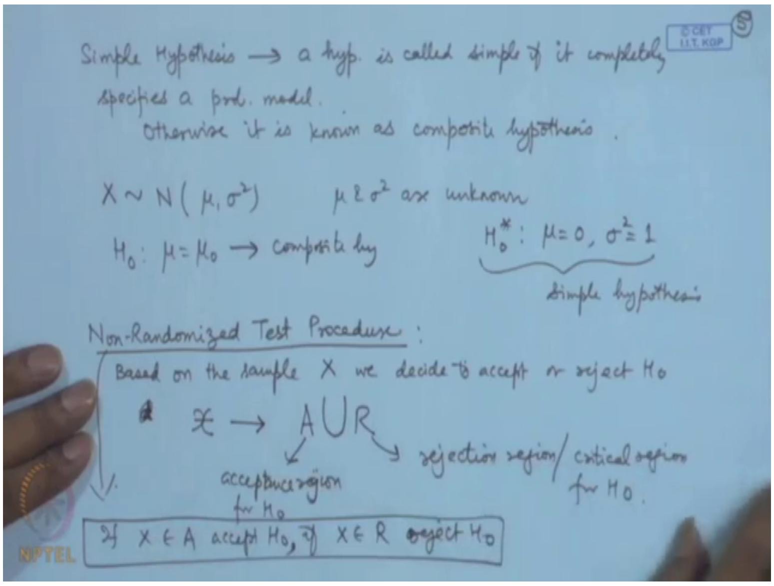
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 32
Testing of Hypothesis: Basic Concepts - II

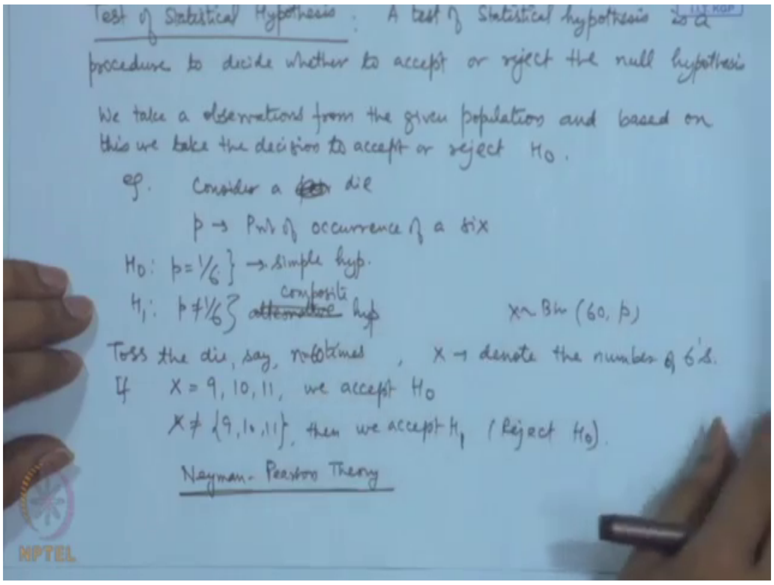
Now, the tests have been derived based on certain classification, the simplest classification that is there that is about a hypothesis which is called a simple hypothesis. So, a hypothesis is called simple a hypothesis is called simple if it completely specifies a probability model.

(Refer Slide Time: 00:29)



Let us consider say this problem here.

(Refer Slide Time: 01:01)



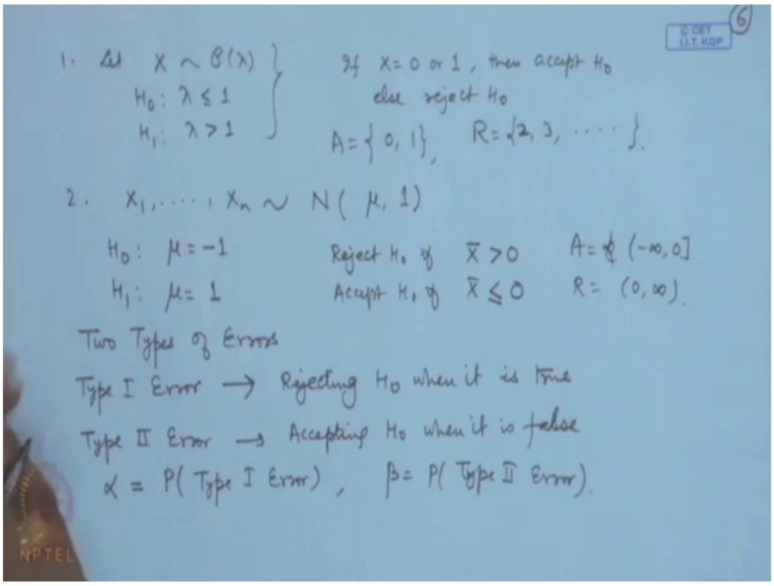
Here if I say p is equal to 1 by 6 then the distribution of X that is the number of 6 X, X follows binomial 60 p. If I say p is equal to 1 by 6 then the distribution becomes completely specified. So, this is actually a simple hypothesis. But if the specification is not complete then it is known as a alternative hypothesis. For example, if you look at p is not equal to 1 by 6 then this is an alternative hypothesis, sorry composite hypothesis, this is known as a composite hypothesis. Otherwise it is known as composite hypothesis.

Let us look at the problems specified earlier. Say let us consider say X follows normal mu sigma square, where mu and sigma square are unknown. If I specify say H naught mu is equal to mu naught, then this is specifies mu, but it does not say anything about sigma naught square therefore, this is a composite hypothesis. Suppose, I give hypothesis mu is equal to 0 sigma square is equal to 1, then it specifies completely the distribution here. So, this is a simple hypothesis.

We specify what is a test procedure? So, a test procedure I will split into 2 portions one of them is called a non-randomized test procedure. So, based on the sample X we decide to accept or reject H naught ok. So that means, a decision rule is nothing, but a non-randomized test procedure is nothing but this procedure. So, for example, you may assign a function say d x.

So, basically what is happening is that if we have the sample space x, it is subdivided into two regions; one is called the acceptance region and another is called the rejection region. So, this is the acceptance region. I am talking and respect of the hypothesis; H naught, acceptance region for H naught and this is called rejection region. In statistical terminology this is also called critical region; critical region for H naught. That means, a non-randomized test procedure is like this if X belongs to A, then you say accept H naught, if X belongs to R we reject H naught. So, this is a non-randomized test procedure. Let me explain through one or two examples here.

(Refer Slide Time: 05:25)



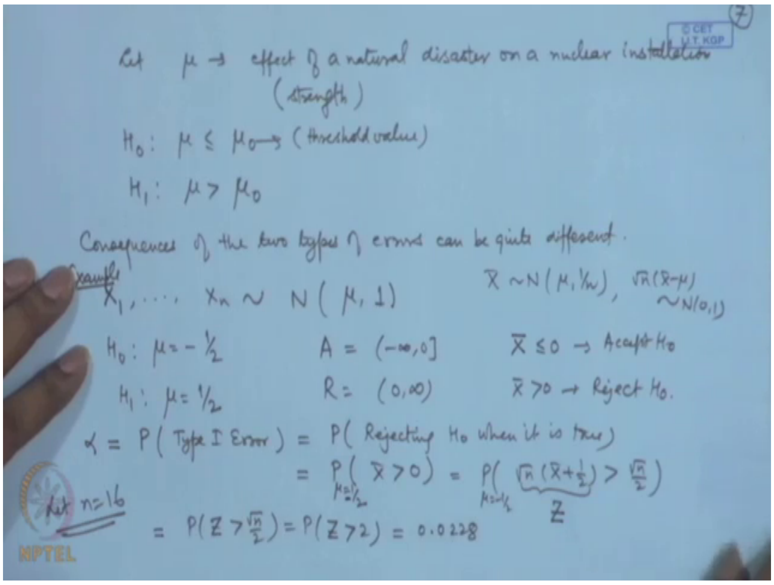
Let us consider say X following Poisson lambda distribution. Our null hypothesis is say lambda less than or equal to 1 and H 1 is say lambda greater than 1. Now, based on the observations if X belongs to 0 or 1, then accept H naught else reject H naught. So, here our acceptance and critical regions are, let us also consider another example say I have a random sample X 1, X 2, X n from a normal distribution with mean mu and variance 1. And we want to test the hypothesis say mu is equal to say minus 1, against say mu is equal to plus 1. Then we may take a test procedure as reject H naught if X bar is greater than 0, accept H naught if X bar is less than or equal to 0. So, here our acceptance region is minus infinity to 0 and the rejection region is 0 to infinity.

Now, when we carry out a test of procedure a test of hypothesis, so we are basically introducing a decision procedure whether based on our sample we should accept a null hypothesis or reject the null hypothesis. Since, our decision is based on the sample there are possibility of errors that means, we might have taken a correct decision. As I mentioned to you that when we say based on our hypothesis procedure that p is equal to say 1 by 6 or p is greater than 0.75 etcetera, it is only an assertion in support of our hypothesis based on the sample. It does not mean that hypothesis is actually true or false.

Therefore, because hypothesis involves the unknown parameter of the distribution or the population which we are not sure what actually the value is, it could be our sample procedure whatever sampling a scheme we have implied based on that whatever sample we have taken, it may be possible that based on that we are taking this decision. Therefore, we are likely to come it two types of errors, they are called type I error and type II error.

So, type I error is rejecting H naught then it is actually true and similarly type II error is accepting H naught when it is false. We use a notation alpha as the probability of type I error and beta is the probability of type II error. Now, the consequences of these two types of errors can be quite different. For example, consider a problem of guessing about say the effect of a natural disaster on a certain construction ok.

(Refer Slide Time: 09:59)



Let us say let say mu denote the effect of. So, this could be power ok, which is measured in power effect of a natural disaster on a nuclear installation. Now, this effect is estimated in terms of say strength ok. We may like to check whether mu is less than or equal to a certain specified number this could be our threshold value or mu is greater than this threshold value, this is the threshold value.

That means if the strength is below this a strength of the natural disaster is below this then, the damage will not be much. However, if it is above this there will be the installation will be demolished. For example, the effect of a tsunami that we observed last year in the Fukushima the nuclear plant and Japan, the effect of the effect of the disaster was such that it basically demolished the nuclear installation leading to a very wide catastrophic effect.

So, if our hypothesis for example, null hypothesis is true and we are actually rejecting it; that means, we will be making arrangements for a much higher scale of natural disaster. Therefore, we are safe in the sense that even if the natural disaster is occurring we are safe; however, it may in tail a very large amount of expenditure and also the maintenance costs. Whereas, if H naught is actually false and we accept it in that case we are making a very serious error as it has happened in the nuclear disaster and Japan. We can consider much simpler situation.

For example, a patient goes to a doctor with certain complaints and certain diagnostic tests are conducted on him, based on the diagnostic tests the doctor concludes that the patient does not have the disease. However, actually he may have the disease, if that is so and the doctor has concluded that the patient does not have the disease he will not give a commensurate medication which may lead to further complications to the patient and he may actually ultimately die also. On the other hand, if the doctor concludes that the patient has the disease when actually he does not have the disease, he may give medicines to treat that disease which may lead to some side effects as well as a financially stress to the patient.

So, the consequences of the two types of errors can be quite different. And this probabilities of type I error and type II error that is alpha and beta they give a measure of the size of the errors here that one may have.

Let us consider the example that we took earlier. Let us look at the relative values of this. So, consider say for example, X 1, X 2 X n following normal mu 1 and our null hypothesis say mu is equal to minus half, H 1 mu is equal to say plus half. Now, we have we take a decision based on X bar and our X bar if it is negative, so we accept the hypothesis the rejection region is 0 to infinity. So, the decision is based on X bar that is

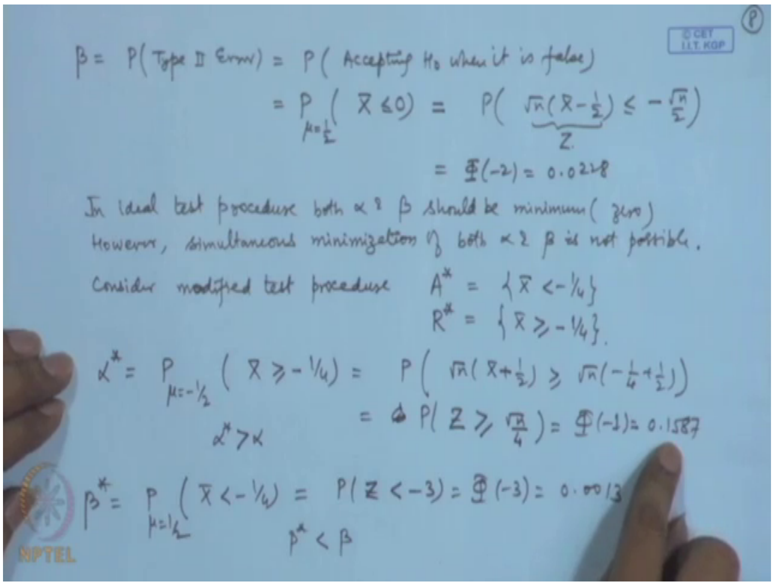
X bar less than or equal to 0 or X bar greater than 0. Here we accept H naught and here we reject H naught.

Now, let us calculate the probabilities here. Alpha that is the probability of type I error, that is equal to probability of rejecting H naught when it is true. That is the probability of the region X bar greater than 0 when it is true means mu is equal to half. Now, here the distribution of X bar is normal mu 1 by N. So, root n X bar minus mu follows normal 0 1.

So, when mu is equal to half sorry then it is true, so here mu is equal to minus half. So, when mu is equal to minus half you will get root n X bar plus half root n X bar plus half greater than root n by 2 that is when mu is equal to minus half. Then this has a standard normal distribution. So, this is equal to probability of Z greater than root n by 2.

Let us for example take say n is equal to 16. If n is equal to 16 this is probability of Z greater than 2 where Z follows normal 0 1, the probability of Z greater than 2 is 0.0228 if we see the tables of the normal distribution. So, the probability of type I error is 0.22; that means, its nearly 2 percent probability of type I error.

(Refer Slide Time: 16:51)



Now, in this case let us also consider beta. Beta is a probability of type II error that is probability of accepting H naught when it is false, that is equal to probability of X bar less than or equal to 0 when it is false means mu is equal to half. That is when H 1 is true, that is here mu is equal to half here. When mu is equal to half we have root n X bar minus half as the standard normal variable since n is equal to 16 this is becoming 2. So, phi of minus 2 that is 0.0228, so, in this case alpha and beta are same.

Now, in ideal test procedure both alpha and beta should be minimum basically there should be 0. But practically speaking this is not possible because if I want to make the probability of type I error as 0; that means, the rejection region should be an empty set with respect to the distribution when the null hypothesis is true.

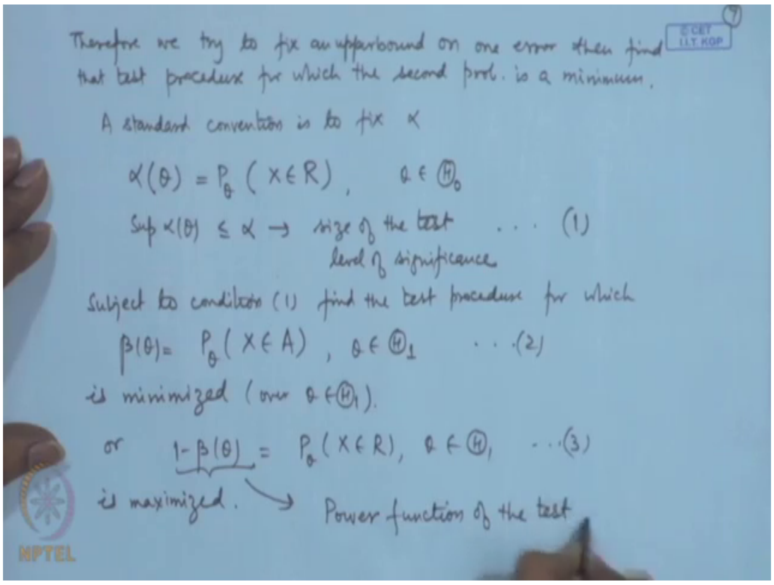
Now, if the set is empty then the probability of accepting H naught that will become almost actually it will become 1, but if there may be a variation in the null alternative hypothesis value then it may be very high value. So, simultaneous minimization of, however, simultaneous minimization of both alpha and beta is not possible. In fact, we can try if we reduce say alpha, then beta will increase if we reduce beta then alpha will increase.

Let us take say for example, we modify consider modified test procedure for the same problem. So, I give the rejection region as say X bar less than minus 1 by 4 and the complementary region that is the rejection region as X bar say greater than or equal to minus 1 by 4. If I take this then our probability of type I error X bar greater than or equal to minus 1 by 4 that is equal to probability of root n X bar plus half greater than or equal to root n minus 1 by 4 plus half, that is equal to probability of Z greater than or equal to root n by 4 that is equal to phi of minus 1 that is 0.1587.

So, you can see here alpha star is greater than alpha. Let us calculate say beta star here that is the probability of X bar less than minus 1 by 4 when mu is equal to plus half. So, are going in the same way this value turns out to be probability of Z less than minus 3 that is phi of minus 3 that is 0.001 3. So, here beta star is actually less than beta. In fact, this is close to 0.1 percent actually 0.001.

So, here we have drastically reduced beta star, but that has increased alpha that is the probability of type I error. Therefore, one has to look for a compromise solution.

(Refer Slide Time: 21:39)



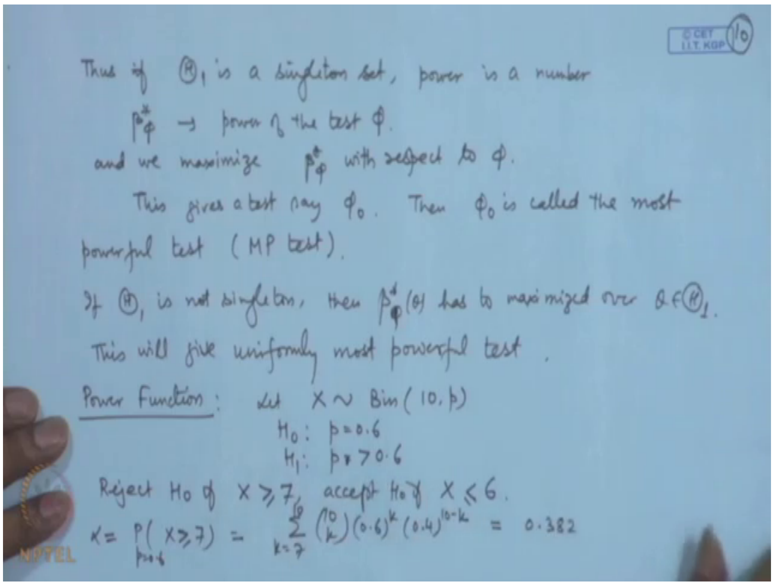
The compromise solution is that, therefore we try to; therefore, we try to fix an upper bound on one error and then find that test procedure for which the second probability is a minimum.

So, a standard convention is to define the hypothesis in such a way that we consider the probability of type I error as more serious and then we fix an upper bound to that. So, a standard convention is to fix alpha. So, the example that I have considered here alpha and beta are 2 numbers, but in general there will be functions of the parameter. So, if they are the functions of the parameter then we need to look at the maximum value. So, for example, alpha will be in general a function of the parameter that is the probability of rejecting H naught when theta belongs to theta naught that is when it is true.

So, we take supremum of alpha theta less than or equal to alpha. This is usually called the size of the test; our level of significance. So, let us put this as condition 1. Then subject to condition 1, find the test procedure for which beta theta that is equal to probability of X belonging to acceptance region, that is accepting H naught when it is false for which this is minimized once again see minimized means this is a function of the parameter. So, minimized means over theta belonging to theta one.

We also say 1 minus beta theta that is called the probability of X belonging to R or theta belonging to theta 1 is maximized. This 1 minus beta theta this is called power function of the test. So, power of the test is defined as 1 minus the probability of type II error. So, then we get the concept of most powerful test and the uniformly most powerful test.

(Refer Slide Time: 25:19)

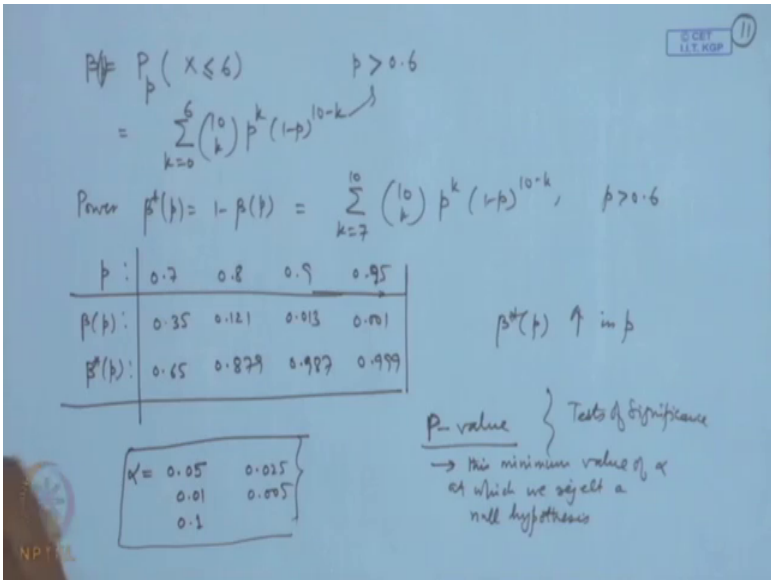


Thus, if theta 1 is a singleton set power is a number. So, we can use the notation say beta star phi to denote the power of the test phi and we maximize beta star phi with respect to phi this is called. So, this gives a test say phi naught. Then phi naught is called the most powerful test, that is MP test.

If theta 1 is not singleton then beta phi star theta has to be maximized over theta belonging to theta 1 this will give uniformly most powerful test. So, I mentioned that terminal Neyman-Pearson theory. So, the Neyman-Pearson theory approaches the testing of hypothesis problem from this viewpoint, that is it solves an optimization problem and gives the solution. So, in the first case they give a simple hypothesis versus simple hypothesis case. So, we are able to get the solution in a standard form and then those procedures are generalized to obtain the solutions in cases where uniformly most powerful test can be derived.

Let me give example of a power function. Let us consider say X follows say binomial 10 p and our hypothesis testing problem is say p is equal to 0.6 against say p is equal to say p greater than 0.6. And let us take the test procedure as reject H naught if X bar X is greater than or equal to 7, accept H naught if X is less than or equal to 6. So, here alpha that is the probability of type I error that is probability of rejecting H naught when it is true that is equal to sigma 10 c k, k is equal to 7 to 10, 0.6 to the power k, 0.4 to the power 10 minus k. From the tables of the binomial distribution one can see this value turns out to be 0.382. So, the probability of type I error is very high using this test procedure.

(Refer Slide Time: 29:13)



Let us look at beta that is the type II error. So, that is probability of accepting H naught when it is false. Now, this means p is greater than 0.6 here. This value is simply 10 c k p to the power k, 1 minus p to the power 10 minus k for k is equal to 0 to 6. Here p is any value greater than 0.6.

If we consider the power. So, this is a function of p here and if we consider the power that is beta star p that is 1 minus beta p then that is equal to sigma k is equal to 7 to 10, 10 c k p to the power k 1 minus p to the power 10 minus k for p greater than 0.6. I have tabulated these values for different values of p. You can see it from the tables of the binomial distribution at 0.7, 0.8, 0.9 and 0.95. This value is 0.35, so this is 0.65, 0.121, 0.879; at 0.9 it is 0.013, this is 0.987 and this is 0.999.

So, you can see here that this power function is actually increasing in p, this is increasing in p, actually it is reaching almost 1 as p nears 1. So, that shows that for this of course, you have the size of the test very high, but even for this size the test is proceeding in the right direction that is the probability of type II error is gradually decreasing and the power function is gradually increasing here.

Now, the next point I mention about that we are saying that fixed the size of the test alpha and then determine a test procedure for which the power function is maximized. Now, when you say fix alpha then the that is the job of the statistician, that he has to fix alpha and for that we have to determine a best procedure. The question is what should be the value of alpha. If you look at by standard textbooks here in the standard textbook the value of alpha and even in the questions that they ask they will fix the values as 0.05, 0.01, 0.1 0.025, 0.005 etcetera. So, these are some of the commonly used values which you can find in the tables of the distributions that are used for testing.

Now, the reason for taking these values is that in those earlier days the tables of the distributions we are calculated manually by using certain computation procedures and then calculators and therefore, it was convenient to have a few tables and therefore, the selected values were taken as like 0.05, 0.01. So, basically this means 5 percent, this means 1 percent, this means 10 percent. And slowly these values became like conventional and a very standardized that people generally use these as the level of significance, but there is nothing sacrosanct about these values. In fact, presently it is more fashionable to use what is called as a P value.

A P value is the value which will assign, so basically this is the minimum value of alpha at which we reject a null hypothesis. This is the minimum value at which, so this is called significance, this is called tests of significance are the P value. We will discuss this procedures a little and afterwards firstly, we will discuss the Neyman-Pearson theory, but this is an alternative way of carrying out the test of hypothesis here.

Now, in the next lecture I will give the Neyman-Pearson fundamental lemma that is the basic result which gives how to find out the most powerful test and later on we will discuss its applications to various distributional models. So, I will be continuing this in the next lecture.