Probability and Statistics

Prof. Dr. Somesh Kumar

Department of Mathematics

Indian Institute of Technology, Kharagpur

Module No. # 01

Lecture No. # 33

Testing of Hypothesis-I

Today, we will take up another aspect of statistical inference. Earlier, I have told that, in a statistical inference, we deal with several kinds of inferences. One is that, when we do not know about the characteristic of a given population, then, we give an estimate for that. That means, we could tell that, this could be the value based on the sample. That is known as point estimation.

Another one is, where we tell, that what an interval with a certain confidence level. That means, we give an interval, such that the probability of inclusion of a given value is a certain value. So, we give a 101 minus alpha percent confidence interval, which has the interpretation that if we do sampling 100 times, then out of that 101 minus alpha percent times, this particular interval will include the true value of the parameter.

However, there is another aspect of statistical inference, where we are in the dark about the value of the parameter or about the distribution and we want to make a guess about that value. For example, we are considering whether the birth and birth rate of the males or females is equal. So, if we consider a birth of a child as Bernoullian trial, and we are considering p as the probability of a birth of a male child, then we want to test whether p is equal to half or p is not equal to half.

In the development of a new drug for a certain disease, suppose there is an existing disease. Suppose, there is a certain drug which is existing and it has a success rate of p, and now, we introduce a new drug in the market, and we want to check whether the effectiveness of that is more than that value p. So, in that case, what we have to do is that

we have to consider a sample, and based on that, we have to give a decision. Such problems are called problems of testing of statistical hypothesis. So, specifically speaking a statistical hypothesis is a statement or you can say assumption about the probability distribution of a random variable or a population.

(Refer Slide Time: 02:56)
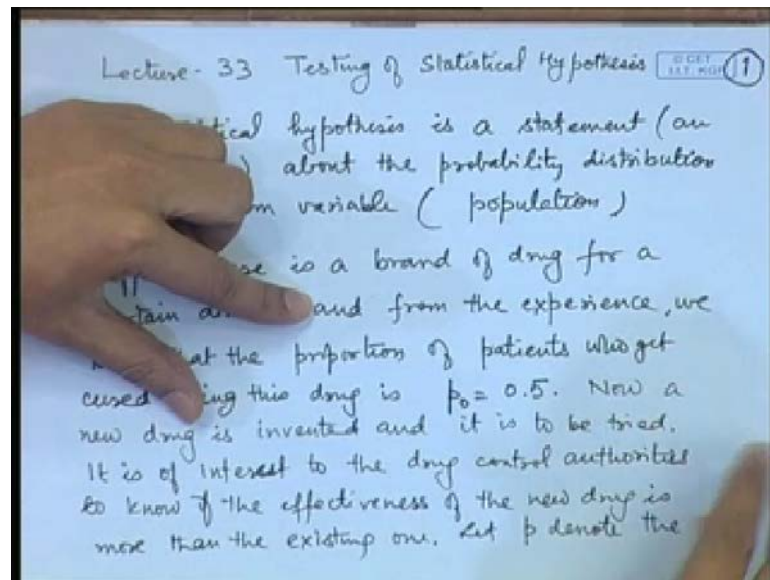
Consider, say, suppose there is a brand of drug for a certain disease, and from the experience, we know that the proportion of patients who get cured using this drug is, say p naught, that is equal to 0.5. That means, the patients who take this medicine, 50 percent of them get cured. Now, a new drug is invented and it is to be tried. It is of interest to the drug control authorities to know if the effectiveness of the new drug is more than the existing one.
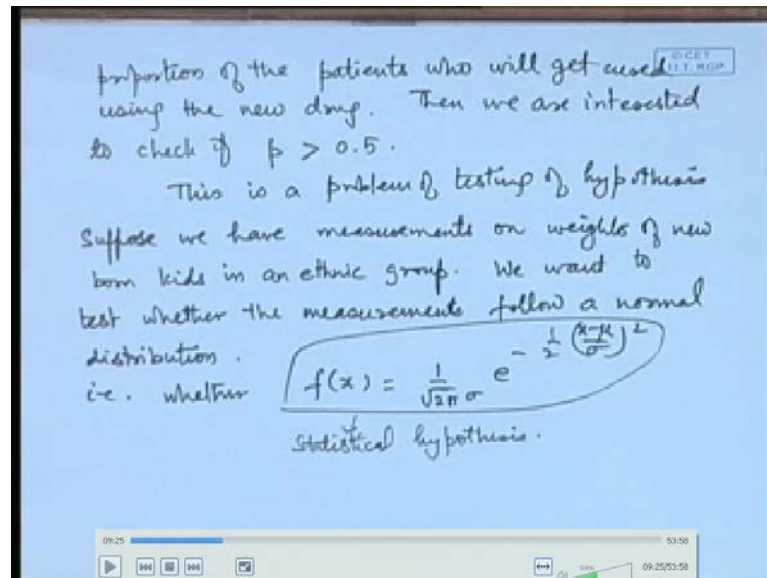
So, what we do? Let p denote the proportion of the patients who will get cured using the new drug. Then, we are interested to check if p is greater than 0.5, because if the new drug does not cure as much as the old drug, then there is no point in introducing the new drug in the market. Suppose, it is having less effectiveness, then there is no use of introducing it in the market. Another way could be, so checking that p is greater than 0.5. This is the problem of testing of hypothesis. So, this is a problem of testing of hypothesis.

Suppose we have, say measurements on weights of newborn kids in an ethnic group. So, we want to test whether the measurements follow a normal distribution. That means, we want to check whether, that is, whether fx is equal to say 1 by root 2 pi sigma e to the power minus 1 by 2x minus mu by sigma square. So, we want to check whether this is true. So, this is again a statistical hypothesis. So, now, how do we go about this problem? The fundamental problem of testing of hypothesis can be seen in the context of that we have to actually tell, whether the hypothesis which we are going to test in tenable or not. How do we do that? As in real problem of statistical inference, that is in point estimation or interval estimation, we will have a random sample with us.

Now, with the help of the random sample, we will like to device a rule to tell that whether this hypothesis is possible or not possible. So, for example, you consider the problem of assigning, whether the new drug is more effective or not. So, the drug trials

are made on patients. Suppose the trials are made on 1000 patients or 100 patients. Now, if we find, that out of the 100 patients, who took the medicine under certain controlled experiment, it turns out that only 25 got recovered using that new drug. Now, that obviously means that the effectiveness of the drug is only 0.25, and here, we want to test whether p is greater than 0.5. So, this hypothesis does not seem to be possible hypothesis.

On the other hand, if it turns out that the number of patients recovered using that medicine is say 70 percent, then there is no reason to disbelieve this proposition. We will feel that the new drug is more effective. So, the problem of testing of hypothesis is to device a procedure on the basis of the random sample to tell whether a given hypothesis is tenable or not.

Now, in the context of this, when we have to device a rule, we are not concerned only about the given hypothesis. What will happen is that, if this hypothesis is not tenable, we have to specify that what else is tenable. If I say, that p is greater than 0.5 is not tenable or not acceptable, then what else is acceptable, or what else is possible. So, we will have to say for example p is less than or equal to 0.05 or p is equal to 0.3 or p is equal to 0.25 etcetera. So, that gives the concept of that in a testing of hypothesis, we should have two hypotheses. If we reject one hypothesis, then it is in the favor of another one. So, that gives rise to the concept of null and alternative hypothesis.

(Refer Slide Time: 11:34)
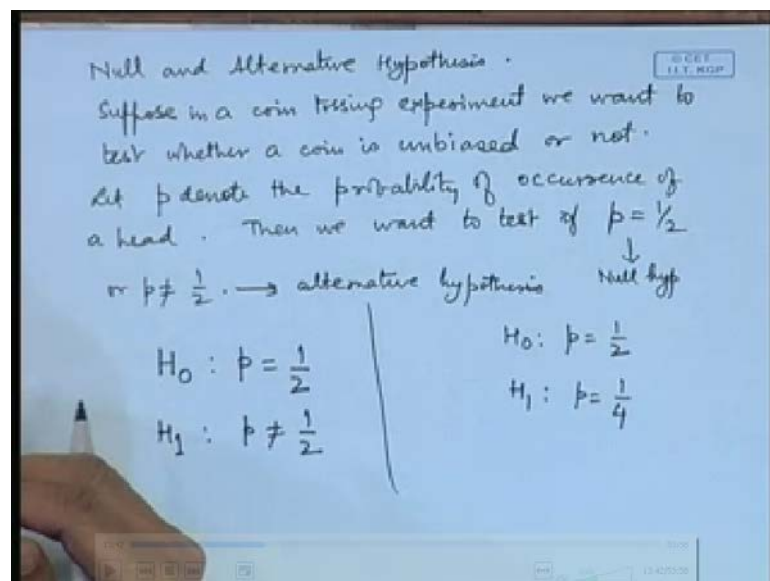
So, suppose in a coin tossing experiment, we want to test whether a coin is unbiased or not, so let p denote the probability of occurrence of a head. Then, we want to test if p is equal to half or p is not equal to half. So, then we can call this hypothesis as the null hypothesis. That means, if it is unbiased, then p is equal to half. If we reject this hypothesis, then we will say it is not unbiased. So, this will be called alternative hypothesis. So, we usually use a notation H naught for the null hypothesis. We say, H naught p is equal to half and against H 1 p is not equal to half. It could also be like H naught p is equal to 1 by 2 versus H 1 p is equal to say 1 by 4.
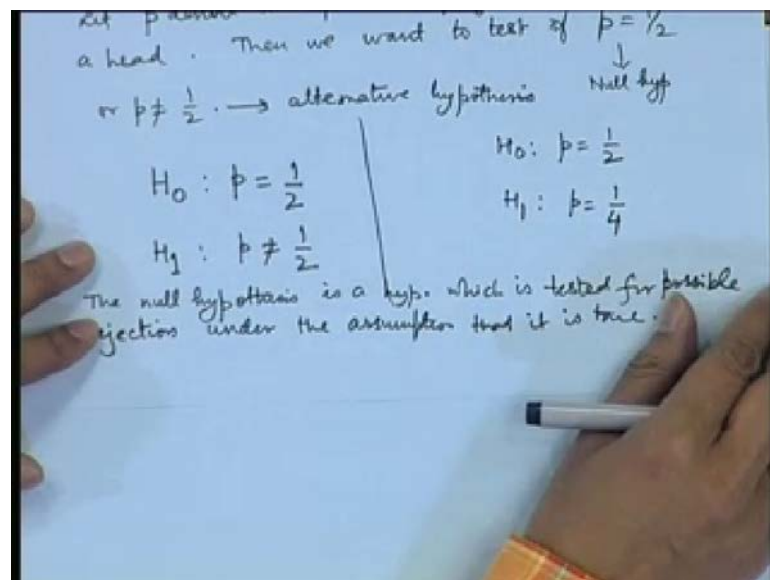
Suppose, we have a strong suspicion that actually the probability of head is only 1 by 4, then we may set up the alternative hypothesis of this form. So, in general, the hypothesis will be framed based on the questions that the experimenter will have and which he actually wants to test in the light of the sample that he is going to have. So, you can say that the null hypothesis is a hypothesis which is tested for possible rejection under the assumption that it is true.
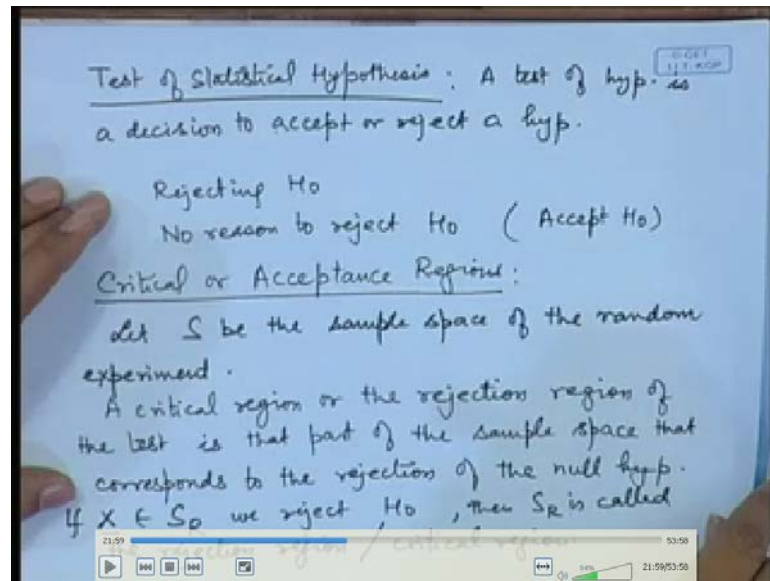
(Refer Slide Time: 14:50)
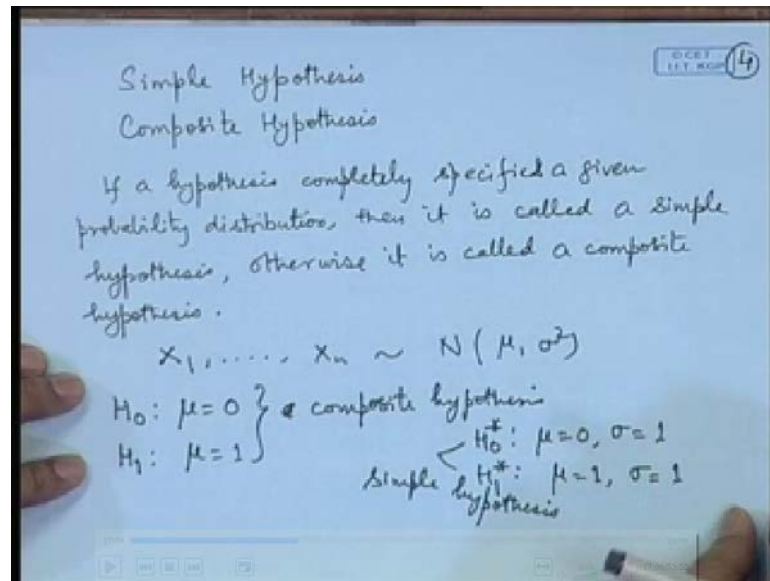
(Audio not available: 14:32-14:45)

So, we have two types of hypothesis. One is, simple hypothesis and another is, composite hypothesis.

If a hypothesis completely specifies a given probability distribution, then it is called a simple hypothesis. Otherwise, it is called a composite hypothesis. So, for example, suppose we know that the data x 1, x 2, x n is from a normal distribution with parameters mu and sigma square, where parameters mu and sigma square may be unknown. Then, if I have a hypothesis, say H naught mu is equal to say 0 and another hypothesis is, say mu is equal to 1, then this is specifying only the parameter mu here and sigma square is still unknown. So, these are composite hypothesis, but suppose I write H naught star mu naught is equal to 0 sigma is equal to 1 against H 1 star mu is equal to 1 and say, sigma is equal to 1, then these are simple hypothesis.

(Audio not available: 17:00-17:07)

So, next we talk about what is a test of statistical hypothesis. So, a test of hypothesis is a decision to accept. Well, decision to accept or reject a hypothesis. Now, here let me specify the practical aspect of it. Since, the decision is based on only a sample, so if the sample does not support the hypothesis, we say that there is no reason to accept the hypothesis or you can say the hypothesis is rejected on the basis of given sample.

As I mentioned earlier, when we are checking the effectiveness of the new drug and we want to test whether the new drug is more effective and suppose, out of 100 patients on which the new drug has been tried, only 25 patients get cured, then we have a very strong reason to reject the null hypothesis that p is greater than half because only one-fourth of the patients are getting cured. On the other hand, if p is equal to 0.7 from the sample, then there is no reason to reject the null hypothesis p is greater than 0.7.

So, since the decision is based on the random sample alone, therefore we do not speak very strongly in favor of accepting a hypothesis. Rather, we say, we reject the hypothesis or we have no reason to reject the hypothesis. So, we use the terminology rejecting H naught or no reason to reject H naught, which of course, in practical term means accepting H naught, but generally, we do not use this word here. So, now a test which we are telling is a decision procedure to accept or reject a hypothesis is based on the sample. So, how it is giving you the decision? Let us see that.

So, we have the concept of a critical or acceptance region. So, since the decision is based on the sample space because we are observing a sample and that sample takes value in a particular sample space. So, let S be the particular sample space of the random experiment. So, on the basis of this, we will give the acceptance region and the rejection region. So, a critical region or the rejection region of the test is that part of the sample space that corresponds to the rejection of the null hypothesis. So, we can use a notation, say S R. So, if our observation X belongs to S R, we reject H naught. Then, S R is called the rejection region or critical region. Obviously, the complementary region of this, that is S R complement, we can call it S A. That is the acceptance region, that is, if x belongs to S A, we do not reject H naught or we can say we accept H naught.

(Refer Slide Time: 22:06)
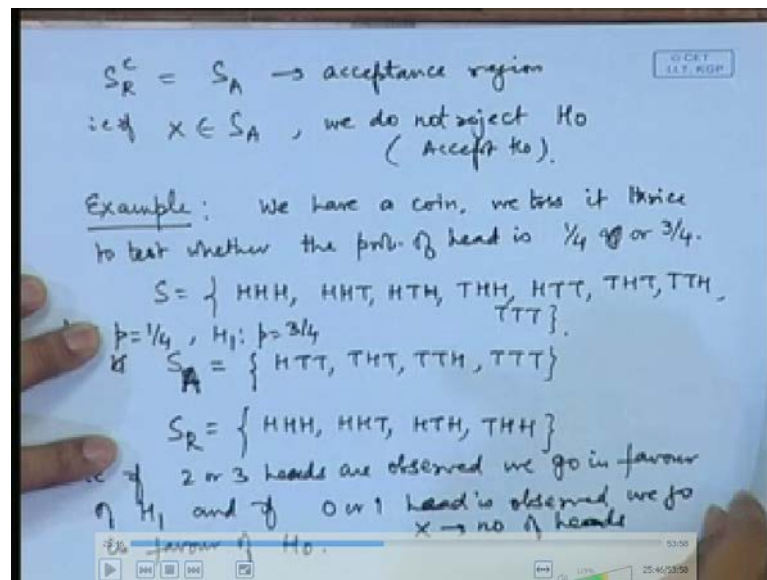
Let us give one example here. <mark>So, we want to test whether,</mark> so we have a coin and we toss it, say thrice to test whether the probability of head is, say 1 by 4 or 3 by 4. So, our sample space consists of all heads or two heads or one head or all tails. Now, we may take a decision like this. So, we can make a decision rule like this S R. So, we have to test the hypothesis here, H naught p is equal to 1 by 4 against H 1, p is equal to 3 by 4. So, p is equal to 1 by 4, means that head has a less probability.

So, if we consider one head or no head, that is, HTT, THT, TTH and TTT. That means more tails are appearing. Then, we consider that, sorry this means that head is having less probability, so H naught is true. So, this will be acceptance region and the

complementation of this could be HHH, HHT, HTH, THH, that is if 2 or 3 heads are observed, we go in favor of H 1 and if 0 or 1 head is observed, we go in favor of H naught.

(Refer Slide Time: 25:49)
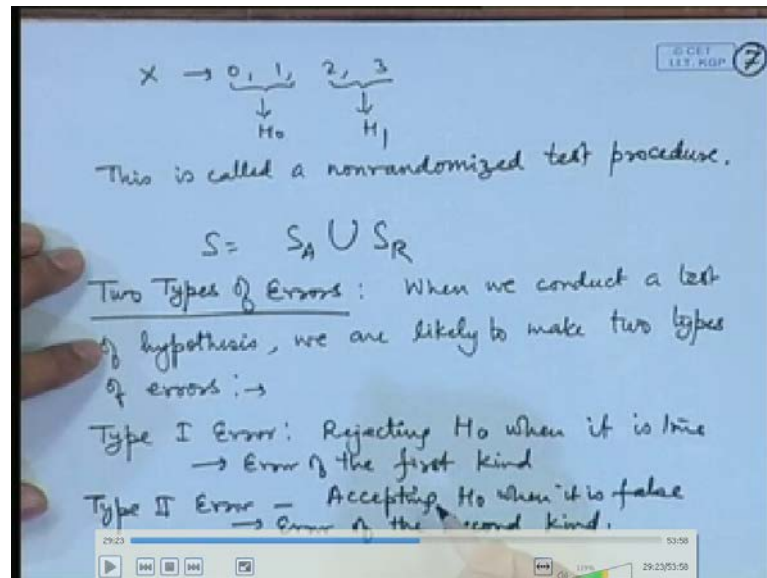
The test procedure seems to be quite simplistic in the nature. Suppose, I am saying X is the number of heads. So, the possible values of X can be 0, 1, 2, 3. We are associating the value 0 and 1 with H naught and 2 and 3 with H 1. So, this is called a non-randomized test procedure for this particular hypothesis testing problem.

So, you can easily see that, this test procedure is splitting the sample space into two portions. 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 complementary regions, two complementary exhaustive regions, such that, one region corresponds to the rejection of the null hypothesis and the other one, corresponds to the rejection of 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.

(Audio not available: 27:40-27:50)

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 patient is suffering from 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 conducts certain tests. For example, he may conduct certain blood test or certain other pathological tests may be conducted. So, when he is conducting those tests, those tests are based on sample.

For example, a blood test involves taking out a drop of blood from your finger or in other pathological test, for example, a urine sample may be taken or some other kind of a 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.
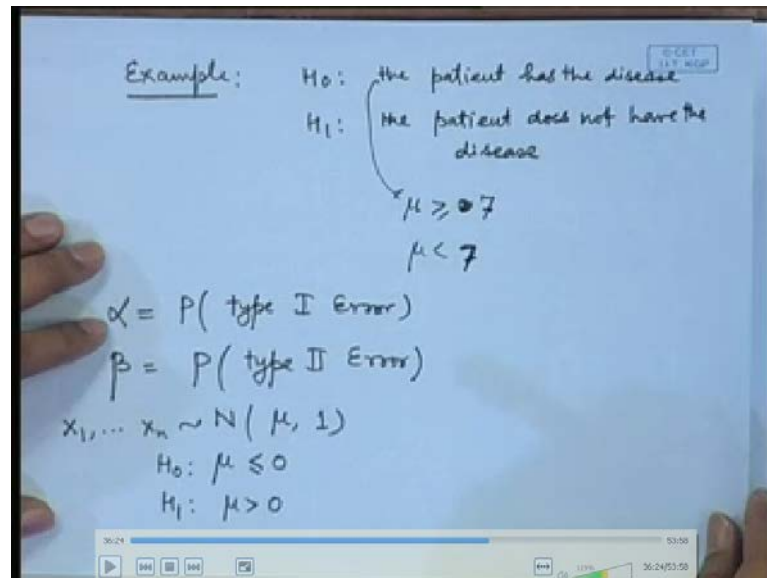
So, H naught may be the patient has the disease and H 1 is that the patient does not have the disease. I have written the hypothesis in the verbal terms, but a statistical hypothesis will mean that this is related to certain parameter. For example, it may be related to certain mean of the measurements. So, you may say that if mu is greater than or equal to say 7, then this may be count of some, say leucocytes or whatever or may be ESR rate. So, if mu is greater than or equal to 7, then we say that the patient has a disease. If mu is less than 7, we will say that the patient does not have a disease. So, this could, these are the likely scenarios that the doctor may be confronted with and the decision is based on the sample which he has taken from the patient.

Now, if based on the sample, the doctor concludes that the patient does not have a disease, whereas actually the patient has a disease, then the consequences can be fatal because if he concludes that he does not have a disease, so he will not give the appropriate medicine and maybe, he will treat with some other related symptoms. The disease may get aggravated and the patient may ultimately die.

So, the consequence of this Type I error that he is rejecting H naught, when actually it is true is disastrous here. Similarly, if you look at the Type II error, that means he accepts H naught when actually H naught is false. That means, that the patient does not have a disease and the doctor concludes that he has a disease. He may give some heavy dosage of medication which may lead to lot of complications and discomfort 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 disastrous than 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 still survive.

So, now the question comes that one has to reduce the possibilities of these errors. So, let me use a notation. We have a standard notation alpha is probability of the Type I error and say, beta is a probability of Type II error. So, for a statistician a good test is the one which keep this 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 the 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 region, when H naught is true and this is the size of the critical, 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 region are the complementary 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 region, but if you reduce the size of the rejection region, the size of the acceptance region will increase and consequently, the probability of Type II error may increase. Therefore, 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 1 population, where I have to take a decision on H naught on regarding, say mu is less than or equal to 0 or say mu is greater than 0. Now, here the hypothesis is composite.

(Refer Slide Time: 36:27)

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. 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 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 I 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 we 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 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: 39:05)
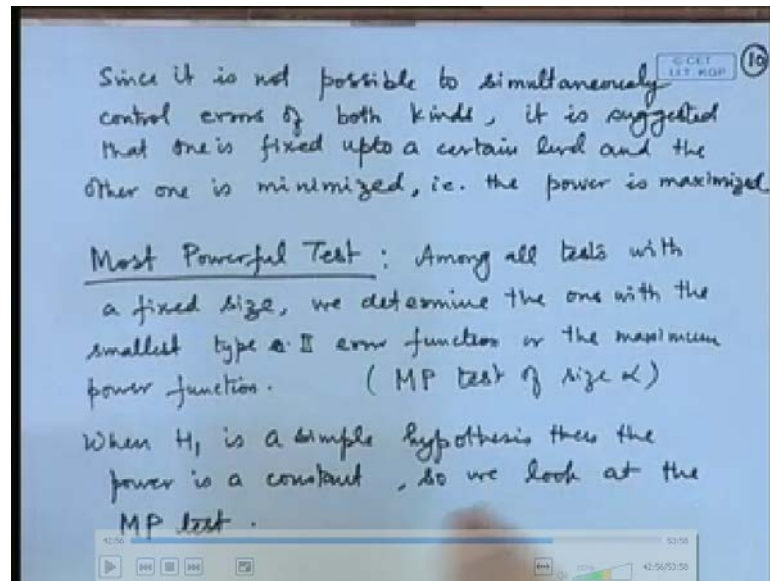
So, in the testing of hypothesis, the usual thing is, since it is not possible to simultaneously control errors of both kinds, it is suggested that one is fixed up to a certain level and the other one is minimized. That means, we should find a test procedure such that, among all the test procedures which have size, a given value, say alpha star, the power function of this test is having the 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.

(Audio not available: 40:36-40:50)

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 the most powerful test or MP test of size, say alpha. Now, if it is a composite hypothesis, then actually power function is dependent upon, when it is false, that means when H 1 is true. So, if H 1 is a simple hypothesis, 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 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: 43:22)
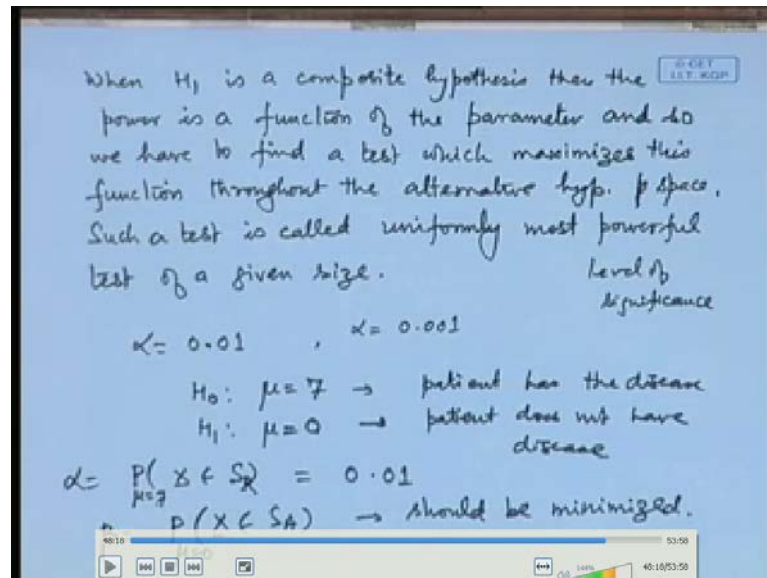
When H 1 is a composite hypothesis, then the power is a function of the parameter. So, we have to find a test which maximizes this function throughout the alternative hypothesis space. Such a test is called uniformly most powerful test of a given size.

So, now let us see the practical situation here. Let us go back to our problem of determining that whether the patient has a disease or not. So, in this case, as we saw that the Type I error is quite disastrous 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 the smallest.

So, 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 a disease and say mu is equal to, it is possible to have only two types of observations here. If mu is equal to 0, then the patient does not have disease. Suppose, the test is devised in such a way that the measurements that are taken only two possible values and when mu is equal to 7, 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 SR X belongs to SR, then mu is equal to 7. We may put this value to be, say 0.01 and now, we find that region, that is, S 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 a theory of finding out optimal test or you can say most powerful test here.

Now, one point about this quantity which I mentioned as the size of the test or it is also used as a terminology level of significance. Now, what is the rationale behind writing down that alpha is equal to 0.01 or alpha is equal to 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 play a 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. Therefore, the tables are calculated for specific parameter values only and therefore, those values were tabulated for values such as 0.01, 0.05, 0.025 etcetera. 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: 49:29)
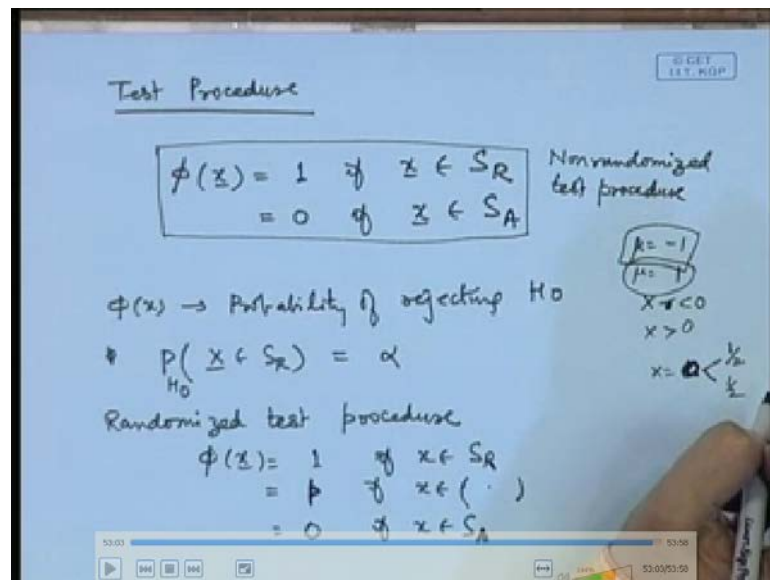
Now, here what we are doing is that, we are specifying a test procedure in such a way, so we may write it like this a test procedure description. What we are doing is, we are saying phi x is equal to 1, if x belongs to S R. It is equal to 0, if x belongs to S A. So, this means the interpretation that we can give phi x is the probability of rejecting H naught. So, 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, probabilities 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 you can say in accordance with the concept that I said that we either reject a null hypothesis or we do not reject the null hypothesis. So, this interpretation is based on that thing.

So, a non-randomized 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 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 maybe changing depending upon the points and when we include few points, it may become more than alpha. If we delete some points, it may become less than 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 that 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 maybe 0 in a 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 test, say a 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 certain level of significance or given size of test. Now, in the forth 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.