Statistical Inference

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

Lecture – 33
Neyman Pearson Fundamental Lemma – I

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 call as null hypothesis, alternative hypothesis, classification of the hypothesis such as simple hypothesis or a composite hypothesis, what is a non-randomized 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 an assertion about the truthfulness or correctness of the hypothesis. It simply means that our sample supports they hypothesis or does not support the hypothesis.

So, the use of the testing procedure should be done with caution they are not absolute truths. 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 than under two categories they are called type I error, that is the probability of the probability of rejecting the null hypothesis when actually this true and beta we called the probability of type II error that is the probability of accepting H naught when it is false.

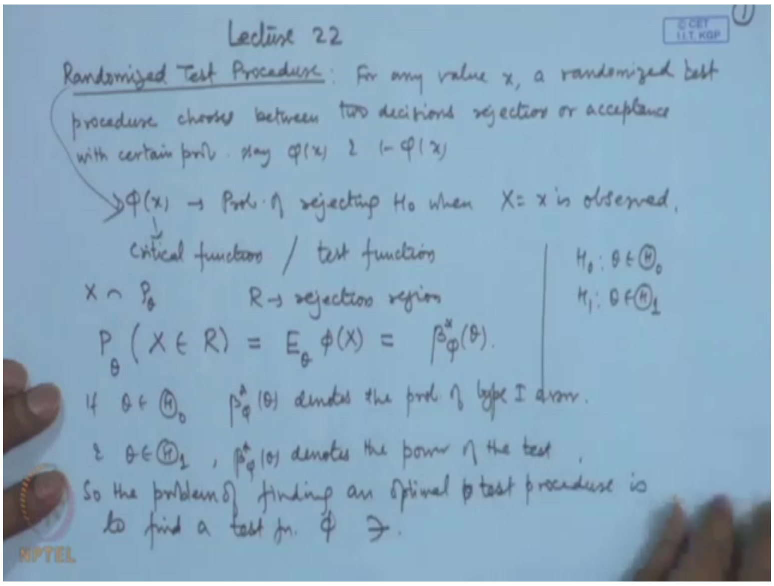
We have seen that the consequences of the 2 types of errors can be quite different and it could be quite disastrous also. And therefore, any reasonable test procedure must control the 2 types of errors and naturally the ideal situation should be that both alpha and beta are actually 0 or you can say 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 I error is taken to be in a more serious way therefore, we fix an upper bound for that. For example, suppose it is a medical problem; that means, the false falsely claiming that the 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 say, in that case with this we trying to find out that test procedure for which beta is actually the minimum or we have introduced a new concept called power that is 1 minus beta, so, power should be maximum. Now, a test when you assign a 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 has a single number in that case it may happen that up to a certain level alpha with the value of the probability of type I error is below alpha and after some stage it becomes greater than alpha; that means, equal to alpha is not achieved. To overcome this situation we can define slightly more general form of the test procedures called randomized test procedures.

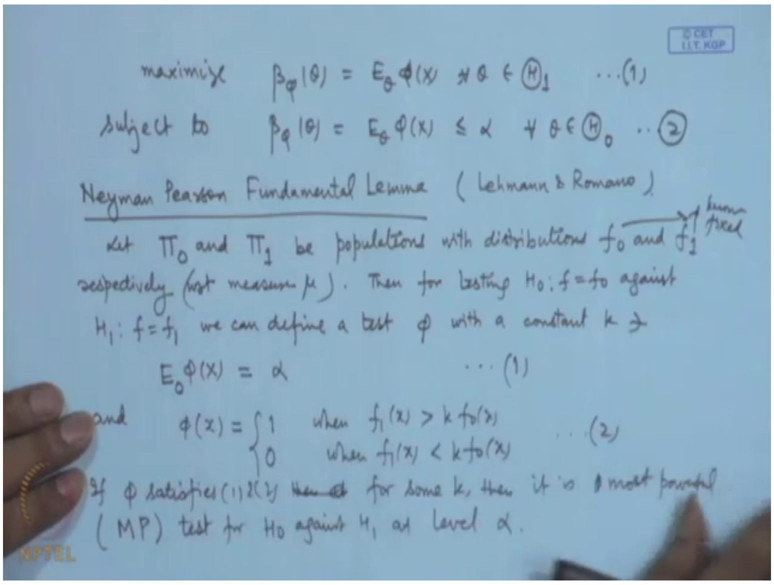
(Refer Slide Time: 04:12)



So, for any value x a randomized procedure chooses between two decisions that is rejection or acceptance with certain probabilities say phi x and 1 minus phi x. So, we are actually saying phi x is the probability of rejecting H naught then 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 star phi theta. Then 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 star phi theta actually denotes the probability of type I error and for theta belonging to theta 1 then beta star phi theta denotes the power of the test. So, the problem of finding an optimal test procedure is it can be stated as to find a test function phi.

(Refer Slide Time: 07:45)



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. Then theta 1 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 Neyman Pearson theorem theory actually specify or you can say solves this problem of hypothesis testing from this point of view.

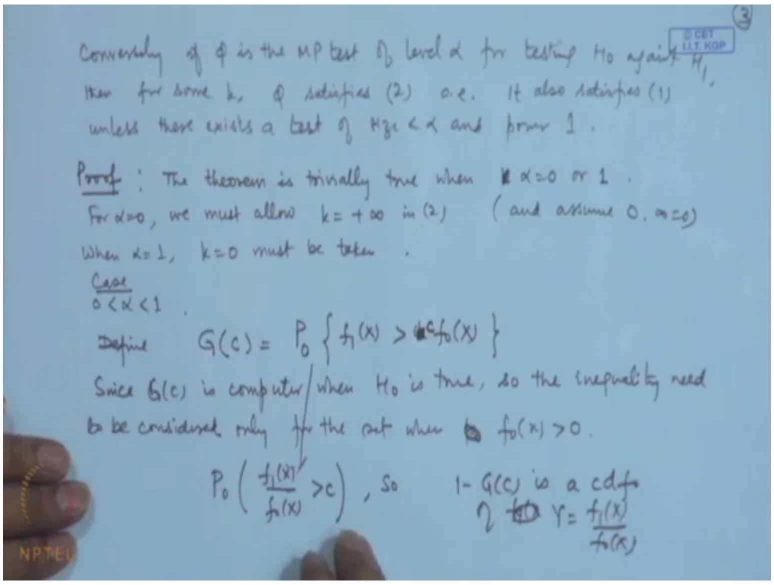
I introduced the first measure result in this direction that is known as the fundamental lemma of Neyman Pearson. This lemma is available in all the statistics tricks. I have considered the statement and the proof from the book of Lehmann and Romano. Let pi naught and pi 1 be populations with distributions 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 the 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 ok. 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 1 x greater than k times f naught x and it is equal to 0 when f 1 x is less than k times f naught x.

So, I have not included the equality here that part we will be defining in the proof that for testing a simple versus simple hypothesis case we can devise 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 the notation MP for H naught against H 1 at level alpha; that means, for the given level this is the most powerful test; that means, the most powerful test must satisfy.

(Refer Slide Time: 13:34)



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 exists a test of size less than alpha and power 1.

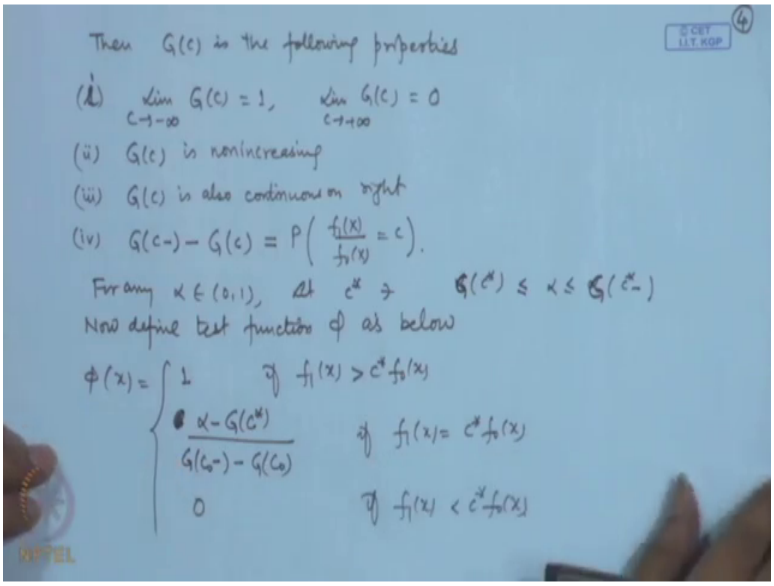
So, this is the exceptional case let us look at the proof of this I have followed the steps similar to Lehmann and Romano. So, let us consider say if I consider alpha is equal to 0. If alpha is equal to 0 is there; that means, we should always accept H naught if we always accept H naught then we can 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 in 2 and also assume that assume 0 and to 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 times 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 cdf of f 1 X let me call it random variable Y f 1 X by f naught X. Now, if it is a cdf it will have certain properties.

(Refer Slide Time: 18:09)



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 use this numbers elsewhere. So, we will call it 1 like this c tends to plus infinity limit of 1 minus G c is 1. 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.

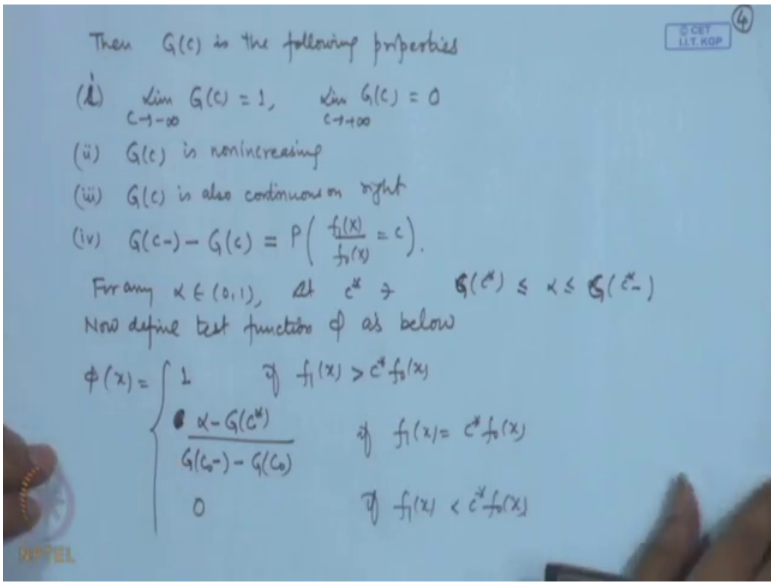
So, these properties follow because 1 minus alpha c is a cumulative distribution function. Further if I consider the left hand limit at G c minus the value at c this is nothing, but the probability of f 1 by f naught is equal to c. Now, for any alpha lying in the interval 0 to 1, let us choose say c 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 than c star f naught x. So, here you note here that in the statement of the lemma we have taken 2 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: 22:35)



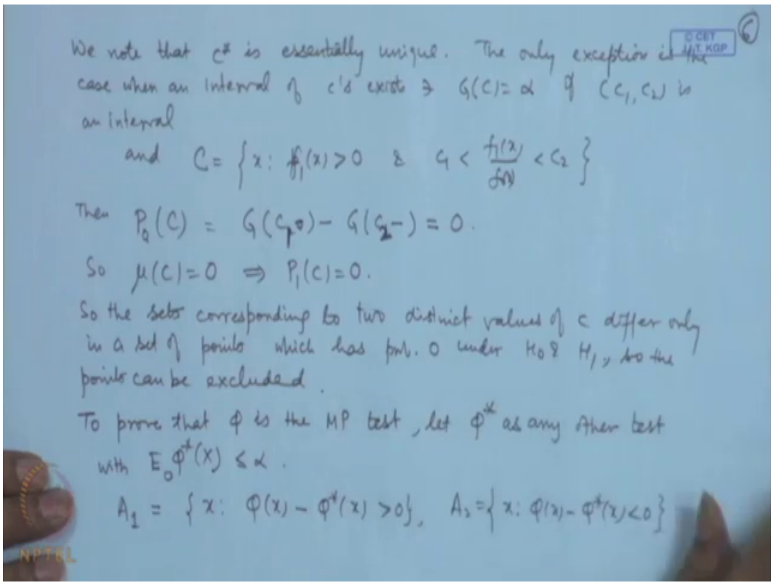
So, let me write this comment here. Note that the expression alpha minus G c star divided by G c naught minus minus G c naught is meaningful if G c naught minus is not equal to G c naught. When G c naught minus is equal to G c naught then probability that f 1 X is equal to c star f naught X that is equal to 0.

So, we do not need this is this expression rather the point f 1 is equal to c star f naught may be included in either phi X equal to 1 or phi X is equal to 0 region and in the continuous case it will not change the probability.

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 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.

(Refer Slide Time: 26:39)



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's exists 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 1 less than f 1 by f naught is less than c 2.

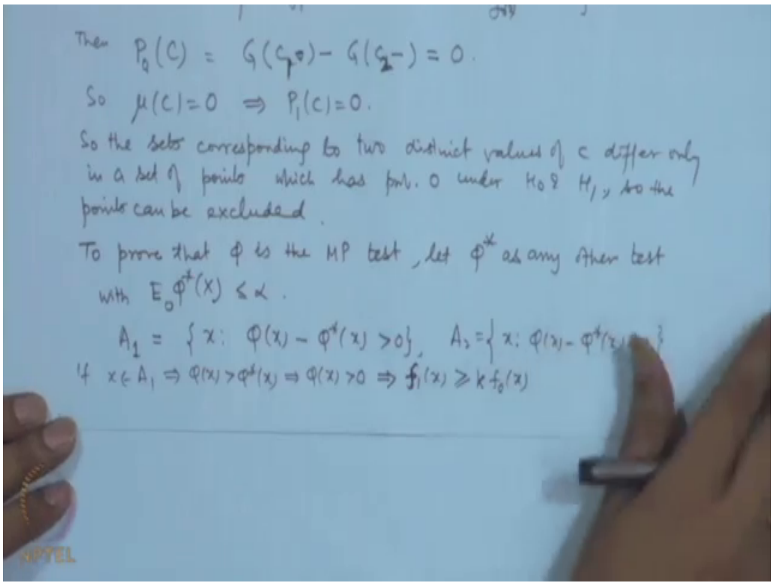
Then, if we consider the probability of the set C under null hypothesis that is equal to G c 2 minus minus G G c 2 minus G c 1 sorry this will be c 1 this will be c 2 minus that is actually equal to 0. So, the measure of the set this will be 0 and this will imply that P 1 C is also 0.

So, the sets corresponding to two 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 the. 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 2 sets a 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 a set such that phi x minus phi star is less than 0. Now, note here the way the sets are defined here.

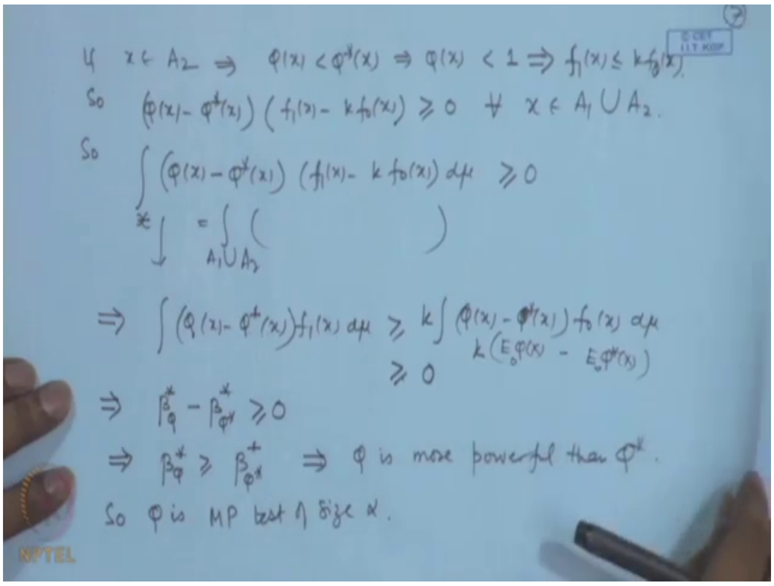
(Refer Slide Time: 30:45)



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 strictly positive. If phi is a strictly positive this implies that P 1 x is f 1 x is greater than or equal to k times 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 equal to c star f naught here. So, this condition will be true here. So, f 1 will be greater than or equal to k times f naught.

(Refer Slide Time: 31:39)



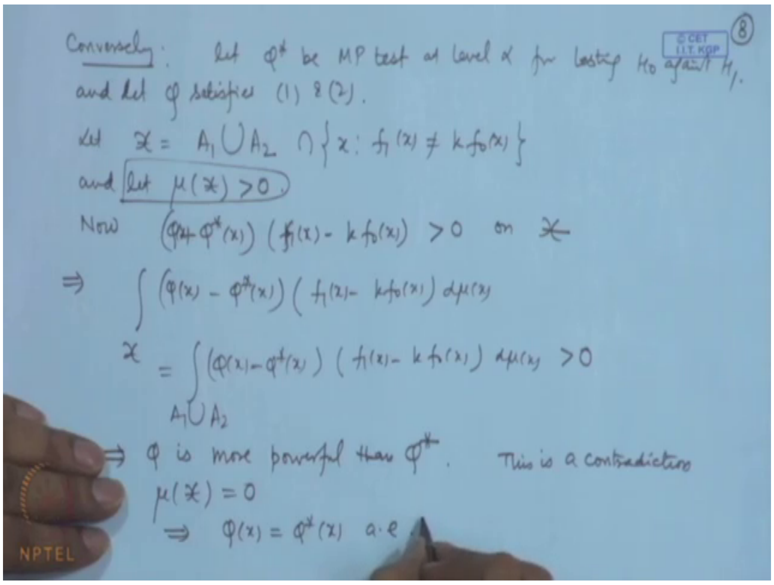
Now, if I take say x belonging to A 2 then this will imply phi x is less than phi star x. So, phi x is less than phi star x now phi star x. So, this implies that phi x must be less than 1, because other region is not possible. You cannot have phi star is equal to 0 and then phi x less than 0. Therefore, 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 whole 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 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, that is the power of the test function phi and this is the power of the function phi star. So, this means that beta phi star is greater than or equal to beta phi star star. So, this means phi is more powerful than phi star. Now, phi star was an arbitrarily chose an test with size alpha we are 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.

(Refer Slide Time: 35:24)



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 1 is different from k times f naught and also assume that the measure of this is positive.

Now, we have already considered that phi minus phi star into p 1 sorry f 1 minus k times 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 these Neyman Pearson fundamentals, so, that we will be covering in the following lecture.