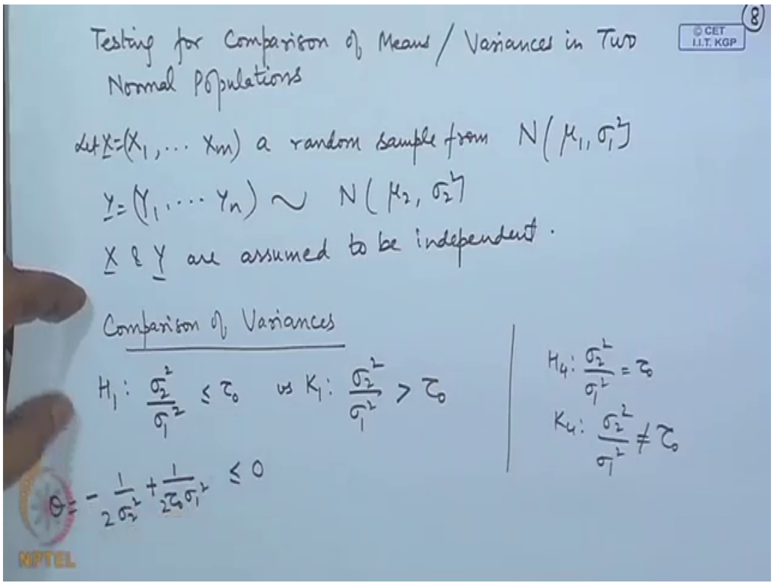
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 50
Unbiased Test for Normal Populations – IV

Now I will consider two sample problems.

(Refer Slide Time: 00:25)

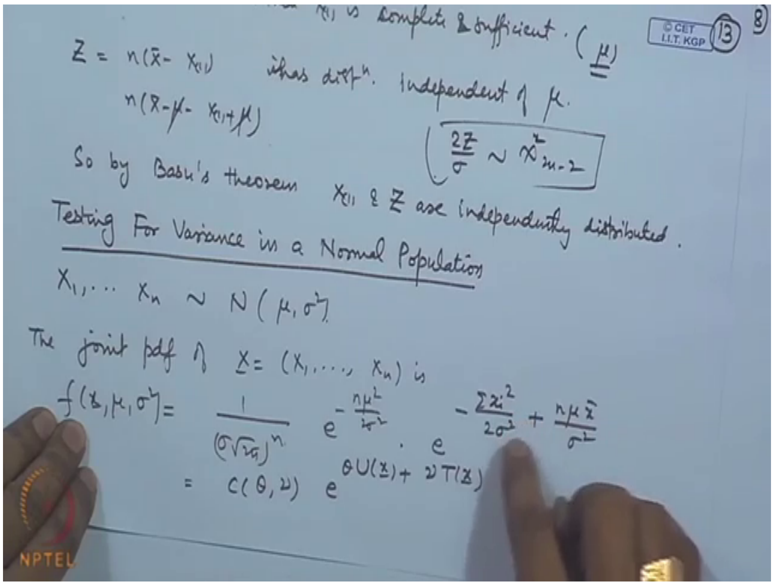


That means we consider testing for comparison of means and variances in two normal populations. Remember, here our aim is to consider the UMP unbiased test. Therefore, we will write the joint density in the form of a multi parameter exponential family. So, let us consider say X 1, X 2, X m, a random sample from normal mu 1 sigma 1 square ; and Y 1, Y 2, Y n a random sample from normal mu 2, sigma 2 square.

So, I assume that X and R are independently collected; the two random samples are independent. Now let us take the first problem, comparison of say variances ; that means, I need to test something like sigma 1 square is equal to sigma 2 square against sigma 1 square is not equal to sigma 2 square. Sigma 1 square is less than or equal to sigma 2 square or sigma 1 square is greater than sigma 2 square etcetera. That is the comparison.

Now, keeping into view that the sigma parameter in the normal distribution, is actually a scale parameter. When we are considering the scale, it will be beneficial if I consider the ratio; that means, the hypothesis sigma 1 square is equal to sigma 2 square. I can express as sigma 2 square by sigma 1 square is equal to 1 or sigma 2 square by sigma 1 square. Sigma 1 square by sigma 2 square is equal to 1 etcetera. Similarly, if I say sigma 1 square is less than or equal to sigma 2 square, I can express it as sigma 2 square by sigma 1 square less than or equal to 1, greater than or equal to 1 etcetera ; that means, I take the ratio. Because in the form of the density, that I have expressed in the normal distribution.

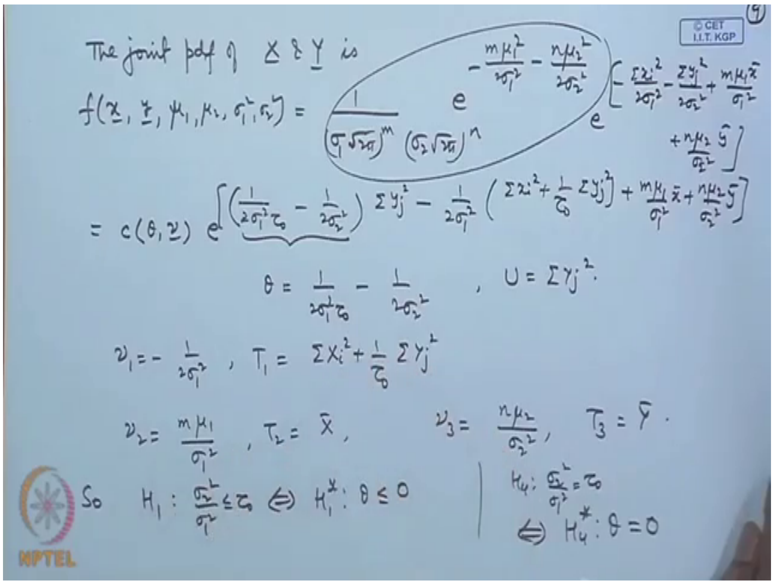
(Refer Slide Time: 03:00)



You can see; here this is appearing in the denominator, so taking the difference will not be easy. In fact, it will be extremely inconvenient. Therefore, if I take the ratio, I will be able to express it in a proper form. So now, let us look at the hypothesis testing problem such as say H 1. I will write slightly general hypothesis, sigma 1 is sigma 2 square by sigma 1 square less than or equal to some say tau naught against say K 1 sigma 2 square by sigma 1 square greater than tau naught and similarly, we may consider say H 4 sigma 2 square by sigma 1 square is equal to tau naught against K 4 sigma 2 square by sigma 1 square is not equal to tau naught.

So, in order to write it in the form of a required desired multi parameter exponential family, let us write down the joint.

(Refer Slide Time: 04:04)



The joint pdf of X and Y; so, that is equal to f x, y mu 1 mu 2 sigma 1 square sigma 2 square, that is equal to so, there will be some terms. Basically, if I write it in full, it will become 1 by sigma 1 root 2 pi to the power m. Sigma 2 root 2 pi to the power n, e to the power minus m mu 1 square by 1 sigma 2 square minus n 2 square by 2 sigma 2 square and then e to the power minus sigma x i square by 2 sigma 1 square minus sigma y j square by 2 sigma 2 square plus m mu 1 x bar by sigma 1 square plus n mu 2 y bar by sigma 2 square here.

Straightforwardly, this is a four parameter exponential family as I have described in the example in one of the previous lecture where I was applying the Gosset's theorem. But the problem here is that, if I straightforwardly write it as with four parameter exponential family, I can test about either sigma 1 square or about sigma 2 square or about mu 1 or about mu 2, which is not our aim.

So, we reparameterize it; so, this coefficients I put it as c of theta, nu and this I write as e to the power 1 by 2 sigma 1 square tau naught minus 1 by 2 sigma 2 square into sigma y j square minus 1 by 2 sigma 1 square sigma x i square plus 1 by tau naught sigma yj square plus m mu 1 by sigma 1 square x bar plus n mu 2 by sigma 2 square y bar.

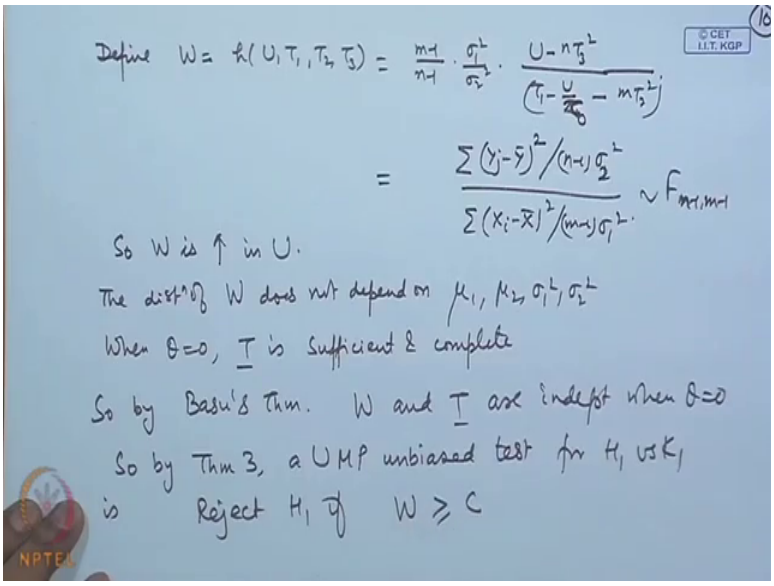
Here, I am taking theta to be 1 by 2 sigma 1 square tau naught minus 1 by 2 sigma 2 square. So, U is equal to sigma Y j square. Nu 1 is equal to minus 1 by 2 sigma 1 square. So, T 1 is equal to sigma X i square plus 1 by tau naught sigma Y j square. Nu 2 is equal to m mu 1 by sigma 1 square. So, T 2 is equal to X bar. Nu 3 is equal to n mu 2 by sigma 2 square so, T 3 is equal to Y bar.

So, if I am considering the testing problem say H 1 that is, sigma 2 square by sigma 1 square less than or equal to tau naught, we can express it as 1 by tau naught sigma 1 square and I divide here. So, 1 by sigma 2 square and I put 2 here. So, this will become plus this will become minus. So, if I read take sigma 2 square in the denominator here and I bring it to the left hand side, this becomes less than or equal to 0.

So, this theta term that I am having this is equivalent to less than or equal to 0. So, this will be equal to theta is equal to 0. So, H 1 sigma 2 square by sigma 1 square less than or equal to tau naught. This is equivalent to say H 1 star theta less than or equal to 0. And similarly, this H 4 that is sigma 2 square by sigma 1 square is equal to tau naught. this is equivalent to H 4 star theta is equal to 0. So, we can derive the UMP unbiased tests.

However, those UMP unbiased tests, if I use theorem 2, there will be conditional tests of U given T 1, T 2, T 3; U given T 1, T 2, T 3 U is this. So, this conditional distribution will be quite complicated even when I am taking say sigma 2 square by sigma 1 square is equal to tau naught. So, this will be quite complicated form here. So what we do? We apply theorem 3. To apply theorem 3, I define function W as an increasing function of U and the distribution of W should be free from T 1, T 2, T 3.

(Refer Slide Time: 09:41)

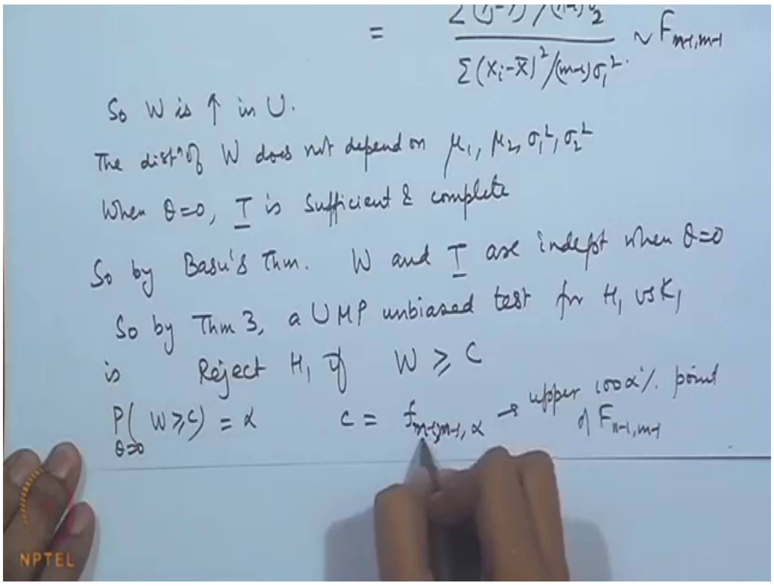


So, we consider like this. Define W is equal to h of U, T 1, T 3 that is equal to m minus 1 by n minus 1 sigma 1 square by sigma 2 square U minus n T 3 square divided by T 1 minus U by delta tau naught minus m T 2 square. It is actually equal to sigma Y j minus Y bar square divided by n minus 1 sigma 2 square divided by sigma X i minus X bar whole square divided by m minus 1 sigma 1 square. So, W is increasing in U. The distribution of W does not depend on mu 1, mu 2, sigma 1 square, sigma 2 square. Actually this is F distribution on m n minus 1 m minus 1 degrees of freedom.

So, since the distribution of W does not depend upon the parameters and when theta is equal to 0 T is sufficient. When theta is equal to 0, T is that is, T, T 1, T 2, T 3 that is sufficient and complete. So, by an application of Basu's theorem W and T are independent when theta is equal to 0. Remember that for testing H 1 and H 4 both, I need the independence when theta is equal to 0, that is satisfied here.

So, by theorem 3 a UMP unbiased test for H 1 versus K 1 is Reject H 1 if W is greater than or equal to C and probability of W greater than or equal to C when theta is equal to 0 that is equal to alpha.

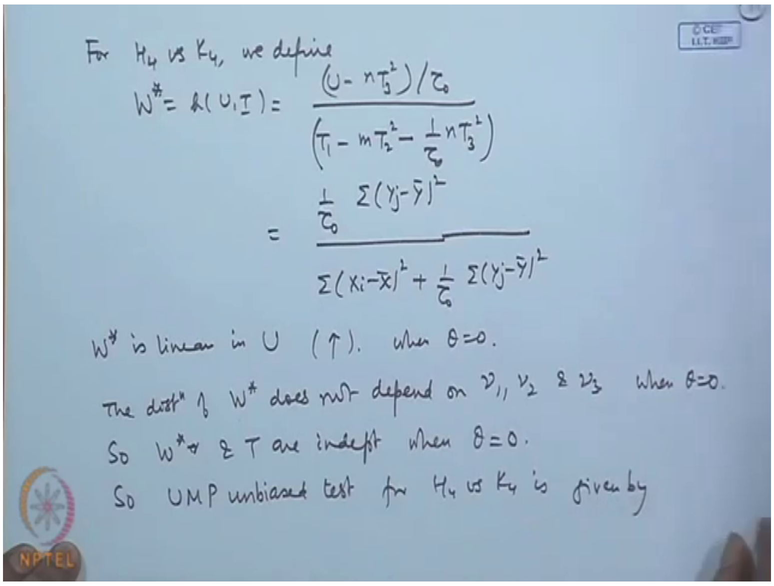
(Refer Slide Time: 12:16)



So, C is given by the point on f distribution, that is upper 100 alpha percent point of F distribution on n minus 1, m minus 1 degrees of freedom.

Now, this should be n minus 1 m minus 1. Firstly, it is n minus 1 in the numerator because the way I have defined here, this function in the numerator the degrees of freedom for chi square is n minus 1 and in the denominator it is m minus 1. So, we are getting an exact test here using the theory of unbiased tests here for testing the comparison of the variances like I have put here sigma 2 square less than or equal to sigma 1 square here. If I consider say sigma 1 square is equal to sigma 2 square versus the inequality then this function will not be useful because this is not linear in tau linear in U. So, we need to define another function which is related to this.

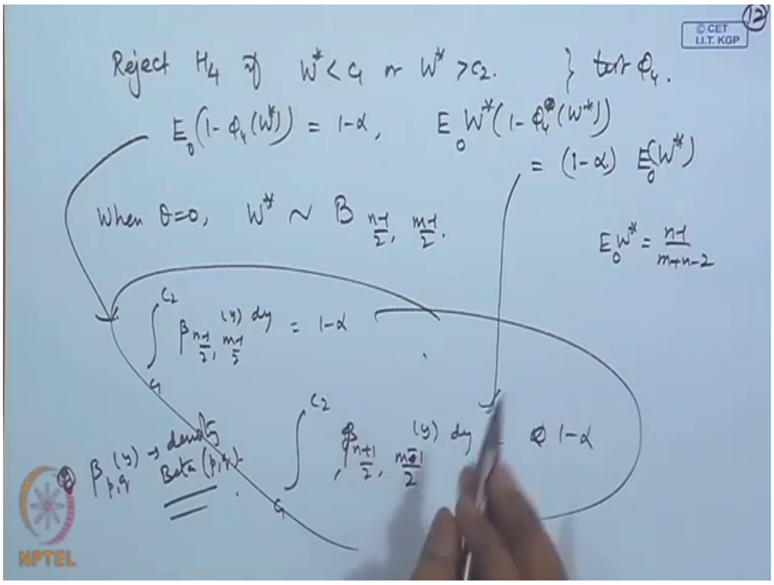
(Refer Slide Time: 13:32)



For H 4 versus K 4, we define say W star is equal to h U, T that is equal to U minus n T 3 square divided by tau naught divided by T 1 minus m T 2 square minus 1 by tau naught n T 3 square. That is equal to 1 by tau naught sigma Yj minus Y bar square divided by sigma X i minus X bar square plus 1 by tau naught sigma Yj minus Y bar square.

Now W star is linear in U and of course it is increasing. Then, theta is equal to 0, the distribution of W star does not depend on nu 1, nu 2 and nu 3. The parameters that I defined here when I wrote it in this particular fashion, nu 1 was minus 1 by 2 sigma 1 square nu 2 was this and nu 3 was n mu 2 by sigma 2 square. So, that is true here that the distribution of this thing does not depend on this when theta is equal to 0. So, W star and T are independent when theta is equal to 0.

(Refer Slide Time: 15:41)



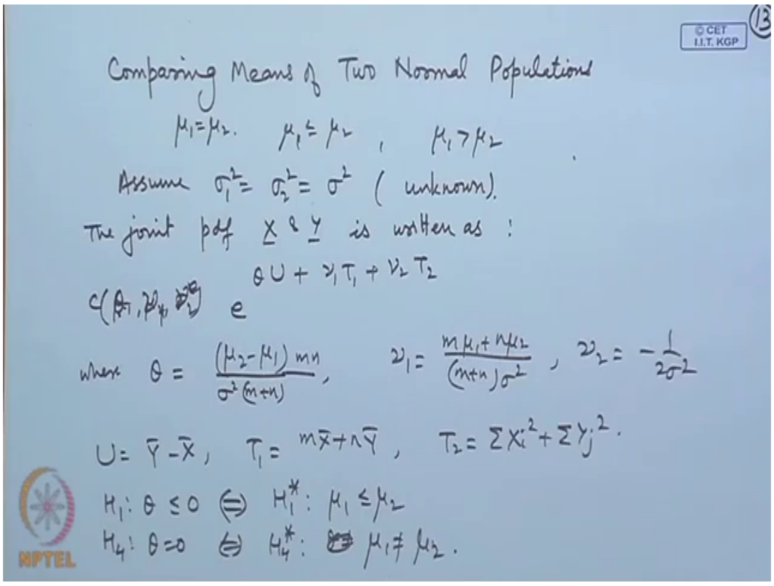
So, UMP unbiased test for H 4 versus K 4 is given by, Reject H 4 if W is less than W star is less than C 1 or W star is greater than C 2. Now to determine C 1 and C 2 the conditions are; let me call this test say phi 4. Then you have expectation of 1 minus phi 4 W star at theta is equal to 0 that is equal to 1 minus alpha and expectation of W 1 minus phi 4 star phi 4 W star that is equal to 1 minus alpha times expectation of the W star at 0.

Now, when theta is equal to 0 the distribution of W star is actually a beta distribution on n minus 1 by 2 and m minus 1 by 2 degrees of freedom. That is, this function when I am assuming sigma 2 square by sigma 1 square is equal to tau naught, then this is actually having because it is actually in the numerator a chi square divided by sum of 2 chi squares and they are independent. So, this is a beta distribution actually, n minus 1 by 2 and m minus 1 by 2.

So, this condition is actually then determined from if I write the density of beta, say beta n minus 1 by 2, m minus 1 by 2 y dy. Then from C 1 to C 2 this should be equal to 1 minus alpha and the second condition, we can simplify because expectation of W star is actually n minus 1 by m plus n minus 2. If we utilize that, then this condition can be transformed to g n plus 1 by 2 m plus 1 by sorry m minus 1 by 2 y dy from C 1 to C 2 is equal to alpha, 1 minus alpha.

Now, this is actually the density of; that means, I am using this beta say pq y, this is density of a beta distribution with parameters p and q. Then these 2 conditions can be used to determine the coefficients C 1 and C 2. For convenience, we are generally taking alpha by 2 point and 1 minus alpha by 2 point on the f m minus 1 n minus 1 degrees of . Now, let me consider comparison of two means of normal populations.

(Refer Slide Time: 18:55)



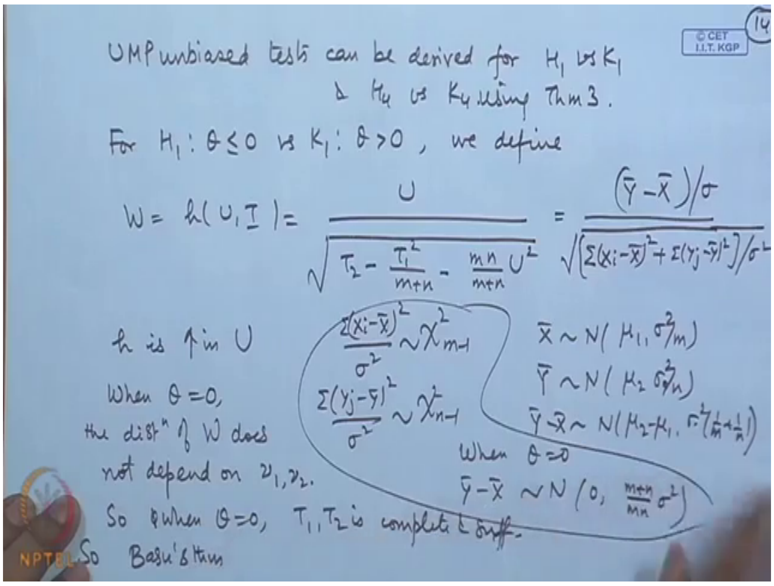
Comparing means of two normal populations; it means we want to test whether mu 1 is equal to mu 2, mu 1 less than or equal to mu 2 or mu 1 is greater than. We assume here that sigma 1 square is equal to sigma 2 square is equal to sigma square, this is of course, unknown here, the joint density function of X and Y as we wrote earlier.,

Now, this has to be written for modified form; that means, when sigma 1 square is equal to sigma 2 square, the terms will be combined. Let me go back to the original form here, in this term sigma 1 square, sigma 2 square are same. So, here sigma x i square and sigma y j square will get combined and this one also will have common denominator so, this term will also get combined. Now when they get combined and I want to get mu 1 minus mu 2 for comparison of mu 1 mu 2. So, we rewrite in the following fashion.

So, let me give this form here is written as because I am not writing the original term again and again. I have already written it several times. So, I am considering this a function of mu 1, mu 2 and sigma square e to the power theta U plus nu 1 T 1 plus nu 2 T 2 where, so this is theta nu now. Nu 1, nu 2 theta nu 1, nu 2, where theta I am defining to be mu 2 minus mu 1 divided by sigma square m plus n into mn. Then, nu 1 is equal to m mu 1 plus n mu 2 divided by m plus n sigma square nu 2 is equal to minus 1 by 2 sigma square. U is equal to Y bar minus X bar, T 1 is equal to mX bar plus nY bar, T 2 is equal to sigma X i square plus sigma Yj square.

So, if I consider the hypothesis H 1 theta less than or equal to 0, this is equivalent to H 1 star. Say mu 1 less than or equal to mu 2. Similarly if I consider theta is equal to 0 this is equivalent to theta naught equal to that is mu 1 not equal to mu 2. So, once again we have the UMP unbiased tests here.

(Refer Slide Time: 22:09)



So, UMP unbiased tests can be derived for H 1 versus K 1 and H 4 versus K 4 using theorem 3.

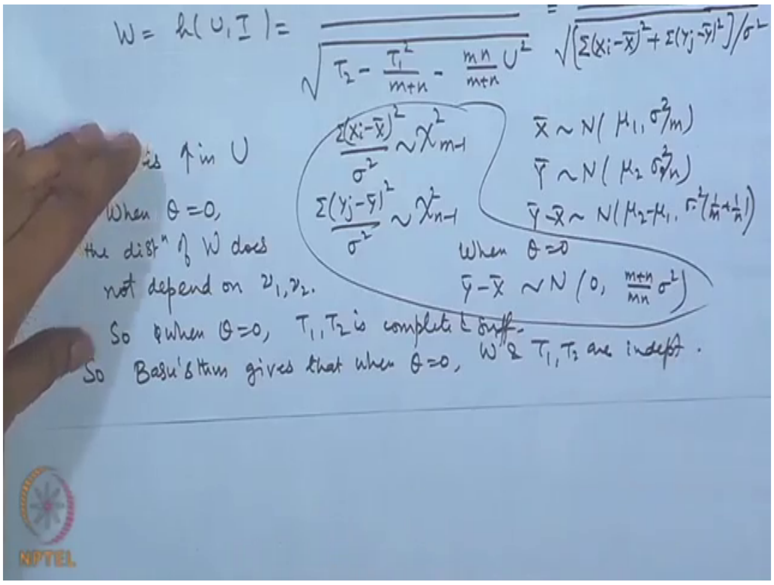
So, if I consider the problem H 1 theta less than or equal to 0 versus K 1 theta greater than 0 we define the function W that is as. U divided by T 2 minus T 1 square by m plus n minus mn by m plus n U square. Now this function is actually Y bar minus X bar divided by square root of sigma X i minus X bar whole square plus sigma Yj minus Y bar whole square;. h is increasing in U that can be easily checked here when theta is equal to 0.

Now, let us look at the distribution theory here. What are the distributions here? X bar will follow normal mu 1 sigma square by m. Y bar follows normal mu 2 sigma square by n. The distribution of Y bar minus X bar is normal mu 2 minus mu 1 sigma square 1 by m plus 1 by n. So, when theta is equal to 0 Y bar minus X bar will follow normal 0, m plus n by mn sigma square.

The distribution of Y bar minus X bar is free from theta also the distributions of sigma X i minus X bar whole square by sigma square, this follows chi square distribution on m minus 1 degrees of freedom. The distribution of sigma Yj minus Y bar whole square by sigma square that follows chi square distribution on n minus 1 degrees of freedom. So, these are all having distributions independent of theta. So when theta is equal to 0, the distribution of W does not depend on; what is nu 1 and nu 2 here? Mu 1 involves mu i's and sigma square nu 2 involve sigma square.

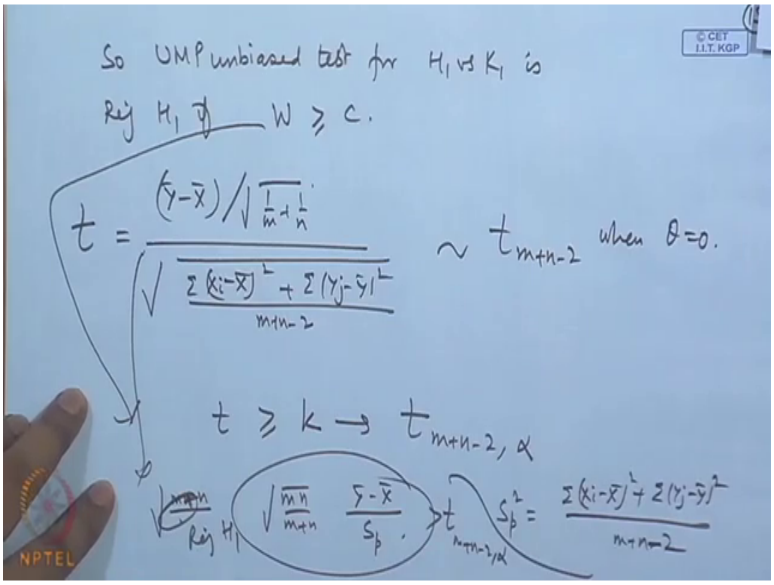
If we consider this function, if I take divided by sigma and here, this thing divided by sigma square, so you can easily see that the distribution of W is also free from sigma. So, this does not depend on nu 1 and nu 2. So, when theta is equal to 0 T 1 T 2 is a complete sufficient statistic here. So, when theta is equal to 0. T 1, T 2 is complete and sufficient. So, Basu's theorem is applicable here.

(Refer Slide Time: 26:20)



Basu's theorem gives, that when theta is equal to 0, W and T 1, T 2 are independent and therefore, we are in a position to write down the UMP.

(Refer Slide Time: 26:36)



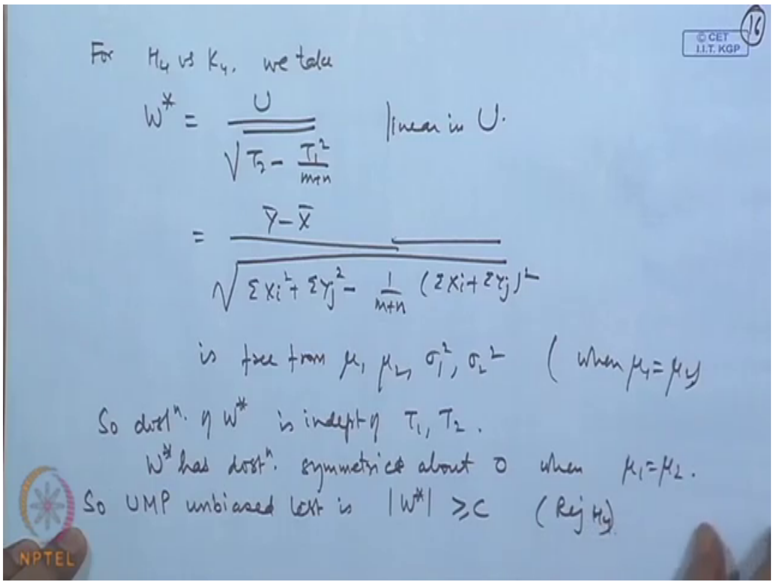
So, UMP unbiased test for H 1 versus K 1 is Reject H 1 if W is greater than or equal to C. Now let us look at the distribution part of W.

So, we have already seen here that Y bar minus X bar this follows normal 0, m plus n by m plus n by mn sigma square when theta is equal to 0. So, we adjust this term here, we get here Y bar minus X bar divided by root 1 by m plus 1 by n divided by square root sigma Xi minus X bar square plus sigma Yj minus Y bar square divided by m plus n minus 2.

Let us define this as say T. Then, this is T distribution on m plus n minus 2 degrees of freedom then theta is equal to 0. Therefore, I can choose this term W as t greater than or equal to k and t will be m plus n minus 2 alpha. The upper handed alpha percent point of the t distribution on m plus n minus 2 degrees of freedom. So you can see here, this is the pooled sample variance formula here that we are getting here.

In fact, we can write this term as root m n m plus n by root mn by m plus n Y bar minus X bar divided by S p, where S p square is actually sigma X i minus X bar whole square plus sigma Y j minus Y bar whole square divided by m plus n minus 2. So, we are comparing the test is Reject H 1 if this is greater than t m plus n minus 2 alpha. We have an exact UMP unbiased test.

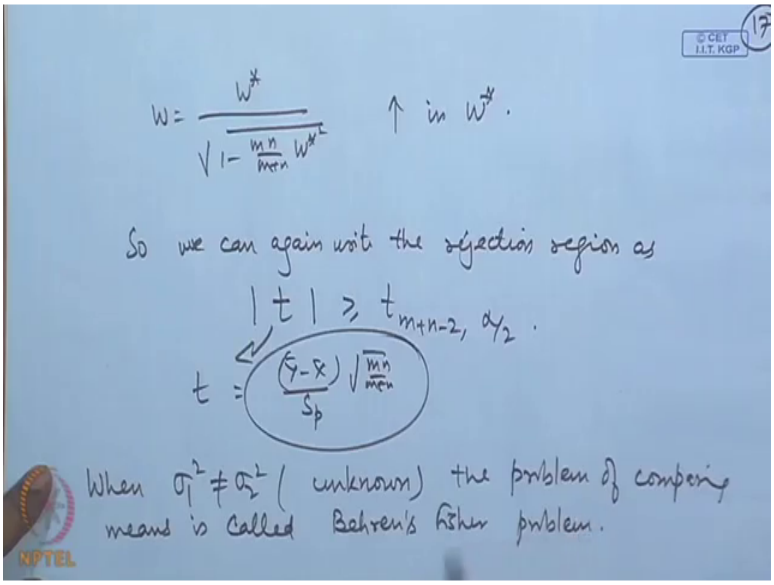
(Refer Slide Time: 29:05)



Now, let us also look at the H 4 problem. For H 4 versus K 4, this function will not be useful because this function is not linear in U. So, I modify. We can consider W star is equal to U divided by square root T 2 minus T 1 square by m plus n. This is linear in U and this function if you see, actually it is Y bar minus X bar divided by square root sigma Xy square plus sigma Yj square minus 1 by m plus n sigma X i plus sigma Yj whole square.

So, this term is free from mu 1, mu 2, sigma 1 square, sigma 2 square when mu 1 is equal to mu 2. So, distribution of W star is independent of T 1, T 2. So, and also W star has distribution symmetric about 0 when mu 1 is equal to mu 2. So, UMP unbiased test is based on W star greater than or equal to C, that is the Rejection.

(Refer Slide Time: 31:05)



Now, what we can do? We can consider say W that is, W star divided by root 1 minus mn by m plus n W star square. This is increasing in W star. So, we can again write the rejection region as modulus of t which I define in the previous case. I am sorry this is t. So, in terms of this t which was having n plus n minus 2 degrees of freedom, we can write here t greater than or equal to t m plus n minus 2 alpha by 2 where this t is actually Y bar minus X bar root mn by m plus n divided by the S p.

In this today's lecture, I have been able to give UMP unbiased tests for parameters testing problem for one sample problem from normal distributions. Also, for the comparison of parameters for two sample problems; that means, when we are having two normal populations then for comparison of the variances I am having UMP unbiased test for the comparison of the means I am having unbiased test. UMP unbiased test, one point which should be noted, when I was considering comparison of the means I have taken sigma 1 square is equal to sigma 2 square.

When I gave the derivation it was clear that it was helping because in the denominator and the exponent in the joint density that sigma square term was getting combined. So, the terms were adjusted ; however, if I do not make this assumption that sigma 1 square is equal to sigma 2 square then, I will not be able to adjust these terms. So, when sigma 1 square is not equal to sigma 2 square, that is unknown the problem of comparing means this is called Behren's Fisher problem.

In fact, it can be shown that there is no exact test, no exact UMP unbiased test for this problem. Approximate tests are available which are actually you can say approximately t distribution is there which are provided by Smith, Satterwhite and Welch etcetera. We are not discussing it here because in this particular course I have actually given you the derivations of the exact test. The theory that I have developed is using the name and Pearson theory.

The concept is started from the simple versus simple hypothesis. We had considered the most powerful test for a fixed size, then we extended that theory to UMP unbiased test. Initially, we consider the families of distribution with monotone likelihood ratio; later on we restricted attention to one parameter exponential family, for I define for a special kind of hypothesis H 1, H 2, H 3 and H 4 for H 1 and H 2 you are having UMP test for H 3 and H 4 we got UMP unbiased test.

Whereas, when we considered multi parameter exponential family, all of the testing problems reduce to we are getting the UMP unbiased test. In the next lectures, I will introduce another method of derivation of the test called Likelihood Ratio Testing Procedures which is also a natural procedure I will be describing those tests in detail in the following lectures.