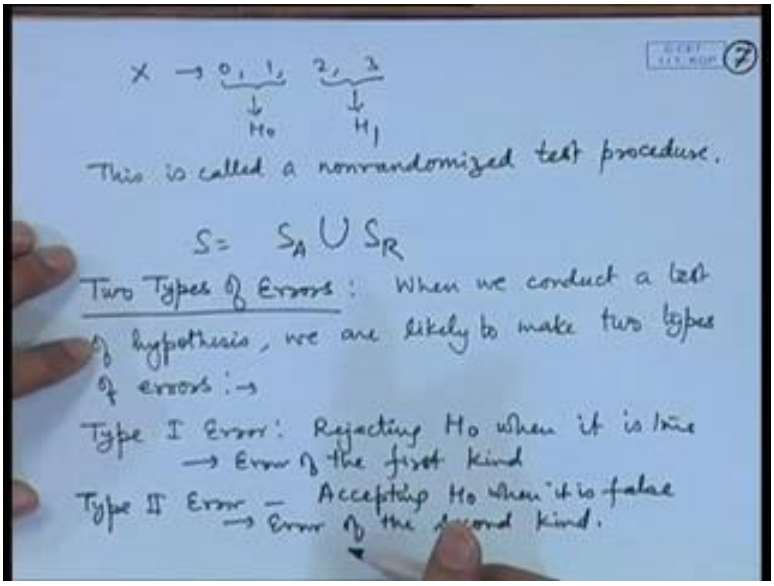
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 66
Two Types of Errors

You can easily see that this test procedure is splitting the sample space into two portions.

(Refer Slide Time: 00:32)



So, S is your S A union S R, that is acceptance region and the critical region or the rejection region. Now while conducting a test of hypothesis you can see that the decision is based on a sample. And based on the sample we are splitting the sample space into two portions, two complimentary regions; two complimentary exhaustive regions such that one region corresponds to the rejection of the null hypothesis and the other one corresponds to the rejection of the alternative hypothesis or you can say acceptance of the null hypothesis.

So, now you can see since we are making a test procedure it is likely that we may make mistakes. The question is that the decision is based on the sample and if the sampling is done in a faulty way or whatever be the reason because each sample has a certain probability of occurrence this may lead to two types of errors. So, when we conduct a test of hypothesis we are likely to make two types of errors. So, we call them type I error and type II error.

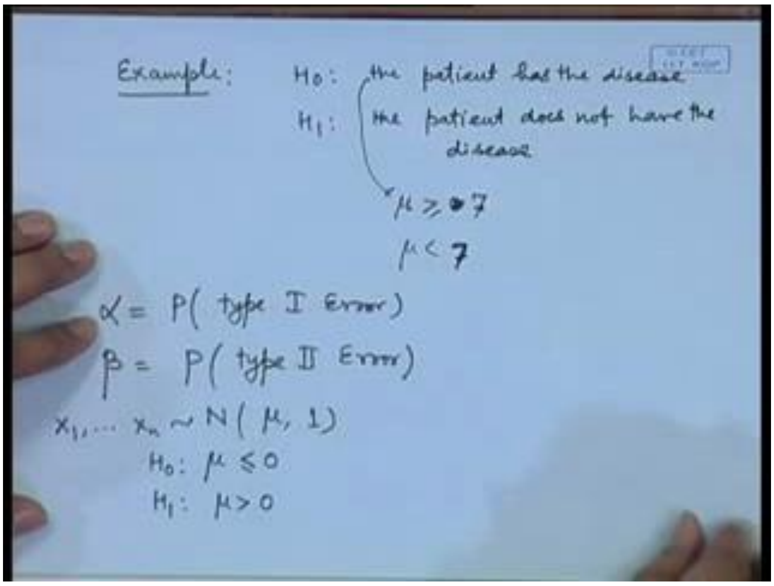
So what is type I error? This could be rejecting. So now, we will use a standard notation H naught is the null hypothesis and H 1 is the alternative hypothesis; so rejecting H naught when it is true. So, this is called error of the first kind. And type II error is accepting H naught when it is false. This is called error of the second kind. Since the decision is based on the sample these two errors are likely to be committed.

The consequences of the two types of errors can be significant. For example, consider a patient who goes to a doctor, and the patient is suffering from a somewhat complicated disease, the doctor has to judge what disease he is having or whether he is having a given disease and for that he conduct certain test. For example, he makes conduct certain blood test or certain other pathological test may be conducted. So, when he is conducting those test, those tests are based on sample.

For example, a blood involves taking out a drop of a blood from your finger or in other pathological test for example a urine sample may be taken or some other kind of say skin test may be there. So, a sample has been taken and on the basis of certain measurements from that sample the doctor has to take a decision whether the patient suffers from a certain disease or not.

Now let us look at the null hypothesis and the alternative hypothesis in this case.

(Refer Slide Time: 04:52)



So, H naught may be the patient has the disease and H 1 is that the patient does not have the disease. I have written the hypothesis in the verbal terms, but a statistical hypothesis will mean that this is related to certain parameter. For example, it may be related to certain mean of the measurement. So, you may say that if mu is greater than or equal to say 7 then this may be count of some (Refer Time: 05:45) or whatever or may be ESR rate. Say if mu is greater than or equal to 7 then we say that the patient has the disease if mu is less than 7 we will say that the patient does not have the disease.

These are the likely scenarios that the doctor may be come fronted with, and the decision is based on a sample which he has taken from the patient. Now if based on the sample the doctor concludes that the patient does not have the disease, whereas actually the patient has the disease then the consequences can be fatal, because if we conclude that he does not have the disease so he will not give the appropriate medicine and maybe he will treat for some other related symptoms. The disease may get aggravated and the patient may ultimately die. So, the consequence of this type I error that is a he is rejecting H naught when actually it is true is disasters here.

Similarly, if you look at the type II error; that means he accepts H naught when actually H naught is falls; that means, the patient does not have the disease and the doctor concludes that he has the disease he may give some heavy dosage of medication which may lead to lot of complications and this comfort for the patient. So, the error of second kind also may lead to difficulties.

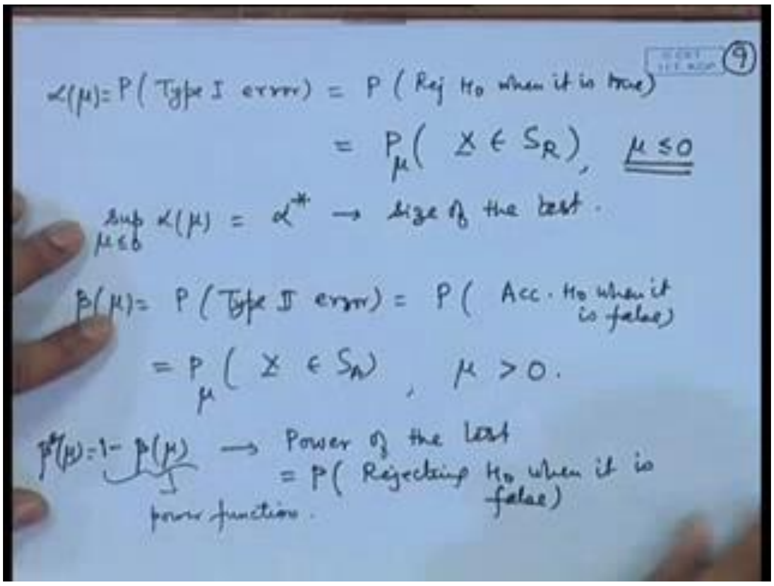
So, the point which I wanted to make here is that both types of errors have different consequences. For example, one consequence may be slightly less disasters then the other. For example in this case the error of the first kind seems to be very complicated, because the patient does not get the medicine and his disease gets aggravated and he may ultimately suffer. In the second case he does not have the disease, but he is given some medication so may be lot of discomfort is there as a consequence of taking the medicines but he may a still survive.

Now the question comes that- one has to reduce the possibilities of these errors. So, let me use the notation. We have the standard notation alpha is probability of the type I error and say beta is the probability of type II error. So, for a statistician a good test is the one which keeps these alpha and beta to a minimum, but as you can see that it may not be possible to control both of the errors. The reason is that the probability of type I error and type II error is based on the sample, because type I error is rejecting H naught that means it is the size of the rejection reason when H naught is true and this is accepting H naught when it is false. So, it is the size of the acceptance region. As we have already devised that acceptance and the rejection regions are the complimentary in nature.

Therefore, if we increase one the other one is decreased. So for example, if we want to reduce the type I error a possible scenario is to reduce the size of the rejection reason, but if you reduce the size of the rejection region the size of the acceptance reason will increase. And consequently the probability of type II error may increase. Therefore, a statistically speaking it is not possible to minimize both alpha and beta simultaneously. Further if we are dealing with the composite hypothesis then alpha and beta both are functions of the parameters. For example, let me take say normal mu 1 population; so we are having a random sample x 1, x 2, x n from a normal mu and one population where I have to take a decision on H naught regarding say mu is less than or equal to 0 or say mu is greater than 0.

Now, here the hypothesis is composite. So, what is the probability of type I error and what is the probability of type II error here. So, it could be in this form.

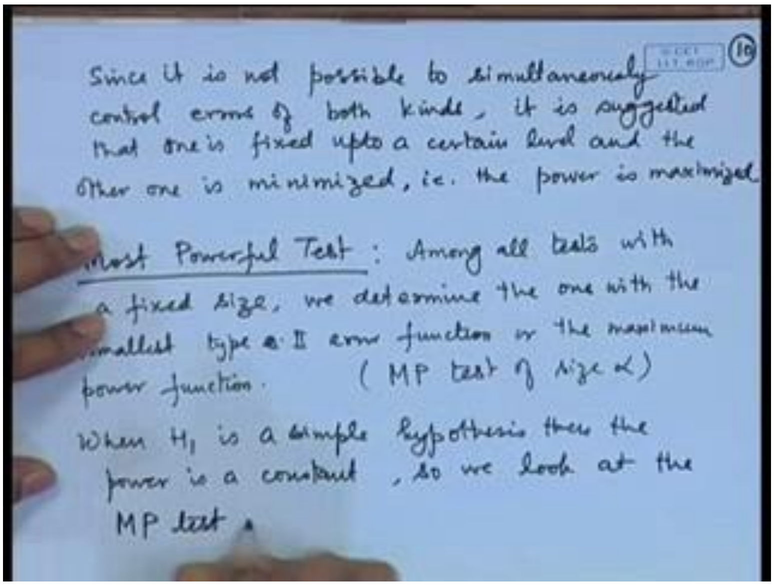
(Refer Slide Time: 10:31)



Probability of type I error is probability of rejecting H naught which may be based on some value of the sample when it is true. So, this could be probability of say x belongs to S R that is x denotes your vector x 1, x 2, x n of observations when it is true; true means the value is mu and mu is less than or equal to 0. So, here the type error is actually a function of mu. So, what one does is that one takes the maximum value of this let us call it say alpha star for mu less than or equal to 0; this is called the size of the test. Similarly, if you look at beta mu that is the probability of type II error that is probability of accepting H naught when it is false so it is the probability of x belonging to S A under mu when mu is greater than 0.

So, we consider 1 minus beta mu let me call it a notation beta is star mu this is called the power of the test. That is probability of rejecting H naught when it is false. So, a procedure is obtained such that we keep the size of the test to a fixed level and we try to minimize the probability of type II error or we maximize the power of the test. So, this is called the power function.

(Refer Slide Time: 13:09)

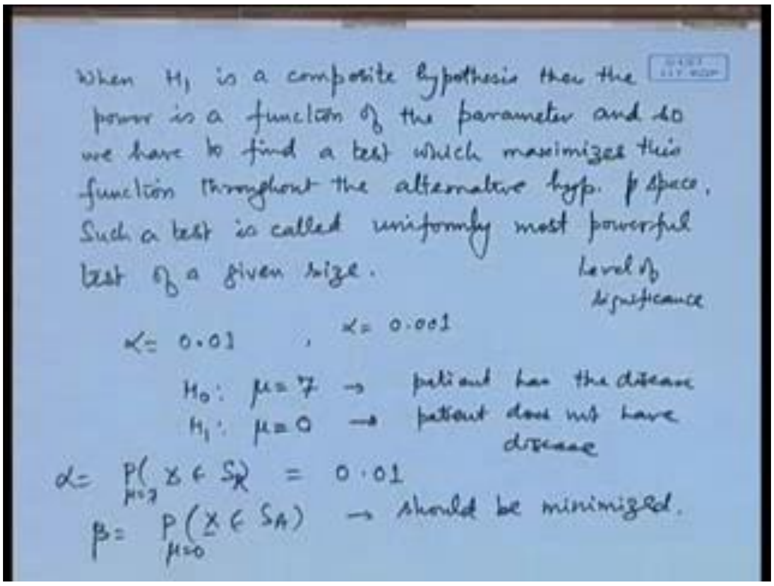


So in the testing of hypothesis the usual thing is- since it is not possible to simultaneously control errors of both kinds, it is suggested that one is fixed up to a certain level and the other one is minimized. That means, we should find a test procedure such that among all the test procedures which have size a given value say alpha star the power function of this test is having the a smallest value.

If it is a simple hypothesis, simple versus simple case then what will happen that alpha mu will be a single value say alpha naught and beta will be a fixed value, so we minimize beta or maximize 1 minus beta. So this is giving the concept of the most powerful test that is the power is maximized. So, most powerful test is among all tests with a fixed size we determine the one with the smallest type II error function or the maximum power function. So, this is called most powerful test are MP test of size say alpha. Now if it is a composite hypothesis then, so actually power function is dependent upon the when it is false; that means, when H 1 is true. So, when H 1 is a simple hypothesis then the power is a constant, so we look at the most powerful test.

However, if H 1 is a composite hypothesis because it is false, so what will happen here that this will be a function of parameter; beta mu and 1 minus beta mu that is beta star mu. So, in that case we have for a fixed size we should have the power maximum at all the parameter points in the rejection region. That means, for the alternative hypothesis where that is it is H naught is false that is when H 1 is true then it gives a concept of uniformly most powerful test.

(Refer Slide Time: 17:26)



When H 1 is a composite hypothesis then the power is a function of the parameter and so we have to find a test which maximizes this function throughout the alternative hypothesis a space. Such a test is called uniformly most powerful test of a given size.

So, now let us see the practical situation here. Let us go back to our problem of determining that whether a patient has a disease or not. So, in this case as we saw that the type I error is quite disasters in consequence; that means a patient may die. So, we may not like to have the probability of type I error to be high. So, we may decide to keep alpha to a very small level, say we may put say alpha is equal to 0.01; that means 1 in a 100 chance of error. Or if we are even more careful we may put 0.001; that means 1 in a 1000 chance of error is 1 in a 1000. So, in that case we will like to find a most powerful test such that the probability of type II error is very smallest.

For example, we may conclude that there is a particular kind of test and it may give the value say mu is equal to say 7 then the patient has the disease- say patient has the disease and say mu is say equal to it is possible to have only two types of observation say here mu is equal to 0 then the patient does not have disease. Suppose the test is devised in such way that the measurements that are taken they may take only two possible values and when mu is equal to say when the patient has the disease, when mu is equal to 0 the patient does not have the disease.

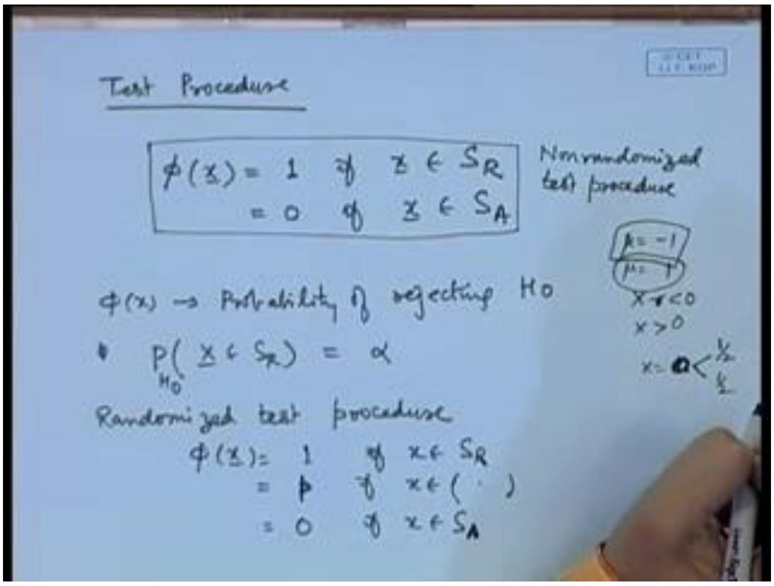
In this particular case the probability of type I error is that is when mu is equal to 7 that is when x belongs to S rejecting means S R x belongs to S R when mu is equal to 7 we may put this value to be say 0.01. And now we find that region that is x belonging to S A when mu is equal to 0. So, this function this value beta should be minimized. That means, we should find a test procedure. So, basically determining a test procedure means that we are fixing S R and S A. So, find that procedure which gives you S R and S A in such a way that this beta is minimized.

So, this is the theory of finding out optimal test or you can say most powerful tests here. Now one point about this quantity which I mentioned as the size of the test or it is also used as the terminology level of significance. Now what is the rationale behind writing down that alpha is equal to 0.01 or 0.001. So, the initial problem of the testing of hypothesis the way it was developed, there the theory of most powerful test was developed in such a way that we fix alpha and then we find the most powerful test. So, in the test procedure then this alpha will a play role.

Now this will involve looking at the probability points of certain distribution such as normal distribution, chi square distribution, t distribution, f distribution etcetera. Now for normal distribution it is fine but for t distribution or chi square etcetera we encounter incomplete gamma functions or incomplete beta functions. And therefore the tables are calculate for a specific parameter values only. And therefore, those values were tabulated for value such as 0.01, 0.05, 0.025 etcetera. And that is why in most of the books you will find these values to be given.

I will also discuss with you the significance testing which is an alternative approach to the hypothesis testing problem, where nowadays because of the computer packages it may not be required to fix the value of alpha. We will come to that problem a little later.

(Refer Slide Time: 23:30)



Now, here what we are doing is that we are a specifying a test procedure in such a wave. So, we may write it like this a test procedures description. What we are doing is we are saying phi x is equal to 1 if x belongs to S R it is equal to 0 if x belongs to S A. So, this means the interpretation that we can give phi x is the probability of rejecting H naught. That means if I say phi x is equal to 1; that means, if x belongs to S R we reject H naught with probability 1 that means we always reject.

Similarly, probability is 0 if x belongs to S A means that we never reject H naught when x belongs to S A. This description of the test function is also in you can say concordance with the concept that I said that we either reject a null hypothesis are we do not reject the null hypothesis. So, this interpretation is based on that thing. So, a non randomize test procedure which I described just now can be described by a test function phi x. However, there may be certain cases, because what we are saying is that the probability of rejecting that is x belongs to S R when H naught is true this is equal to alpha.

Now in a given situation when we are devising a test it may turn out that this value is not exactly equal to alpha for any value here, because that may be changing depending upon the points and when we include few points it may become more then alpha, if we delete some points it may become less then alpha. In that case we may adopt a procedure called a Randomized test procedure, where we may take say phi x is equal to 1 if x belongs to S R it is equal to say some value let me call it p if x belongs to some value and 0 if x belongs to S A. So, what has happened here? That means, this S R and S A these are not exhaustive. So, we have a space here where we reject with probability p and accept with probability 1 minus p. So, this may be needed for example in a situation like say we are having testing say mu is equal to minus 1 against mu is equal to 1 and our test is based on x; so if I say x is negative then you go in favor of this, if x is positive you go in favor of this.

Now, what happens if x is equal to 0? Although the probability of this may be 0 in continuous distribution, but in the discrete distribution this may be occurring with the positive probability. In that case we may decide to conduct an additional experiment and decide on the basis of that is. Say coin tossing, say with probability half we accept with probability half we reject. So, this is called a randomized test procedure. So, sometimes we may have to adopt a randomized test procedure in order to achieve a given level of significance or given size of test.

Now, in the fourth coming lecture I will be discussing the procedure for obtaining the most powerful tests and then we will look at the applications for various distributions.