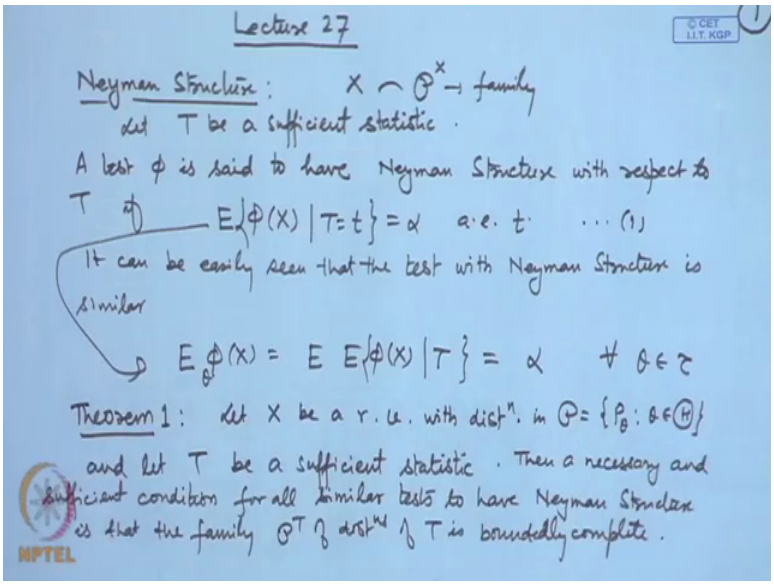
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 43
UMP Unbiased Tests – III

Yesterday we have discussed the case that when we have two sided alternative hypothesis, then the UMP test does not exist we have also shown through an example. And then we said that if we have a condition called similarity similar test; level alpha similar test then for distributions in the exponential family we have UMP unbiased test, when the alternative hypothesis is two sided. And we have demonstrated the test for normal distribution.

Now, we will further discuss this thing, normally we have seen that when we are discussing exponential families; there is a concept of sufficient statistics then there is a concept of nuisance parameters etcetera. We will today show that we can incorporate these concepts to derive the UMP unbiased test when we are dealing with the multi parameter exponential families.

(Refer Slide Time: 01:25)



So, in this regard first of all I introduce Neyman structure. So, what is a test with a Neyman structure. So, as usual we have a random variable or a random vector X; which is having a distribution and we say and the family of distributions is X ok.

Let T be a sufficient statistic here; let T be a sufficient statistic. So, we say that a test phi is said to have Neyman structure with respect to T; if the conditional expectation of phi X given t is equal to alpha almost everywhere t. Now, what does this condition represent? See if I consider simply expectation of phi X; then this is nothing, but the power function.

For theta in the null hypothesis parameter set, this denotes the probability of type 1 error and for theta belonging to the alternative hypothesis said this represents the power of the test. Now, if we consider expectation of phi X given T; if T is a sufficient statistic then what does it mean? This term will be independent of theta. Now what does it mean that; on every value of T is equal to t; which we can call orbits of T on every orbit of T this will have power is equal to alpha.

So, this is a much more stronger condition and here we can say of course, you can observe that let me call it to 1. It can be easily seen that the test with Neyman structure is similar. Because if I consider another expectation here; expectation of phi X that is equal to expectation of expectation phi X given T, now this is equal to alpha. So, for all theta belonging to tau that is the boundary of the null and alternative hypothesis set.

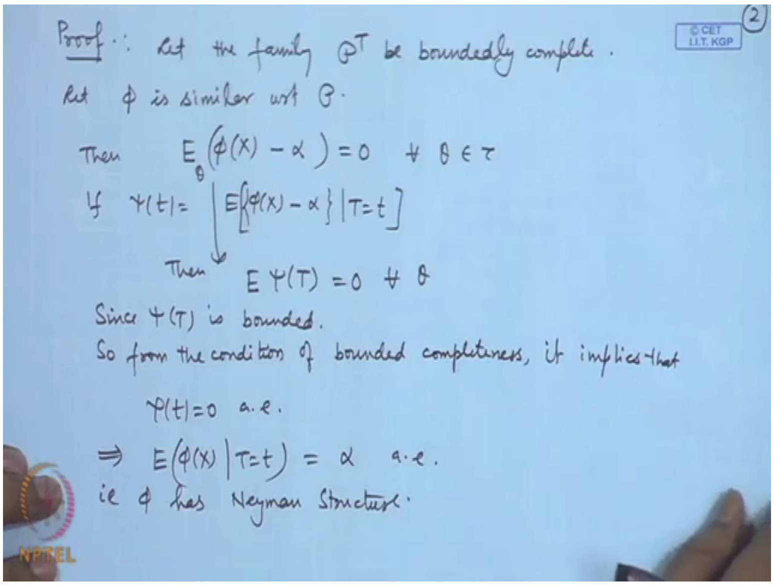
So, here what we are saying is that the condition of the power or you can say similarity is brought down to the level of the sufficient statistic that is the orbits of the sufficient statistic. So, now many times what happens that it is easier to obtain the most powerful test or the UMP test; among the test which is having Neyman structure and then since for every T it is most powerful; so overall it will be most powerful.

So, here frequently another idea that is used is the completeness idea. The distribution the definition of the completeness and examples of the complete family of distributions and the complete statistics; we have discussed earlier in the point estimation in connection with the derivation of the uniformly most powerful in connection with the uniformly minimum variance unbiased estimation etcetera.

So, I will not be repeating those steps again; I just advise the students to go back to the lectures on point estimation and again revise the concept of the completeness. Here what I will do? I will try to incorporate or you can say use the concept of completeness in deriving the UMP unbiased test. And in particular the Boscoes theorem is also used here, the Boscoes theorem is regarding the independence of two statistics if one statistic is sufficient and boundedly complete and another statistic is having a distribution free from parameters then the two statistics are independent.

So, these things will be there, I am not going to repeat these steps here; the students are advised to refer to my earlier lectures which are related to the completeness. Now I will give a result here; let me give some numbering here theorem 1; let X be a random variable or random vector with distribution in P. And let T be a sufficient statistic; then a necessary and sufficient condition for all similar tests to have Neyman structure is that the family say P T of distributions of T is boundedly complete; note here that full completeness is not required here, bounded completeness is enough here. Let me give the proof of this here.

(Refer Slide Time: 08:13)

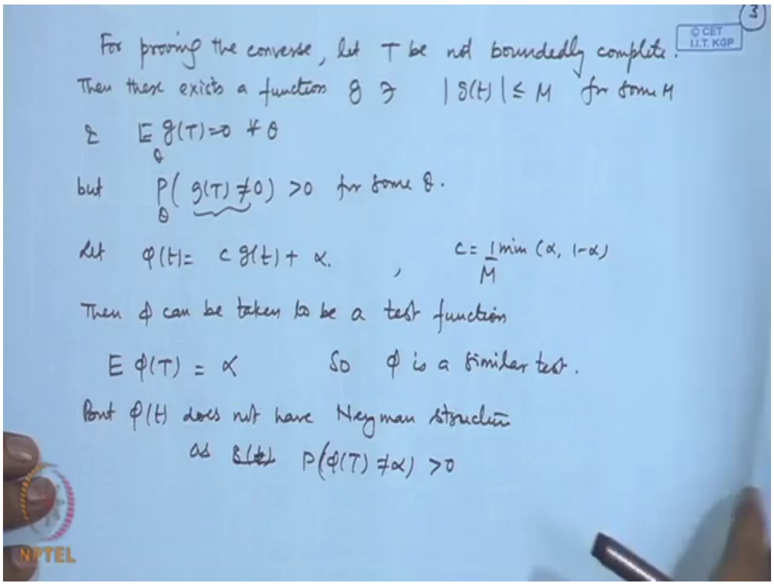


So, let the family P T be boundedly complete and let us assume that phi is similar. So, if it is similar we are able to write down that expectation of phi X minus alpha is equal to 0 for all theta belonging to; if it is similar then we can say it is equal to this for every belonging to tau.

Now, if we are having psi t is equal to expectation of phi X minus alpha given T is equal to t; let me give this notation here. Then what we are saying is this statement will imply expectation of psi T is equal to 0 for all theta. Now this is a test function; so this lies between 0 to 1 because this is simply denoting the probability of rejecting h naught. And alpha is the number between 0 and 1 this is also the probability level we are fixing. So, this psi T; psi T is bounded; so from the condition of let me just revise the theorem of similarity here; that we are having beta phi theta is equal to alpha that was equal to expectation of phi X is equal to alpha.

So, from the condition of bounded completeness; it implies that psi t is equal to 0 almost everywhere which implies that expectation of phi X given T is equal to t it is equal to alpha almost everywhere; that means, the test has phi has Neyman structure.

(Refer Slide Time: 11:19)

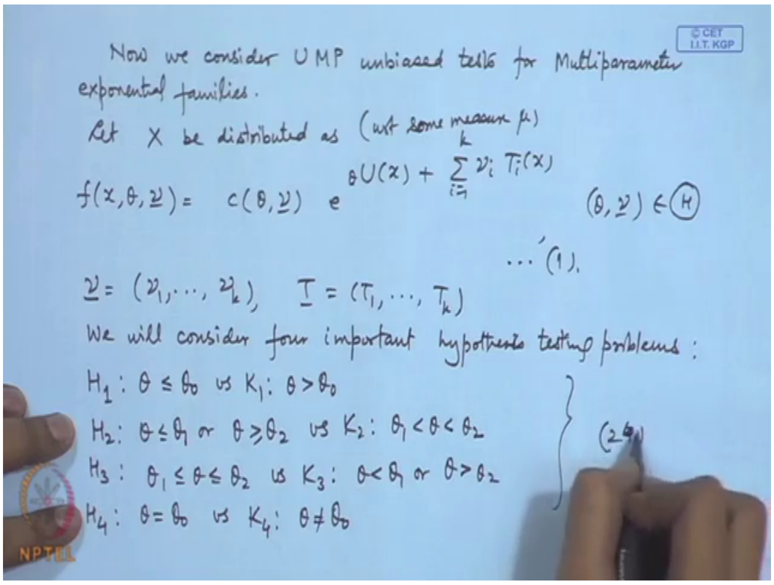


Now, let us look at the converse; for proving the converse, let T be not boundedly complete. Then there exist a function say g; such that g is bounded and expectation of g T is equal to 0 for all theta, but probability that g T is not equal to 0 is positive for some theta.

Now, let us assume this phi t to be equal to a constant times g t plus some alpha. And here c I am choosing to be minimum of alpha and 1 minus alpha divided by M; then what can we say about phi? See c is the minimum of this thing alpha 1 minus alpha by M and g T is bounded by M. So, this phi t becomes a test function; then phi t can be taken to be a test function; also what is expectation of phi T? Since expectation of g t is 0 this is simply equal to alpha; so phi is a similar test.

But phi t does not have Neyman structure because I am assuming that g t is not equal to 0 as g t so; that means, we are assuming that probability that phi T is not equal to alpha is positive. So, phi t does not have Neyman structure. If I assume that it is not bounded a complete then phi T does not have Neyman structure. So, the converse part is also proved; that means, T should be boundedly complete. Now, these results are useful for deriving the UMP unbiased tests for multi parameter exponential families.

(Refer Slide Time: 14:03)



So, let us consider now; now we consider UMP unbiased tests for multiparameter exponential families. So, let us consider the multiparameter exponential family as let X be distributed as; so we are writing f x theta and some nu that is equal to c theta nu E to the power theta U x plus sigma nu i T i x i is equal to 1 to k.

Now this is with respect to some major mu because we may deal with the discrete or continuous or mixed distribution. So, this is a general form of the probability density, here you have theta U x plus sigma nu i T i and this theta nu this belongs to some parameter space say script theta let me call it 1. We will use the abbreviated notation nu for nu 1 nu 2 nu k; T for T 1, T 2, T k.

And now if we remember yesterday's reference I have introduced four important type of hypothesis. Let me repeat them here we will consider four important hypothesis testing problems. So, we will follow the notation that I introduced yesterday, H 1 theta less than or equal to theta naught versus K 1 theta greater than theta naught. H 2 theta less than or equal to theta 1 or theta greater than or equal to theta 2 versus K 2; theta 1 less than theta less than theta 2; H 3 theta 1 less than or equal to theta less than or equal to theta 2 versus

K 3 theta less than theta 1 or theta greater than theta 2, H 4 theta is equal to theta naught versus K 4 theta is not equal to theta naught.

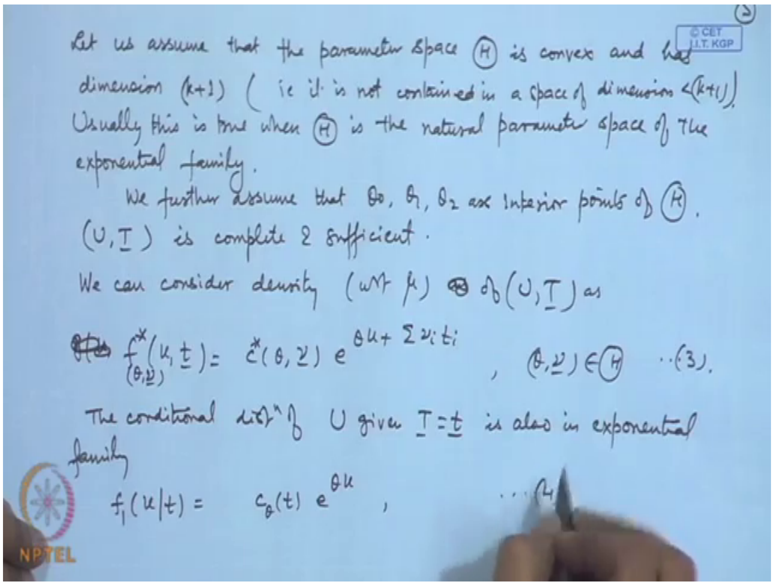
So, let me give reference 2 to all these four important types of hypothesis. In the case of one parameter exponential family we have shown that UMP tests exist for H 1 and H 2 and UMP unbiased test exists for H 3 and H 4, but now we are dealing with multi parameter exponential family.

Here I am writing theta is one of the parameters, but there are other parameters also like nu 1 nu 2 nu k. These are termed usually as nuisance parameters; for example, if I write down the normal distribution with parameters mu and sigma square. Then in the exponent I will be able to write E to the power minus X square by 2 sigma square plus theta plus mu by sigma square X.

So, if I have N observation then it will become sigma X i square and sigma X i there. So, I will have two parameters from mu and sigma square I can write mu by sigma square and minus 1 by 2 sigma square. So, either of them can be considered as theta and other one can be considered as nu. So, this is an example of a two parameter exponential family.

The one which I have written here; this is a k plus 1 parameter exponential family; now we make certain assumption on the parameter space also.

(Refer Slide Time: 18:35)



Let us assume that the parameter space theta is convex and has dimension k plus 1. Now this assumption is required if you remember the result for the k plus 1 parameter exponential family; when we have this type of thing then the parameter space if it contains k plus 1 dimensional rectangle, then U T 1 T 2 T k is a complete and sufficient statistic.

Sufficiency is of course, clear from the factorization theorem, but this will also be complete. Therefore, this assumption that the dimension of this parameter space is full that is required; that means, we are not assuming. So, we are saying that it is not contained in a space of dimension less than k plus 1.

Usually this is true when theta is the natural parameter space of the exponential family we have seen one example where we are dealing with the two normal distributions and the means were same. When the means became same the dimension became one less and therefore, the completeness was lost there.

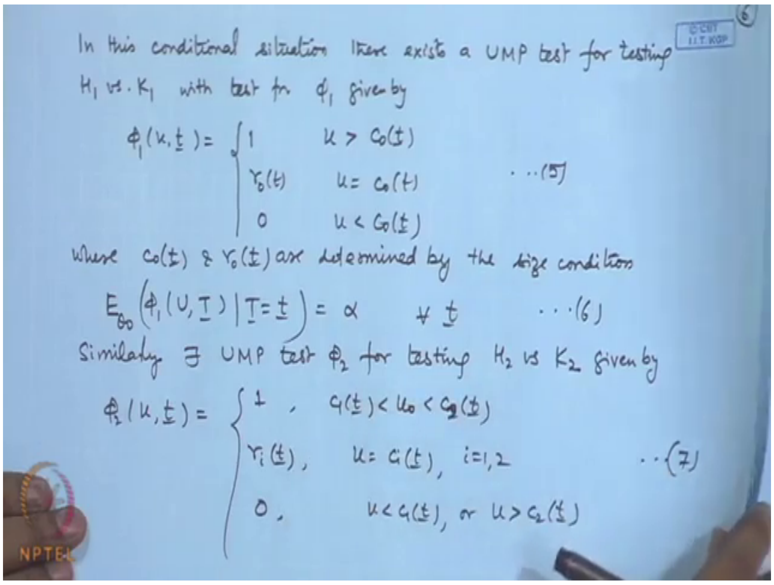
Then when we are dealing with the testing problems; we have mentioned certain points like theta naught theta 1 theta 2. So, we assumed that these are in the interior; that means, there are points which are less or more than these. So, we further assume that theta naught theta 1 theta 2 are interior points of theta. So, U T this is complete and sufficient. So, we can restrict attention to density with respect to measure mu as of U T; c theta nu E to the power theta u plus sigma nu i t i.

So, this constant may change here I will put here c star earlier I have written c in the case of f density. So, here I change it to c star and of course, the parameters theta and nu are occurring here; theta and nu belongs to theta. The conditional no distribution of U given T is equal to t is also in exponential family.

So, I can write the notation here say f 1; u given t that is equal to say c theta t e to the power theta u and some coefficient will come. Now if you look at this here t has become fixed here; so this is nothing, but a one parameter exponential family. In the one parameter exponential family if I am considering the tests H 1 and H 2; I have UMP test and for H 3 and H 4 I have UMP unbiased tests.

Let me relate these things here note here that there will be a little modification in the coefficients here because the densities will be with respect to different measures. Here we have started with mu then we are dealing with x; when we are dealing with nu and T then the measure gets little bit modified. So, I have changed here c star and when we are considering the conditional distribution of u given T; I further modified this coefficient. So, the measure will be accordingly whatever variable we are considering here.

(Refer Slide Time: 24:21)

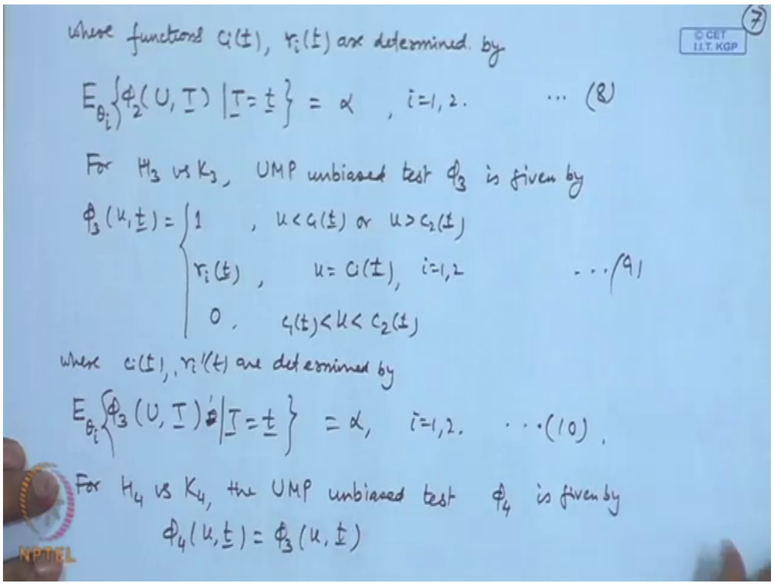


So, in this conditional situation there exists a UMP test for testing H 1 versus K 1 with test function phi 1 given by; it is 1, when u is greater than some coefficient say c naught, but this may depend upon t, this is gamma naught t even u is equal to c naught t, it is 0 u is less than c naught t; where c naught t and gamma naught t are determined by the condition.

Expectation of phi 1 U T given T is equal to t; it is equal to alpha for all t. So, here you can see the modification from the original one, in the original we are considering simply one parameter exponential family and therefore, the distribute the test was 1; if u is greater than c naught gamma naught if u is equal to c naught and 0 if u is less than c naught. But now there is a dependence on t and this size condition is also conditional now.

In a similar way if we are considering H 2; similarly there exists UMP test say phi 2 for testing H 2 versus K 2 given by phi 2 u t is equal to; let me describe these things in detail so that it is clear the dependence on t. And once again the constant from the function c 1, c 2, gamma 1, gamma 2 are determined by c i t and gamma i t are determined by phi of; phi 2 of U T given T is equal to t, this is equal to alpha for i is equal to 1, 2.

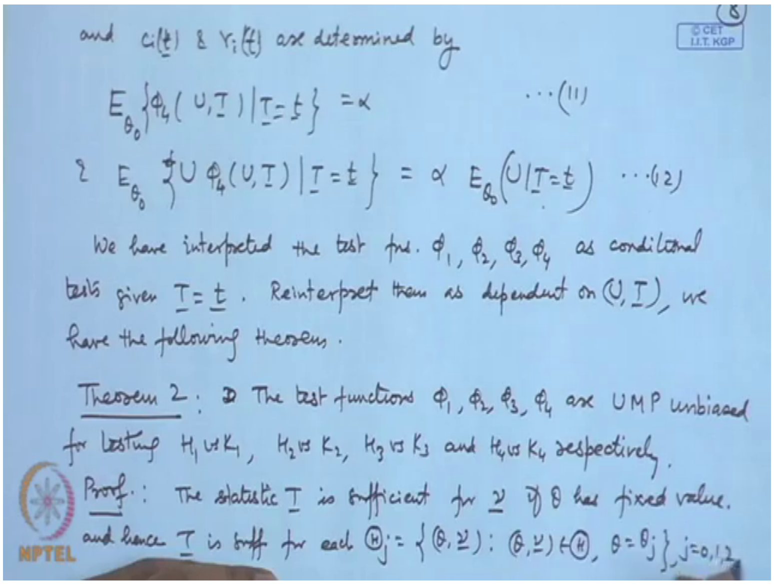
(Refer Slide Time: 27:33)



Now, for H 3 problem and X 3 versus K 3 problem and H 4 verses K 4 problem; we have seen in the one parameter exponential family we had UMP unbiased tests. So, if we consider the conditional here, conditional distribution and that is u given t; then for this again we will have the UMP unbiased test that will be the conditional test here.

So, for H 3 versus K 3 UMP unbiased test phi 3 is given by that is equal to 1 for u less than c 1 T or u greater than c 2 T it is equal to gamma I T if u is equal to c i t for i is equal to 1, 2 and it is equal to 0; when u is lying between c 1 and c 2, where once again these things are determined by expectation of theta i phi 3 U T is it given T is equal to t; this is equal to alpha for i is equal to 1, 2. For H 4 versus K 4 the UMP unbiased test that will be phi 4; actually phi 4 will be same as phi 3.

(Refer Slide Time: 30:41)



And ci's and gamma i's they are determined by expectation of theta naught phi 4 U T; given T is equal to t, it is equal to alpha and now you can see here that the size conditions are all conditional.

So, if I take the expectations; I will get the conditions without the conditional here. So, what we can say; we have interpreted the test functions phi 1, phi 2, phi 3, phi 4 as conditional tests given T is equal to t.

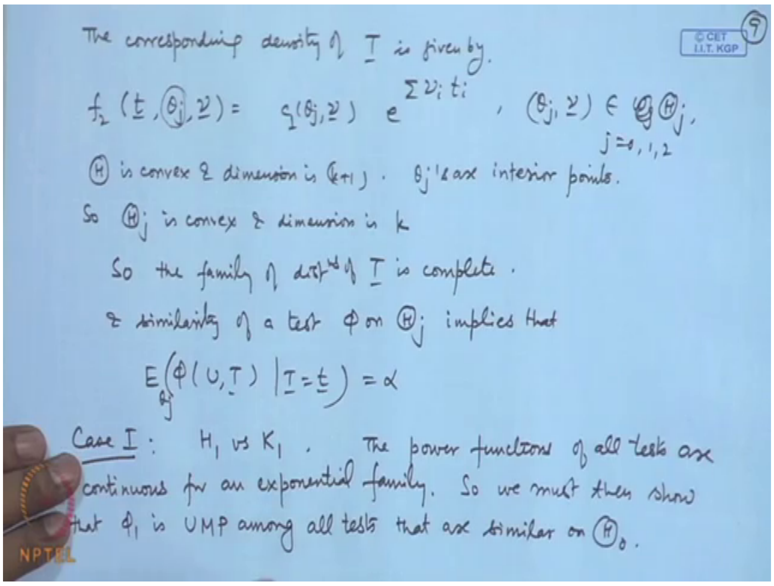
Now, we reinterpret them as dependent on U T we have the following theorem. So, I will call it theorem 2 say, this is regarding the UMP test here. Note here one point I have given the test functions phi 1 and phi 2 the conditional test says UMP and the phi 3 and phi 4 as UMP unbiased. But when I consider them as unconditional test all of these tests will become UMP unbiased; so, the statement is in the given theorem here. So, the test functions phi 1 phi 2; so phi 1 is defined by these two conditions, phi 2 is defined by these conditions, phi 3 is defined by these conditions etcetera.

So, the test functions phi 1, phi 2, phi 3, phi 4 are UMP unbiased for testing H 1 versus K 1, H 2 versus K 2, H 3 versus K 3 and H 4 versus K 4 respectively, under the given set up. That means, the joint distribution of the initial random variables was a multi parameter exponential family. In fact, it was a k plus 1 dimensional distribution and the distribution of u and t; the sufficient statistics were also in the exponential family.

In that case we will have these as UMP unbiased tests. Let me sketch a proof of this of course, for the tale proof you may look at the book of Neyman here. The statistic T is sufficient for nu if theta has fixed value. So, this you can easily see; if I am writing down the distribution in this one if I fix theta, then this part will become random variable here it is dependent upon the variable only you will have only e to the power sigma nu i T i x that will show that T 1, T 2, T k is sufficient for the parameters nu 1 nu 2 nu k.

So, and hence we can say that T is sufficient for each; this is we will call subsets of the parameter spaces, theta nu where theta nu belongs to script theta and theta has been fixed sum. So, this is for j is equal to 0 1, 2; now these points we have considered because in all these tests we are having the cutoff points in the hypothesis as theta naught theta 1 and theta 2. So, at least for those points the sufficiency of T is maintained here.

(Refer Slide Time: 36:27)



The corresponding density of T is given by; so we can use some notation say f 2 for the distribution of t and of course, theta j will be coming it is fixed here; nu that is equal to c theta j nu, e to the power sigma nu i, t i. So, this will be some coefficient let me put it c 1 here where theta j nu this belongs to omega sorry theta j for j is equal to 0, 1, 2. Now, we have assumed theta is convex that is assumed and dimension is k plus 1 and we have assumed that theta j's are interior points; so this theta j is convex and dimension is k.

So, basically what we are done is we have taken one hyperplane there theta is equal to theta j there. So, the family of distributions of T; so the family of distributions of T is complete and similarity of a test phi on theta j; this will imply that expectation of phi U T; given T is equal to t, that will be equal to alpha for theta j. Alright; so this is the general description so far.

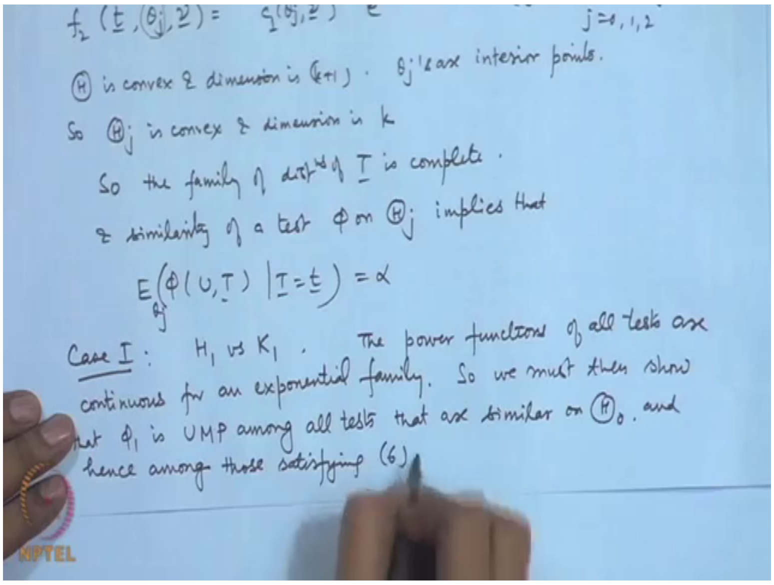
We have derived the conditional tests as the UMP tests; now in the theorem I am claiming that for the unconditional problem, the tests phi 1 phi 2 phi 3 phi 4 UMP unbiased. So, in order to prove this one we will take help of the theorem 1 which I have given today; that means, the test with the Neyman structure. And the result which I have given for the similar tests in the previous lecture; so we will use both of these results here.

Now first thing that we notice here is the structure of the multi parameter exponential family here. So, far the fifth two values of theta as theta naught theta 1 or theta 2; T is a complete and sufficient statistic here. So, if I have a test function phi to be similar then we should have expectation of phi U T; given T is equal to t is equal to alpha. So, now, let us consider this phi 1 phi 2 phi 3 and phi 4 separately. So, let us take case 1 that is the testing problem H 1 versus K 1 ok.

So, another point that studies lemma which we want to use the power functions of the test functions whatever we are considering must be continuous. Since we are dealing with the exponential families the power functions are basically bounded; therefore, integrable functions and therefore, the expectations of the test functions must be continuous.

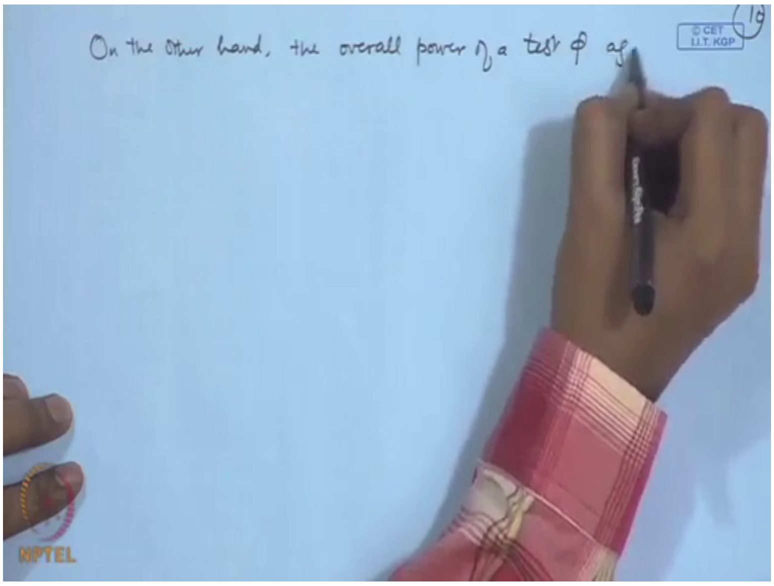
So, let me give a general statement the power functions of all tests are continuous for an exponential family. So, we must then show that phi 1 is UMP among all tests that are similar on theta naught.

(Refer Slide Time: 41:19)



And hence among those which are satisfying condition 6; the condition 6 let me repeat here this the condition for the Neyman structure this condition here.

(Refer Slide Time: 41:51)



On the other hand, the overall power of a test phi; I will be carrying out in the next lecture.