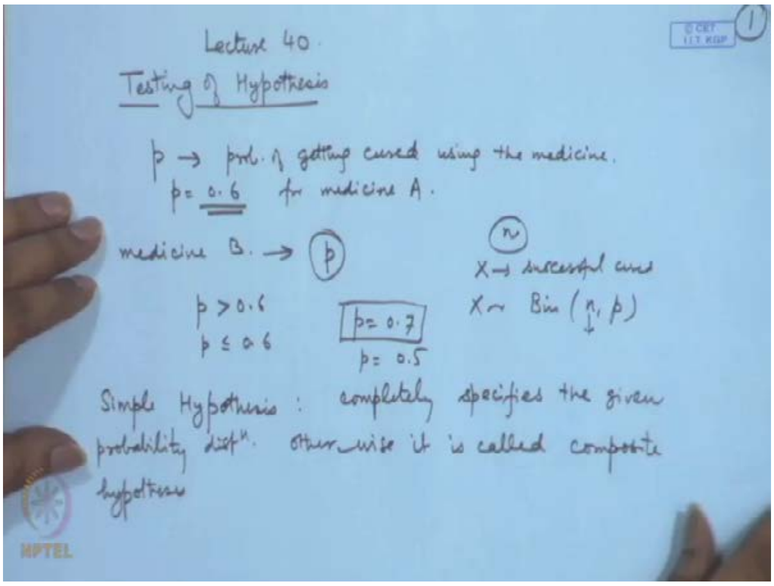
Advanced Engineering Mathematics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture No. # 41
Basic Concepts of Testing of Hypothesis

Today I will consider the third problem related to this statistical inference. Are here we have discussed, the problem of point estimation in which we specify a value for the unknown parameter of the population. Another one in the previous lecture I have introduced that is called the problem of interval estimation, where in place of giving a single value; we give an interval for the specified parameter. The other type of problem of inference occurs, when we want to test whether our parameter value satisfies certain condition. For example, if we are considering the cure rate using a certain medicine, then how many patients get cure? Suppose, there is a medicine which we have been using, and it it is known that approximately 60 percent of the patients get cure for that.

(Refer Slide Time: 01:22)

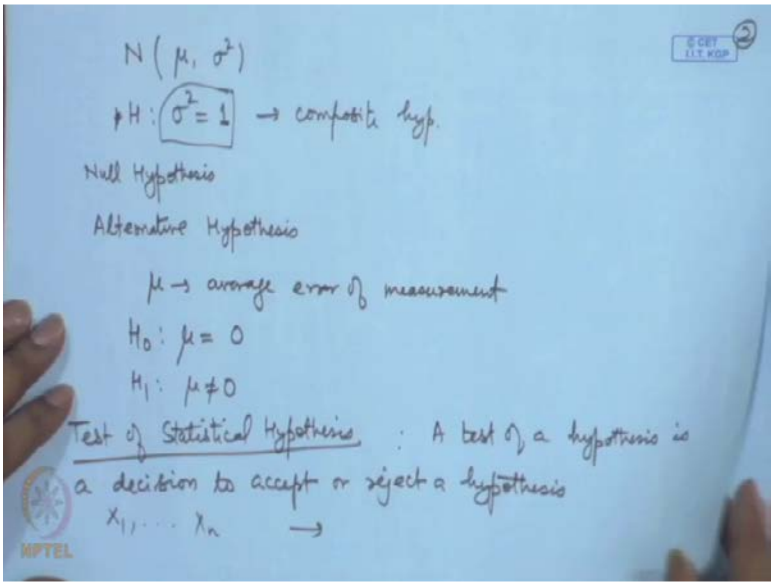


So, if we use the parameter p to denote the success rate or the probability of getting cure using the medicine, now a new medicine is introduced in the market. So, here it is given that p is equal to 0.6 for say medicine A. Now, a new medicine B is introduced, and the success rate with this is say p. Then we want to know, whether p is greater than 0.6 or p is less than or equal to 0.6. Suddenly this is a problem of interest to the under manufacturer, and the person who is going to market it. Now, here the problem of inferences based on certain assumption about the probability distribution, because if we consider say n patients to which the medicine has been given, and x is the number of successful cures successful cures.

Then we can say that x follows binomial n, p. And then we want to test the hypothesis, whether p is greater than 0.6 or p is less than or equal to 0.6 or say p is equal to 0.7 or p is 0.5, etcetera. We can make any a statement of this nature, this is called a hypothesis. Now, the they can be several I have written several a statements here, p is equal to 0.7, p is equal to 0.5, p is greater than 0.6, p is less than or equal to 0.6. We classify this hypothesis into different forms.

If we consider here n is known and p is unknown, and if we write a statement like p is equal to 0.7, then the probability distribution is completely specified. Whereas, if we write p is greater than 0.6, then the probability distribution is not completely specified; it only says p is some value greater than 0.6. So, this gives a simple hypothesis - in a simple hypothesis completely specifies the given probability distribution, otherwise it is called composite hypothesis. So, you consider that the testing of hypothesis problem can be classified as the problem to test as simple hypothesis or to test composite hypothesis. See, we may have a situation like this.

(Refer Slide Time: 04:49)



Suppose we are considering normal distribution with parameter mu, and sigma square. We want to test whether sigma square is equal to 1. Now, I if I say sigma square is equal to 1 this is not actually a simple hypothesis, because this is only specifying the value of sigma square; mu is not a specified here. So, this is actually a composite hypothesis; usually a notation H or H naught, etcetera is use to denote a hypothesis statement. For example, here we may write H naught, here we may write here we may write H 1, here we may write say H 2, here we may write say H 3, etcetera.

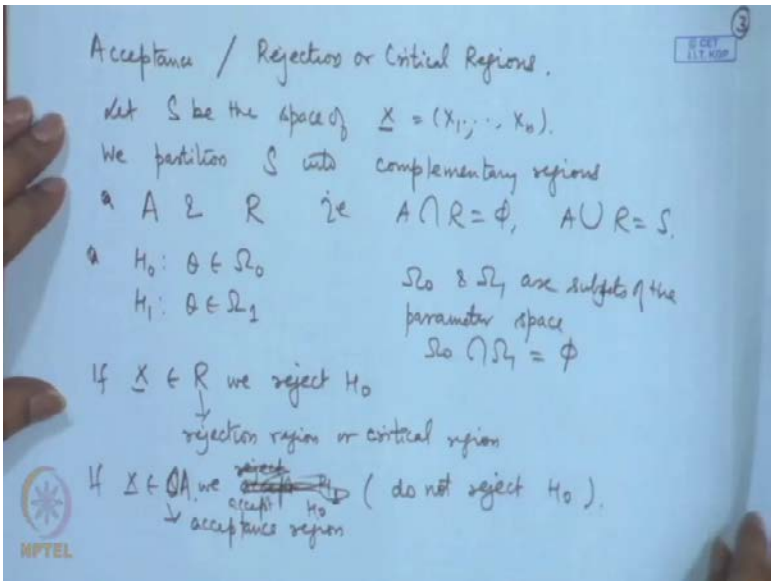
So, a hypothesis usually we use a notation H for denoting the hypothesis. Now, there is in another classification of the nature of the hypothesis, we say something is a null hypothesis, and there is an alternative hypothesis. Now, what is an null hypothesis? We frame this problem here that we want to test, whether the new medicine is more effective, then the previously used medicine. So, we have framed a hypothesis whether p is greater than 0.6, and then we say whether it is more effective R naught. So, R naught a statement that is corresponding to p is less than or equal to 0.6, we may also say here we may put greater than or equal to, and here we may put less. Of course, it does not matter here or we may might a put one hypothesis as p is equal to 0.7 or p is equal to 0.5.

So, we may call this as H naught, and this as H 1. Now, one initial hypothesis which we want to check, that is certain original assumption about the parameter of the population, that we call or we term it as a null hypothesis. Now, when the null hypothesis not likely to be true, in that case something else is going to be true; that we designate as an alternative hypothesis.

So, we have. So, we may have something like, we are considering certain we are measuring certain thing. Now, that measuring device may introduce errors in the measurements, and then we want to check whether the average errors are eliminated or not. So, if my mu is the average error of measurement, tell me may like to check whether this average is 0 or it is not 0; if it is 0, then we will say our measuring device performs consistently or it is performing well.

Now, when we introduce the problem of testing of hypothesis, ultimately we have to check whether the hypothesis is consistent with our given data or not. When we test with respect to our given data, it is called the test of hypothesis, test of statistical hypothesis. So, a test of a hypothesis is a decision to accept or reject a hypothesis. So, for example, we take a sample X 1, X 2, X n; based on that we make a decision basically what we will do? We will partition the entire sample space into two regions, when the value our sample point belongs to one portion of the sample space. We will say accept one hypothesis, if it is in the other case we will say accept the other hypothesis.

(Refer Slide Time: 09:33)



Based on that we frame what is called acceptance, and rejection or critical regions. Let S be the space of X is equal to X 1, X 2, X n; we partition S into complementary regions say A and R that is A intersection R is phi, and A union R is equal to S. Suppose, we are considering the hypothesis testing problem H naught theta belonging to say omega naught, and H 1 theta belonging to omega 1. Where omega naught, and omega 1 are subsets of the parameter space, and certainly omega naught intersection omega 1 must be equal to 5, they must be (( )).

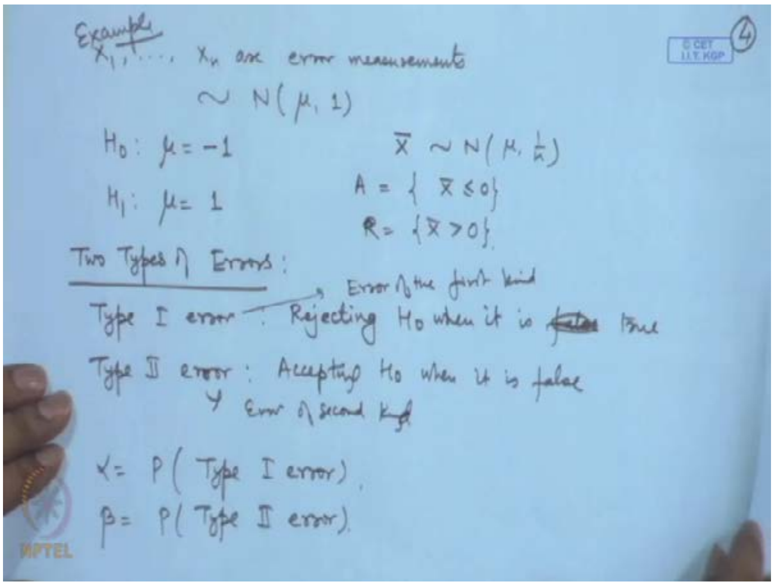
If say X belongs to R, we reject H naught. So, R is called the rejection region or critical region; if X belongs to S, we accept H 1 or we say do not reject H naught.

(No audio from 11:45 to 12:00)

This is called the sorry this is A, this is called the acceptance region. So, we actually I am written wrongly we accept H naught, then X belongs to A, we accept H naught or we say do not reject H naught. Here, I would like to mention one point, the decision to accept or reject hypothesis is based on the sample.

So, when we observe a sample based on which, if we can say our null hypothesis is not consistent with the value of the test statistics or value of the criteria that we are deciding; then we simply say that the H naught or the null hypothesis is rejected. However, when our hypothesis turns have to be consistent with the criteria for that is criteria, in that case we do not used our that we accepted is not where other say that there is no sufficient evidence to reject H naught. So, roughly speaking we say that H naught is accepted, but we do not say it that way, this is something like what we say in a testing of hypothesis problem.

(Refer Slide Time; 13:40)

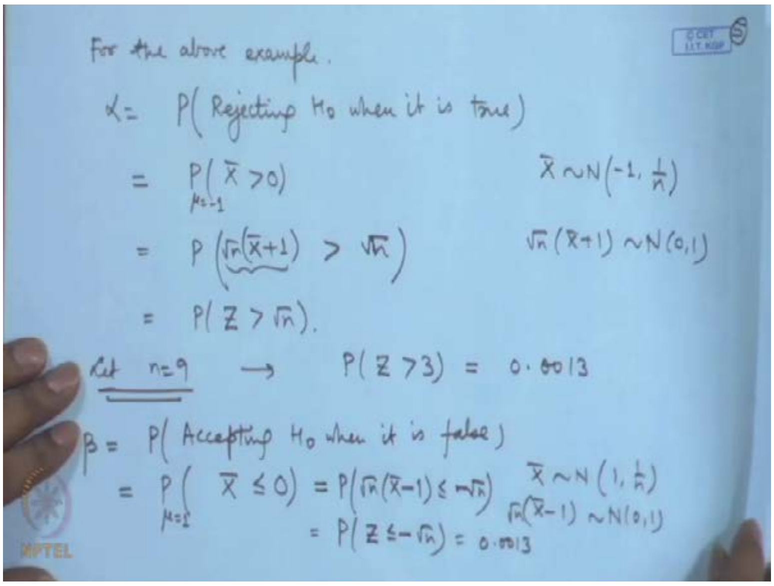


Let me introduce the one problem; let us consider that same where our X 1, X ,2 X n are error measurements, and we are assuming that they follow say normal mu, one distribution; here, I am assuming that sigma square is equal to 1. We want to test whether mu is equal to say minus 1 against mu is equal to plus 1; it is something like saying that whether the is (( )) are negatively waving or they are waving positively. Let us consider, here we can consider X bar - X bar follows normal mu, 1 by n; we can make a decision based on X bar.

So, we give the acceptance region as X bar less than or equal to 0, and the rejection region as X bar greater than 0. Naturally our decision is based on sample, and therefore there is a chance of errors. There are two types of errors that are committed: One is called type one error, that is rejecting H naught - when it is false, when it is true. And type two error accepting H naught, when it is false. So, this is called the error of the first kind, and this is called the error of the second kind.

Usually we use a notation alpha is equal to the probability of type one error, and beta is probability of the type two error.

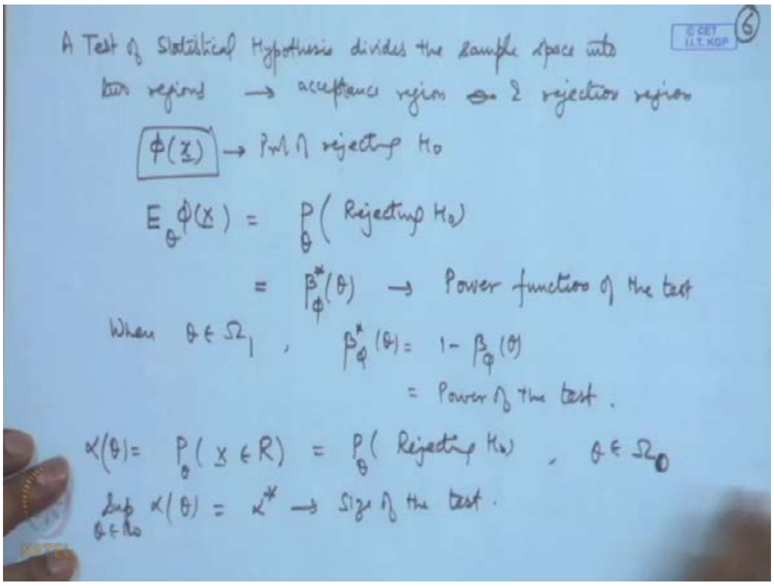
(Refer Slide Time: 16:33)



Let me calculate these probabilities for this example, for the above example alpha that is the probability of type one error, that is probability of rejecting H naught, when it is true. Now, when are we rejecting, we are rejecting when X bar is greater than 0. So, probability of X bar is greater than 0, when it is true - true means mu is equal to minus 1; now when mu is equal to minus 1 X bar follows normal minus 1, 1 by n. So, this value, then you can write as probability of X bar plus 1 into root n greater than 1 root n itself. So, this will have root n X bar plus 1, this will follow normal 0, 1. So, this value turns out to be probability of Z greater than root n, suppose n is equal to 9 here; then this value will turn out to be probability of Z greater than 3; that is 0.0013 in the case of a standard normal distribution.

So, the probability of type one error is 0.0013 here. Let us consider say beta here, that is the probability of type two 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 that is mu is equal to 1. So, now, when mu is equal to 1, X bar follows normal 1, 1 by n; that means, X bar minus 1 into root n will follow normal 0, 1. So, this is probability of root n x bar minus 1 less than or equal to minus root n, that is equal to probability of Z less than or equal to minus root n. Now, in this particular case when I have taken n to be 9; this is again 0.0013. So, in this case both probability of type one error, and probability of type two error; they are 0.0013 here.

(Refer Slide Time: 19:28)



So, a what is a test of statistical hypothesis? A test of statistical hypothesis divides the sample space into two regions, that is the acceptance region, and the rejection region

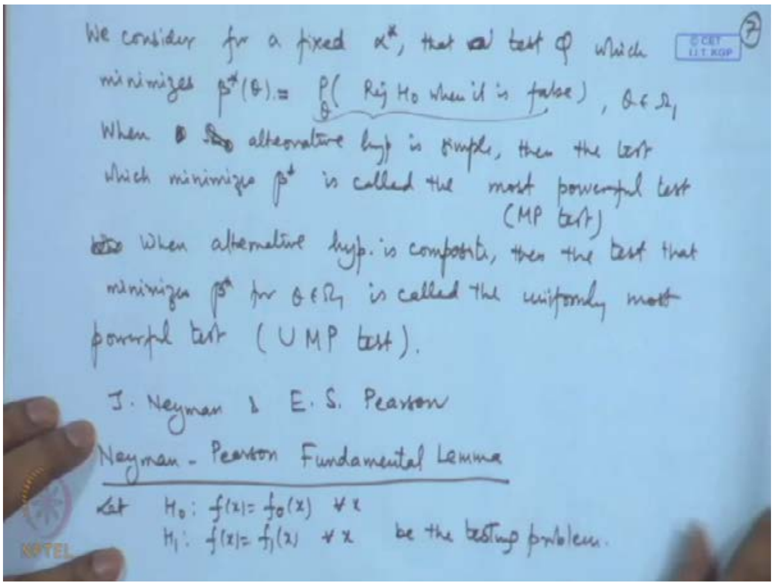
(No audio from 19:55 to 20:12)

So, based on this what we are doing? We are actually determining a function say phi X, this is called a test function, this we can call probability of rejecting H naught. So, we are actually assigning a function here and, we are calculating expected value of phi X that is equal to probability of rejecting H naught, when theta is a parameter value. Now remember here, when theta is belonging to the null hypothesis space, then this is the coming the probability of type one error, and when theta belongs to the alternative hypothesis set, then this is becoming 1 minus the probability of type two error.

So, we give it a new name, we call it beta star of theta; this is called the power function of the test. So, phi here… So, when theta belongs omega 1, then beta star phi theta is actually 1 minus beta phi theta, this is called the power of the test. Now, you can notice here, that the probabilities alpha and beta that have been calculated, they are based on the complementary regions; here X bar is greater than 0, and X bar is less than or equal to 0, so they are based on the complementary regions.

In an ideal test both alpha, and beta must be equal to 0, but the simultaneous minimization is not possible. If we decrease alpha say, then beta will increase; if we increase alpha, then beta will decrease. Therefore, as a compromise solution what we do 6, one of the errors and then we try to minimize the other one; this gives as the concept of the most powerful test. So, what we do, we fix let us called alpha theta, that is equal to probability of type one error, that is when X bar belongs to the rejection region; that is the probability of rejecting H naught, when theta belongs to the null hypothesis set. So, consider the maximum value of this; let us call it alpha star, this is called the size of the test.

(Refer Slide Time: 23:26)



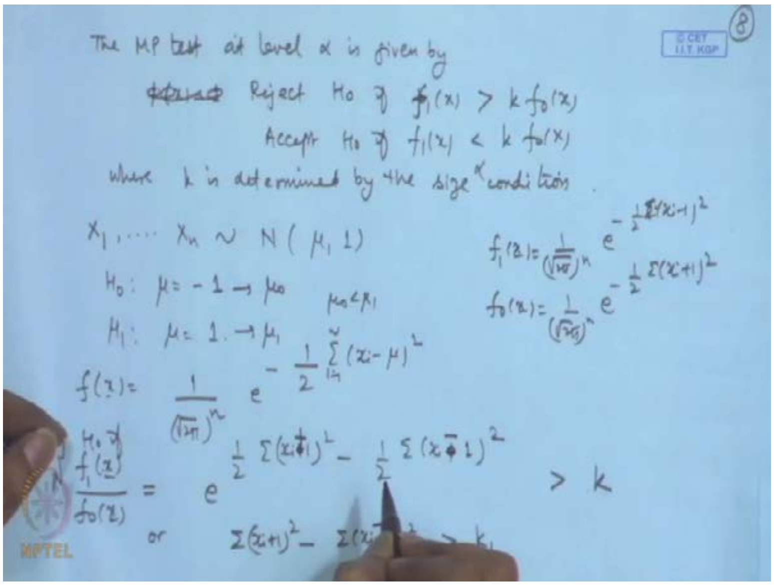
So, what we do, that we consider we consider for a fixed alpha star that value that test phi which minimizes beta star theta. So, when omega naught, see this beta star theta is actually probability of rejecting H naught, when it is false; when theta belongs to omega one here. So, this is called the power of the test. So, when alternative hypothesis is simple, then the test which minimizes beta star is called the most powerful test or MP test. Otherwise when when alternative hypothesis is composite, then the test that minimizes beta star for theta belonging to omega 1 is called the uniformly most powerful test, that is we called U M P test.

So, in a given problem our aim is to find out the most powerful test or the uniformly most powerful test. This problem of finding out the most powerful test has been solve by statistician, and a solution which is given for the simple versus simple case was introduced in 1920's by Jorgen Neyman, and E.S Pearson, and the result is famously called Neyman Pearson Fundamental Lemme. This result which was introduce for testing simple versus simple case was later on extended to the case of composite hypothesis also, and the solutions for the cases we had the uniformly most powerful test or the uniformly most powerful test within a certain class of test, we are also derived.

In this particular lecture, I will be only giving the Neyman Pearson Fundamental Lemme, and then based on this how various test have been derived. So, we will give the form of the test for testing the parameters of the normal population only, and we will not discuss the general methodology for finding out the test; there are other methodology also such as the likely hood ratio test etcetera. However, we will not be concentrating on those things in this particular lecture here. Let me introduce the name part of the Neyman Pearson Lemme.

So, let us consider say hypothesis H naught, say f(x) is equal to f naught x, against H 1 f(x) is equal to f 1(x). Of course, this means that the density, whether the density is f naught or the density is f 1. So, this is the testing problem.

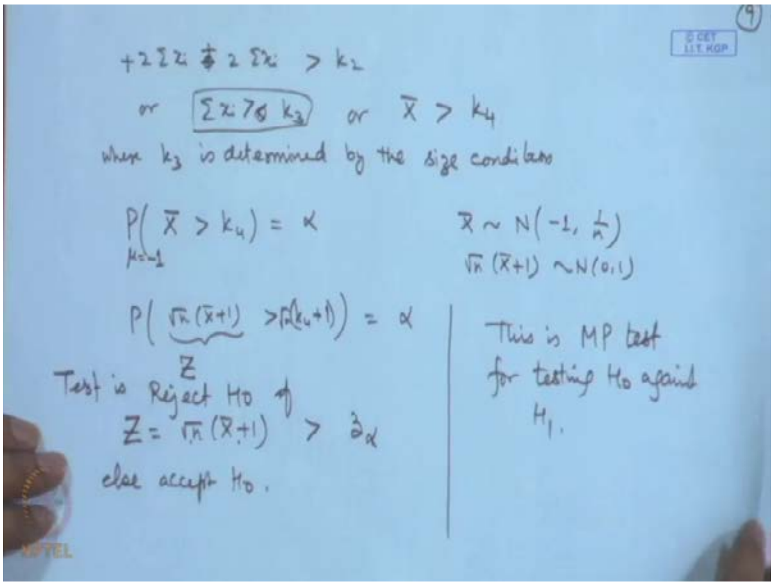
(Refer Slide Time: 28:10)



Then the Neyman Pearson Lemme says that the most powerful test at level alpha is given by phi x is equal to 1, other we can say reject H naught; if p 1(x) is sorry f 1(x) is greater than K times f naught x accept H naught; if f 1(x) is less than K times f naught x, where K is determinant by the size condition - size alpha condition. Let me solve one problem. Let us consider the normal example X 1, X 2, X n following normal mu, 1; and our hypothesis testing problem is whether mu is equal to minus 1, against whether mu is equal to plus 1. Let us write down the density function here, f(x) is equal to 1 by root 2 phi to the power n e to the power minus 1 by 2 sigma x i minus mu square i is equal to 1 to n.

So, if we consider f 1(x) by f naught x that is equal to e to the power half sigma x i plus 1 square minus half sigma x i minus… f 1 is corresponding to mu is equal to 1. So, this will be with the minus here, and this will be with the plus sign here. So, we are saying it is greater than K, now we can simplify this, if we take the logarithm here; this will imply sigma x i minus 1 square minus sigma x i plus 1 square greater than some constant say K 1.

(Refer Slide Time: 30:57)



Now this can be further simplified, we can write it as minus twice sigma x i plus twice again minus twice sigma x i greater than some k 2 or we consider sigma x i is less than k 3. Let us check the a steps here, f(x) is equal to 1 by root 2 phi to the power n e to the power minus 1 by 2 sigma x i minus mu square. So, if we write f 1(x) here, f 1 will become equal to 1 by root 2 phi to the power n e to the power minus 1 by 2 sigma x i minus 1 square, and f naught will become equal to 1 by root 2 phi to the power n e to the power minus 1 by 2 sigma x i plus 1 square.

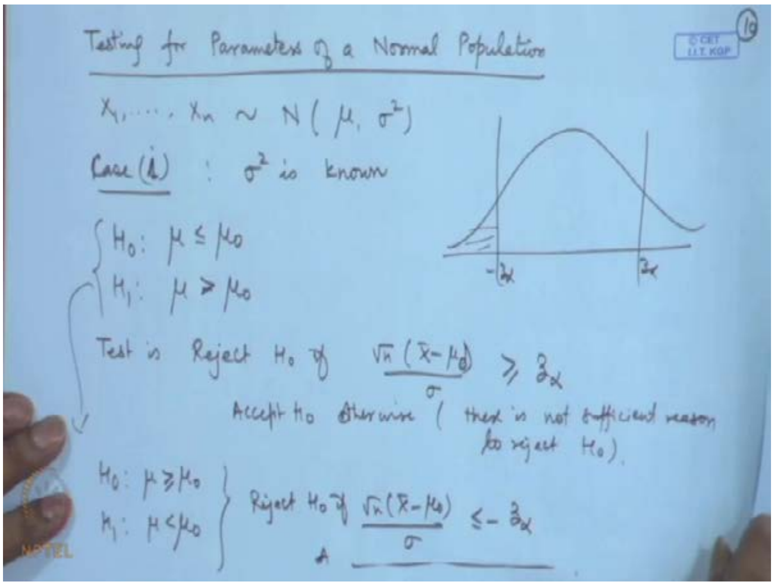
So, if we consider f 1 by f naught here, then this will become plus here, and this will become minus here; this is plus, this is minus here. So, here we will get plus, and this will become become plus. So, this is greater than k 3, where k 3 is determinant by the size condition. So, here we we make use of the fact that probability of… So, we can write it as X bar greater than say k 4. So, X bar greater than k 4, this should be equal to alpha, when mu is equal to minus 1 that is the probability of rejecting H naught; this is the rejection region. We are saying reject H naught, if if this statement is true.

Now, again X bar follows normal minus 1, 1 by n. So, we will get root n X bar minus plus 1 following normal 0, 1. So, this is statement can then be written as root n X bar plus 1 greater than k 4 plus 1 into root n, that is equal to alpha. So, this is nothing but Z. So, we can write Z is equal to root n X bar plus 1 greater than z alpha. So, the test is reject H naught, if root n X bar plus 1 is greater than Z alpha else accept H naught. So, this is nothing but the most powerful test for testing H naught, against H 1. Now note here, in the null hypothesis we are saying mu is equal to minus 1, and in the alternative hypothesis mu is equal to 1.

So, actually roughly speaking you can say that in the alternative hypothesis, the average value is higher than the value in the null hypothesis. And therefore, the rejection region also reflects that, we are rejecting actually for higher value of X bar, because X bar is actually an estimated for mu here. It is actually a consistent, and (( )) estimated; and here it turns out that X bar is greater than we reject H naught. Likewise in case of in place of here minus 1, and mu 1 if I had written here say mu naught, and here I had written mu 1; where mu naught is less than mu 1, that form of the test is statistics would have been the same. Except for the fact that here one would have been replaced by X bar minus mu naught.

So in fact, this observation let to the solution of the composite hypothesis also; that means, in place of say mu is equal to mu naught against mu is equal to mu 1, if we consider say mu less than or equal to mu naught against mu greater than mu 1, etcetera. Then the uniformly most powerful tests are found out.

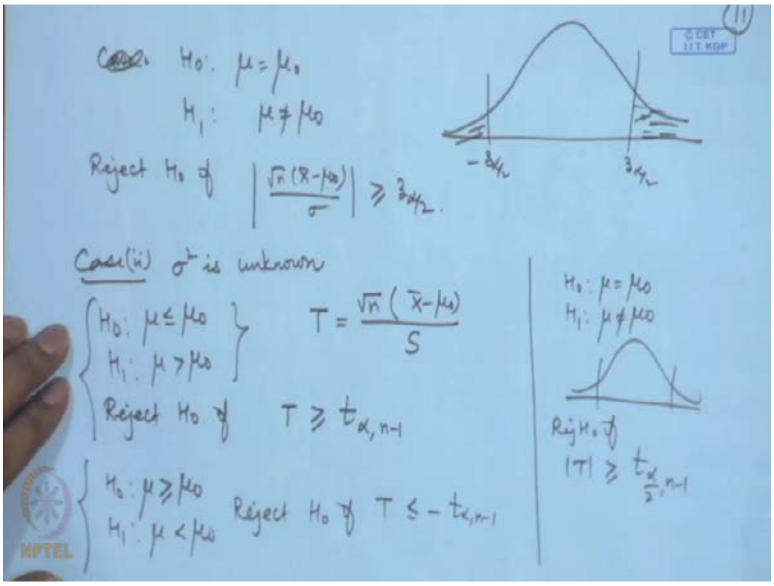
(Refer Slide Time: 35:32)



And let me introduce those testing problems here testing for parameters of a normal population. So, X 1, X 2, X n follows normal mu sigma square distribution here. We are considering say case one when say sigma square is known, we are considering say hypothesis say mu is less than or equal to mu naught, against mu is greater than mu naught. Here the test is then reject H naught, if the root n X bar minus mu by sigma mu naught by sigma is greater than or equal to Z alpha; the alternative region will become the acceptance region; accept H naught otherwise or you can say that there is no region, not sufficient region to reject H naught.

See, this case could have been that in the null hypothesis we have greater than or equal to this, and in the alternative hypothesis we have the… When if you look at the curve here of the normal distribution, then this is z alpha; here you will have minus z alpha. So, if it is (( )) side, the smaller side then the test will get reversed. You will say reject H naught, if root n X bar minus mu naught by sigma is less than or equal to minus z alpha, and the other region is complementary region will be the acceptance region.

(Refer Slide Time: 37:58)

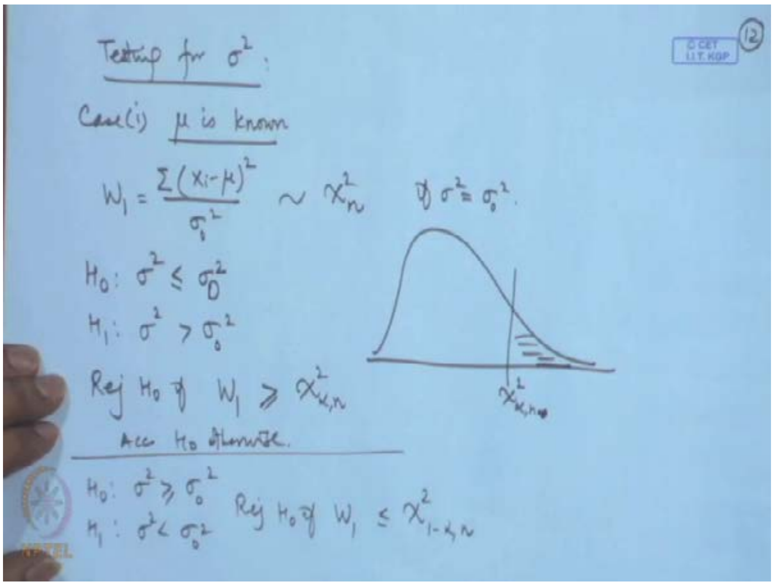


Now, we may also have the case when the alternative hypothesis is on both the sides; for example, you may discuss mu is equal to mu naught or say mu is not equal to mu naught. In that case on both the regions, we will be considering rejection. So, if we consider the reason of the alpha into two portion, then this will come alpha by 2, and minus alpha by 2 here and alpha by 2 here. So, we will then consider reject H naught, if modules of root n X bar minus mu naught by sigma is greater than or equal to z alpha by 2.

Let us also consider the case, when sigma square is unknown when sigma square is unknown; here our decision making will be dependent upon square root n X bar minus mu naught by S, let us call T. Then we will have the test statistics as reject H naught, if T is greater than or equal to T alpha n minus 1. On the other hand, if we are considering say H naught mu is greater than or equal to mu naught, against H 1 mu less than mu naught; in that case we will say reject H naught, if T is less than or equal to minus t alpha n minus 1. However, if I have H naught mu is equal to mu naught, against H 1 mu naught equal to mu naught, then as in the normal distribution T distribution is also symmetric distribution. So, you will have reject H naught, if modules of T greater than or equal to T alpha by 2 n minus 1.

(No audio from 40:18 to 40:44)

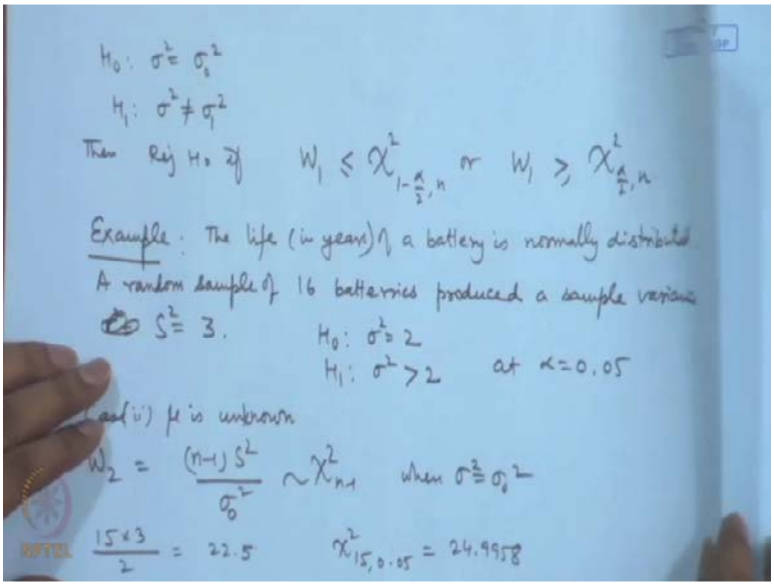
(Refer Slide Time: 40:52)



Let us also consider the test for sigma square; testing for sigma square here, and again we may have two cases: mu is known, if mu is known then we can consider sigma x i minus mu square. And we can consider this divided by sigma naught square, let us call it say W 1, then this follows chi square on n degrees of freedom when sigma square is equal to sigma naught square. So, if we consider the hypothesis says sigma square is less than or equal to sigma 1 square sigma naught square, against H 1 sigma square greater sigma naught square, then our rejection region will be… If we consider the chi square curve here, then on the larger side chi square alpha n minus 1.

So, here n reject H naught, if W 1 is greater than or equal to chi square alpha on n degrees of freedom, accept H naught otherwise. Similarly, we can consider the complementary region, and here we will say reject H naught if W 1 is less than or equal to chi square 1 minus alpha n.

(Refer Slide Time: 42:43)



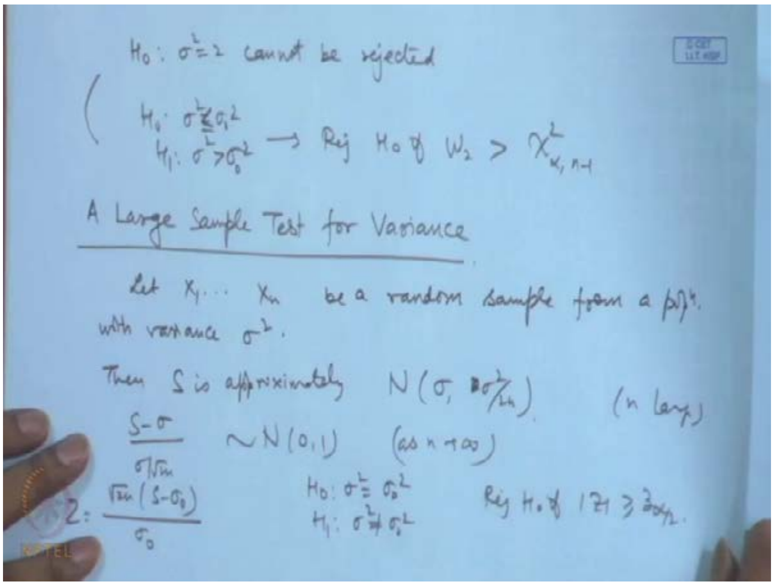
In the case of two sided region, suppose we are considering sigma square is equal to sigma naught square, against sigma square is not equal to sigma 1 square. Then reject H naught, if W 1 is less than or equal to chi square 1 minus alpha by 2 n or W 1 is greater than or equal to chi square alpha by 2 n.

(No audio from 43:05 to 43:38)

The life in years of a battery is normally distributed. A random sample of say 16 batteries produced a sample variance say sigma square is equal S square is equal to 3. We want to test the hypothesis sigma square is equal to 2, against say H 1 sigma square is greater than 2 at alpha is equal to 0.05 level of significance. Here, we apply the method here described that we calculate W, and that is sigma here.

Mu is not known here. So, we have to consider the procedure for mu unknown, let me give the procedure for mu unknown. Case two mu is unknown, when mu is unknown then we consider the quantity W 2 is equal to n minus 1 S square by sigma square, there follows chi square n minus 1. So, when sigma square is equal to sigma naught square. So, the test will be based on this; that is rejection regions, and acceptance regions. So, using this we calculate 15 into 3 by 2 that is equal to 22.5, and the value of chi square on 15 degrees of freedom and 0.05 is equal to 24.9958.

(Refer Slide Time; 46:01)



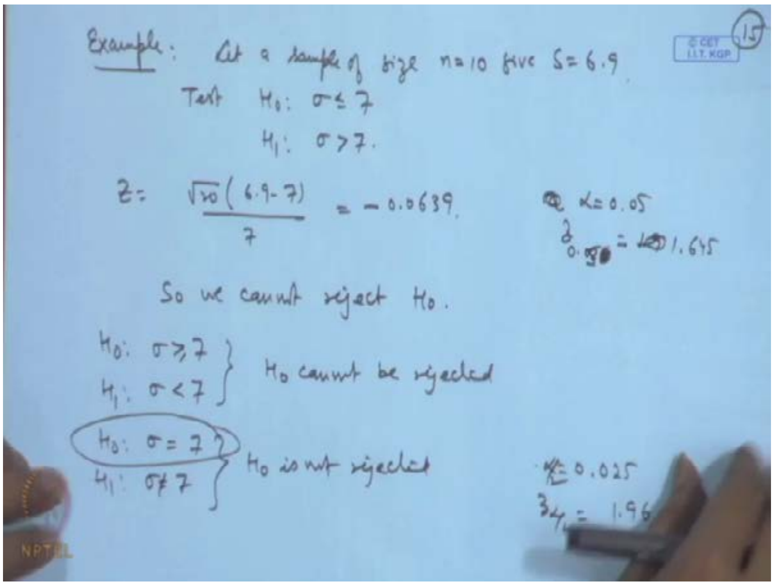
So, here we say that H naught sigma square 2 cannot be rejected, because here the general rejection region for H naught sigma square is equal to sigma naught square, against H 1 sigma square is greater than sigma naught square. This rejection region would have been reject H naught, if W 2 is greater than chi square alpha n minus 1. So, here we cannot reject the null hypothesis.

(No audio from 46:34 to 46:49)

We also we will large sample test for variance. So, let X 1, X 2, X n be a random sample from a population with variance say sigma square, then S is the approximately normal sigma 2 sigma square by 2 n, for n large. So, S minus sigma by sigma by root 2 n this is approximately normal 0, 1 as n tends to infinity.

So, we can consider here, the test based on root 2 n S minus sigma naught by sigma naught S, let me call z, then if I consider the hypothesis sigma square is equal to sigma naught square, against sigma square naught equal to sigma naught square. Then, we can consider reject H naught, if modules z is greater than or equal to z alpha by 2.

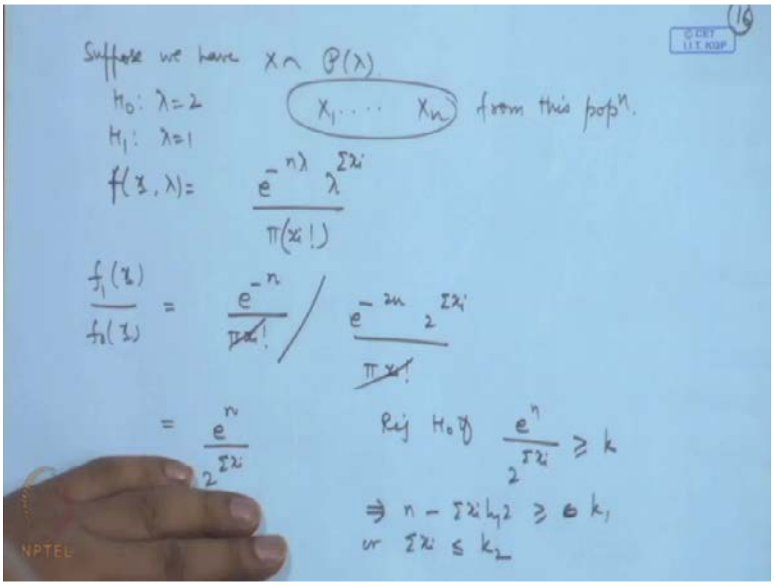
(Refer Slide Time: 48:20)



Let me consider one example for this problem: Let a sample of size n is equal to 10 give S is equal to 6.9, we want to test the hypothesis whether sigma is less than or equal to 7 or sigma is greater than 7. So, z equal to root 20 6.9 minus 7 by 7 that is minus 0.0639; if we consider say alpha is equal to say 0.05, then z of 0.025 is equal to 1 point… z is equal to 0.05, that is equal to 1.645. So, we cannot reject H naught. If we consider say yours of the say H naught sigma greater than or equal to 7, against H 1 sigma less than 7; in that case, again H naught cannot be rejected. See, if we consider H naught sigma is equal to 7, against H 1 sigma is equal not equal to 7; here you will consider two sided alpha by 2. So, z alpha by 2 is equal to 1.96. So, again H naught is not rejected. So, actually we can say that sigma is approximately 7 here.

(No audio from 50:12 to 50:29)

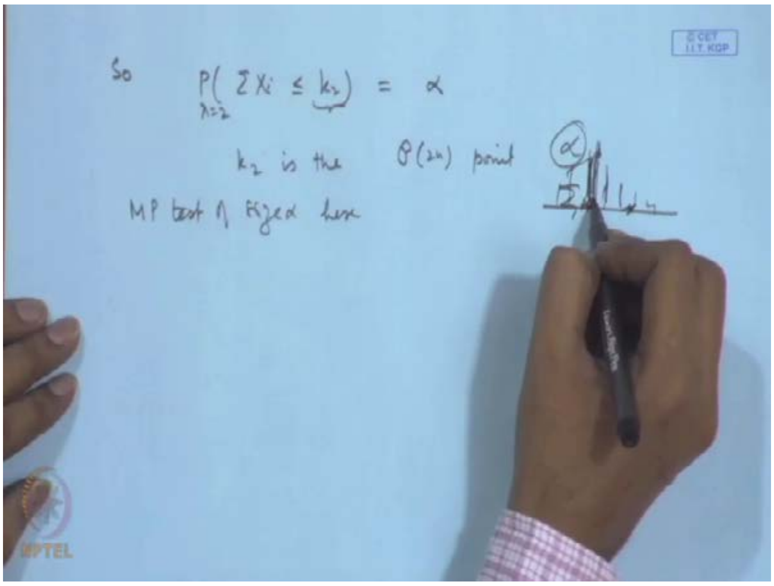
(Refer Slide Time: 50:49)



Let me give 1 or 2 more example here. Suppose, we have X following Possion lambda distribution, we want to test whether lambda is equal to 2, against lambda is equal to 1. And we have a random sample X 1, X 2, X n from this population. So, we consider here, the density function f(x) lambda that is equal to e to the power minus n lambda lambda to the power sigma x i by product x i factorial. If we consider f one by f naught that is equal to e to the power minus n by product x i factorial divided by e to the power minus 2 n 2 to the power sigma x i divide by product of x i factorial. Now, this term cancels out, your left with e to the power n, and and divided by 2 to the power sigma x i.

So, rejection region is if e to the power n divided by 2 to the power sigma x i is greater than or equal to k. So, here you can write it as n minus sigma x i log 2 greater than or equal to log k let; me write it as k 1 or sigma x i is less than or equal to some k 2. Once again, here you notice that in the alternative hypothesis the value of lambda is smaller than the value in the null hypothesis. And therefore, the rejection region is for the lower value. here you notice here that sigma x i follows Poisson, and lambda that is 2 n when H naught is true.

(Refer Slide Time: 53:02)



So, probability of sigma x i less than or equal to k 2, when lambda is equal to 2 that is equal to alpha; that means, k 2 is the Poisson 2 n point; that is the point on the Poisson 2 n distribution, that is the probability up to this is equal to alpha, up to this the probability some up to alpha here.

So, this will be the most powerful test of size alpha here. Now, here you may notice in Poisson is a discrete distribution; there may be a case that up to a certain point the value alpha will be achieved, and after that the value will exceed alpha. So that means, at one point the value will be below alpha, and after that the point value will be above alpha.

So, at that point we do the randomization, and we I locate that point with certain probability say gamma, here you reject and 1 minus gamma you accept. You choose the gamma in such a way that the total actually becomes equal to alpha here. In the next lecture, I will consider the two sample problems for the normal distributions, and we will consider testing about the equality of means, the equality of variances, etcetera. And we will consider certain examples based on that, that will be the part of the next lecture.