Statistical Inference

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

Lecture No. # 22

Neyman Pearson Fundamental Lemma

In the previous lecture, I have introduced certain basic concepts about the testing of a statistical hypothesis. It included the specification of the hypothesis, which we called as null hypothesis, alternative hypothesis, and classification of the hypothesis such as simple hypothesis or a composite hypothesis. What is a non and minced test procedure that is based on the sample, we take a decision to accept or reject the null hypothesis. I also cautioned that by accepting or rejecting a hypothesis based on a sample does not mean, and as action about the truthfulness or correctness of the hypothesis.

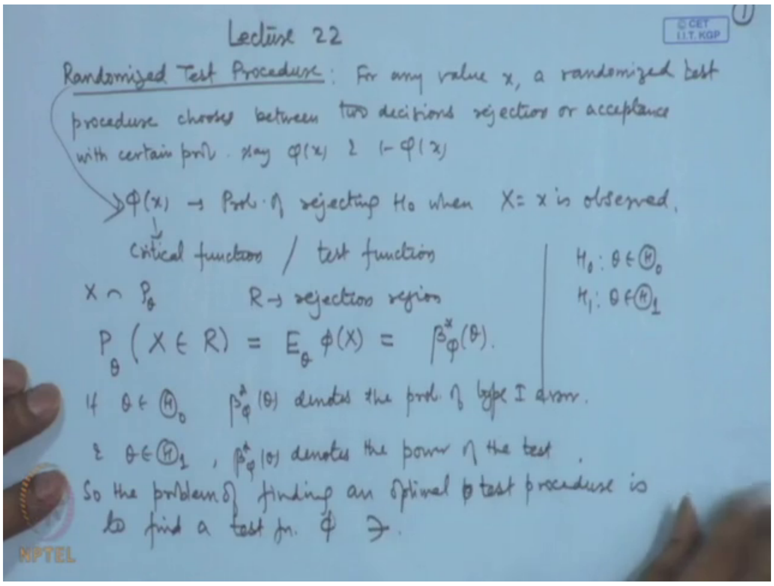
It simply means that our sample supports a hypothesis or does not support the hypothesis. So, the use of the testing procedure should be done with caution, they are not absolute truth. Now the question is how to derive a good test procedure, I mentioned that there are possibilities of the error, and we can actually cross classify broadly them under two categories they are called type one error, that is the probability of the probability of rejecting the null hypothesis when actually it is true, and beta we actually called the probability of type two error that is the probability of accepting h naught when it is false.

We have seen that the consequences of the two types of errors can be quite different and it could be quite disasters also. And therefore, any reasonable test procedure must control the two types of errors. And naturally the ideal situation should be that both has find beta are actually 0 or you can say both are to their minimum level, but there is a problem in this approach we cannot actually do this, that is we cannot simultaneously minimize alpha and beta.

Therefore, a practical solution is thought that if we know we can frame the hypothesis testing problem in such a way, that the type one error is taken to be in a more serious way therefore, we fix an upper bound to that. For example, suppose it is a medical problem; that means the false falsely claiming that disease is not there it is a very serious issue. So, probability of this we can fix a 1 in 100 something like 0.1 percent. 1 percent or 0.1 percent 1 in 1000 says. In that case with this we try to find out that test procedure for which beta is actually the the minimum or we have introduce a new concept called power that is 1 minus beta. So, power should be maximum.

Now, a test when you assign the rejection region then the probability of the rejection region under the null hypothesis we are saying it should be equal to some number alpha or it should be less than or equal to a number alpha. Now it may happen in particular when we are dealing with the discrete distributions, that and we may consider it as a single number, in that case it may happen that up to a certain level alpha with the value of the probability of type one error is below alpha and after some stage it becomes greater than alpha; that means, equal to alpha is not achieved. To overcome the situation we can define slightly more general form of the test procedures called randomized test procedures.

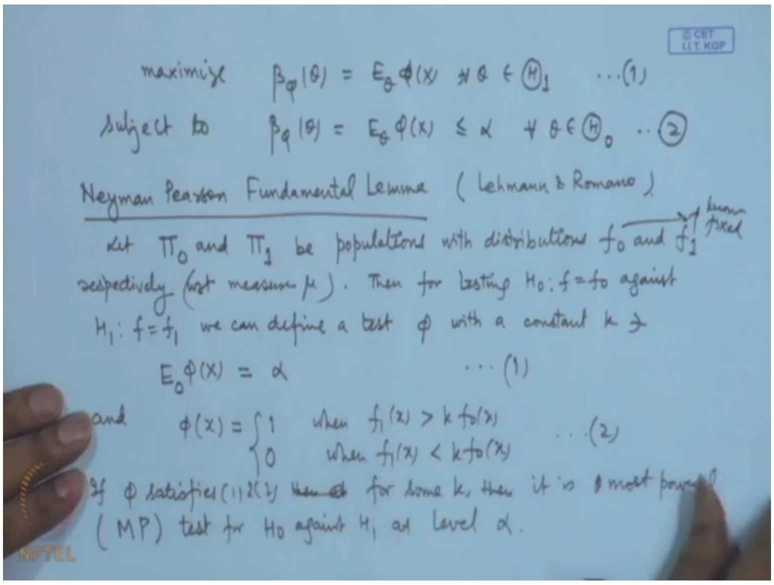
(Refer Slide Time: 04:12)



So, for any value x, a randomized procedure chooses between two decisions that is rejection r acceptance with certain probabilities say phi x and 1 minus phi x. So, we are actually saying phi x is the probability of a rejecting H naught, when x is equal to x is observed; this is called a randomized test procedure. So, this is also called critical function or a test function

So, let us say x follows P theta and say R is the rejection region. So, probability of X belonging to R probability of rejecting h naught when theta is the true value it is actually expectation of phi x, which we use a notation say beta is star phi theta, when theta belongs to theta naught, our general hypothesis framework let me again specify theta belongs to theta naught h 1 theta belongs to theta 1. So, if theta belongs to theta naught this beta is star phi theta actually denotes the probability of type one error and for theta belonging to theta 1 then beta is star phi theta denotes the power of the test.

(Refer Slide Time: 07:44)



So, the problem of finding an optimal test procedure is it can be stated as to find a test function phi. Such that maximize beta phi theta for theta belonging to theta 1, subject to the condition that beta phi theta is equal to expectation theta phi x is less than or equal to alpha for theta belonging to theta naught this is called the size condition. And this is maximization of the power. When theta one is a singleton one, this will give a most powerful test and otherwise this will give the uniformly most powerful test.

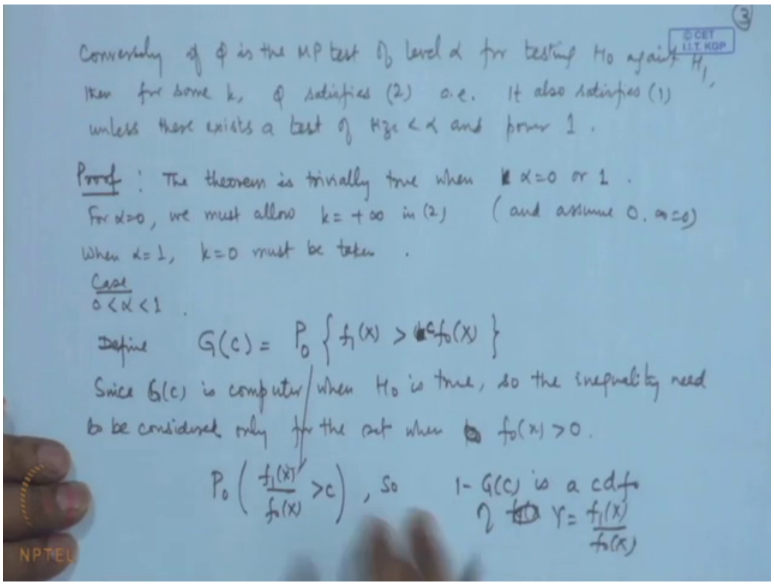
Now, in this model in you can easily see that our solution is dependent upon the alternative hypothesis. So, that is why I was mentioning that this approach has an important component that is we specify apart from specifying a null hypothesis, we must also specify an alternative hypothesis and that is what is happening in this particular situation here. So, this Nyman Pearson theorem theory actually specifies or consists solves this problem of hypothesis testing from this point of view.

I introduce the first measure result in this direction that is known as the fundamental lemma of Nyman Pearson. This lemma is available in all the statistics test, I have considered the statement and the proof from the book of Lehmann and Romano. Let pi naught and pi 1 be populations with distributions say f naught and f 1 respectively. And certainly we have to assume a probability measure with respect to which these will be the probability mass function or probability density function.

So, let me say with respect to measure mu, then we have then for testing h naught that is f is equal to f naught against the alternative h 1 f is equal to f 1 that is the simple versus simple hypothesis case. So, these f naught and f 1 are known these are fixed. So, for this hypothesis problem we can define a test phi with a constant k such that expectation of phi x under the null hypothesis, I will denote E naught is equal to alpha and the form of the phi x is equal to 1, when f one x greater tha n k times f naught x and that is equal to 0 when f one x is less than k times f naught x

So, I have not included the quality here, that part will be defining in the proof, that for testing a simple versus simple hypothesis case we can device a test function, which will achieve the exact level of significance or exact size, and the form is of this that is if f 1 by f naught is greater than k or f 1 by f naught is less than k. If phi satisfies 1 and 2 then for some k then it is most powerful which we use a notation M p for h naught against h 1 at level alpha.

(Refer Slide Time: 13:33)

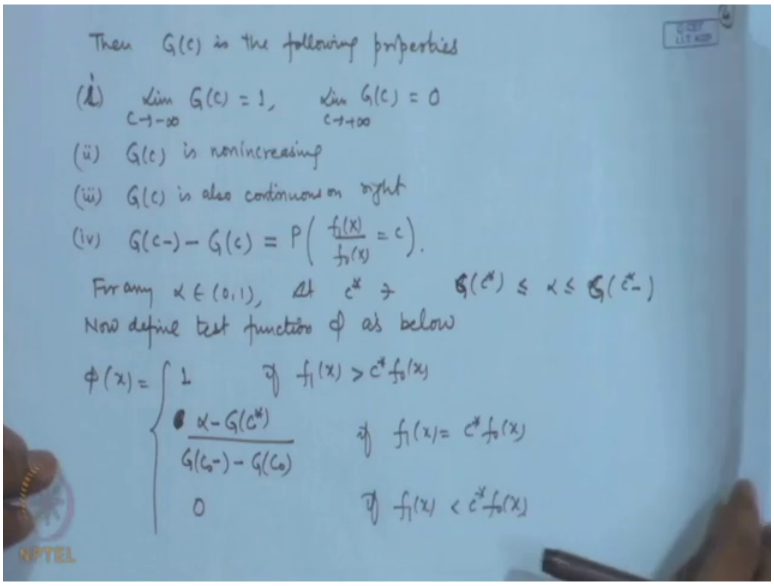


That means, for the given level this is the most powerful test; that means, the most powerful test must satisfy conversely. Conversely if phi is the most powerful test of level alpha for testing h naught against h 1, then for some k phi satisfies the condition 2 almost everywhere. It also satisfies 1 unless there exist a test of size less than alpha and power 1. So, this is the exceptional case, let us look at the proof of this either followed the step similar to Lehmann and Romano.

So, let us consider say if I consider alpha is equal to 0 if alpha equal to zero is there; that means, we should always accept h naught. If we always accept h naught then we can take k to be infinity as a convention. If I take alpha is equal to 1 then we should always reject and then we can take k as equal to 0 therefore, these two cases are trivially true. The theorem is trivially true when k when alpha is equal to 0 or 1. So, we are saying that for alpha is equal to 0 we must allow k is equal to plus infinity into and also assume that assume 0 into infinity is equal to 0. When alpha is equal to 1 then k is equal to 0 must be taken.

So, now we are considering the case when alpha is strictly between 0 and 1. Let us define a quantity a function as say G of c which is the probability of say f 1 x greater than k time's f naught x sorry c times. So, this is under h naught since G c is computed when h naught is true. So, the inequality need to be considered only for the set when p naught sorry f naught is positive. In that case this is actually becoming probability that f 1 x by f naught x is greater than c. So, 1 minus G c is a c d f of f 1 x let me call it random variable y f one x by f naught x.

(Refer Slide Time: 18:08)



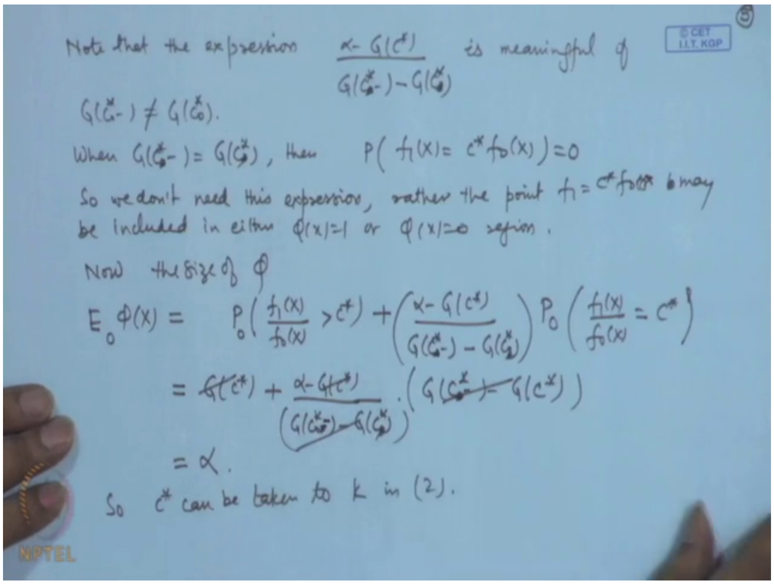
Now, if it is a c d f it will have certain properties, then G c has the following properties. So, for example, we know that limit of 1 minus G c as c tends to minus infinity this should be 0. So, limit of G c as c tends to minus infinity that will become 1 similarly, limit of G c let me not call it this 1, because we have used this number say elsewhere. So, we call it like this. c tends to plus infinity limit of 1 minus G c is one. So, this will become 0 then 1 minus G c is a non decreasing function. So, G c will be non-increasing and 1 minus G c is continuous on right. So, G c is also continuous on right.

So, these properties follow, because 1 minus alpha c is a cumulative distribution function. Further if I consider the left hand limit at G c minus the value at c this is nothing, but the probability of f 1 by f naught is equal to c. Now for any alpha line in the interval 0 to 1 let us choose say c star such that alpha c star less than or equal to alpha less than or equal to alpha c star minus. So, in the case of continuous this will be equal otherwise, this need not be equal in that case we may choose any value which is in between.

Now, define test function phi as below. So, we define phi x is equal to 1, if f 1 x is greater than c star f naught x it is equal to sorry this is G c. So, this is alpha minus G c star divided by G c naught minus minus G c naught. If f 1 is equal to c star f naught x. So, this is the randomization part, because in the discrete case there may be a positive probability of this thing. So, there we are assigning a value and it is equal to 0 if f 1 x is less then c star f naught x.

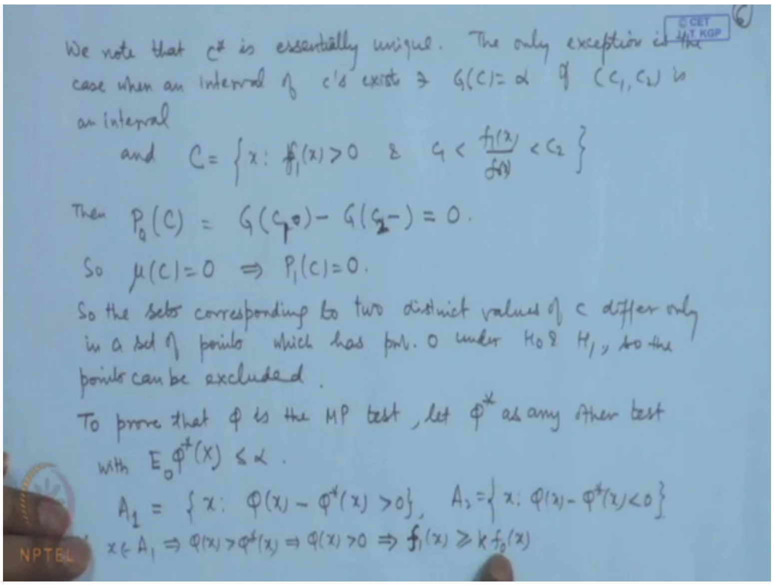
So, here you note here that in the statement of the lemma. We have taken two parts 1 and 0 and these parts you can see they are matching here. So, this k is equal to c star here the unspecified portion that is f 1 is equal to c star f naught x. Now we have a specified here. So, now if you have the situation that f 1 is actually if for example, G c naught minus is equal to G c naught; that means, if it is continuous then this expression will actually become meaningless in that case we do not have to define this. So, we do not have to consider this when it is continuous.

(Refer Slide Time: 22:34)



So, let me write this comment here, note that the expression alpha minus G c star divided by G c naught minus minus G c naught is meaningful. If G c naught minus is naught equal to G c naught when G c naught minus is equal to G c naught then probability that f 1 x is equal to c star f naught x that is equal to 0. So, we do not need this case this expression rather the point f 1 is equal to c star f naught may be included in either in phi x equal to 1 or phi x equal to 0 region and in the continuous case, it will not change the probability.

(Refer Slide Time: 26:38)



Now, let us consider the size of the test that is expectation of phi x. So, this is equal to probability of f 1 x by f naught x greater than c star under h naught plus alpha minus G c star divided by G c naught minus minus G c naught p naught f 1 x by f naught x is equal to c star plus other portion will be 0. So, that is equal to now this is nothing, but G c star plus alpha minus G c star by G c naught minus minus G c naught into this portion is nothing, but once again G c naught G c star minus G c star. So, these are stars here there is a mistake here, this should be star this is a star similarly, this should be star this should be star here, naturally this cancels out and then this cancels out. So, this is equal to alpha.

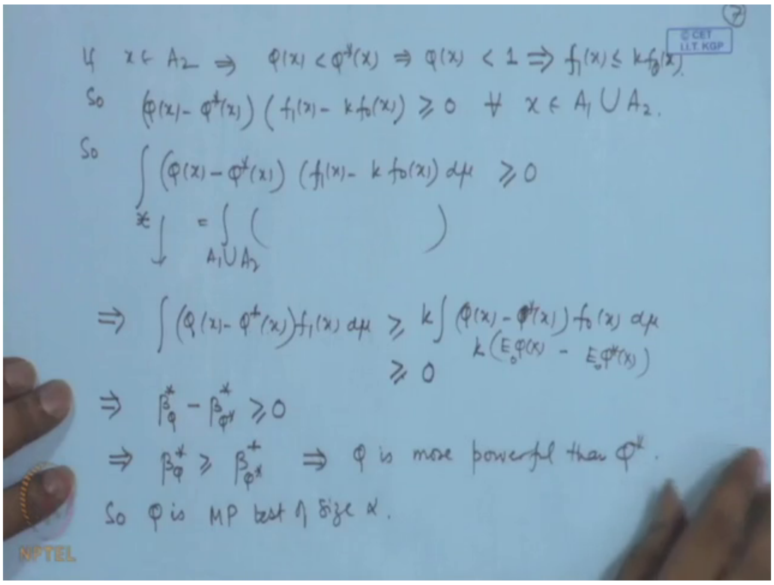
So, c star can be taken to be k in expression 2. Now another point is regarding the choice of c star I mentioned that when we have the continuous case then it is equal. So, there is a unique value, but in the discrete case there may be a possibility that there is more than one value, but all those values will give the same option here. So, there will not be any change in the ultimate solution. Let me just give a comment about this, we note that c star is essentially unique the only exception is the case when an interval of c is exist such that G c is equal to alpha if c 1 to c 2 is an interval of this nature.

And we consider the set c to be the set of all those values where f 1 is greater than 0 and c one less than f 1 by f naught is less than c 2. Then if we consider the probability of the set c under null hypothesis, that is equal to G c 2 minus minus, G g c 2 minus G c 1 sorry this will be c 1, this will be c 2 minus that is actually equal to 0. So, the measure of the set this will be 0 and this will imply that p 1 c is also 0. So, the sets corresponding to to distinct values of c differ only in a set of points, which has probability 0 under h naught and h 1; so, the points can be excluded from the sample space. So, that takes care of this non unique part here.

Now, let us consider. So, we what we have done we are able to construct a test function phi which satisfies condition 1 and 2. Now what we are saying is that this will be actually the most powerful test. So, to prove that to prove that phi is the most powerful test. Let us consider phi star as any other test with expectation of phi star, less than or equal to alpha under h naught. Now let us consider the two sets say a. Let us call the sets A 1 as the set of all those points such that phi minus phi star is positive and say A 2 is the set such that phi x minus phi star is less than 0.

Now, note here the way the sets are defined here, if x belongs to A 1 then phi x is greater than phi star. So, this implies that phi x is greater than phi star. Now these are this is also the probability; that means, phi is a strictly positive. If phi is a strictly positive this implies that p 1 x is f 1 x is greater than or equals to k time's f naught x. Now why this is true, because that is the way we have defined, phi is positive in these two regions. So, f 1 is greater than or equals to c star f naught here.

(Refer Slide Time: 31:38)

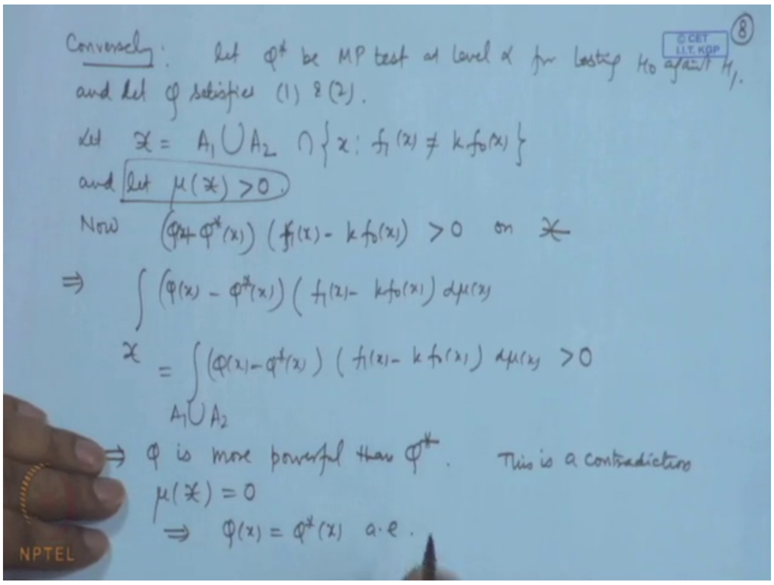


So, this condition will be true here. So, f 1 will be greater than or equal to k times f naught. Now if I take say x belonging to A 2 then this will imply phi x is less than phi star x. So, phi x is less than phi star x; now phi star x. So, this implies that phi x must be less than 1, because the other region is not possible we cannot have phi star is equal to 0 and then phi x less than 0 therefore, phi x must be less than 1, but this is about the region that f 1 x is less than or equal to k times f naught x.

Because in this portion and this portion we have f 1 less than or equal to c star f naught therefore, what we have concluded here that phi x minus phi star x multiplied by f 1 x minus k times f naught x is greater than or equal to 0 for all x belonging to A 1 union A 2. Now this implies that if I take the expectation or the integral f 1 x minus k times f naught x d mu this will be greater than or equal to 0 over the over the space, but over the whole space it is same as over A 1 union A 2 of the same thing.

So, what we are concluding then this implies that phi x minus phi star x f 1 x is greater than or equal to phi x minus phi star x, k times f naught x d mu. Now this is actually greater than or equal to 0, because this is nothing, but expectation of phi under h naught and this is expectation of phi is star under h naught into k. So, this we have assumed that this is less than or equal to alpha and this is equal to alpha. So, this will be greater than or equal to 0; now the left hand side is nothing but beta phi star minus beta phi star star; that is greater than or equal to 0.

(Refer Slide Time: 35:23)



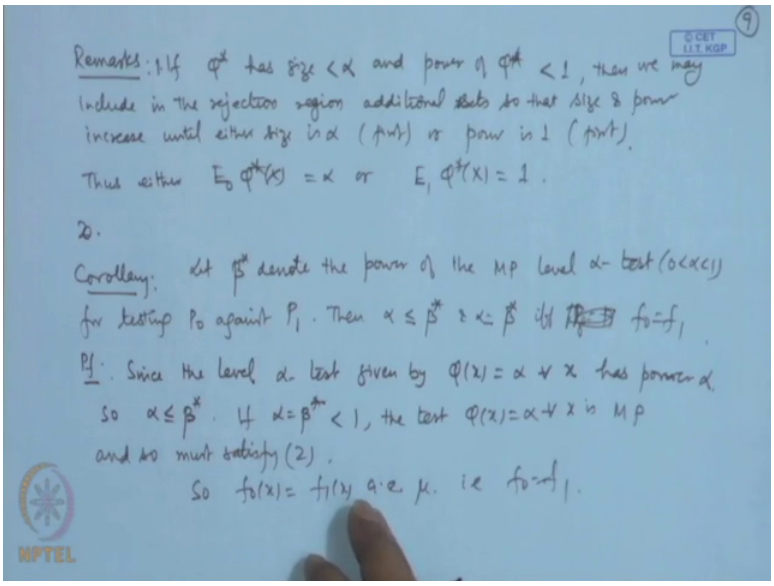
That is a power of the test function phi and this is a power of the function phi star. So, this means, that beta phi star is greater than or equal to beta phi star. So, this means, phi is more powerful than phi star. Now phi star was an arbitrarily chosen test with size alpha we had assumed expectation of phi star less than or equal to alpha. So, phi is most powerful test of size alpha, now let us prove the converse part of this here. So, let phi star be another most powerful test at level alpha for testing h naught against h 1 that we have a specified here.

And let us consider that phi satisfies 1 and 2 conditions. Let us consider say x as the A 1 union A 2 intersection with the set of those values where f one is different from k times f naught, and also assume that the measure of this is positive. Now you already consider that phi minus phi star into p 1 sorry f 1 minus k time's f naught is greater than 0 on x. So, this will imply that integral of phi x minus phi star into f 1 x minus k times f naught x that is equal to A 1 union A 2 phi x minus phi star x f 1 x minus k times f naught x is greater than 0.

Now, this condition implies that phi is more powerful than phi star. So, this is a contradiction, because we assume that phi star is most powerful. So, what does it mean, the only possibility is that this assumption is not correct. So, we should have mu x equal to 0, if mu x is equal to

0 then this means, that phi and phi star are same almost everywhere. So, this proves this Nyman Pearson fundamental lemma. Now let us look at further remarks on this.

(Refer Slide Time: 38:25)



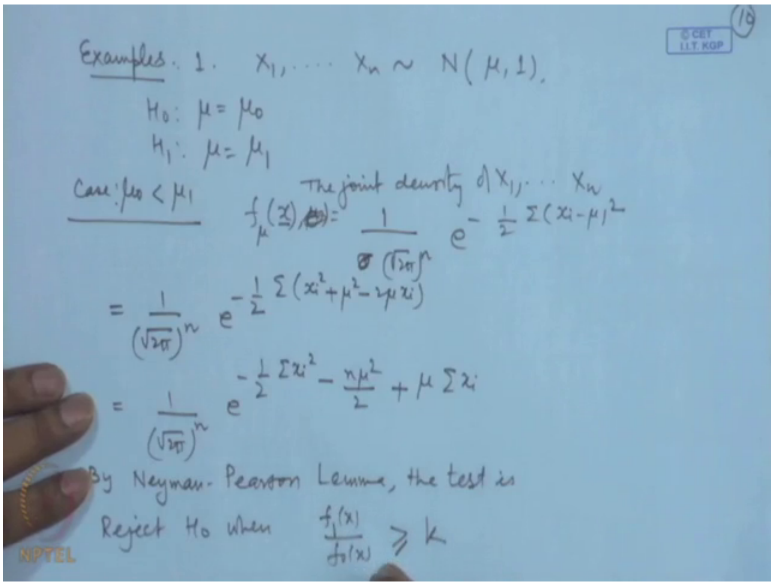
If phi star has size less than alpha and power of phi star is say less than 1. Then we may include in the rejection region see power and the size both are related to the probability of the rejection region. So, we may include additional points, additional points or you can say additional space or additional set. So, that size and power increase until either size is alpha first or power is one first. So, thus we will have either expectation of phi star equal to alpha or expectation of phi star x is equal to 1.

So, another thing that we have noticed here that except the point f 1 x is equal to k f naught. At other points the uniqueness of the in the definition of phi is there, and on the set where f one is equal to k times f naught here, there is a chance of shifting, because of the value that we are having there, that is the difference that we are having there G c minus this point here, because the c star may not be chosen uniquely and therefore, we can define arbitrarily.

However it means that it the size is still alpha. So, therefore, this does not make any difference here. Let me give examples here and of course, we have the following corollary. Let beta denote the power of the beta star denote the power of the most powerful level alpha test. For testing p naught against p 1 then alpha is less than or equal to beta and alpha is equal to beta, if and only if p naught is equal to that is pi naught is equal to pi or we can say f naught is equal to f 1.

Since the level alpha test given by phi x is equal to alpha for all x, this will have power alpha. So, alpha should be less than or equal to beta, because this is one of the test with power alpha and beta is the most powerful test power. If alpha is equal to beta star that is less than 1 then the test phi x is equal to alpha for all x is most powerful. And so, must satisfy 2, because of the necessity converse part of the Nyman Pearson lemma. So, f naught will be equal to f one almost everywhere mu that is the two densities are same.

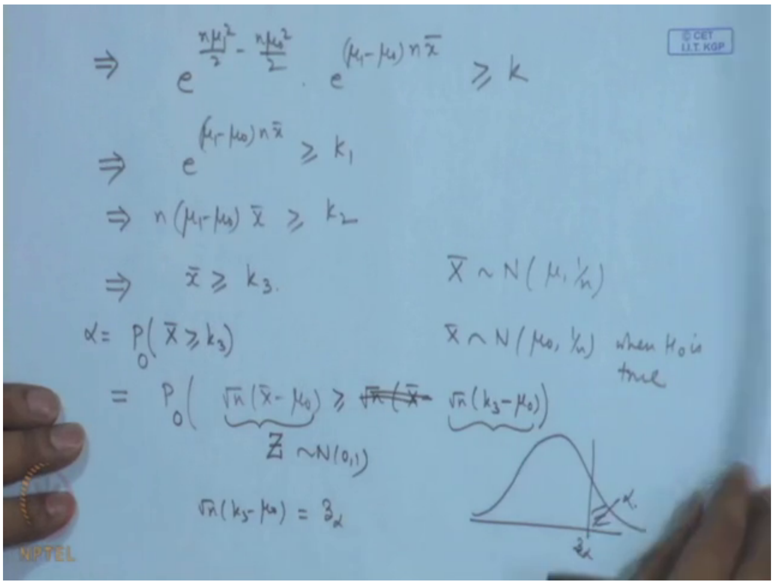
(Refer Slide Time: 43:26)



Now, let me give applications of this Nyman Pearson lemma in deriving the tests for simple versus simple hypothesis case. Let us start with say x 1, x 2, x n is a random sample from a normal mu 1 distribution, we are testing the hypothesis say mu is equal to mu naught against say mu is equal to mu 1. Let us take the case say mu naught is less than mu 1. So, let us write down the joint distribution of x 1, x 2, x n the joint density of x 1, x 2, x n.

So, that is that mu that we are calculating. So, that is equal to 1 by sigma is 1 root 2 pi to the power n e to the power minus 1 by 2 sigma x i minus mu square. Now this step we can simplify 1 by root 2 pi to the power n e to the power minus 1 by 2 sigma x i square plus mu square minus 2 mu x i, that is equal to one by root two pi to the power n e to the power minus one by two sigma x i square minus n mu square by two plus mu sigma x i.

(Refer Slide Time: 46:05)

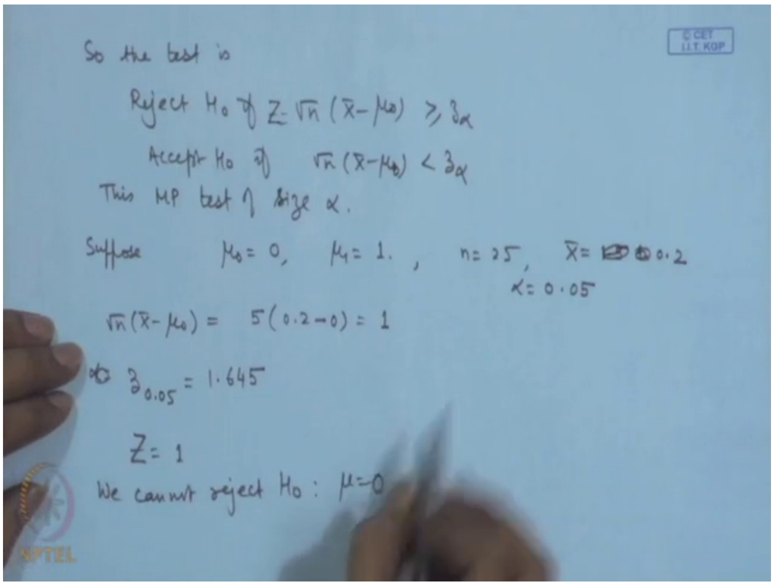


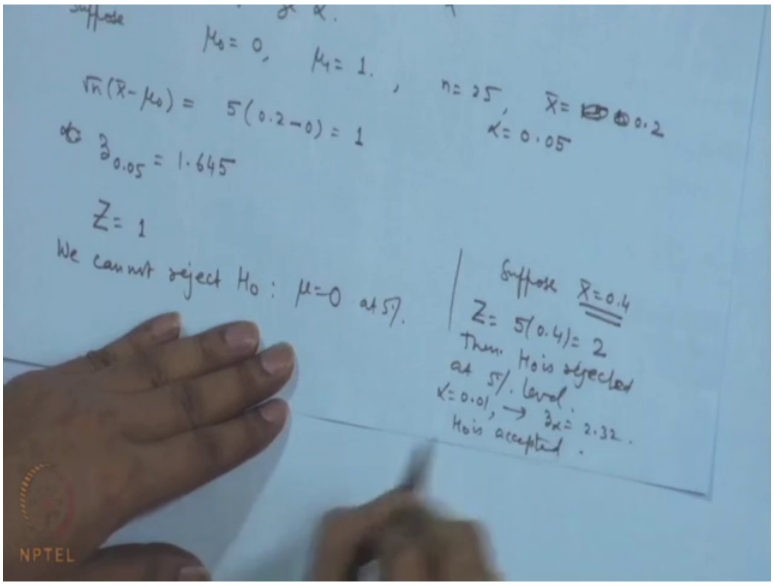
So, if I write down by Nyman Pearson lemma the test is reject h naught when f 1 x by f naught x is greater than k. We will put greater than or equal to or greater it will not make any difference in this case, because the distribution of x is continuous here. We are dealing with the normal distribution. So, the middle part of the phi function which is we given in the Nyman Pearson lemma is not required in this case. So, this condition is equivalent to now here we are having f 1 here mu 1 is coming and in f naught we have mu naught. So, when we write the ratio this terms get cancelled out.

We are getting e to the power n mu 1 square by 2 minus n mu naught square by 2 e to the power mu 1 minus mu naught sigma x , I we can write as n x bar greater than or equal to k 2. Now this is all constant mu 1 and mu naught are fixed constants. So, this is equivalent to e to the power mu 1 minus mu naught n x bar greater than or equal to some k one this is equivalent to saying if I take log on both the sides I get mu 1 minus mu naught n x bar greater than or equal to some k 2. Now this is equivalent to now I have here mu 1 minus mu naught positive.

So, this is equivalent to saying x bar greater than or equal to k 3. Now the distribution of x bar is normal mu 1 by n. So, when we consider alpha that is the probability of rejecting h naught then here x bar follows normal mu naught 1 by n when h naught is true. So, this can be written as probability root n x bar minus mu naught, greater than or equal to root n x bar minus sorry root n k 3 minus mu naught, when mu is equal to mu naught, when mu is equal to mu naught this is following a standard normal distribution.

(Refer Slide Time: 48:40)



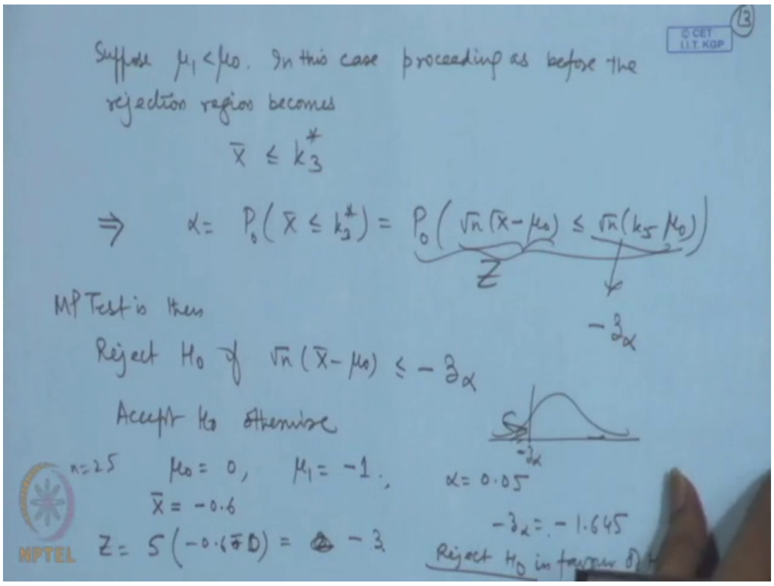


So, these values root n k 3 minus mu naught is nothing, but z alpha value. Where z alpha denotes the upper 100 alpha percent point on the standard normal distribution. So, the test is reducing to. So, the test is reject h naught if root n x bar minus mu naught is greater than or equal to z alpha, and accept h naught or you can say reject h 1 or do not do not reject h naught if root n x bar minus mu naught is less than z alpha inclusion of equality in this case or this case does not make any difference, because of the continuity the property of the equality will be actually 0.

So, this is the most powerful test of size alpha. Let us take an practical example, here suppose I take say mu naught is equal to 0, mu 1 is equal to say 1 and say n is equal to say 25. And in a given problem suppose my x bar is equal to 1.5 or say 1 point x bar is say equal to 0.2. In that case let us calculate this quantity root n x bar minus mu naught that is equal to 5.2 minus 0 that is equal to 1. And let us consider say alpha is equal to say 0.05 if I take alpha is equal to 0.05 then z of 0.5 is equal to 1.645.

So, we are getting here this let me call this value as z, we are getting z as 1 and z alpha value is 1.645. So, we cannot reject h naught that is mu is equal to 0, if x bar is equal to 0.2. On the other hand suppose here, I would have got say x bar is equal to 0.4 in that case z value would be equal to 5 into 0.4 that is equal to 2 in that case this value will be higher. So, then h naught is rejected at 5 percent, but if I change the level of significance it may still be possible to accept this.

(Refer Slide Time: 52:16)



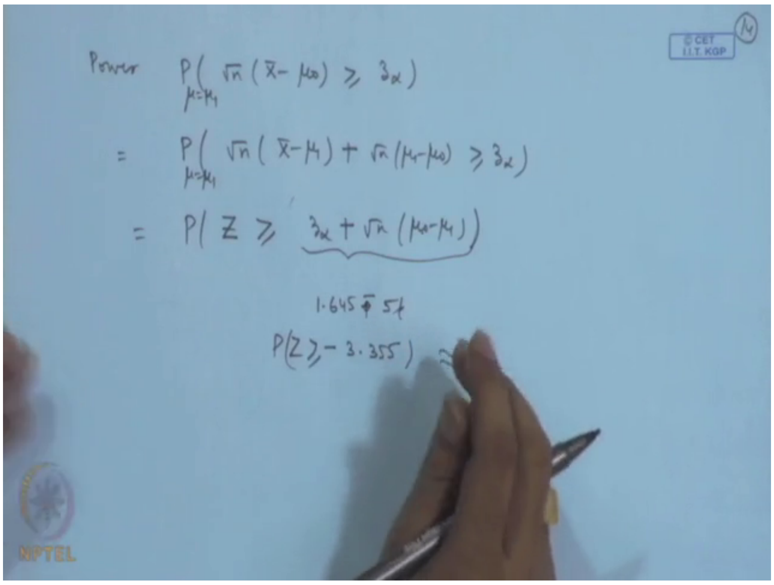
For example, if I take say alpha is equal to say 0.01 if I take this then I will get z alpha is equal to 2.32, and here I will get h naught is accepted. Also notice here that I have considered mu 1 greater than mu naught suppose I consider mu 1 less than mu naught. Suppose mu 1 is less than mu naught. In that case from here if you consider the condition it will change to x bar less than or equal to k 3, because if mu 1 minus mu naught is negative then the region will get reversed. In this case proceeding as before the rejection region becomes x bar less than or equal to k 3 let me call it k 3 star.

So, in that case if I consider alpha and this is then reduced to root n x bar minus mu naught less than or equal to root n k 3 minus mu naught. Now if we are putting this is equal to alpha, then this is z here and this will become minus z alpha, because this point here minus z alpha where the probability lower 100 alpha percent point here. So, the test is then this is the most powerful test, reject h naught if root n x bar minus mu naught is less than or equal to minus z alpha accept h naught or do not reject h naught otherwise.

For example, here if I take say mu naught is equal to say 0 mu 1 is equal to say minus 1. And let us take say alpha is equal to once again 0.05. Suppose the observed value of x bar turns out to be say, point minus 0.6 then what will happen here this value let us call it z and n is equal to 25. So, this is equal to 5 x bar is minus 0.6 plus 1 that is equal to 2 here. So, once again you note sorry mu naught is 0. So, that is minus 3 and minus z alpha is equal to minus 1.645. So, here you are observing that this value is smaller than this. So, we will reject h naught. So, we will reject h naught in favor of h 1.

Now, the problem that I have discussed can be easily seen to have wider ramifications and of course, in both the cases we can calculate the power of the test also. What will be the power of the test for example, in this case let us see power what will be the power here it is the probability of rejecting root n x bar minus mu 1 greater than or equal to z 0.05, because this point has already been decided here. So, this is 1.645 that is probability of 25 sorry 5 x bar is 0.2minus 1 greater than or equal to 1.645 sorry this is not the correct calculation let me do it again here.

(Refer Slide Time: 56:34)



Probability of root n x bar minus mu naught greater than or equal to z alpha under mu is equal to mu 1. So, that is equal to and mu is equal to mu 1 this does not have the a standard normal distribution rather we need shifting here, x bar minus mu 1 plus root n mu 1 minus mu naught greater than or equal to z alpha, for mu is equal to mu 1. That is equal to probability of z greater than or equal to z alpha, plus root n mu naught minus mu 1. So, this can be again evaluated for example, in this particular case we had taken z alpha is equal to 1.645 plus 5 times mu naught minus mu 1 is minus 1. So, this is minus 5. So, that is equal to minus 3.355 that is probability z greater than or equal to this that is nearly 1.

So, the power of this test in this particular case is almost 1. So, it is good because in the normal distribution case this probability is almost 1. In the next class I will consider further applications of the Nyman Pearson lemma to derive the most powerful tests in the simple versus simple hypothesis case. And then we will further extend these results to cover the case when the hypothesis may become composite. And we will be discussing then certain results or certain conditions on the density functions which will keep the results for those distributions. So, that too I will be covering in the following lecture.