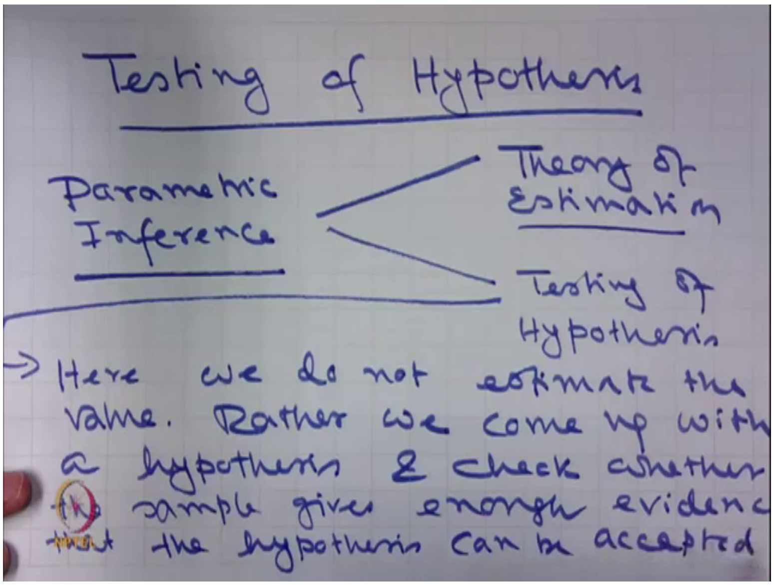
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 19
Statistical Inference

Welcome students to the MOOC's lecture on Statistical Inference. This is lecture number 19. As I said at the end of the last lecture that today what I am going to start is called Testing of Hypothesis.

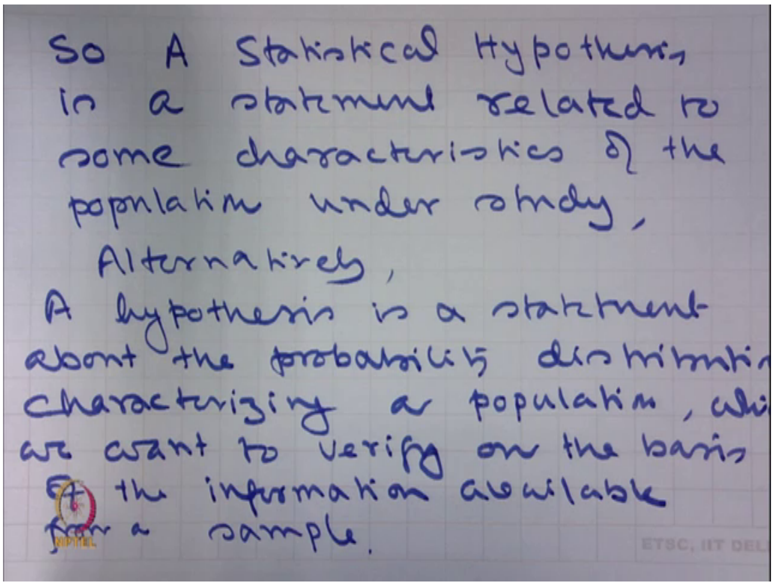
(Refer Slide Time: 00:47)



At the very beginning I said that parametric inference has two forms; theory of estimation and testing of hypothesis. It is parametric means we have an idea of the distribution. Only thing we want to know from the sample that: what are the possible values of the parameter of the distribution. In theory of estimation, we have seen we try to obtain those values either as one specific value which we call point estimation where we have learnt the method of moments and method of maximum likelihood. And also, we have learnt how to estimate a confidence interval, so that we are confident which gives us the probability that the parameter of the distribution will lie within this interval with very high probability say 95 percent or 99 percent.

At least in this class, we have dealt with problems associated with these two probabilities. Testing of hypothesis is slightly different. Here we do not estimate the value rather we come up with a hypothesis and check whether the sample gives enough evidence that the hypothesis can be accepted or perhaps it can be rejected also if the sample does not give enough evidence in support of the acceptance.

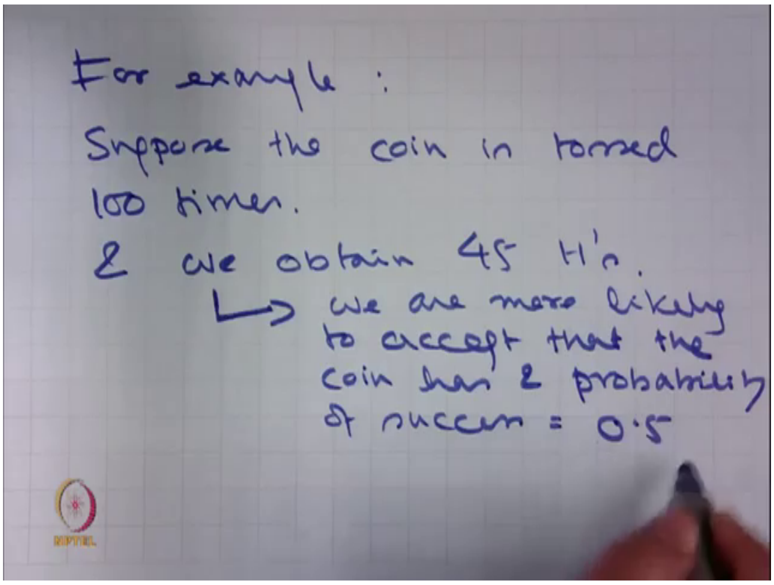
(Refer Slide Time: 04:02)



So, a statistical hypothesis is a statement related to some characteristics of the population under study. Or alternatively we can say that hypothesis is a statement about the probability distribution characterizing a population which we want to verify on the basis of the information available from a sample.

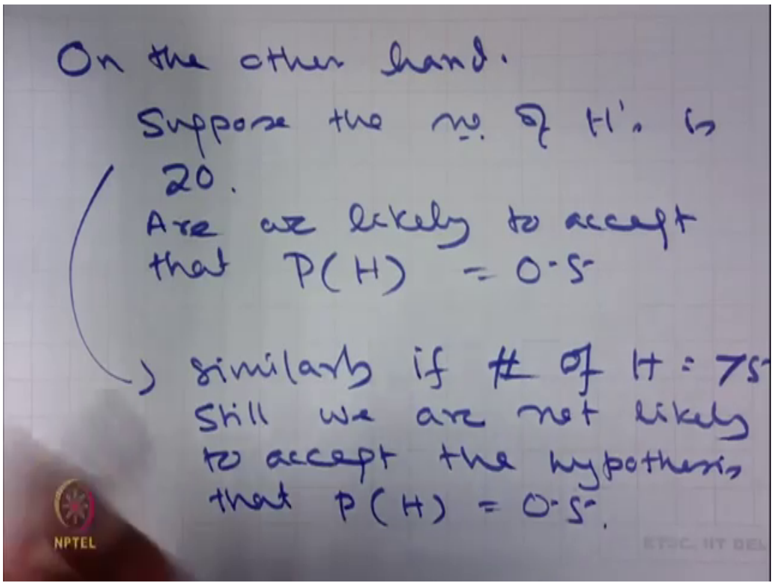
So, let me explain this. Suppose there is a coin and our hypothesis is that it is probability of obtaining a head is 0.5, we want to verify that. So, what we do? We toss the coin certain number of times and we have already decided that if the number of heads shows certain property, then we are going to accept the hypothesis that indeed that coin has the probability of a head or probability of a success to be 0.5. And otherwise, you are going to say that the coin does not have the probability of getting a head with 0.5.

(Refer Slide Time: 07:31)



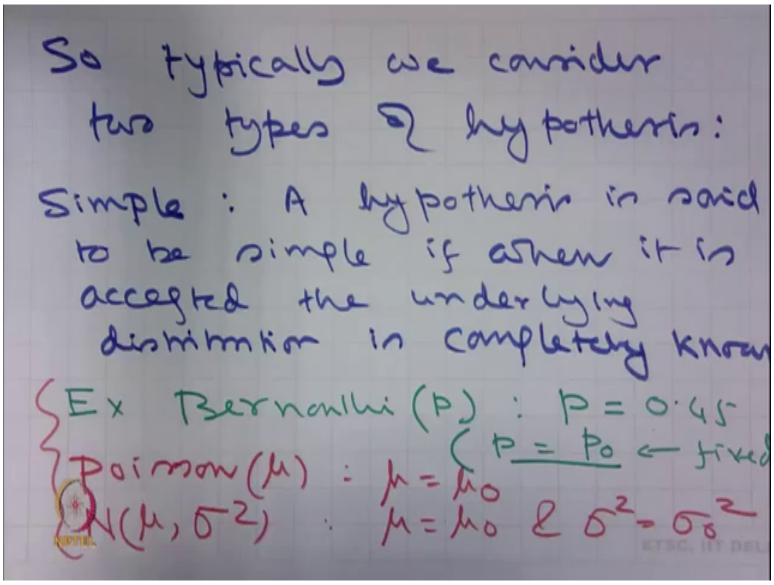
So, for example, suppose the coin is tossed 100 times and we obtain 45 heads, we are more likely to accept that the coin has a probability of success is equal to 0.5.

(Refer Slide Time: 08:35)



On the other hand, suppose the number of head is say 20; are we likely to accept that probability of a head is equal to 0.5 very unlikely. Similarly, if number of head is equal to say 75, still we are not likely to accept the hypothesis that probability of head is equal to 0.5. So, the testing of hypothesis is all about designing the scheme such that based on the sample evidence, we can accept the hypothesis or we reject the hypothesis.

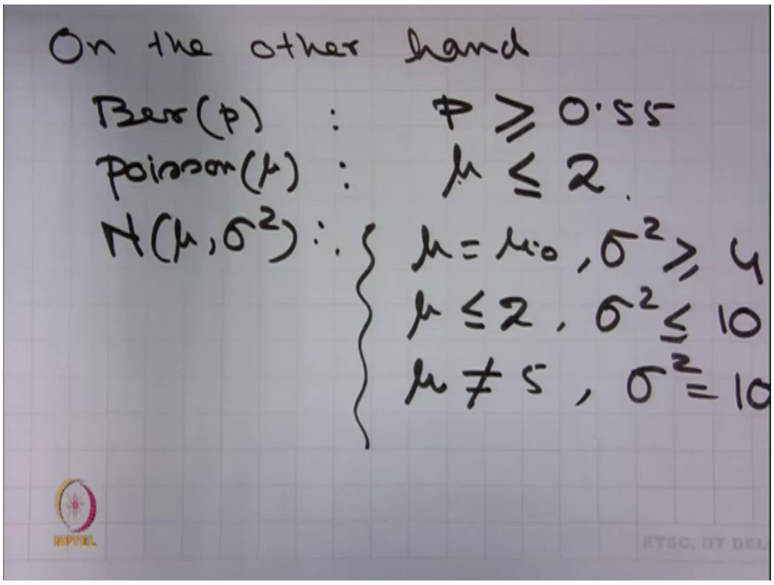
(Refer Slide Time: 10:13)



So, typically we consider two types of hypothesis. Simple hypothesis is said to be simple if when it is accepted the underlined distribution is completely known. Example Bernoulli p and hypothesis is p is equal to 0.45 or more generally p is equal to p 0 which is fixed, so that if we accept that p is equal to p 0, then we know that distribution completely.

Similarly say Poisson with mu and our hypothesis is mu is equal to mu naught, where mu naught is some fixed value and if the sample evidence establishes that mu is equal to mu naught can be accepted, then we know the distribution completely. When you are looking at normal mu comma sigma square hypothesis like mu is equal to mu naught and sigma square is equal to sigma naught square. So, these are all simple hypothesis, so that if the hypothesis is accepted, then we know the distribution completely.

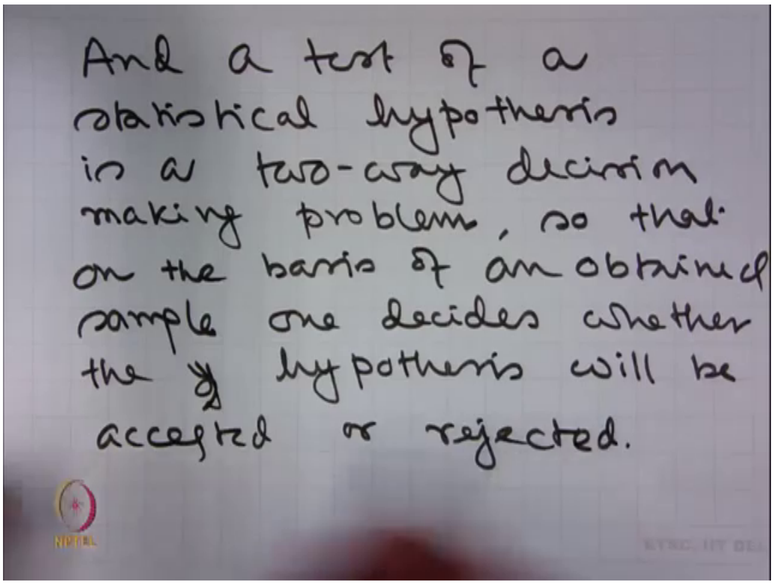
(Refer Slide Time: 13:09)



On the other hand, suppose we consider Bernoulli p and our hypothesis is p is greater than equal to 0.55 or if it is a Poisson distribution with mu and our hypothesis is mu is less than equal to 2. And normal mu, sigma square and we can have hypothesis like mu is equal to mu naught sigma square is greater than equal to 4 or suppose we have mu less than equal to say 2 and sigma square less than or equal to 10.

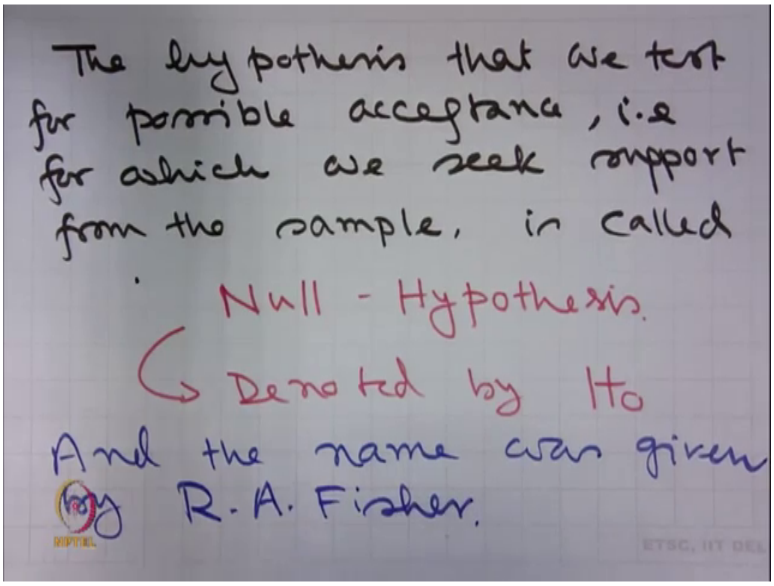
Or we can have something like mu not equal to 5 and sigma square is equal to 10 say. In all these cases, you can see that even if the hypothesis is accepted, we do not get the distribution completely because, there will be a family of distributions each of which will actually fall into this category satisfying this hypothesis.

(Refer Slide Time: 14:56)



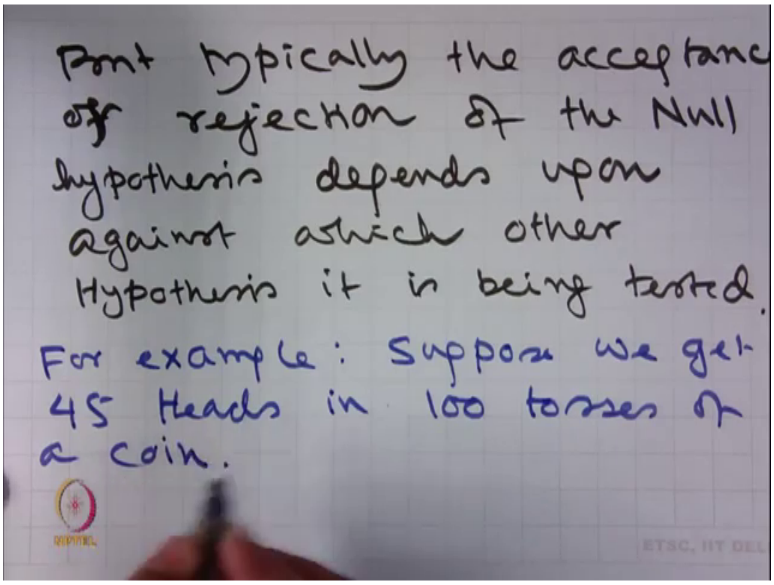
So, such hypothesis is called composite hypothesis and a test of a statistical hypothesis is a two way decision making problem, so that on the basis of this obtained sample, one decides whether the hypothesis will be accepted or rejected.

(Refer Slide Time: 16:28)



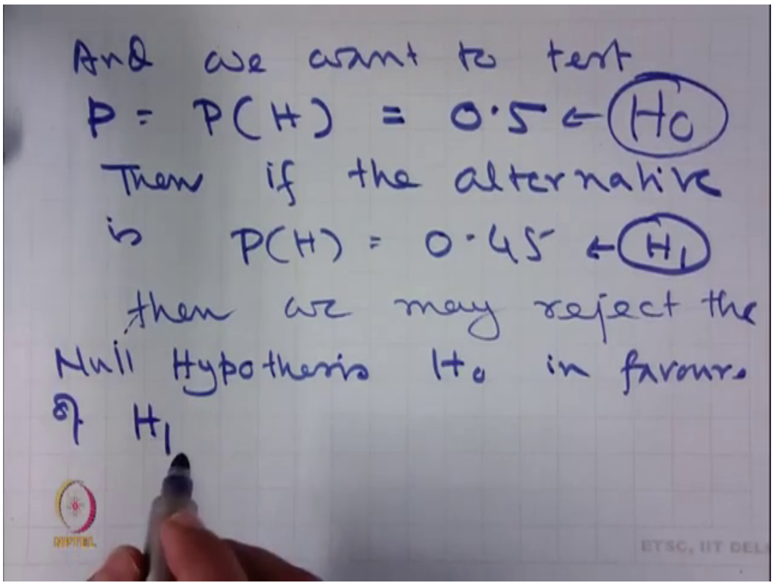
So, to proceed further the hypothesis that we test for possible acceptance that is for which we seek support from the sample is called Null Hypothesis. It is denoted by H naught and the name given by R.A. Fisher.

(Refer Slide Time: 17:57)



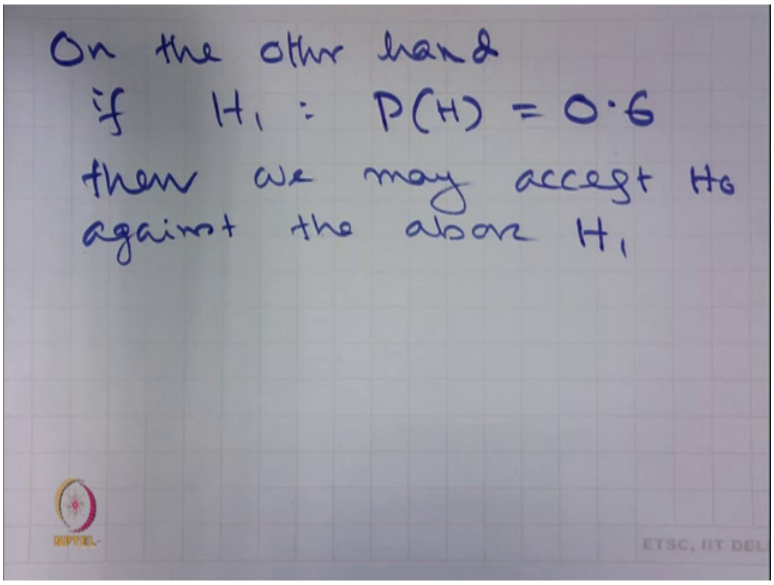
But, typically the acceptance or rejection of the null hypothesis depends upon against which other hypothesis it is being tested. For example, suppose we get 45 heads in 100 tosses of a coin.

(Refer Slide Time: 19:32)



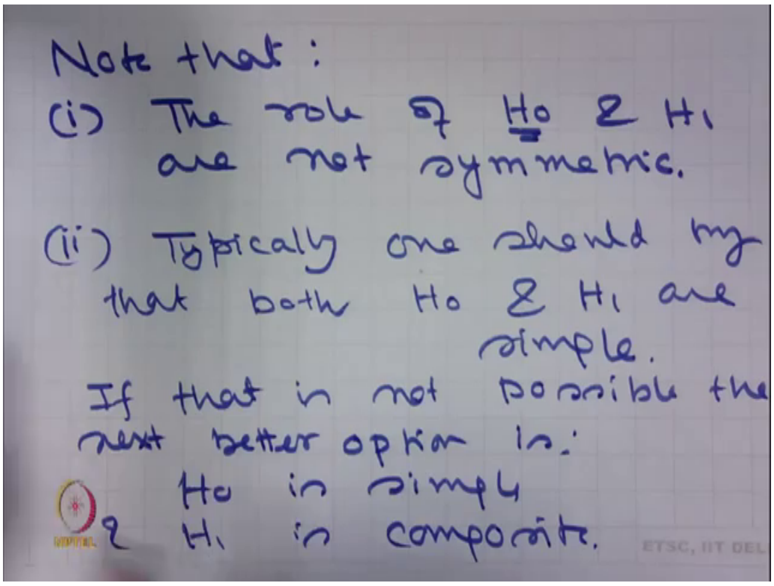
And we want to test that P is equal to probability of head is equal to 0.5, then if the alternative is probability of head is equal to 0.45, then we may reject the null hypothesis. This is the null hypothesis H naught in favor of this hypothesis.

(Refer Slide Time: 20:54)



On the other hand if H1 is that probability of H is equal to 0.6, then we may accept H naught against the above H1. Therefore, in testing of hypothesis it is not only the null hypothesis or H naught, one has to see: what is the alternative hypothesis because the acceptance and rejection will depend upon the alternative hypothesis as well.

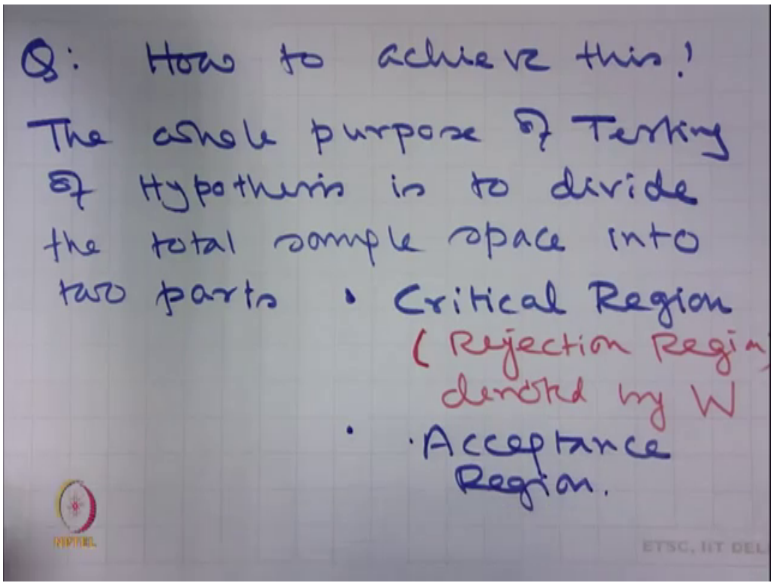
(Refer Slide Time: 21:56)



Note that the role of H naught and H1 are not symmetric. Our focus is on H naught and we want to see if H naught is accepted or rejected when tested against H1, that is the focus of testing of hypothesis. And, second point is that typically one should try that both

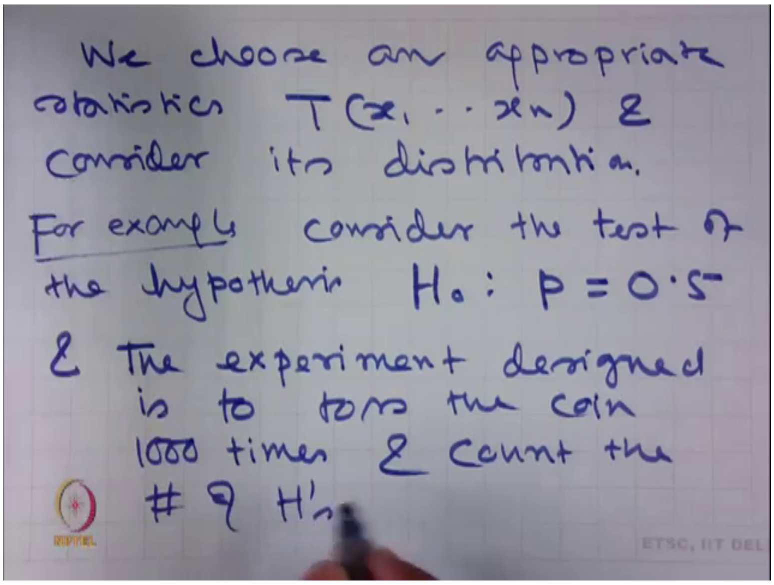
H naught and H1 are simple. If that is not possible, the next better option is H naught is simple and H1 is composite because if H naught is composite, then by accepting that we do not really learn much about the characteristic of the population, question.

(Refer Slide Time: 23:51)



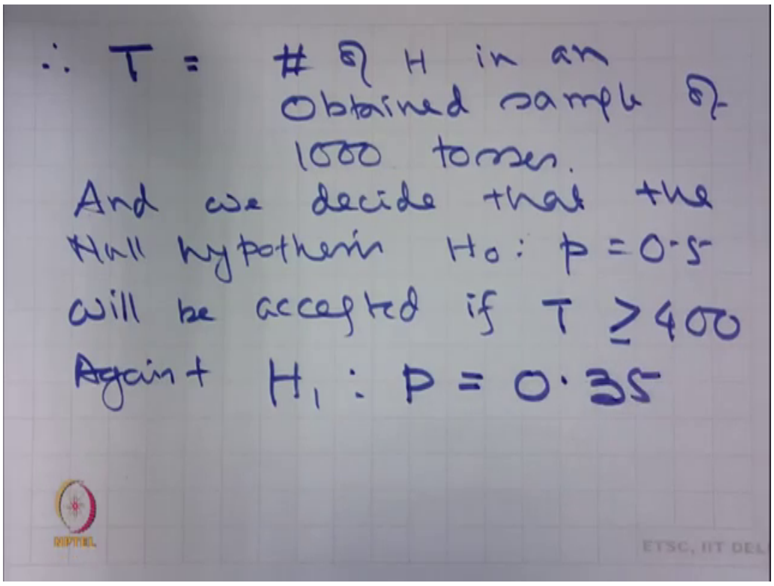
How to achieve this? So, the whole purpose of testing of hypothesis is to divide the total sample space into two parts. One is called the critical region or rejection region and we will denote it by W and the other one is the acceptance region that is.

(Refer Slide Time: 25:27)



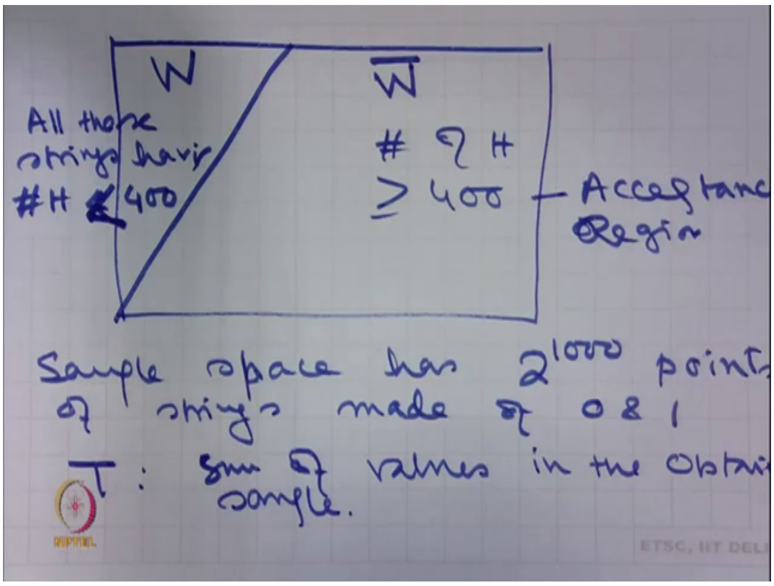
We choose an appropriate statistics T x 1 x 2 x n and consider its distribution. For example, consider the test of the hypothesis H naught P is equal to 0.5 and the experiment designed is to toss the coin 1000 times and count the number of heads.

(Refer Slide Time: 27:09)



Therefore, T is equal to number of H in an obtained sample of 1000 tosses and we decide that the null hypothesis that is H naught P is equal to 0.5 will be accepted, if T is greater than equal to 400 against H1 that let us assume that a simple alternative that P is equal to 0.35.

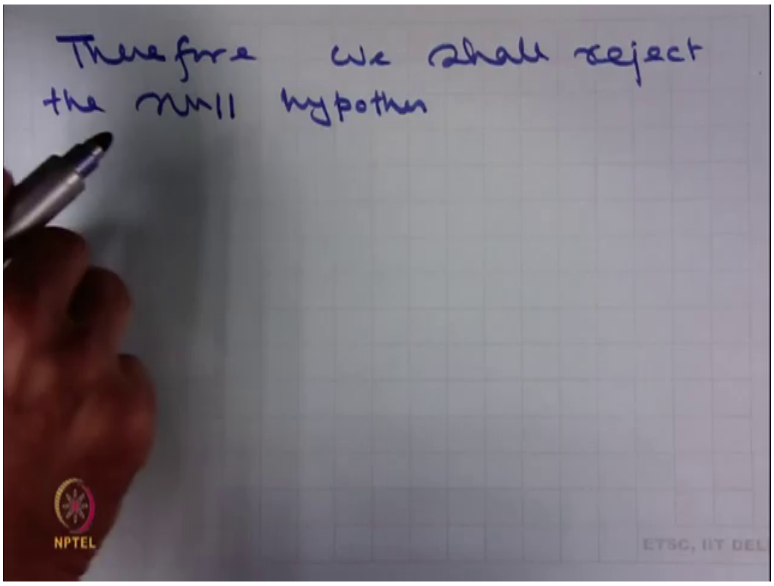
(Refer Slide Time: 28:32)



What is the sample space? Sample space has 2 to the power 1000 points of strings made of 0 and 1, because each toss ends up in either 0 or 1 depending upon whether it is a tail or it is the head. So, 1000 process means 2 to the power 1000 points and suppose this is my W, that means all those streams having number of head is less than equal to 400. So, it is less than 400. So, if the number of heads obtained is less than 100, we are going to reject the null hypothesis or if the number of heads is greater than or equal to 400, then we are going to accept the null hypothesis.

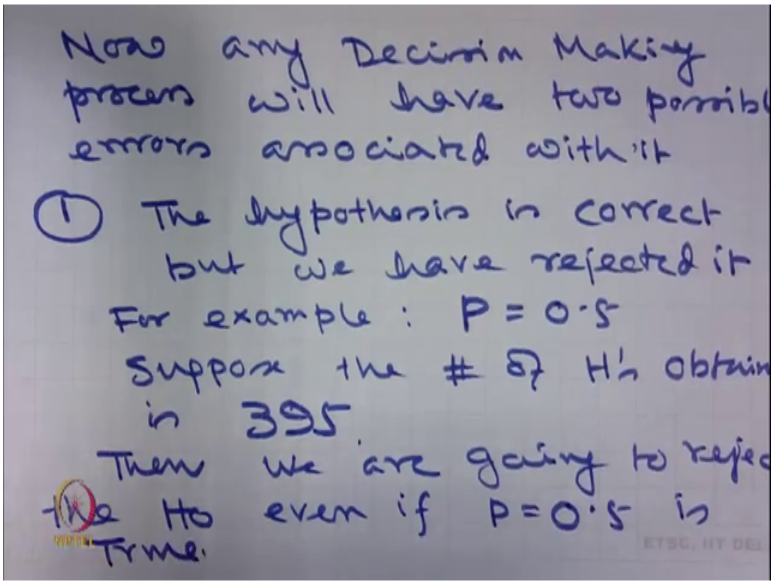
So, this is called W complement. So, what is our T? T is equal to sum of values in the obtained sample. What is the rejection criteria? The rejection criteria is that the number has to be less than 400, otherwise you are going to accept it.

(Refer Slide Time: 30:53)



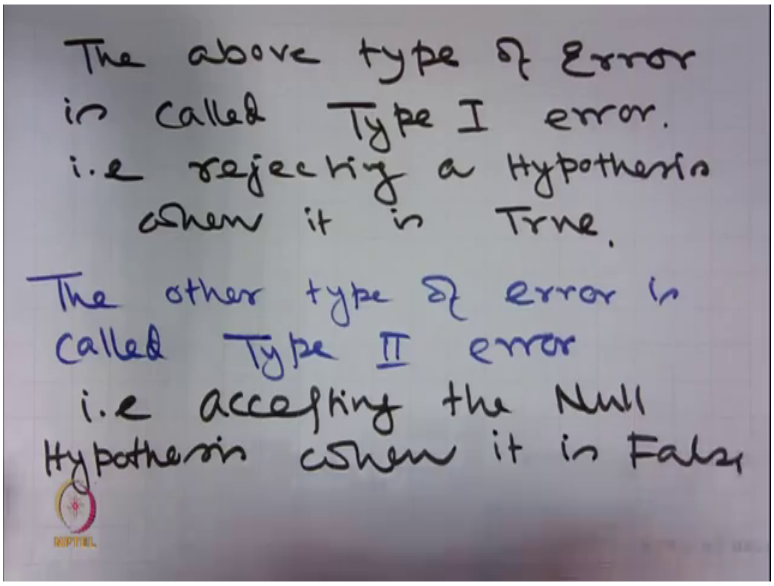
Therefore, we shall reject the null hypothesis.

(Refer Slide Time: 31:24)



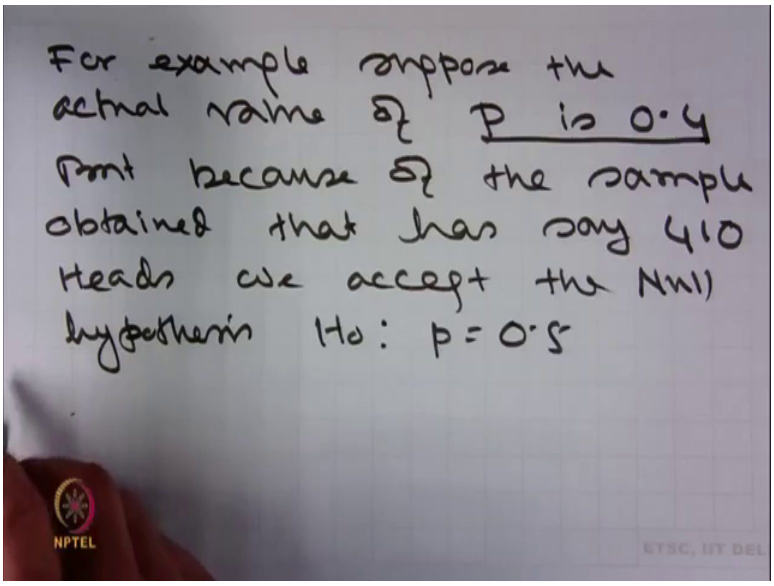
Now, any decision any decision making process, we will have two possible errors associated with it. What are these? The hypothesis is correct, but we have rejected it. For example, P is equal to 0.5 with respect to the above example of tossing the coin 1000 times and suppose the number of heads obtained is 395, then we are going to reject H naught even if P is equal to 0.5 is true.

(Refer Slide Time: 33:49)



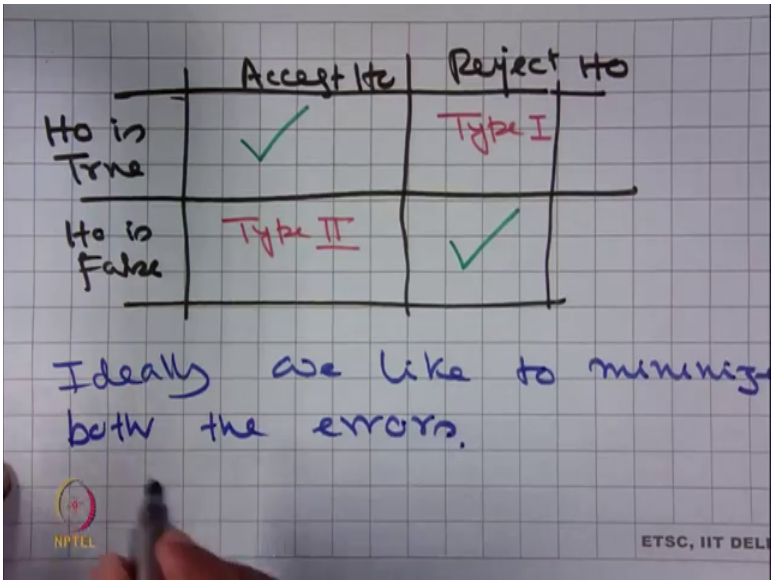
The above type of error is called type I error that is rejecting a hypothesis when it is true. The other type of error is called type II error that is accepting the null hypothesis when it is false.

(Refer Slide Time: 35:23)



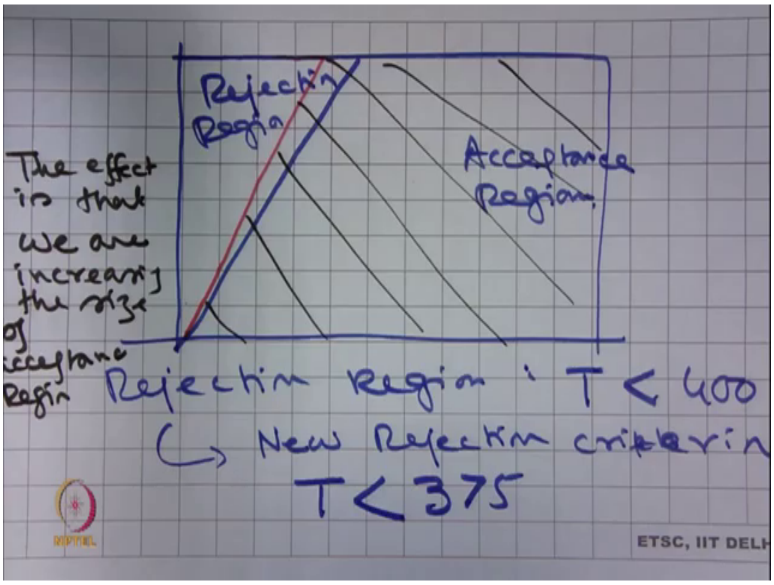
For example suppose the actual value of P is 0.4, but because of the sample obtained that has say 410 heads, we accept the null hypothesis H naught P is equal to 0.5 because that is our acceptance and rejection criterion although the null hypothesis is false.

(Refer Slide Time: 36:44)



So, you get these type of table accept H naught reject H naught. These are two actions when H naught is true H naught is false. So, when H naught is true and we have accepted that, then we are making a correct decision. When H naught is false and you are rejecting H naught, we are making another correct decision, but H naught is true, but we are rejecting we are committing type I error. If H naught is false, but we are accepting, then we are committing a type II error. Ideally we like to minimize both the errors. Why? Because the acceptance region and the rejection region are decided beforehand.

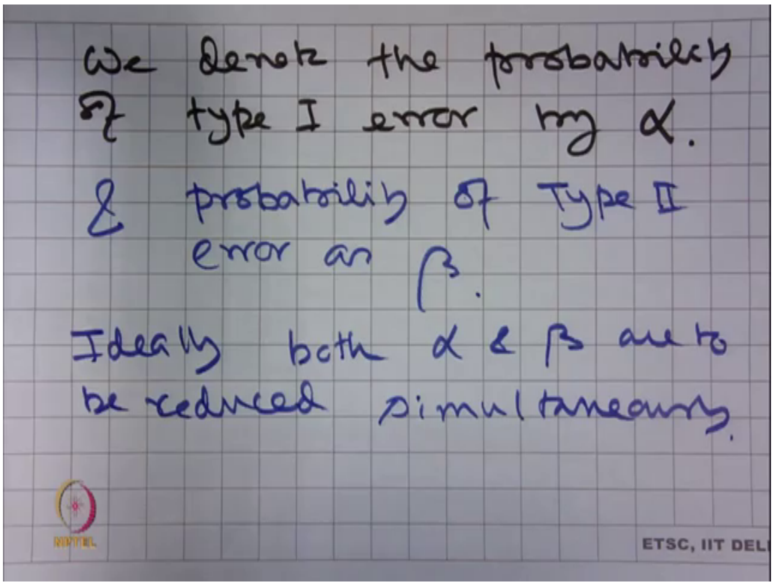
(Refer Slide Time: 38:39)



So, suppose this is my sample space, this is the rejection region, this is the acceptance region. We want to reduce the extent of type I error, then what we will do is, we will make the rejection region smaller, so that probability of rejecting a null hypothesis when it is true is less say for example, I was talking about that rejection region is that T less than 400 in order to reduce type I error. Suppose we make it new rejection criteria that T is less than 375, that means if number of heads is greater than 375, we are going to accept that the coin is having 0.5 probability of getting a head. What is the effect? The effect is that we are increasing the size of acceptance region, right. As you can see now the acceptance region is bigger.

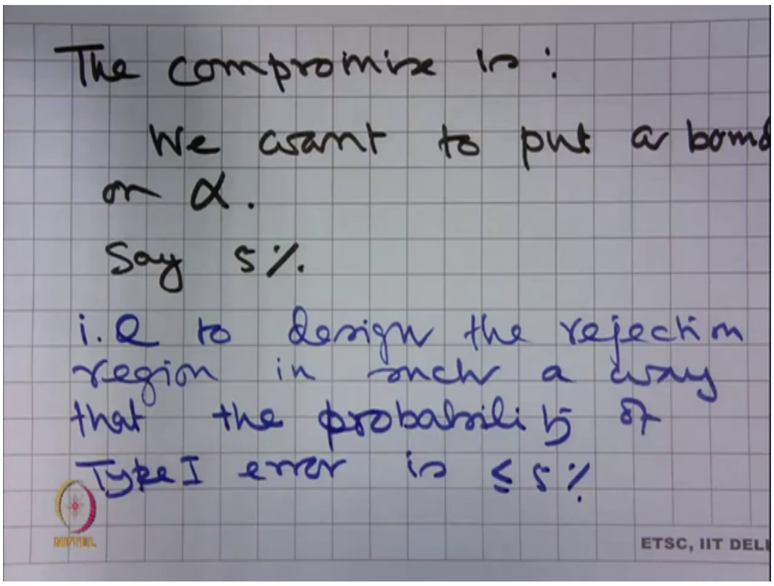
So, when the null hypothesis is not correct or actual value, if value of P is not equal to 0.5, now we have higher chance of accepting the null hypothesis say for example, if the number of heads is 390. In earlier cases we would have rejected it, but now we are going to accept it. I hope the concept is clear. Therefore, what it suggests that if you want to reduce the type I error, we are going to increase the probability of type II error and vice versa.

(Refer Slide Time: 41:57)



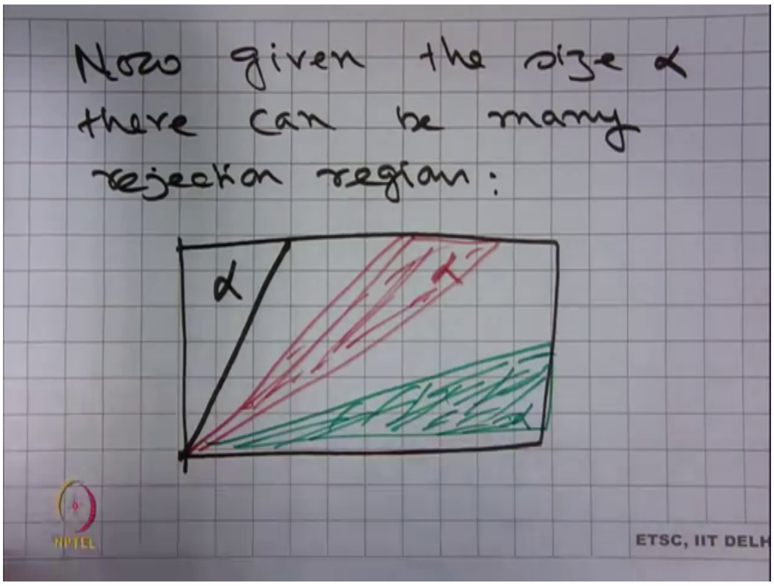
We denote the probability of type I error by alpha and probability of type II error as beta ideally both alpha and beta are to be reduced simultaneously, but as we have just seen that is very difficult. So, what we do? We have to come to a compromise. What is the compromise?

(Refer Slide Time: 43:19)



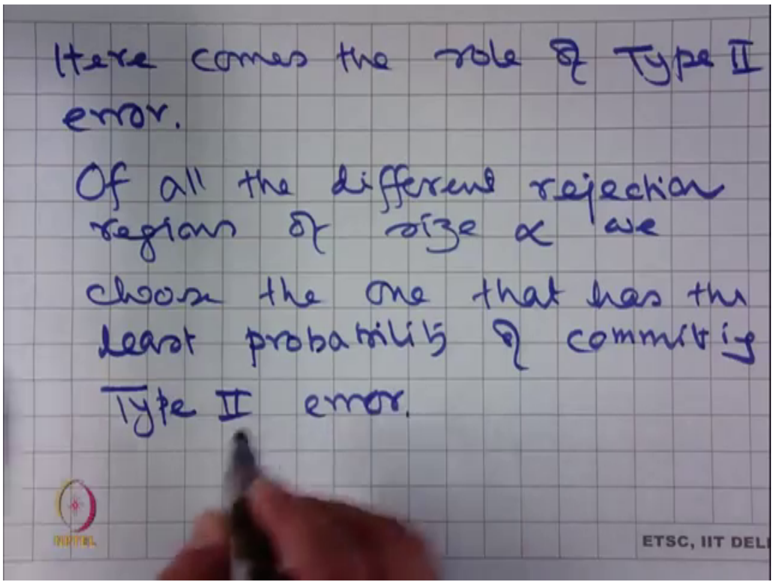
The compromise is we want to put a bound on alpha say 5 percent. That means, that we have to design the rejection in such a way that the rejection region in such a way that the probability of type I error is less than or equal to 5 percent. That means, only in 5 percent of cases even if the null hypothesis is true, we are going to reject that null hypothesis not more than 5 percent of cases.

(Refer Slide Time: 44:54)



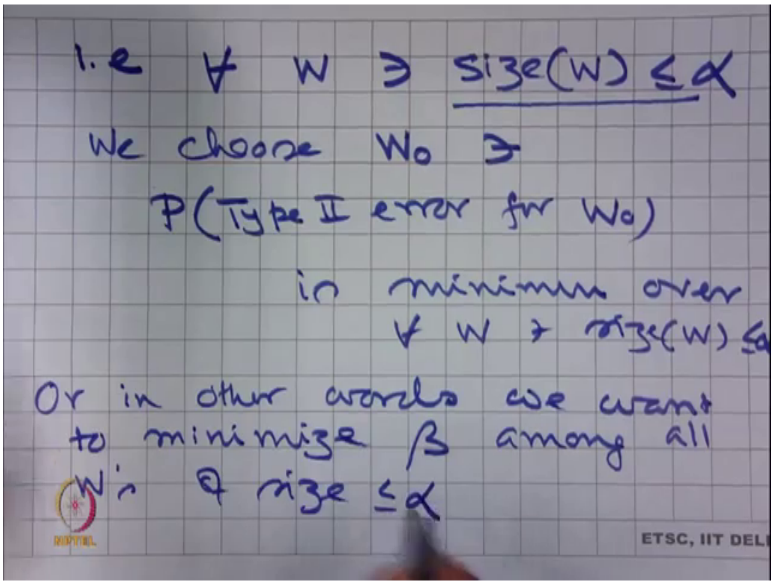
Now, given the size alpha there can be many rejection regions say this is one rejection region of size alpha, suppose this is another rejection region of size alpha and suppose this is another rejection region of size alpha. The question is which one of these we should consider for ultimate decision making. So, all 3 are alpha and which one we should take, that is the question.

(Refer Slide Time: 46:11)



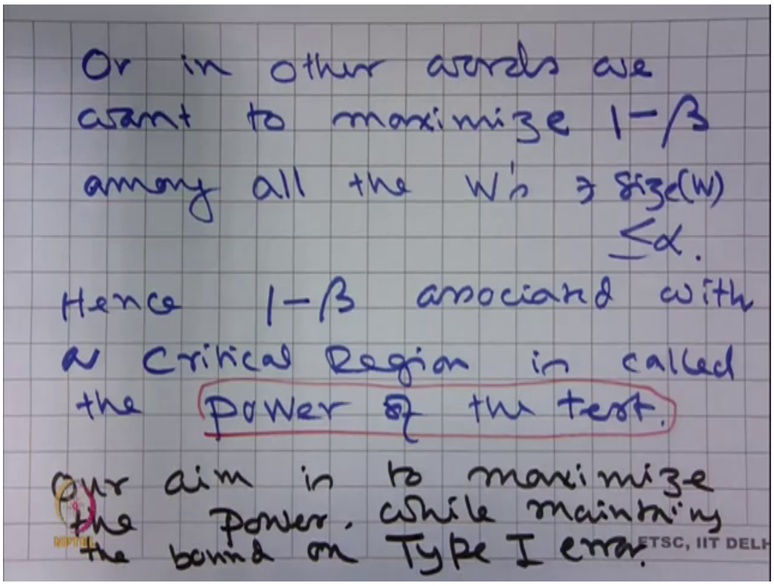
Here comes the role of type II error. Of all the different rejection regions of size alpha, we choose the one that has the least probability of committing type II error that is for all W such that size of W is less than or equal to alpha that is the size of the critical region is less than equal to alpha that is the probability of committing type I error is less than equal to alpha. We choose W0 such that probability of type II error for W naught is minimum.

(Refer Slide Time: 47:24)



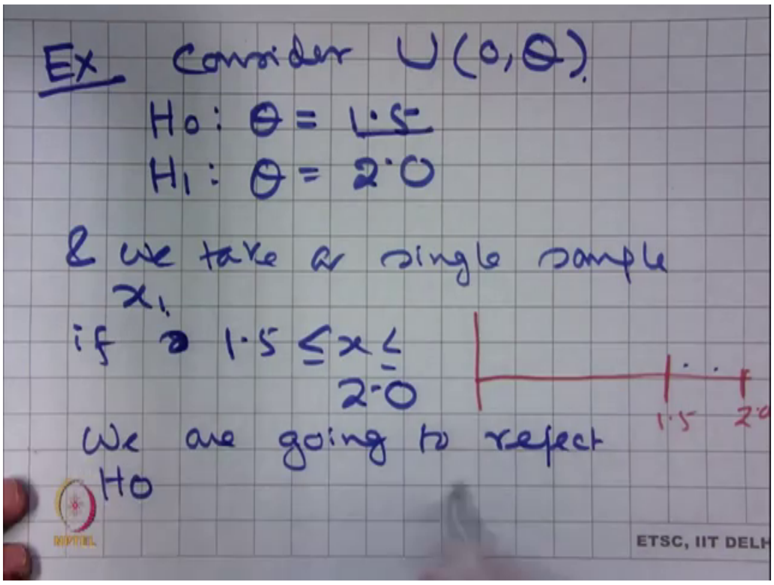
Over all W such that size of W is less than or equal to alpha or in other words, we want to minimize beta among all W's of size less than or equal to alpha.

(Refer Slide Time: 49:14)



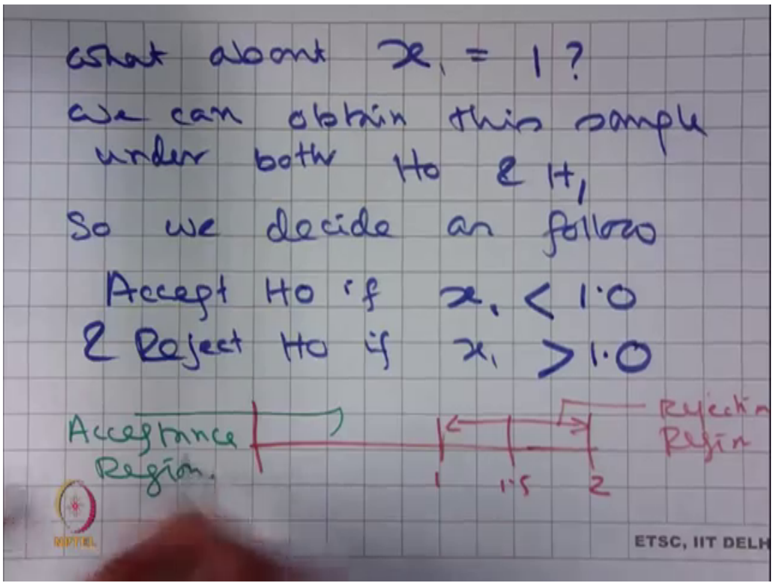
In other words, we want to maximize 1 minus beta among all the W's such that size of W is less than equal to alpha. Hence 1 minus beta associated with a critical region is called the power of the test and our aim is to maximize the power while maintaining the bound on type I error. Before proceeding any further, I give you an example.

(Refer Slide Time: 51:12)



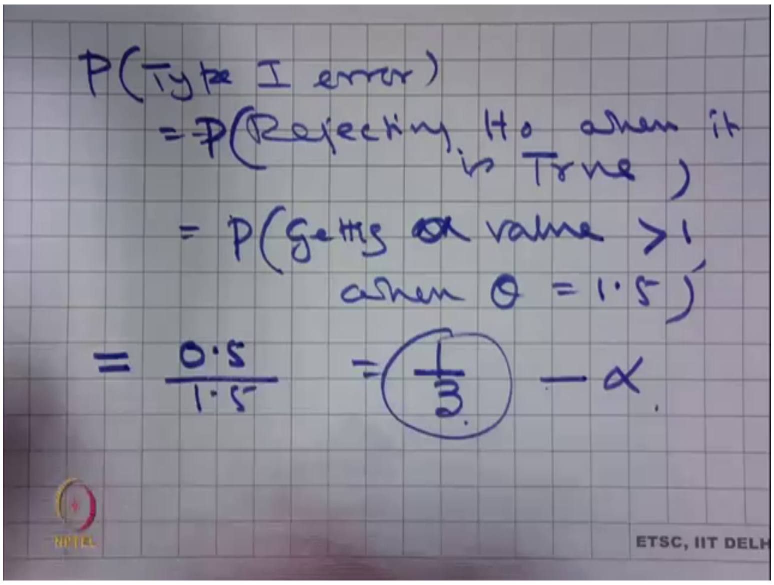
Consider uniform 0 theta. Suppose H naught is theta is equal to 1.5 and H1 is theta is equal to 2.0 and we take a single sample x 1. So, what is the situation? We take a value from uniform 0 theta. Our aim is whether to accept that theta is equal to 1.5 or we reject it in favor of theta is equal to 2.0. Obviously, if the sample falls in the interval 1.5 to 2.0, we are going to reject naught because under H naught, the theta is 1.5, but what about some value say 1.

(Refer Slide Time: 53:06)



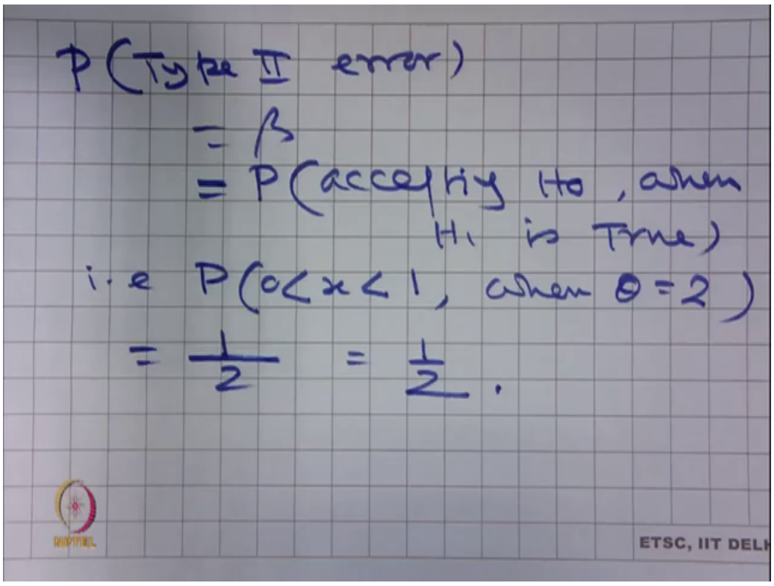
We can obtain this sample under both H naught and H1. So, we decide as follows. Accept H naught if x 1 is less than 1.0 and reject H naught if x 1 is greater than 1.0 that is if this is 1, this is 1.5 and this is 2. So, this is my rejection region and this is the acceptance region. So, what are the probabilities of the errors?

(Refer Slide Time: 54:47)



So, probability of type I error is equal to rejecting H naught when it is true is equal to say probability of getting a value greater than 1 when theta is equal to 1.5 is equal to 0.5 because the value of x will lie between 1 to 1.5 upon 1.5 is equal to 1 by 3. So, this is my alpha.

(Refer Slide Time: 55:52)



What is the probability of type II error? It is equal to beta is equal to probability of accepting H naught when H1 is true that is probability 0 less than x less than 1 when theta is equal to 2 is equal to 1 upon 2 is equal to half.

So, I hope the concept of type I error and type II error is well understood. In the next classes, I shall do some problems on testing of hypothesis and also, I will conclude my lecture by focusing on an important theorem with respect to testing of hypothesis which is called Neyman Pearson Lemma. Ok students thank you so much. See you in the next class.

Thank you.