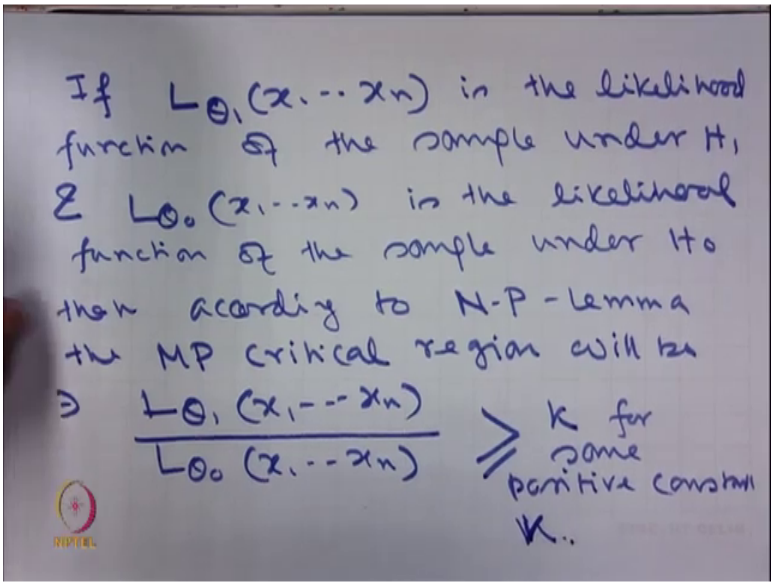
So, that based on that, we can make the decision of rejecting or accepting the null hypothesis and the corresponding test is most powerful. In this respect the most important result is called Neyman-Pearson Lemma. Today I will give you the Lemma and solve a few problems based on the Neyman-Pearson Lemma. In the next class I shall give you the proof of the Lemma so that you can understand how such a result has come.

(Refer Slide Time: 23:33)



Neyman-Pearson Lemma: suppose we are testing two simple hypothesis H naught theta is equal to theta naught versus H 1 theta is equal to theta 1 and we have taken a sample x 1, x 2, x n.

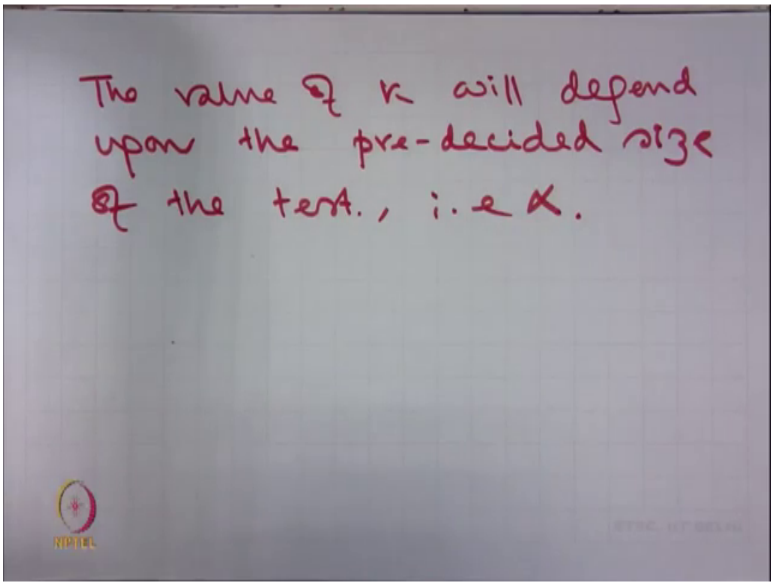
(Refer Slide Time: 24:55)



If L theta 1 x 1, x 2, x n is the likelihood function of the sample under H 1 and L theta naught x 1, x 2, x n is the likelihood function of the sample under H naught. Then according to Neyman-Pearson Lemma the most powerful critical region will be such that

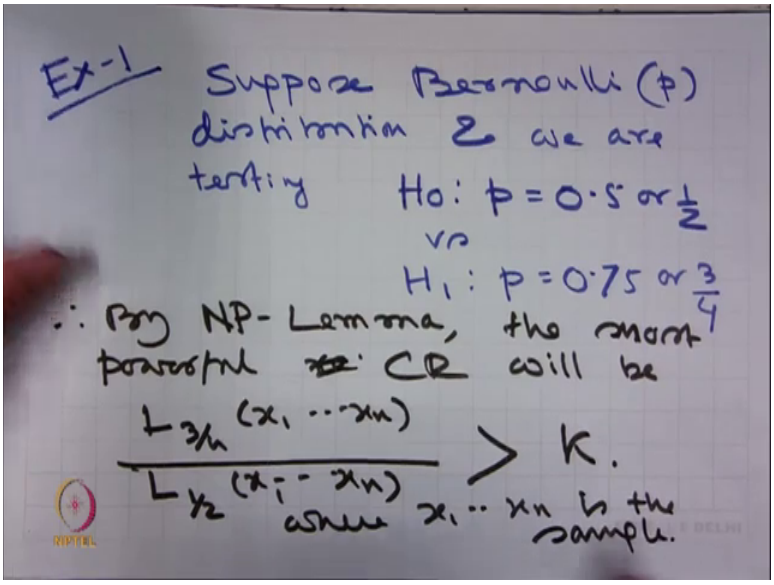
L theta one of x 1, x 2, x n upon L theta 0, x 1, x 2, x n is greater than K for some positive it is a greater than equal to K for some positive constant K.

(Refer Slide Time: 26:52)



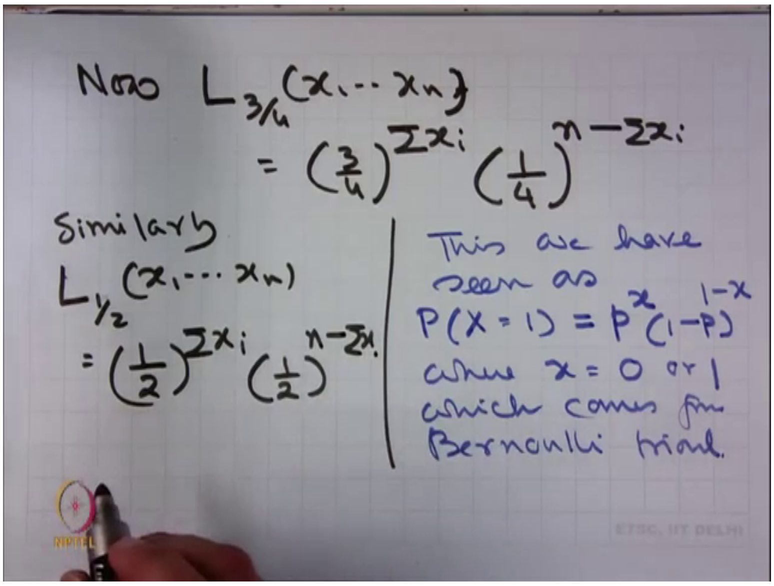
The value of K will depend upon the pre decided size of the test that is alpha. So, at the beginning as a statistician one will decide upon the size alpha whether it is 5 percent or 1 percent or 100 percent. And based on that we should get a constant such that L theta 1 of x 1, x 2, x n upon L theta naught of x 1, x 2, x n has to be greater than that constant. Before proving that as I said in today’s class I will solve a few problems using the above Neyman-Pearson Lemma.

(Refer Slide Time: 28:10)



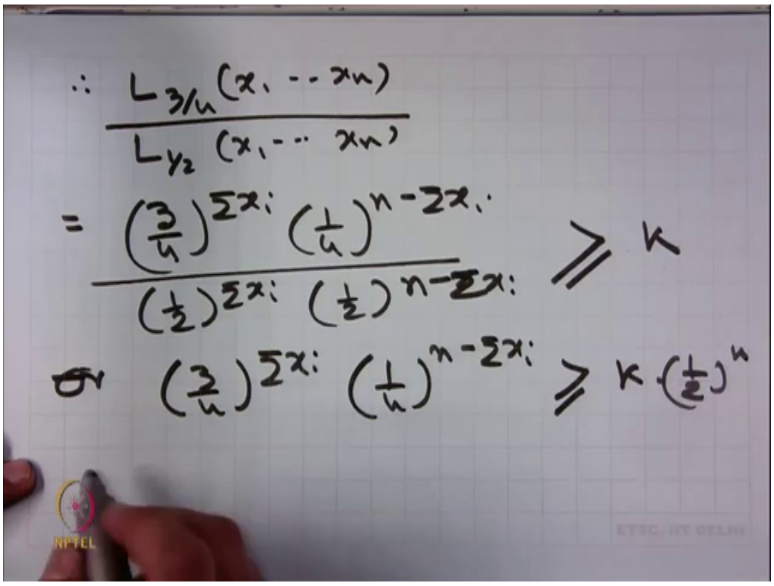
Let us consider the first example suppose we are tossing a coin. So, we are looking at Bernoulli p distribution and we are testing say H naught p is equal to 0.5 or half versus H 1 p is equal to 0.75 or 3 by 4. Therefore, by Neyman-Pearson Lemma, the most powerful region critical region will be L say 3 by 4 of x 1, x 2, x n upon L half of x 1, x 2, x n is greater than K where x 1, x 2, x n is the sample.

(Refer Slide Time: 30:16)



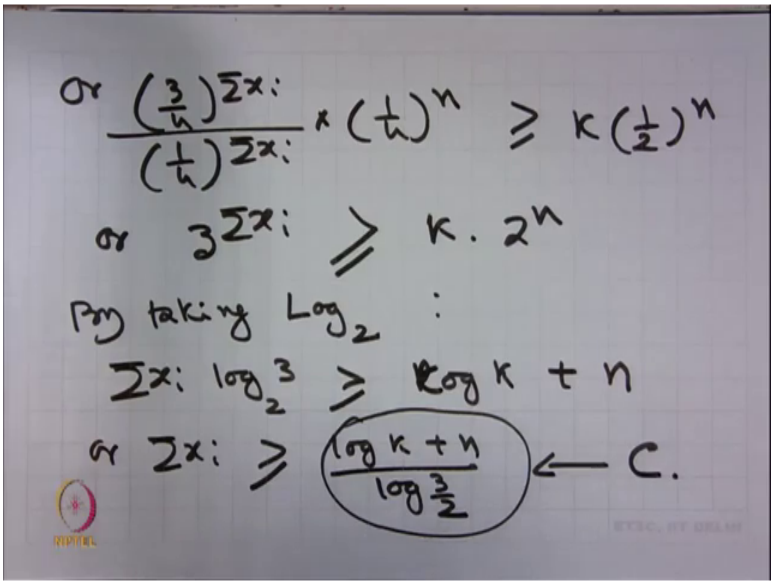
Now, L 3 by 4 of x 1, x 2, x n is equal to 3 by 4 to the power sigma x i into 1 by 4 to the power n minus sigma x i. This we have seen as probability X is equal to 1 is equal to P to the power x 1 minus P to the power 1 minus x where x is equal to 0 or 1 which comes from Bernoulli trial. Similarly, L half of x 1, x 2, x n is equal to half to the power sigma x i into half to the power n minus sigma x i.

(Refer Slide Time: 31:59)



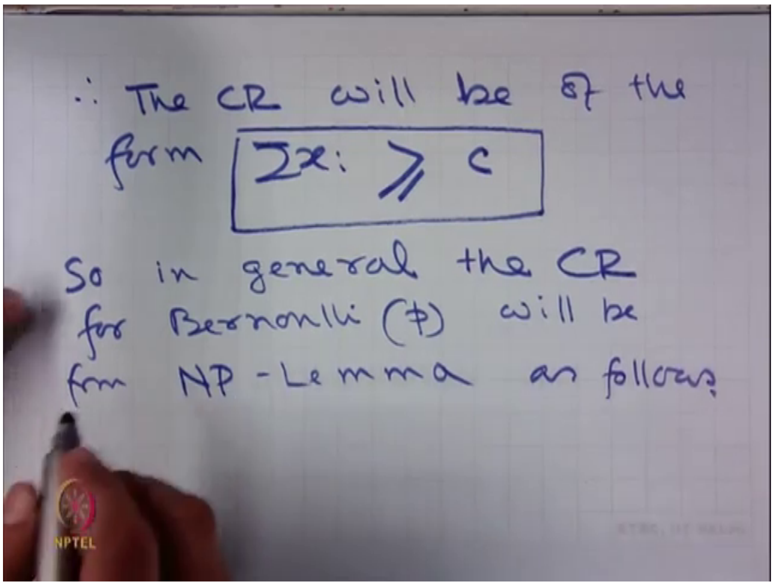
Therefore, L 3 by 4 of x 1, x 2, x n upon L half of x 1, x 2, x n is equal to 3 by 4 to the power sigma x i into 1 by 4 to the power n minus sigma x i into half to the power sigma x i into half to the power n minus sigma x i. And that has to be greater than equal to K. Or 3 by 4 to the power sigma x i 1 by 4 to the power n minus sigma x i is greater than equal to K times half to the power n.

(Refer Slide Time: 33:04)



Or 3 by 4 to the power sigma x i upon 1 by 4 to the power sigma x i into 1 by 4 to the power n is greater than equal to K into half to the power n. Or 3 to the power sigma x i is greater than equal to K times 2 to the power n. By taking log base 2, we get sigma x i log 3 2 the base 2 is greater than equal to log K plus n or sigma x i is greater than equal to log K plus n upon log 3 by 2. This is a constant, let me call it C.

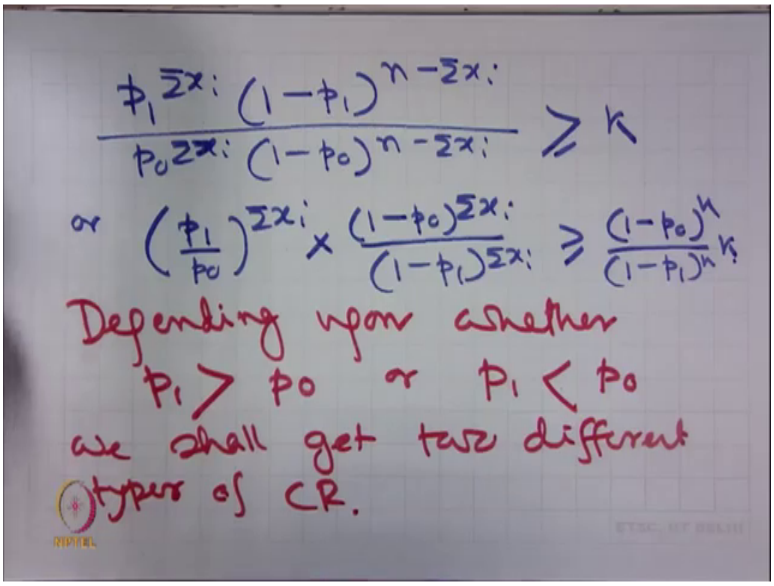
(Refer Slide Time: 34:38)



Therefore, the critical region will be of the form sigma x i is greater than equal to a constant C. Or in other words, we will reject the null hypothesis that the probability of a head for this coin is half, we will reject that in favor of that the probability of getting a head is 3 by 4, if sigma x i or the number of heads is greater than some constant; which is very intuitive because the alternative here is 3 by 4 which is bigger than the null hypothesis value that is half.

So, in general the critical region for Bernoulli p will be from NP Lemma as follows.

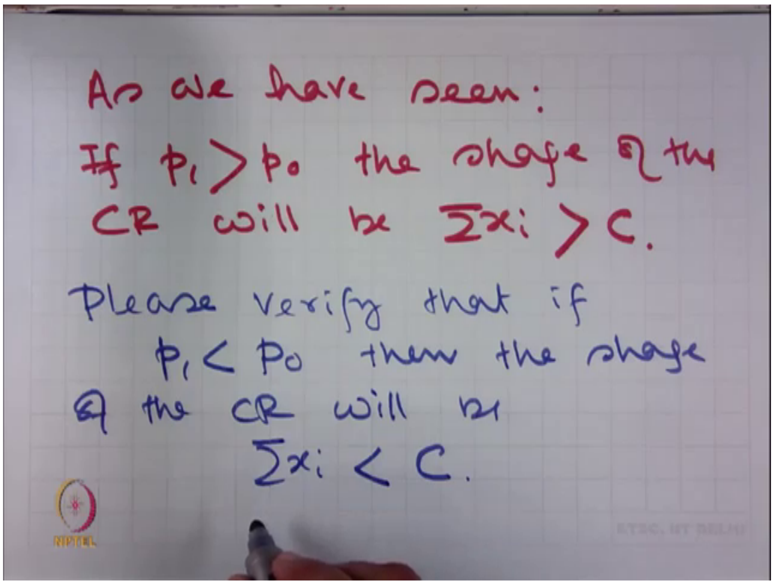
(Refer Slide Time: 36:15)



P 1 to the power sigma x i into 1 minus p 1 to the power n minus sigma x i upon p 0 to the power sigma x i into 1 minus p 0 n minus sigma x i is greater than equal to K. Or, p 1 upon p 0 sigma x i into 1 minus p 0 sigma x i upon 1 minus p 1 sigma x i is greater than equal to 1 minus p 0 to the power n upon 1 minus p 1 to the power n times K.

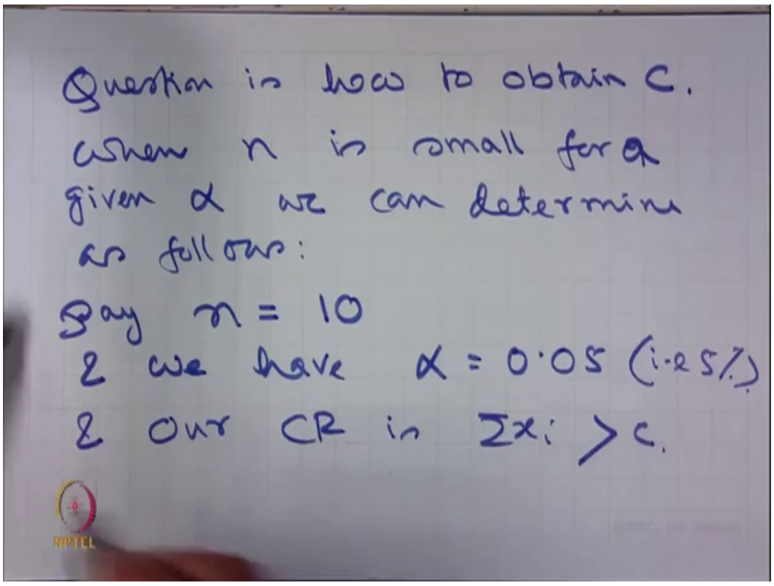
This is the general structure. So, depending upon whether p 1 is greater than p 0 or p 1 is less than p 0 we shall get two different types of Critical Region as we have seen.

(Refer Slide Time: 38:03)



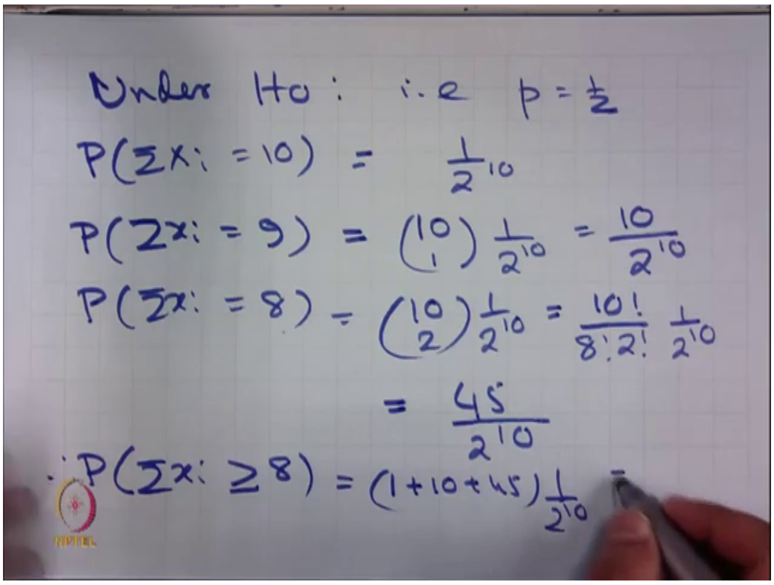
If p 1 is greater than p 0 the shape of the critical region will be sigma x i is greater than some constant. I will leave it on you, you verify that if p 1 is less than p 0, then the shape of the critical region will be sigma x i is less than some constant C.

(Refer Slide Time: 39:09)



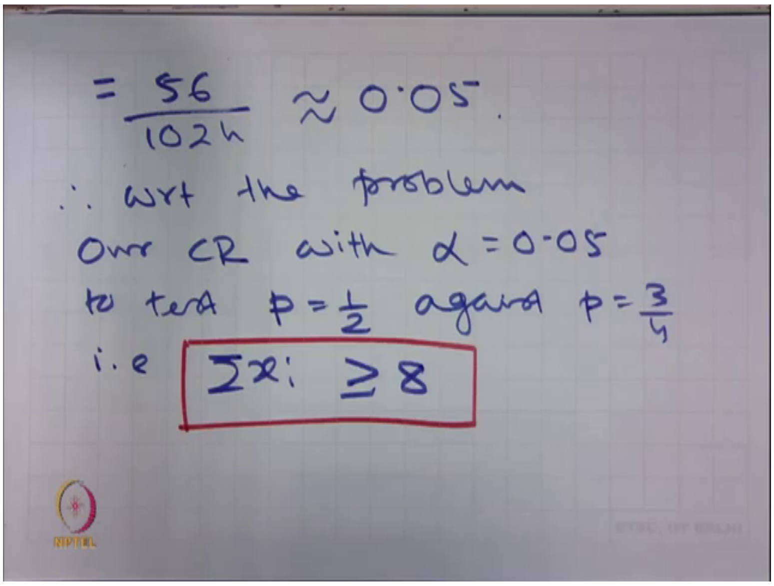
The question is how to obtain C? When n is small for a given alpha, we can determine this easily as follows say n is equal to 10. And we have alpha is equal to say 0.05 that is 5 percent and with respect to the earlier problem our CR is sigma x i is greater than sum C.

(Refer Slide Time: 40:34)



Now, under H naught that is p is equal to half probabilities sigma x i is equal to 10 is equal to half to the power 10 probability sigma x i is equal to 9 is equal to 10 C 1 half to the power 10 is equal to 10 divided by 2 to the power 10 probabilities sigma x i is equal to 8 is equal to 10 C 2 half to the power 10 is equal to factorial 10 upon factorial 8 factorial 2 into half to the power 10 is equal to 45 upon 2 to the power 10.

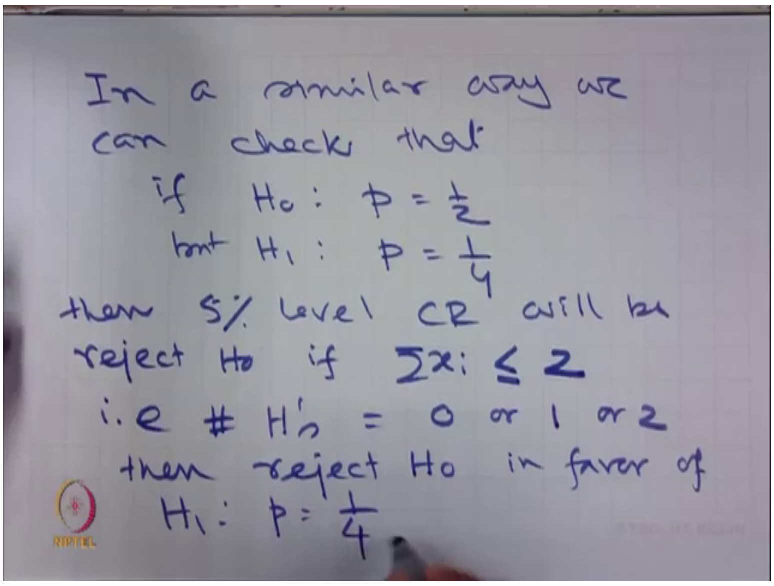
(Refer Slide Time: 41:59)



Therefore, probability sigma x i greater than equal to 8 is equal to 1 plus 10 plus 45 up to the power 10 is equal to 56 upon 1024 which is roughly equal to 0.05.

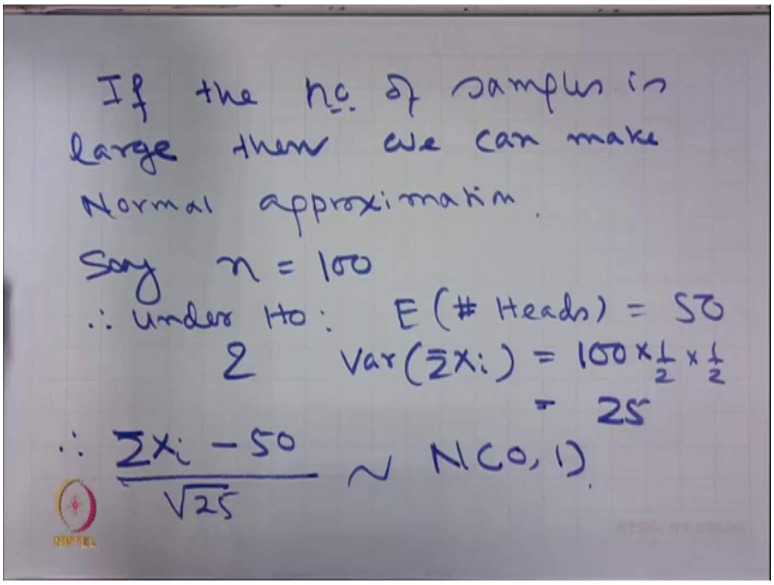
Therefore, with respect to the problem our critical region with alpha is equal to 0.05 to test p is equal to half against p is equal to 3 by 4 is sigma x i is greater than equal to 8. Or if number of it is obtained is 8 or more, then we are going to reject the null hypothesis that the coin is unbiased otherwise we are going to accept that this coin is unbiased when testing against p is equal to 3 by 4.

(Refer Slide Time: 43:18)



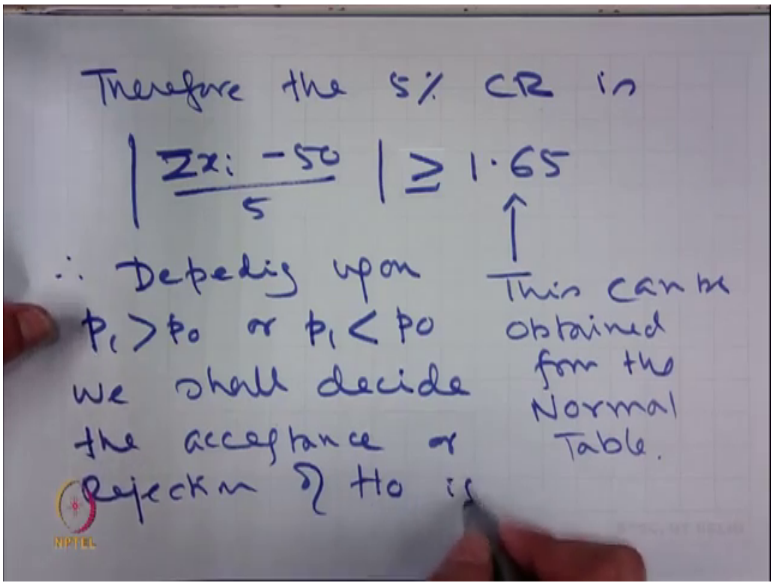
In a similar way, we can check that if H naught is p is equal to half. But H 1 is p is equal to 1 by 4, then the 5 percent level critical region will be reject H naught, if sigma x i is less than equal to 2 that is number of heads is equal to 0 or 1 or 2 then reject H 1 reject H naught in favor of H 1 p is equal to 1 by 4.

(Refer Slide Time: 44:39)



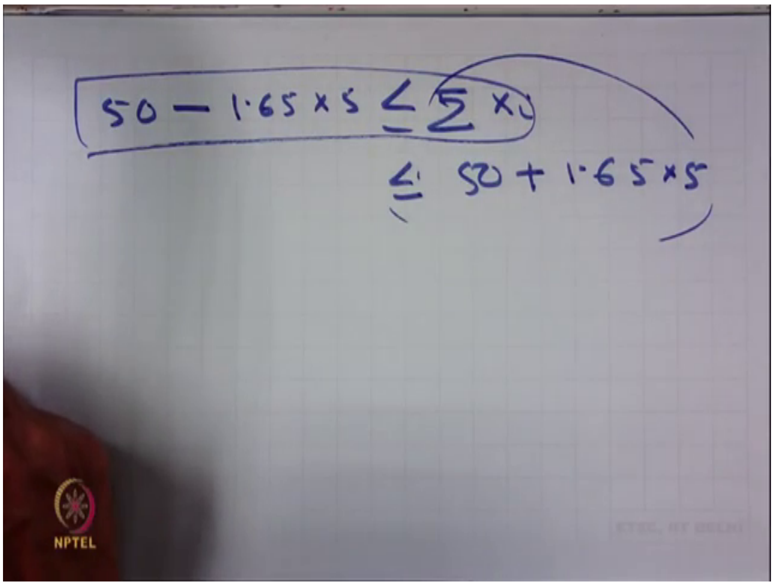
If the number of samples is large, then we can make normal approximation say n is equal to 100. Therefore, under H naught expected number of heads is equal to 50 and variance of sigma x i is equal to 100 into half into half is equal to 25. Therefore, sigma x i minus 50 upon root over 25 may be approximated as normal with 0 1.

(Refer Slide Time: 46:06)



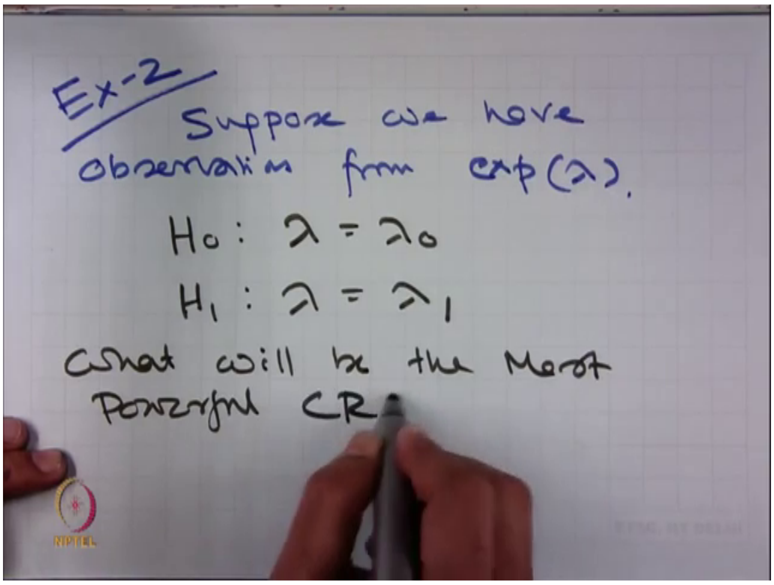
Therefore, the 5 percent critical region is modulus of sigma x i minus 50 upon 5 is greater than equal to 1.65. This is, this can be obtained from the normal table. Therefore, depending upon p 1 is greater than p 0 or p 1 is less than p 0 we shall decide the acceptance or rejection of H naught.

(Refer Slide Time: 47:25)



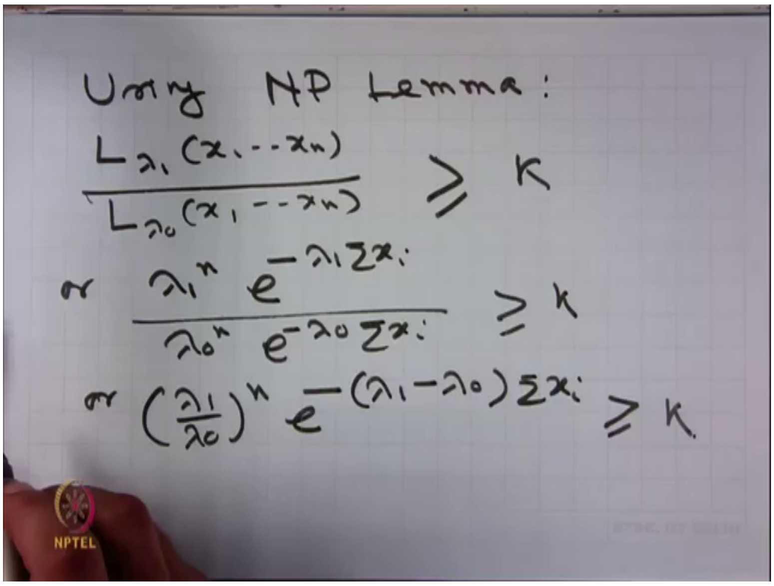
If 50 minus 1.65 into 5 is less than equal to sigma x i is less than equal to 50 plus 1.65 into 5 and we know that if p 1 is greater than p 0, then this should be the condition. If p 1 is less than p 0 then this should be the condition. Therefore the shape of the critical region will be obtained accordingly. Let us now consider a second example.

(Refer Slide Time: 48:06)



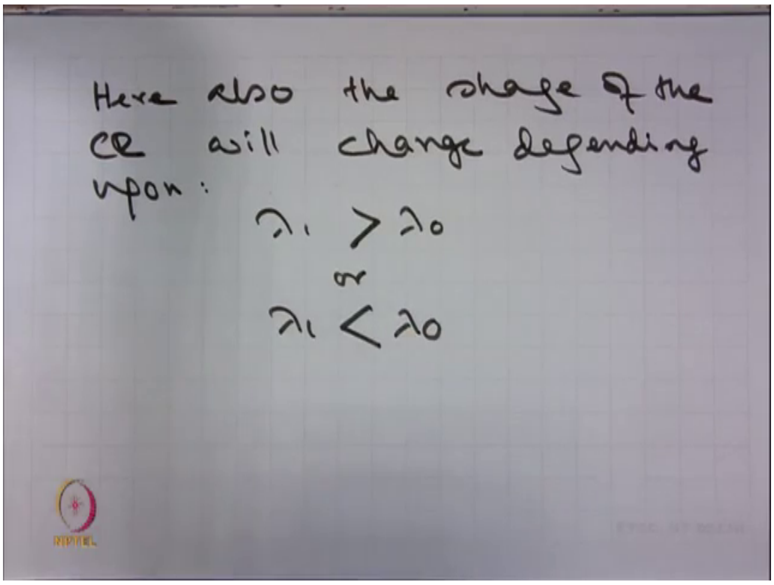
Suppose, we have observations from exponential lambda and our H naught is lambda is equal to lambda naught and H 1 is lambda is equal to lambda 1, what will be the most powerful critical region?

(Refer Slide Time: 49:05)



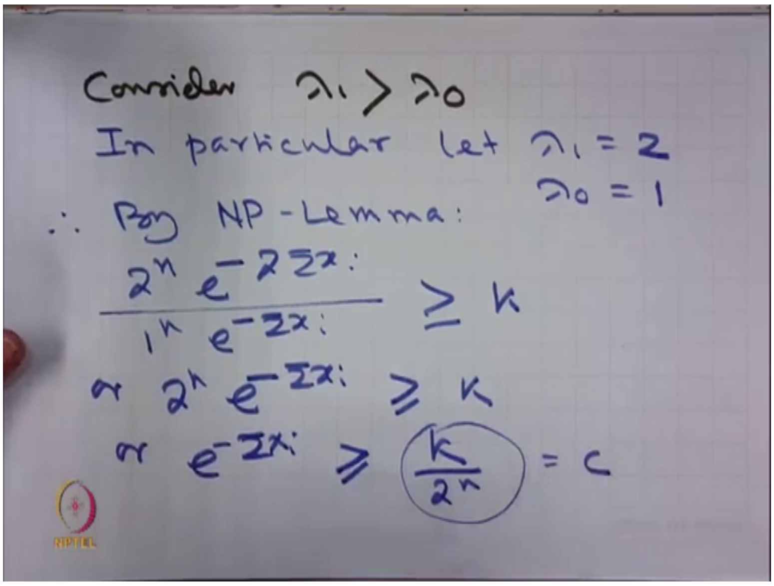
Using, Neyman-Pearson Lemma L lambda 1 of x 1, x 2, x n L lambda 0 of x 1, x 2, x n is greater than equal to some positive constant K or lambda L to the power n e to the power minus lambda 1 sigma x i. Upon lambda 0 to the power n e to the power minus lambda 0sigma x i is greater than equal to K or lambda L 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: 50:21)



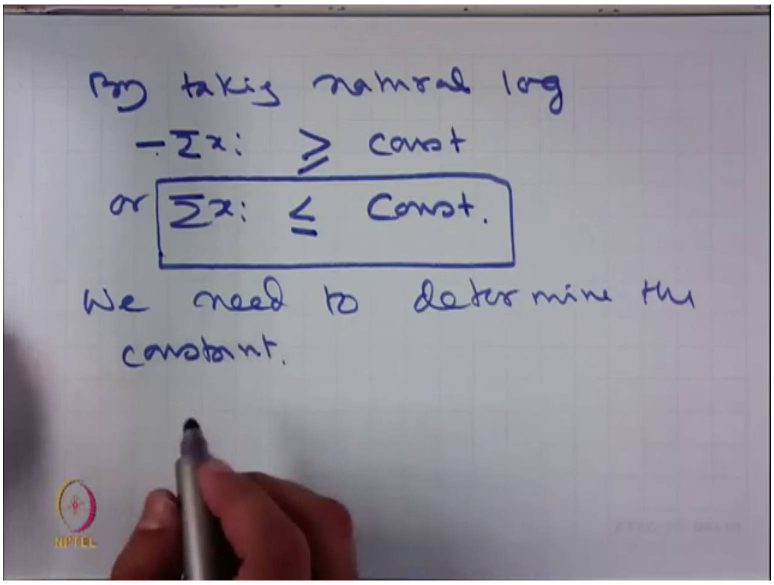
Here, also shape of the critical region will change depending upon lambda 1 is greater than lambda 0 or lambda 1 is less than lambda 0.

(Refer Slide Time: 50:54)



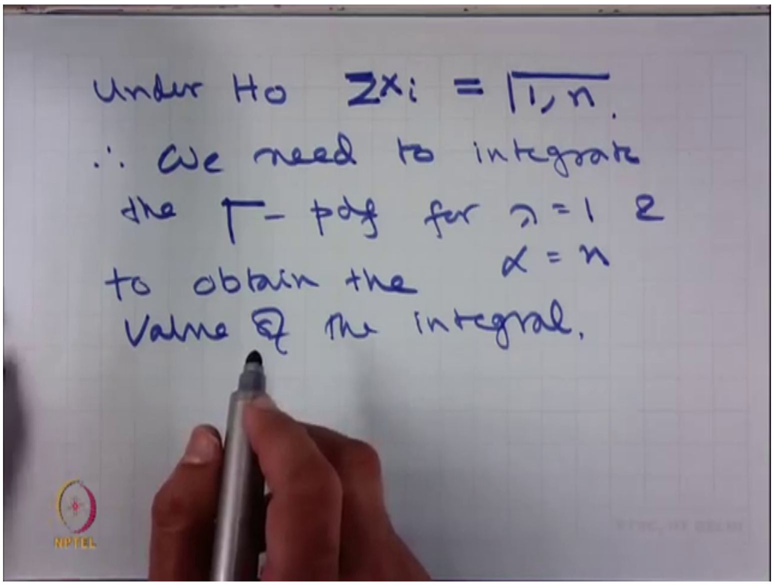
Consider lambda 1 is greater than lambda 0 in particular let lambda 1 is equal to 2 lambda 0 is equal to 1. Therefore, by NP Lemma 2 to the power n e to the power minus 2 sigma x i upon 1 to the power n e to the power minus sigma x i is greater than equal to K or 2 to the power n e to the power minus sigma x i is greater than equal to K or e to the power minus sigma x i is greater than equal to K upon 2 to the power n.

(Refer Slide Time: 52:14)



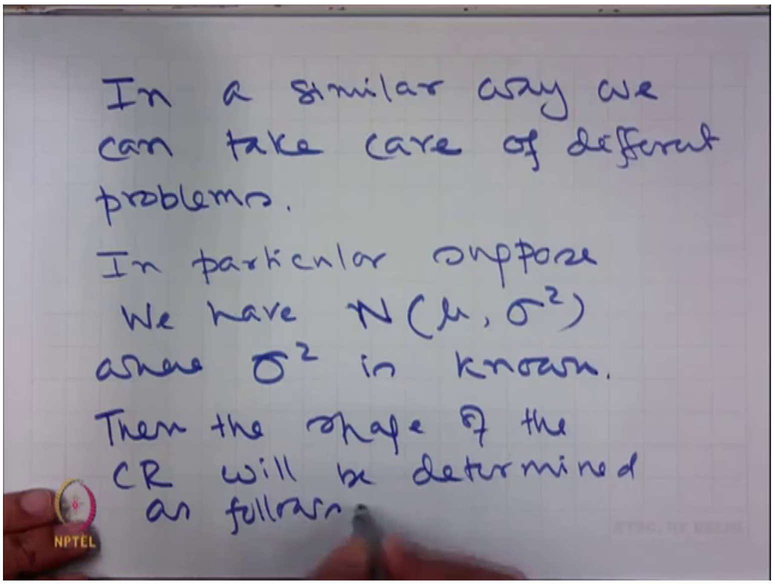
Let us call it a constant equal to C by taking natural log minus sigma x i is greater than equal to some constant or sigma x i is less than equal to some constant because this is negative there is a change in the inequality. Now, we need to determine the constant.

(Refer Slide Time: 53:10)



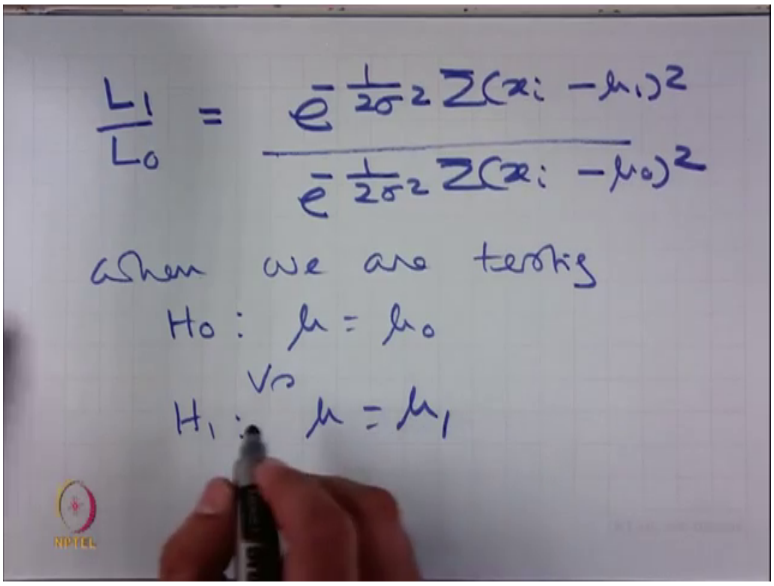
Under H naught sigma x i is equal to gamma 1 comma n. Therefore, we need to integrate or the gamma pdf for lambda is equal to 1 and alpha is equal to n to obtain the value of the integral.

(Refer Slide Time: 54:13)



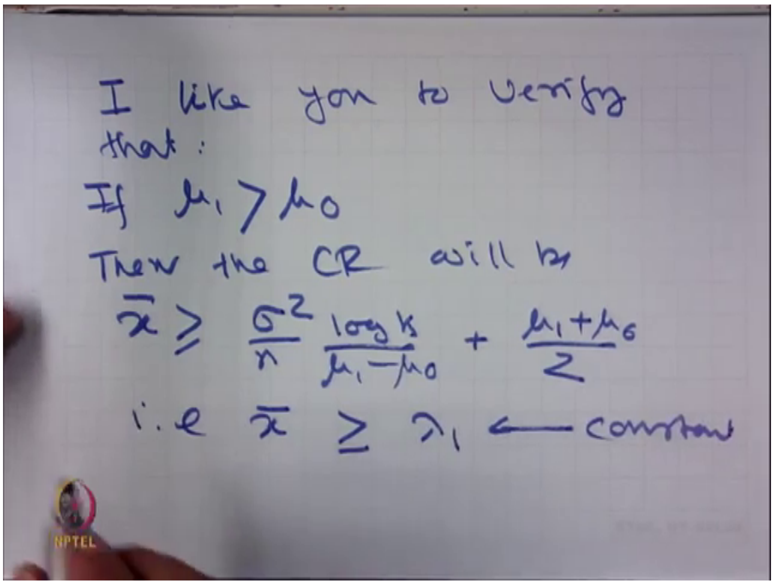
In a similar way we can take care of different problems. In particular, suppose we have normal mu comma sigma square where sigma square is known, then the shape of the critical region will be determined as follows.

(Refer Slide Time: 55:34)



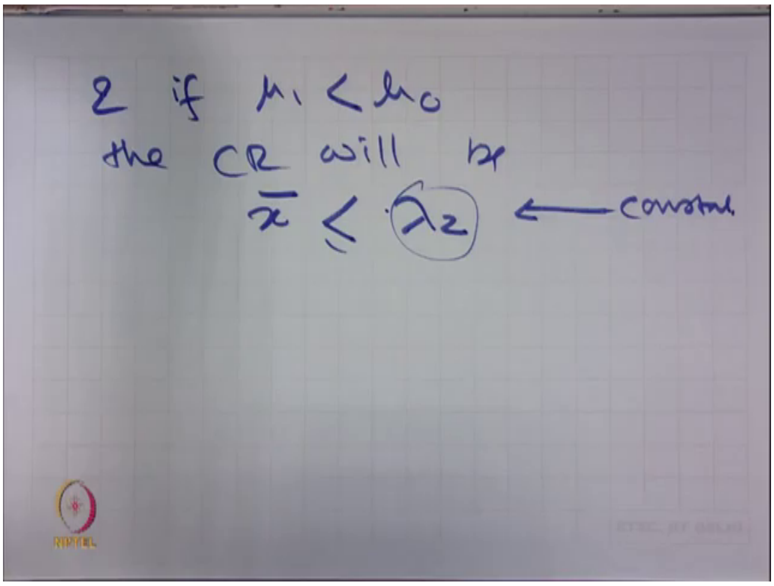
So, L 1 upon L 0 is equal to e to the power minus 1 upon 2 sigma square into sigma x i minus mu 1 whole square upon e to the power minus 1 upon 2 sigma square sigma x i minus mu naught whole square. When, we are testing H naught mu is equal to mu naught versus H 1 mu is equal to mu 1.

(Refer Slide Time: 56:20)



I like you to verify that, If mu 1 is greater than mu naught, then the critical region will be x bar is greater than equal to sigma square by n log of K mu 1 minus mi naught plus mu 1 plus mu naught by 2. That is x bar has to be greater than equal to lambda 1 which is a constant.

(Refer Slide Time: 57:18)



If mu 1 is less than mu naught, the CR will be x bar less than equal to lambda 2 constant. And we can obtained the value of lambda 1 or lambda 2 by using the normal table.

So, this is how we decide whether to accept a simple null hypothesis against a simple alternative hypothesis where the distributions are known. If the statistic is such that we know it is pdf and we get its table then we can obtain the value from there otherwise we will have to integrate the pdf or you have to numerically obtain the value of the threshold. So, that depending upon whether the statistics is greater than that or less than that we can take a decision of accepting or rejecting the null hypothesis.

Friends, I stop here today. In the next class I shall prove the Neyman-Pearson Lemma and also solve a few problems to understand the method of testing of hypothesis.

Thank you.