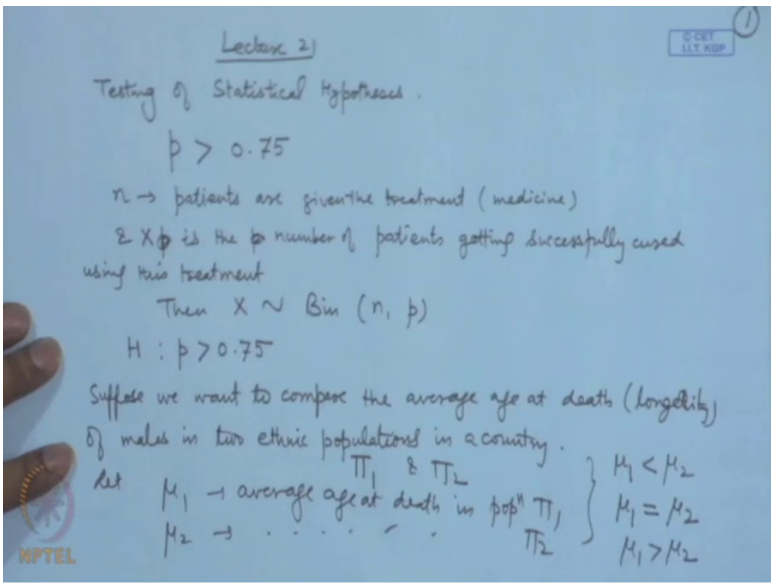
Statistical Inference

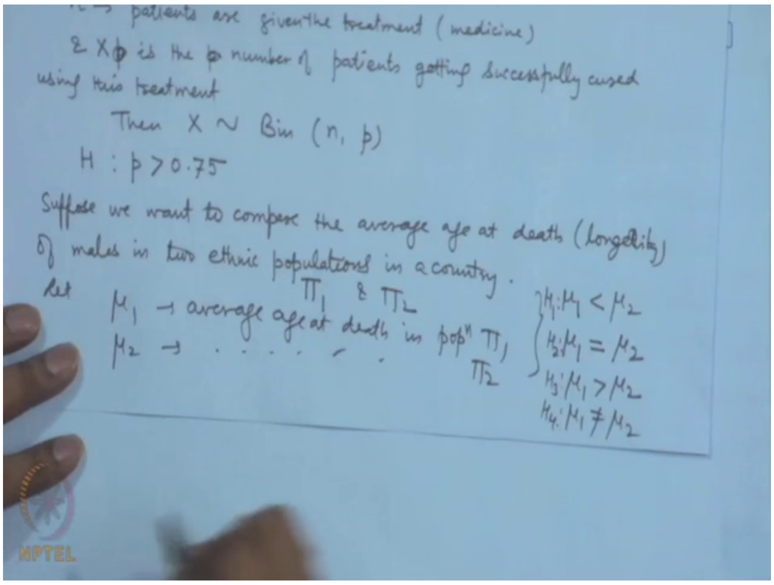
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Module No. # 01
Lecture No. # 21
Testing of Hypotheses: Basic Concepts

In the previous lectures I have discussed the problem of point estimation. In the beginning when we introduce the problem of a statistical inference, we mentioned that there are two broad division of this topic. One is where we want to guess the value of a parameter or parametric function; this is called the problem of estimation. And, in the estimation, we had the problem of point estimation where we specify a value which we call as an estimator or sometimes we assign an interval, which is called as an interval estimate or confidence interval. However, there may be situations when we want to test some statement about the parametric function. So, for example, there is a medicine for treatment of a certain disease and we know that the success rate of this medicine is say 3 4, like 0.75 or 75 percent of the patients who take this medicine they get cured.

(Refer Slide Time: 01:56)

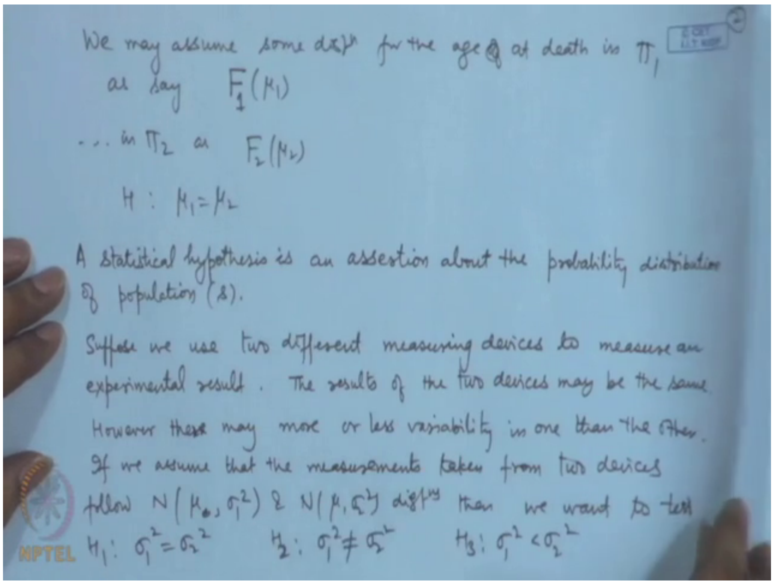




Now, a drug manufacturer introduces and improvisation over this medicine and introduces, he wants to introduce the medicine in the market. Now, certainly he will be interested to know whether the new medicine is more effective. Suppose, I say p is the proportion of the patients who get cured using this new medicine, then the question that obviously one has to ask, whether p is greater than 0.75. So, now this problem now can formulate in statistical wave in this following fashion. Suppose, in general n patients are there, n patients are given the treatment or medicine and p is the, so, observed proportion X is the number of patients getting successfully cured using this treatment. Then, we can say that X follows binomial n p distribution and we want to test the hypothesis, we use the notation H for a hypothesis, p is greater than 0.75

Similarly, let me state another problem, suppose we want to compare the average age at death that is called longevity of males in two ethnic populations in a country. So, let us say there are two ethnic populations, let me call this population say pi 1 and pi 2. And, let mu 1 denotes the average age at death in population pi 1 and mu 2 is the average at death in population pi 2. Then, we are interested in testing hypothesis say whether mu 1 is less than mu 2, we may like to test whether mu 1 is equal to mu 2, we may like to test whether mu 1 is greater than mu 2, we may like to test whether mu 1 is not equal to mu 2. Let me give the some names to be hypothesis, say H 1 H 2 H 3 H 4.

(Refer Slide Time: 05:06)



Now, to put it in the statistical frame work, we will assume that the ages at death in the two populations follow certain distribution, we may assume here. So, we may assume some distribution for the age at death in pi 1 as say let me call it f 1 mu 1 and same thing assume in pi 2 you assume as say f two mu 2. And then we are testing the hypotheses whether mu 1 is equal to mu 2 or mu 1 less than mu 2 or mu 1 not equal to mu 2 etcetera.

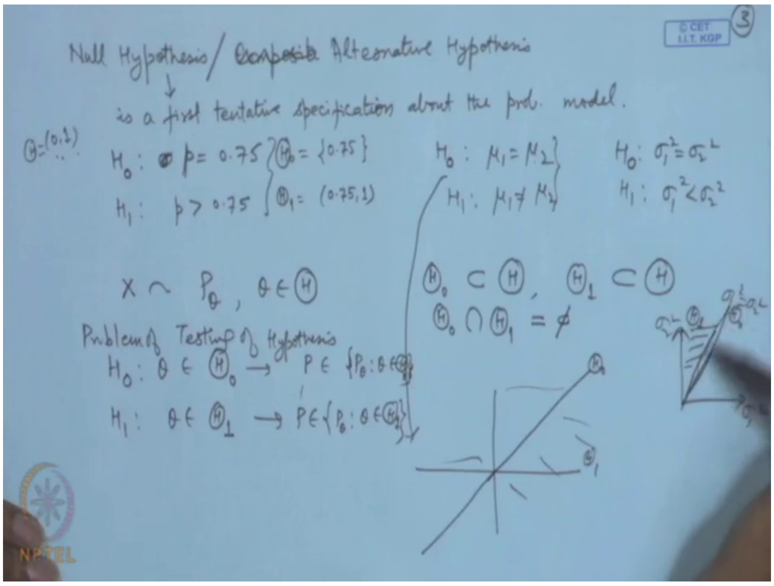
Suppose, so, basically what we are doing, what is statistical hypotheses then, a hypotheses is treatment about the parameter of a population for example, here I am considering binomial distribution and we are making an statement that p is greater than 0.75. Similarly, here we may be considering two populations say f 1 and f 2 of course, we may take them as normal populations are exponential populations are gamma populations. And then we are checking above an a statement about the parameters a we are making an statement about the relationship between the parameters here.

So, broadly speaking then we can say that a statistical hypotheses is an assertion about the; actually, here we have given the example where we are making statement about the parameter. But we may also consider something like this we want to find out the distribution of incomes. So, we may like to check whether the distribution of incomes follows a pareto distribution or the distribution of the incomes follow some other distribution. So, in that case we were checking the form of the distribution itself.

So, in general a statistical hypotheses is an assertion about the probability distribution of population. So, it could be one population it could be more than one population also. Let us consider suppose we use two different measuring devices to measure an experiment experimental result. So, some outcome is there from the experiment and we are using two different measuring devices.

Now, the results of the two devices may be the same; however, there may be more or less variability in one than the other. So, if we assume that the measurements take in from two devices follow say normal mu sigma 1 square and normal mu sigma 2 square distributions then we want to test say sigma 1 square is equal to sigma 2 square are we may like to test say sigma 1 is square is not equal to sigma 2 is square are we like to test say whether sigma 1 is square less than sigma 2 square etcetera. So, this are statements about the parameters of the distribution.

(Refer Slide Time: 10:16)



Now, we introduce what is called a null hypotheses are a composite hypotheses, sorry, alternative hypotheses. Now, in the example that I have described here, for example, we wanted to check whether the average longevity are average age at death in the two ethnic population is the same. Now, this is statement may not be true this may be true. Similarly, in the problem of introducing a new medicine in the market for curing a certain decease, we introduced the hypotheses whether p is greater than 0.75

Now, it may be possible that p is greater than 0.75 are it could be that p is less than or equal to 0.75 or we may say p is equal to 0.75 or p is equal to 0.95 or p is equal to 0.9. So, when we make and initial assumption about the probability distribution that we call as a null hypotheses. So, we make take a decision whether two except the null hypotheses are two reject the null hypotheses.

So, a null hypotheses is the, you can say tentative are the first specification about the probability model. So, a null hypotheses is a first are you can say tentative is specification about the probability model. So, for example, we may say H mu p is equal to 0.75, we may say H mu 1 is equal to mu 2 we may say H sigma 1 is square is equal to sigma 2 square.

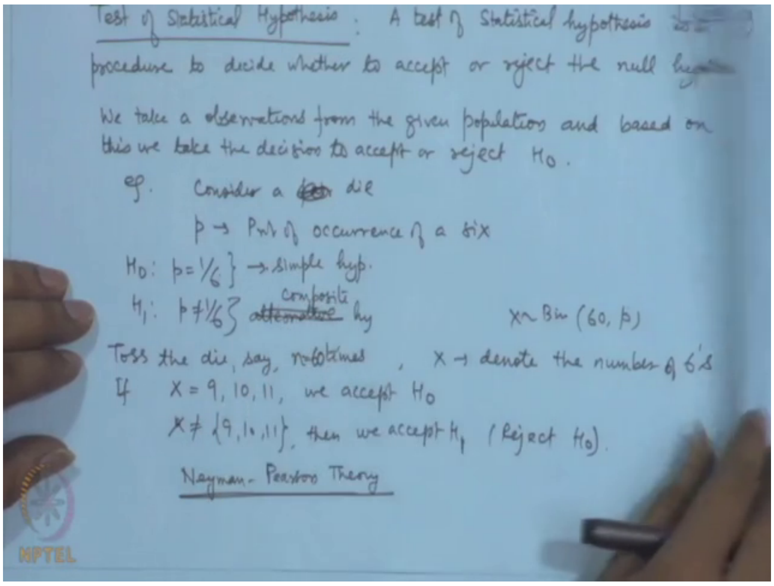
So, this is called a null hypotheses usual notation in the standard books is used H naught here. Now, when we specify a null hypotheses we also is specify an alternative hypotheses. To check that if this null hypotheses is rejected then what is other possibility? That is specification of another hypotheses in contrast to the null hypotheses that is called an alternative hypotheses. For example, in this problem we may say p is greater than 0.75, here we may say mu 1 is not equal to mu 2, here we may say sigma 1 square is a less than sigma 2 square.

So, certainly the generally specification can be like this suppose we are considering probability modal p theta were theta belongs to the parameter is space theta. Then let us consider two subsets let me call theta naught as a subset of theta and theta 1 as a subset of theta, such that theta naught and theta 1 they are disjoint. Then we may setup the null hypotheses and the alternative hypotheses as; so this is problem of testing of hypotheses. For example, in this case you are theta naught will become equal to the single ten set 0.75 and theta 1 will become the set 0.7521. In this case this is the parameter is space here is a two dimensionally space mu 1 is equal to mu 2 is the line. So, the theta naught is this line and all other place this is actually theta 1. In this particular case, we are dealing with the positive half sigma 1 is square is equal to sigma 2 is square is this and sigma 1 is square less than sigma 2 is square is this. So, is this is theta naught and this is theta 1.

So, the null hypotheses and the alternative hypotheses is specified the parameters to be in to distinct parameter is spaces. Basically, we are specifying the familiar, we can also say if p is the probability distribution, then we can say it belongs to the familiar p theta, theta belonging to theta naught or here we can say p belongs to p theta, theta belonging to theta 1. So, the null hypotheses and alternative hypotheses they usually divide the parameter is space in 2 to complementary regions.

Now, the way I have explained here it is not necessary that the two regions exhaust the full possibility. For example, in this case, the union of theta naught theta one that is not the full parameter space, for example, the binomial distribution theta is actually 01, but the union of these two is not this. In this particular case the union will be equal to the full type two dimensionally space. Once again in the third problem the union of these two is only half plain here, half of the first quadrant. So, we made take the null and alternative hypotheses as completely complementary regions are sometimes they will be only disjoints, but need not be completely complementary; that means, they may not exhaust the fully space. Then what is a test of statistical hypotheses?

(Refer Slide Time: 16:39)



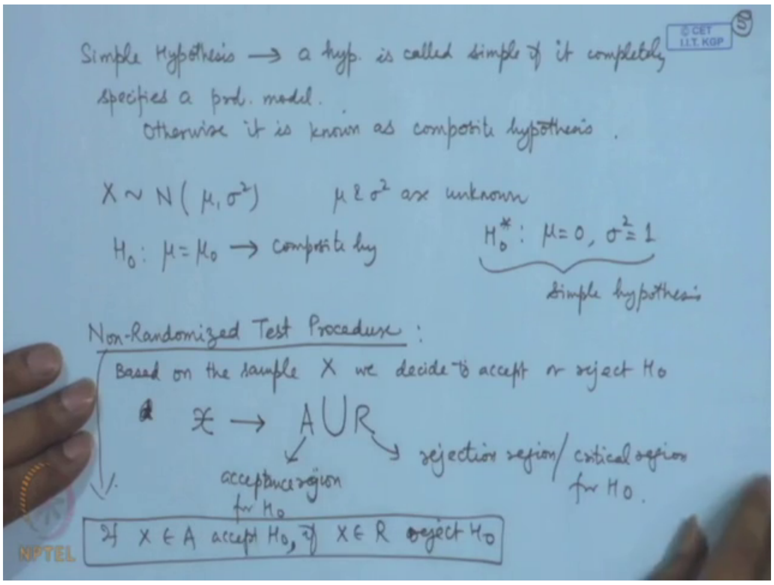
So, a test of statistical hypotheses is a procedure to decide whether to except or reject the null hypotheses. So, usually, what we do? We take a random sample, we take observations from the given population and based on this we take the decision to except or reject H naught. See for example, consider say a die, suppose, we are considering; consider a fair die consider a die we do not know whether it is fair or naught and consider say p is the probability of occurrence of a six. Can we would like to check whether p is equal to 1 by six or not. Then we may take a decision based on certain experiment suppose, we conduct the experiment we toss the die say a certain number of times say n times. Let me put n is equal to some number here, suppose, I put 60 then we may say that if and here let X denote the number of 6's.

So, if X is say 9 10 11 we accept H naught and if X is naught equal to this numbers, then we accept H 1 are you can say reject H naught. Then this is a test procedure, because what we are doing is based on the sample we are taking a decision whether to accept H naught or to reject H naught. Here I would like to clarify, when we say that we accept H naught or we reject H naught then this statements do naught have the validity of a truth. For example, if I say that X equal to 9 10 11 then we accept H naught. It does not mean that p is equal to 1 by 6. It simply means that based on our test procedure we are in favor of hypotheses H naught. Similarly, if I observe that X is different from 9 10 and 11 then we conclude here that H 1 is more possible, it does not mean that H 1 is actually true. It means that based on the sample we do not feel that H naught is a correct hypotheses; we feel that H 1 is more possible.

So, acceptance are rejection of a hypotheses based on the sample has only a circumstantial validity; that means, based on the observations we are making a decision whether to accept or reject a given hypotheses, these are naught the assertion about the truth full ness about these hypotheses. So, this is that is y this is called only a is statistical test or statistical hypotheses test.

Now, the theory that we are going to discussing the beginning that is known as the Neyman Pearson theory of testing as hypotheses, there are some alternating theory is also, but I am going to discuss this particular theory. In this theory it is essential to specify a null hypotheses and an alternative hypotheses, there are some other approaches where only the hypotheses for which we are initial interested will be enough.

(Refer Slide Time: 22:30)

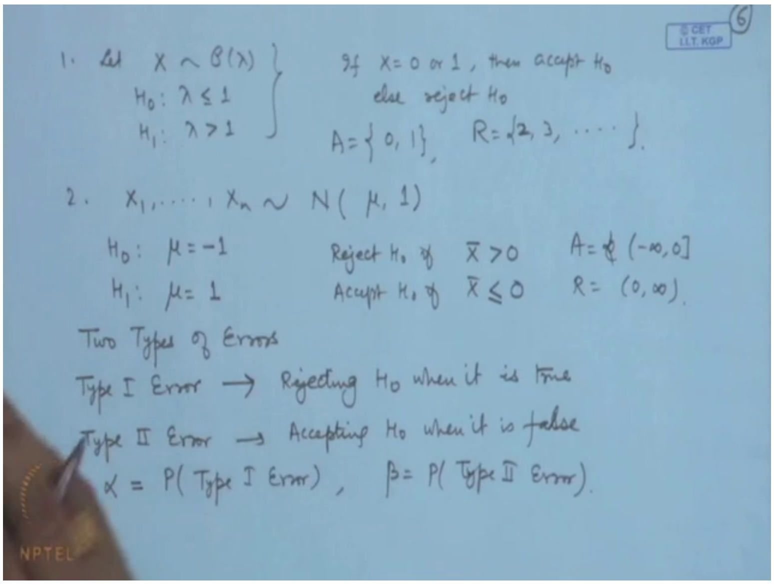


Now, the test have been derived based on certain classification. The simplest classification that is there that is about hypotheses which is called a simple hypotheses. So, hypotheses is called simple if it completely specifies a probability model. Let us consider say this problem here, here if I say p is equal to 1 by 6, then the distribution of X that is the number of 6's X follows by nominal 60 p. If I say p is equal to 1 by 6 then the distribution becomes completely specified so, this is actually a simple hypotheses. But if the specification is not complete then it is known as alternative hypotheses. For example, if you look at p is not equal to 1 by 6 then this is n alternative hypotheses, sorry, composite hypotheses this is known as a composite hypotheses. Otherwise, it is known as composite hypotheses.

Let us look at the problems is specified earlier say; let us consider say X follows normal mu sigma is square, where mu and sigma is square are unknown. If I specifies 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 hypotheses. Suppose, I give hypotheses in mu is equal to 0 sigma is square is equal to 1 then it is specifies completely the distribution here so, this is a simple hypotheses. (No audio from 24:59 to 25:09) We specify what is a test procedure? So, a test procedure I will explained into two portions one of them a scrolled a nonrandomized test procedure.

So, based on the sample X we decide to accept or reject H naught so that means, a decision rule is nothing but a nonrandomized test procedure is nothing but this procedure so, for example, you may assign a functions say d x. So, basically what is happening is that if we have the sample is space X it is sub divided into two regions, one is call the acceptance region and another is called the rejection region. So, this is the acceptance region, I am talking and respect of the hypotheses H naught, acceptance region for H naught and this is called rejection region; in statistical terminology this is also called critical region for H naught; that means, a nonrandomized test procedure is like this if X belongs to a then you say accept H naught if X belongs to R v reject H naught.

(Refer Slide Time: 27:26)

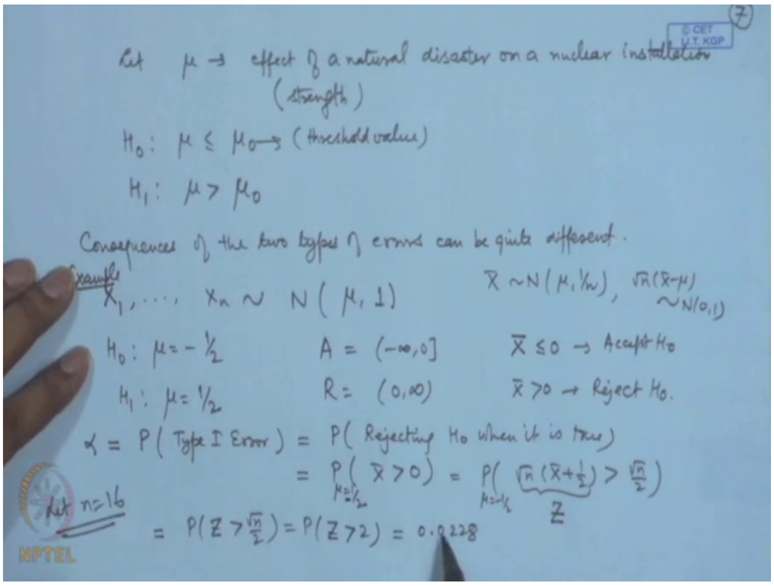


So, this is a nonrandomized test procedure. Let me explain through one or two examples here. Let us consider say X following Poisson lambda distribution our null hypotheses is say lambda less than or equal to 1 and H 1 is say lambda greater than 1. Now, based on the observations if X belongs to 0 or 1, then except H naught else reject H naught. So, here our acceptance and critical regions are (No audio from 28:03 to 28:12) 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 hypotheses 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 r 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 test of procedure a test of hypotheses so, we are basically introducing a decision procedure, whether based on our sample we should accept a null hypotheses are reject the null hypotheses. Since, our decision is based on the sample there are possibility of errors; that means, we might have taken a correct decision. As I mention to you that then we say based on our hypotheses procedure that p is equal to say 1 by 6 are p is greater than 0.75 etcetera, it is only an assertion in support of our hypotheses based on the sample. It does not mean that hypotheses actually true or false.

Therefore, because a hypotheses involves the unknown parameter of the distribution are the population which we are not sure what actually the value is. It could be our sample procedure whatever sampling is 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 one error and type two error. So, type one error is rejecting H naught when it is actually, true. And similarly, type two error is accepting H naught when it is false.

(Refer Slide Time: 31:59)



We use a notation alpha as the probability of type one error and beta as the probability of type two 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 certain construction. Let us say a let say mu denote the effect of; so, this could be power which is measured in power effect of a natural disaster on a nuclear installation.

Now, this effect is estimated in terms of say is strength. We may like to check whether mu is less than are 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 a strength is below this, a strength of natural disaster below this then, the damage will naught be much and however, if it is above this there will be the installation will be demolish. For example, the effect of a Tsunami that, we observed last year in the Fukushima nuclear plant and Japan, the effect of the disaster was such that it basically demolish the nuclear installation lead in to a very wide cut of strophic effect.

So, if our hypotheses for example, null hypotheses is true and we are actually rejecting it that means, we will be making arrangements for a much higher is 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 cost. 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.

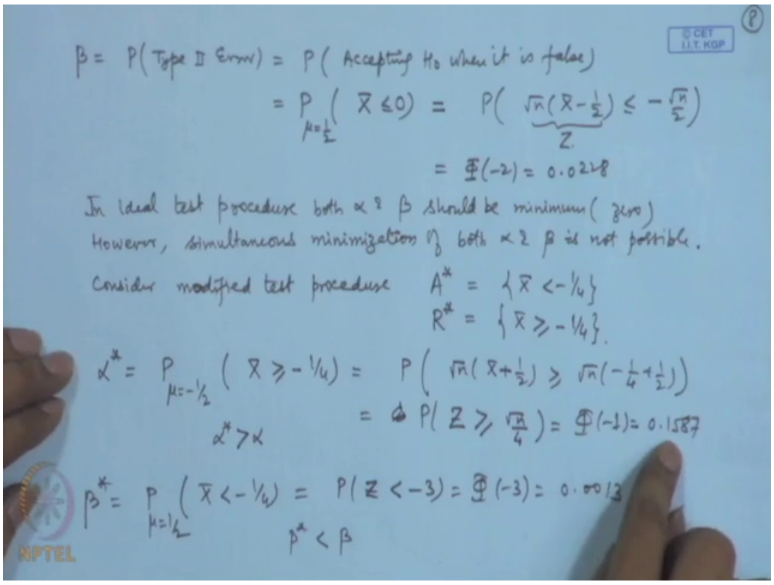
For example, a patient was to doctor with certain complaints and certain diagnostics are conducted on here, based on the diagnostic test the doctor concludes that the patient does not have the decease. However, actually he may have the decease, if that is so and the doctor as concluded that the patient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

So, the consequences of the two types of errors can be quite different and this probabilities of type one error and type two error that is alpha and beta. They give a measure of the the size of the errors here, that one may have a, let us consider the examples at we to curlier let us look at the relative values of this. So, consider so for example, X 1 X 2 X n following normal mu 1 and our null hypotheses 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 except the hypotheses are rejection region is 0 to infinity. So, the decision is based on X bar that is X bar less than or equal to 0 are 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 one 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 when 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 as a standard normal distribution. So, this is equal to probability of z greater than root n by 2. Let us for example, takes a 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 you see the tables of the normal distribution.

(Refer Slide Time: 38:53)



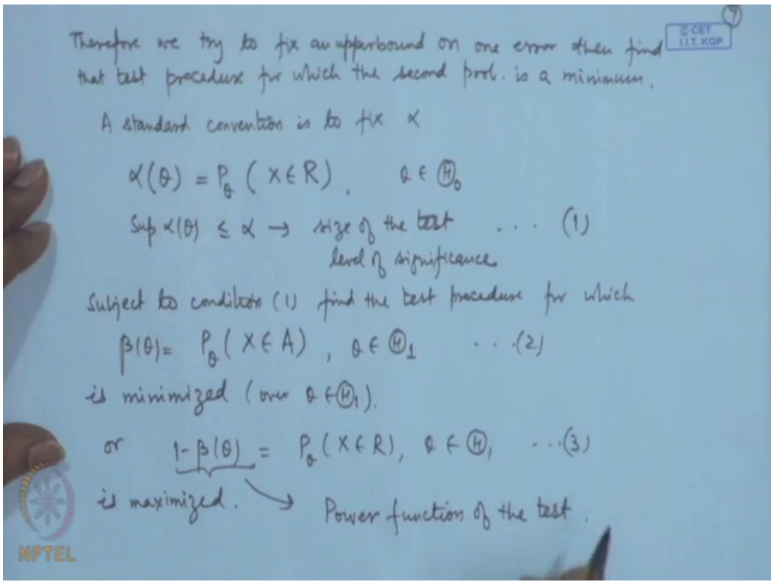
So, the probability of type one error is 0.0228; that means, it is nearly 2 percent probability of type one error. Now, in this case, let us also consider beta, beta is the probability of type two error, that is probability of accepting H naught, when it is falls that is equal to probability of X bar less than or equal to 0, when it is falls 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 r same.

Now, in ideal test procedure alpha and beta should be minimum basically, they should be 0. But practically speaking this is not possible, because if I want to make the probability of type one error as 0 that means, the rejection region should be an empty set with respect to the distribution, when the null hypotheses 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 and the null alternative hypotheses 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 a rejection region as X bar say greater than r equal to minus 1 by 4. If I take this then, our probability of type one 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 is star is greater than alpha. Let us calculate say beta is 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 0.0013.

(Refer Slide Time: 43:40)

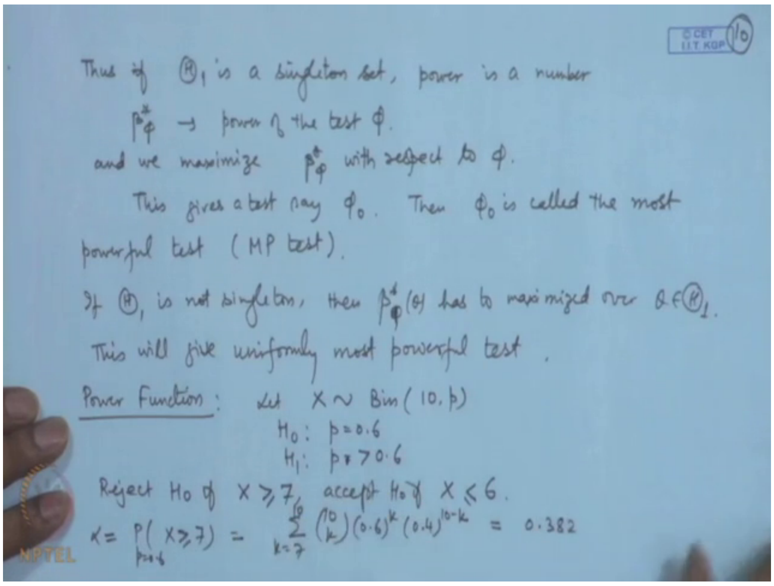


So, here beta is star is actually less than beta in fact this is close to 0.1 percent actually, 0.001. So, here we have drastically reduced beta is star, but that has increased alpha, that is the probability of type one error. Therefore, one has to look for compromise solution, the compromise solution is that; 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 hypotheses in such a way that, we consider the probability of type one error has 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 two numbers, but in general they will be functions of the parameter. So, if they are the function 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 probability of rejecting H naught when theta belongs to theta naught that is when it is true. So, we take super of alpha theta less than are equal to alpha, this is usually called the size of the test or 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 a accepting H naught when it is falls, for which this is minimized. Once, again see minimize means this is a function of the parameters so, minimized means over theta belonging to theta 1. We also say 1 minus beta theta that is called the probability of X belonging to r for 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 two error.

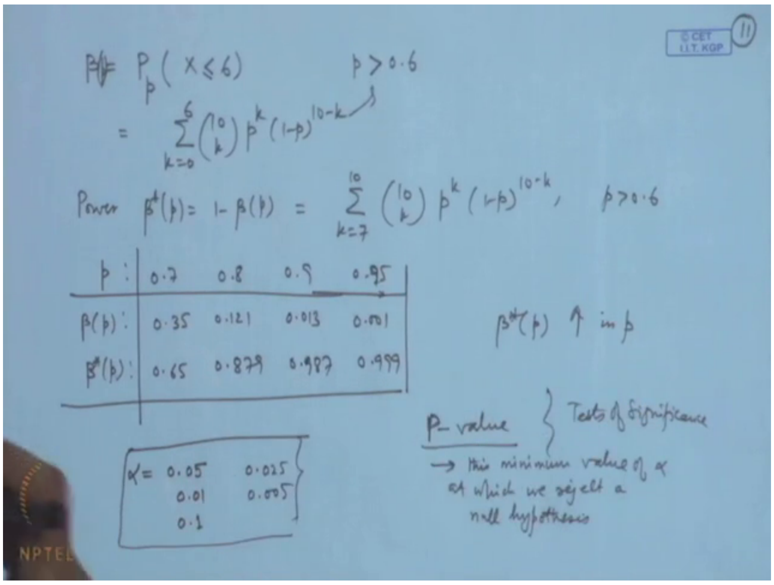
(Refer Slide Time: 47:18)



So, then we get the concept of most power full test and the uniformly most power full test, thus if theta 1 is a singleton set power is a number. So, we can use a notation say beta is star phi to denote the power of the test phi and we maximize beta is star phi with respect to phi this is called. So, this gives a test say phi naught, then phi naught is called the most power full test, that is m p test. If theta 1 is not single ton, then beta phi star theta has to be maximized over theta belonging to theta 1, this will give uniformly most power full test.

So, I mentioned the term and Neyman Pearson theory. So, the Neyman Pearson theory approaches the testing of hypotheses problem from this view point that is it solves an optimization problem and gives a solution. So, in the first case they give a simple hypotheses verses simple hypotheses case so, we are able to get the solution in a standard form. And then, those procedures are generalize to obtain the solution n cases where uniformly most powerful tests can be derived, let me give example of a power function. (No Audio From: 49:50 to 50:00)

(Refer Slide Time: 51:14)



Let us consider say X follows say binomial 10 p and our hypotheses 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 at reject H naught if X bar X is greater than or equal to 7, accept H naught if X is less than r equal to 6. So, here alpha that is the probability of type one error that is probability of rejecting the 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 a from the tables of the binomial distribution one can see this value trans out to be 0.382. So, the probability of type one error is very high using the stress procedure, let us look at beta that is the type two error so that is probability of accepting H naught when it is false now this means p is greater than 0.6 here.

This values 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 and e 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 is 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 this values for different values of p, you can see it from the tables of the binomial distribution at 0.7 0.8 0.9and 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 is 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 and the right direction that is the probability of. Type two error is gradually decreasing and the power function is gradually increasing here. The next point I mention about that, we are saying that fix the size of the test alpha and then, determinate test procedure for which the power function is maximized.

Now, when you say fix alpha then the that is a job of the statistician, that he has to fix alpha and for that, we are determine a best procedure. The question is what should be the value of alpha, if we look at the standard text books here, in the standard text books a value of alpha and even in the questions that they ask they will fixed values as 0.5 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 distribution that are used for testing. Now, the reason for taking this values is that in those earlier days, the tables of the distribution were 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 are taken as the like 0.05 0.01.

So, basically this means 5 percent this means 1 percent this means 10 percent and slowly this values became like conventional and very standardize that people generally use these as the have significance. But there is nothing sacrosanct about this 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 hypotheses, 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 little I mean after words first we will discuss the Neyman Pearson theory, but this is an alternative way of currying out the test of hypotheses here.

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