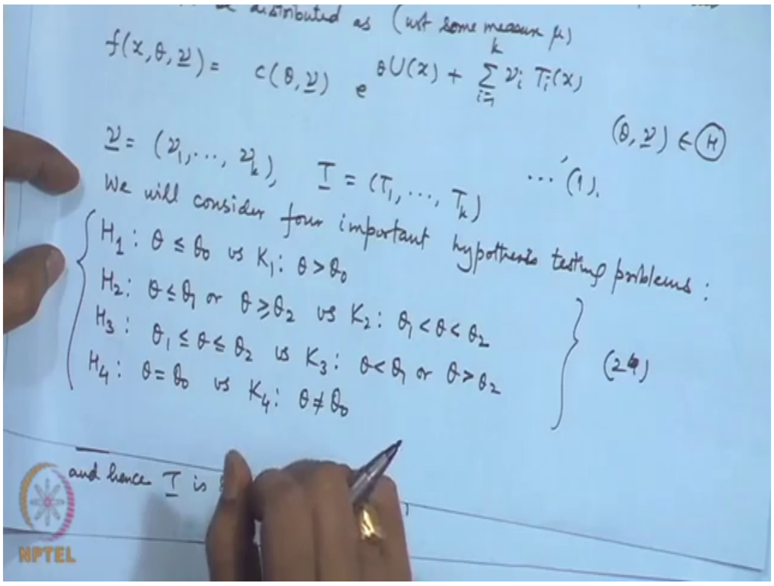
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 45
Applications of UMP Unbiased Tests – I

So, in the previous lecture I have described in detail how for the multiparameter exponential families for the four types of hypothesis testing problems H 1 versus K 1, H 2 versus K 2, H 3 versus K 3 and H 4 versus K 4 we have UMP unbiased test.

(Refer Slide Time: 00:45)

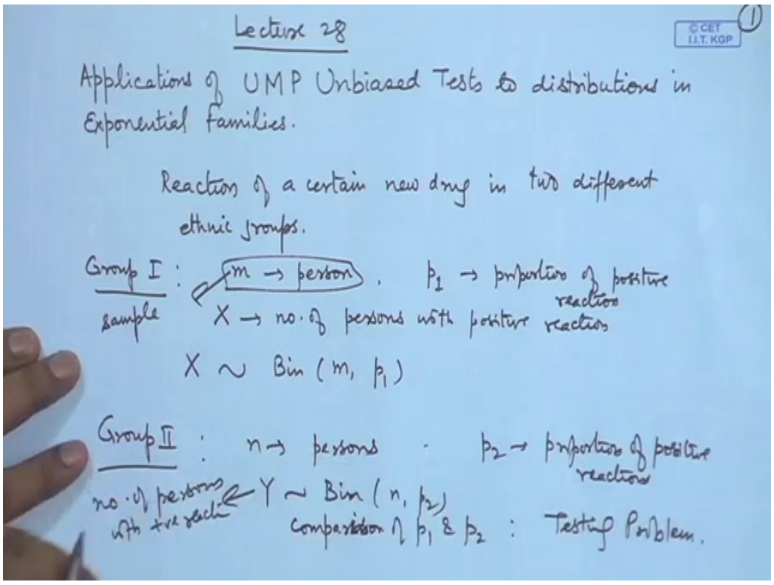


So, the hypothesis testing problems let me show again H 1 theta less than or equal to theta naught against theta 1 theta greater than theta naught. So, this is a one sided testing problem if a UMP unbiased test is here for dual problem also it will exist. Now for H 2 we have for H 3 and for H 4 H 4 is actually the null hypothesis point hypothesis whereas, the alternative hypothesis is two sided.

For all the cases UMP unbiased test can be found; the method that I described was, firstly, we derived the conditional test and then we derive the unconditional test. The conditional test could have been for H 1 and H 2 UMP test and for H 3 and H 4 there were UMP unbiased, but when we considered unconditional then all the test became UMP unbiased.

So, I will demonstrate certain popular testing problems for example, you are considering parameters of binomial distributions, Poisson's distributions, normal distributions; so these are all you can say popular distributions. Suppose I have two proportions; so, it could be like we are having reaction to certain measure or reaction to certain drug.

(Refer Slide Time: 02:13)

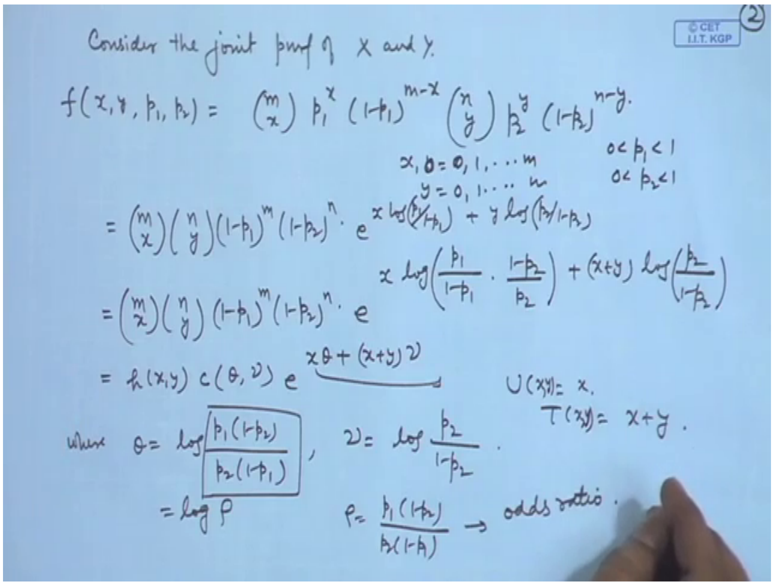


So, let us consider reaction of a certain new drug in two different ethnic groups ok. So, suppose for the first group; group I; we considered that suppose m persons were tested and p 1 is say the proportion of positive reaction say. So, we took the observation and X is the observation number of persons with positive reaction in the sample; if I am considering a sampling of m persons; then here we can write the model as X follows binomial m p 1.

Similarly in the group II, you can consider say n persons are there a sample of n persons is considered and p 2 is the proportion of say positive reactions and then Y follows binomial n p 2; where y is the number of persons out of n with positive reactions. So, the problem could be compare comparison of p 1 and p 2; this could be a testing problem. For example, we may like to check whether in the two populations the people are you can say proportion of the people with positive reactions is the same or not ok. So, we may like to test whether p 1 is equal to p 2 or p 1 is less than p 2 or p 1 is greater than p 2 or simply p 1 is not equal to p 2 etcetera various type of comparison a statements can be given.

Let us see we will demonstrate here that using this procedure that I developed in the last class it is applicable here and we will be able to derive an exact UMP unbiased test for these problems.

(Refer Slide Time: 05:01)



So, consider the joint probability mass function of X and Y. So, x, y, p 1, p 2 that is equal to m c x p 1 to the power x 1 minus p 1 to the power m minus x n c y p 2 to the power y 1 minus p 2 to the power n minus y. Here p 1 p 2 both lie between 0 and 1 and this x and y values; x is from 0 1 to m and y is equal to 0 1 to n. Now we have written the distribution in this particular fashion it does not look like in the exponential family form. Our aim is to express it in the form of multiparameter exponential family that I described of this nature; we will show it here that it can be done.

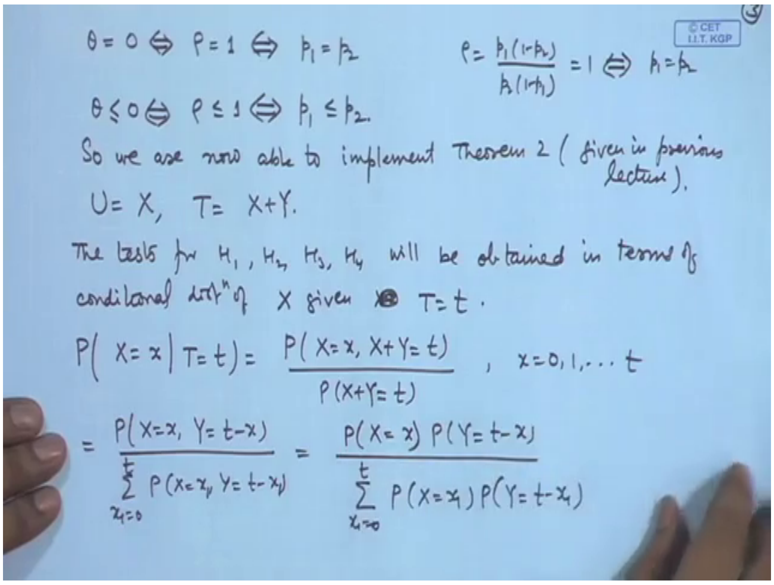
So, we write it as say m c x n c y 1 minus p 1 to the power m 1 minus p 2 to the power n; e to the power x log p 1 divided by 1 minus p 1 plus y log p 2 divided by 1 minus p 2. Now, this you can easily see this is a two parameter exponential family here x and this is you can say theta and this you can say nu and it is a that form. However, if I write this particular form I can test either about p 1 or about p 2 whereas, our original m was to test about a comparison between p 1 and p 2.

So, here little bit of you can say reparameterization is required what we do we write it in this form; m c x n c y 1 minus p 1 to the power m 1 minus p 2 to the power n; e to the power x log of p 1 by 1 minus p 1 1 minus p 2 by p 2 plus x plus y; log of p 2 by 1 minus p 2. Let me call this as sum h of x y; this function is c of our parameters we will rename this thing. So, I am calling it theta nu e to the power x theta plus x plus y nu where this theta is actually log of p 1 into 1 minus p 2 divided by p 2 into 1 minus p 1; nu is equal to log of p 2 divided by 1 minus p 2. And then this U x we can take to be x and T x we can take to be; so U x y; T x y that is equal to x plus y.

We give some names here this quantity; this particular quantity we can call it log of sum rho for example, rho is equal to p 1 into 1 minus p 2 divided by p 2 into 1 minus p 1. So, this rho this is also called odds ratio; that means, p 1 by 1 minus p 1 is odds in favor of events happening with x and p 2 by 1 minus p 2 is the odds in favor of events happening with y.

So, therefore, this ratio is also called odds ratio and we are able to actually write down the parameter theta in terms of this. Notice this manipulation is actually helping us to write a function which is involving both p 1 and p 2. Now let us see whether this is actually helpful in testing a hypothesis regarding comparison between p 1 and p 2.

(Refer Slide Time: 09:39)



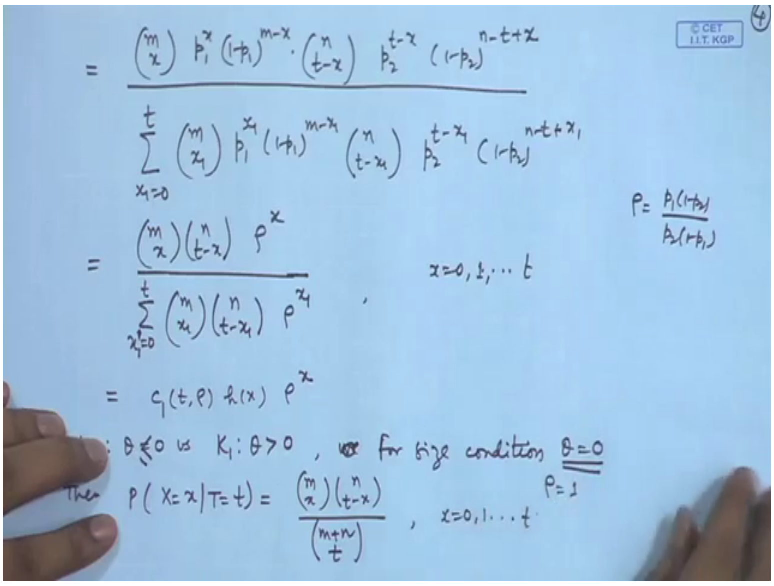
Now, let us see if I consider say theta is equal to 0 theta is equal to 0 is equivalent to same rho is equal to 1 and rho is equal to 1 is equivalent to same p 1 is equal to p 2 because rho is this quantity. So, if I say rho that is equal to p 1 into 1 minus p 2 divided by p 2 into 1 minus p 1 is equal to 1. Then if I write down and expand p 1 p 2 will cancel out you will get p 1 is equal to p 2.

So, if I am considering equality of the two proportions; it is equivalent to hypothesis in terms of my parameter which is appearing here. And if I am looking at the point null hypothesis H 4 that is theta is equal to theta naught, then theta is equal to 0 is the corresponding hypothesis here; let us also see others. If I say theta less than or equal to 0 then this is equivalent to rho less than or equal to 1 and this is equivalent to p 1 less than or equal to p 2. So, the hypothesis of the nature H; H 1, H 2 and H 3 can also be written in the form of this.

So, we are now able to implement theorem 2 given in previous lecture. And I have already written that U is actually equal to X and T is equal to X plus Y. The tests for H 1, H 2, H 3, H 4 will be obtained in terms of conditional distribution of X given sun X plus given T is equal to t; so we derive this distribution now. Probability of say X equal to x given T is equal to t that is probability of X is equal to x X plus Y is equal to t divided by probability of X plus Y is equal to t; for x is equal to 0 1 to t.

This we can write as probability X is equal to x, Y is equal to t minus x divided by probability X equal to x, Y is equal to t minus x; let me write here x 1 and x 1 is equal to 0 to t. Since X and Y are independent here this becomes a product here; product of probability X equal to x probability Y is equal to t minus x divided by sigma probability of X equal to x 1 probability of Y is equal to t minus x 1; x 1 is equal to 0 to t.

(Refer Slide Time: 13:09)



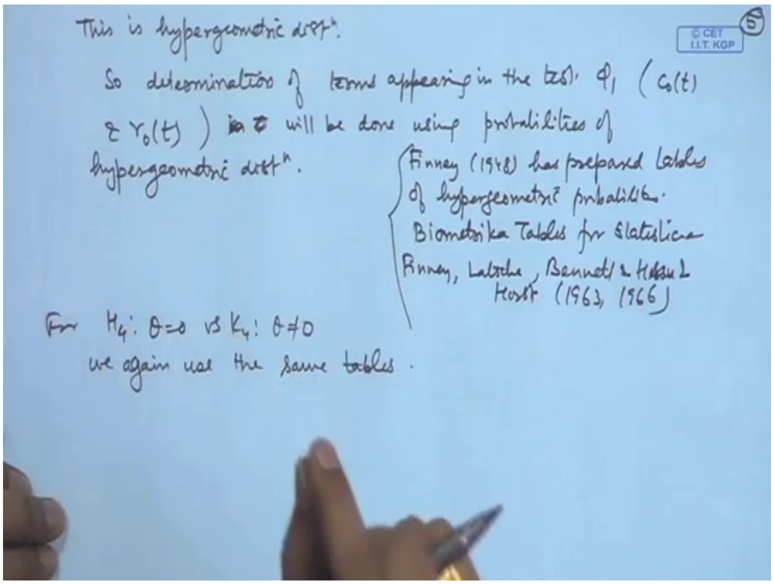
And the distributions of X and Y are known; so we can substitute these values and we will get this is equal to m c x p 1 to the power x 1 minus p 1 to the power m minus x; n c t minus x p 2 to the power p minus x, 1 minus p 2 to the power n minus t plus x; p 1 to the power x 1 1 minus p 1 to the power m minus x 1, n c t minus x 1 t minus x 1 mi minus p 2 the power n minus t 1 plus x 1; x 1 is equal to 0 to t.

Now, here certain terms will get cancelled out and we can write it in a slightly simplified form. So, we can use the notation that rho; rho as p 1 into 1 minus p 2 divided by p 2 into 1 minus p 1. So, if we use this notation this can be reduced to m c x; n c t minus x rho to the power x divided by sigma x prime x 1 is equal to 0 to t; m c x 1, m c t minus x 1 rho to the power x 1; x varies from 0 1 to t.

Now, if you look at this term here x 1 will be is the variable; so, this will be added up actually. So, this becomes finally, a function of t and rho, so this is a function of say t and rho and you are getting a function of x here; let me call it h x and then rho to the power x here. Suppose we consider say H 1 that is theta is equal to 0 versus theta less than or equal to 0 versus K 1 theta is greater than 0.

So, for size condition theta will be equal to 0; if I take theta is equal to 0 that is rho is equal to 1, then this term will go away. If this term goes away the denominator will be added up as m plus n c t; then this probability of X is equal to x given T is equal to t; that is becoming simply m c x; n c t minus x divided by m plus n c t; this is for x equal to 0 1 to t. And you can notice here that this is nothing, but a hypergeometric distribution; hypergeometric distribution.

(Refer Slide Time: 16:27)



So, determination of terms appearing in the test phi 1; so, let me take up the previous lecture slide we have given the exact form there in the theorem 2. The form of phi 1 if u is greater than or equal to c naught it is gamma naught u is equal to c naught and 0 if u is less than c naught and c naught and gamma naught were determined from this size condition.

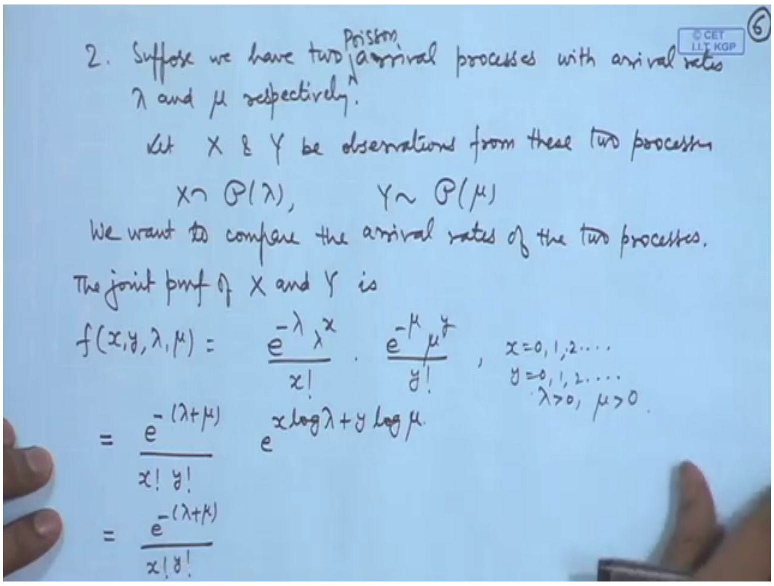
Now, here the conditional distribution of x given T has been obtained as the hypergeometric distribution; that means, c naught t that is c naught t and gamma naught t that is in we have given equation number 6, but I do not have to refer it to again. So, this is done using probabilities of hypergeometric distribution; Finney in 1948 has prepared tables of hypergeometric probabilities.

And they have also been published in biometrika tables for statisticians and also by tables of Finney, Latscha, Bennett and Hsu and Horst in 1963, 1966. So, these people have given the tables of the hypergeometric probabilities or you can say the critical points of the hypergeometric distribution which can be used in the determination of this.

For H 4 that is theta is equal to 0 versus K 4 theta naught equal to 0; we again use the same tables that is the distribution that will be used for theta is equal to 0; that will again become hypergeometric. So, here we have seen a problem of comparison of proportions in two binomial distributions can be solved using this theory of UMP unbiased tests here.

It is very interesting that we are able to get an exact test; that means, if we are fixing the level of significance then we are having an exact decision making procedure available here whether to accept say H 1 or reject H 1 or accept H 4 or reject H 4 etcetera; that means, whether the proportions are same or one is less than the other etcetera. I will give one more example here.

(Refer Slide Time: 20:27)



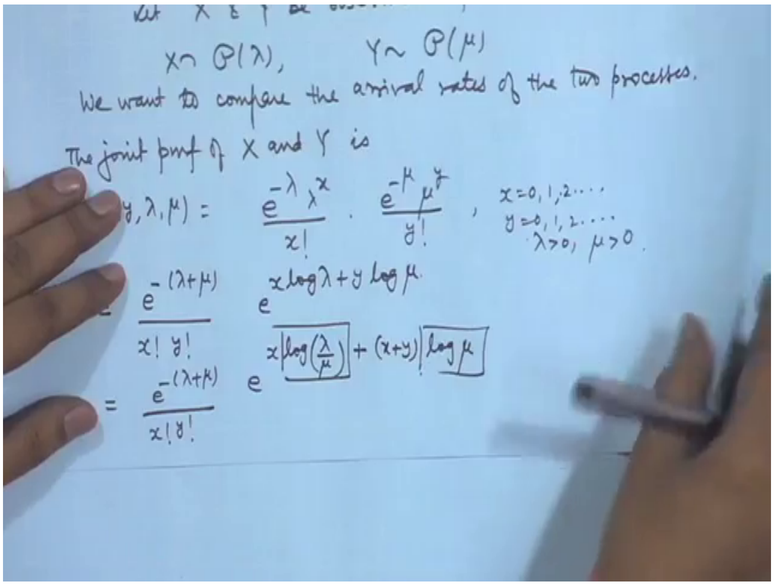
Suppose we have we have say two arrival processes and they are Poisson arrival prices with arrival rates say lambda and mu respectively. So, it could be like we are looking at the arrival rates at two service counters; where there are large number of service provider; so we may be looking at those things. We may be looking at the traffic through say a internet service provider we may be looking at the number of packets arriving etcetera there can be various such applications where we may have two arrival processes.

Let X and Y be observations from these two processes ok; that means, I am assuming say X follows Poisson lambda Y follows Poisson mu; that means, we have fixed the time or the area etcetera in which this is; that means, they are on the same scales that we are considering. We want to compare the arrival rates of the two processes; so, we will show here again that we can put up the problem in a mutliparameter exponential family and then we can simplify the things.

So, the joint probability mass function of X and Y that is equal to e to the power minus lambda lambda to the power x by x factorial; e to the power minus mu nu to the power y by y factorial x is equal to 0, 1, 2 and so on, y is equal to 0, 1, 2 and so on; lambda and mu both are positive. So, this we write as e to the power minus lambda plus mu divided by x factorial y factorial; e to the power x log lambda plus y log mu.

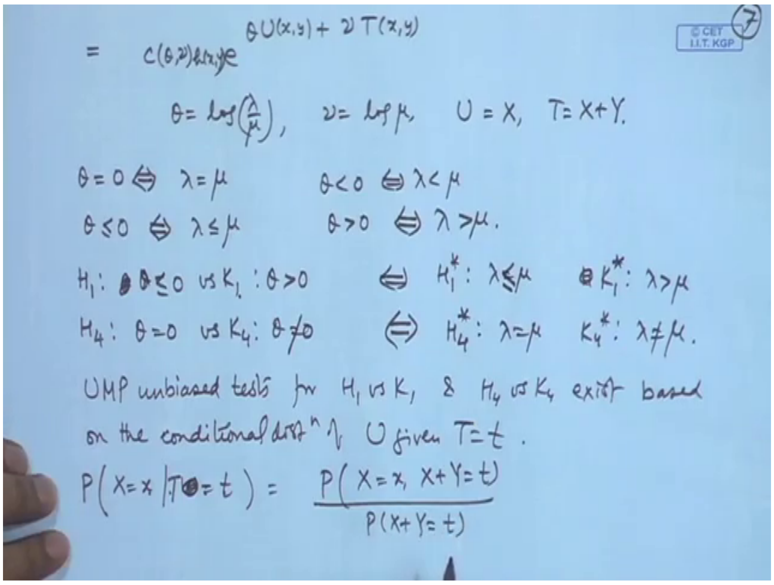
Now, this is a two parameter exponential family I can consider this as theta and this as mu this as u and this as t. However, writing down like this does not help me to test about comparison between lambda and mu; this will help us to test about lambda or testing about mu; however, that is not our problem of interest. So, like in the binomial problem we make some reparameterization or rewriting of this term here.

(Refer Slide Time: 23:59)



So, we write it as e to the power minus lambda plus mu divided by x factorial y factorial; e to the power x log lambda by mu; that means, I have subtracted x log mu; so I add that thing. So, this becomes x plus y log of mu. So, now you can see this entire thing has become in the same form I can consider this as some theta; this as nu, this as u and this as t.

(Refer Slide Time: 24:29)

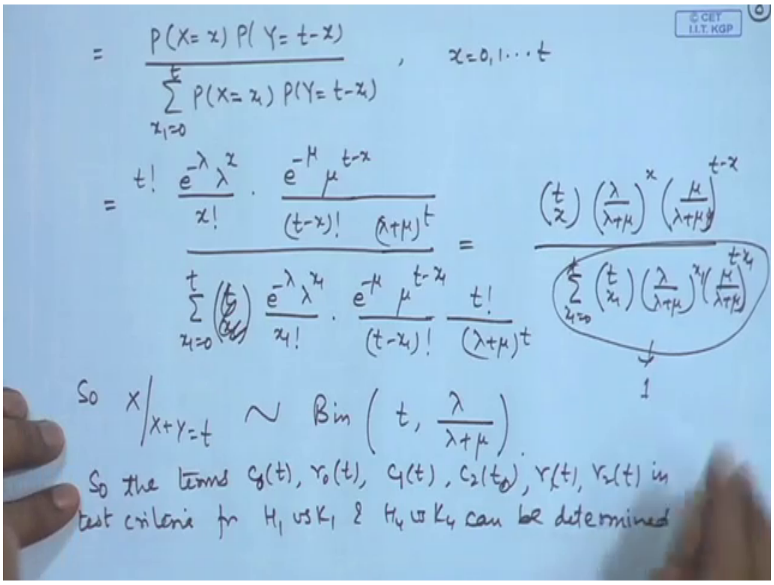


So, I write the whole thing as e to the power theta U x y plus nu T x y into a function of theta and nu and a function of x y. So, here theta is log of lambda by mu; nu is log of mu U is equal to X and T is equal to X plus Y. So, if I say theta is equal to 0; this is equivalent to saying lambda is equal to mu. Similarly if I say theta is less than or equal to 0; this is equivalent to hypothesis lambda is less than or equal to mu. If I say theta less than 0 that is equivalent to same lambda is less than mu; if I say theta greater than 0, this is equivalent to saying lambda is greater than mu.

Therefore we have equivalent of H 1, H 2, H 3, H 4 in terms of theta equal to 0, theta less than or equal to something theta greater than or equal to something etcetera. In particular if I consider say H 1 that is lambda sorry theta is less than or equal to 0 versus K 1; theta is greater than 0 this is equivalent to testing lambda is equal to less than or equal to mu versus lambda is greater than mu. And similarly H 4 that is theta is equal to 0 versus K 4 theta is equal not equal to 0; this is equivalent to testing H 4 star, lambda is equal to mu against K 4 star lambda is not equal to mu.

Now for both of this hypothesis; UMP unbiased tests for H 1 versus K 1 and H 4 versus K 4 exist based on the conditional distribution of U given T. So, let us calculate that thing like in the previous problem actually the statements few of the statements will actually be the same X is equal to x given T is equal to t. So, that is equal to probability X is equal to x; X plus Y is equal to t divided by probability of X plus Y is equal to t.

(Refer Slide Time: 27:37)

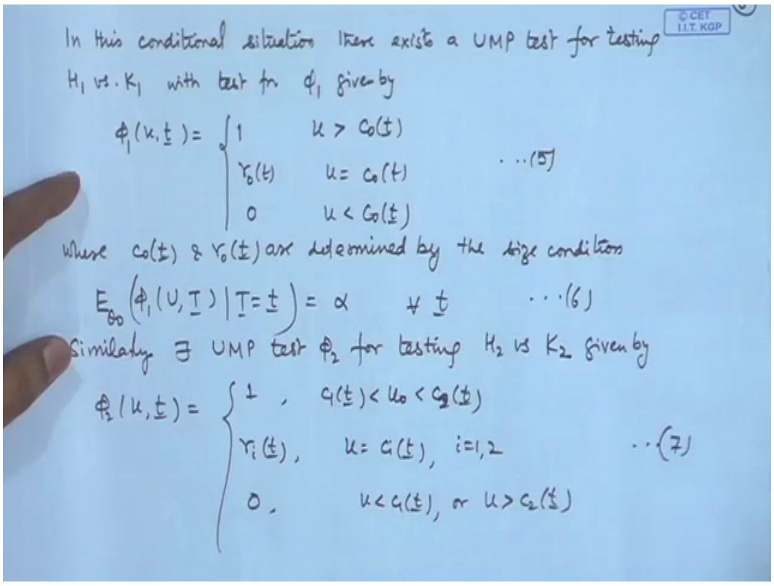


So, this is equal to as before probability of Y is equal to t minus x divided by probability of; and once again we can write down the full description sigma probability of say X equal to x 1; probability Y is equal to t minus x 1; x 1 is equal to 0 to t. So, here x can go from 0 1 to t only. So, this is e to the power minus lambda lambda to the power x by x factorial; e to the power minus mu, mu to the power t minus x by t minus x factorial; this divided by sigma t c x 1 sorry this is e to the power minus lambda lambda to the power x 1 by x 1 factorial and e to the power minus mu mu to the power t minus x 1 by t minus x 1 factorial x 1 is equal to 0 to t.

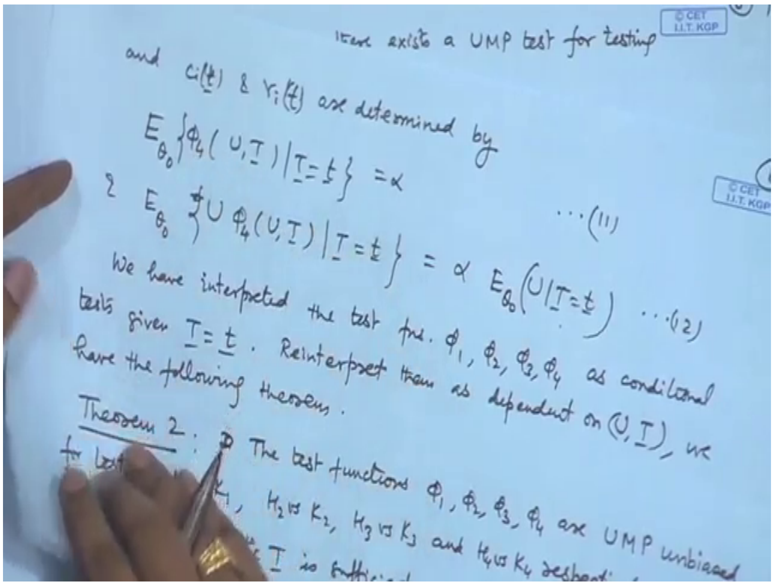
Now, here some of the terms will get cancelled out especially e to the power minus lambda plus mu that gets cancelled out in the numerator and the denominator and we can multiply by say t factorial here and t factorial here. So, I get t c x and lambda by lambda plus mu to the power x and mu by lambda plus mu to the power t minus x. What I have done? I have considered division by lambda plus mu to the power t; lambda plus mu to the power t. So, this term can be adjusted here and we get this here and similarly the denominator will give me t c x 1; lambda by lambda plus mu to the power x mu by lambda plus mu to the power t minus x 1 where x 1 is from 0 to t.

Now, if you notice the denominator here this is actually equal to 1 because this is nothing, but a sum of the binomial terms here you are getting something like p to the power x and q to the power n minus x here and sigma m c x from 0 to a. So, this term is one; so this is actually reducing to a binomial distribution. So, the conditional distribution of X given X plus Y; this is binomial with parameter t and lambda by lambda plus mu.

(Refer Slide Time: 30:31)

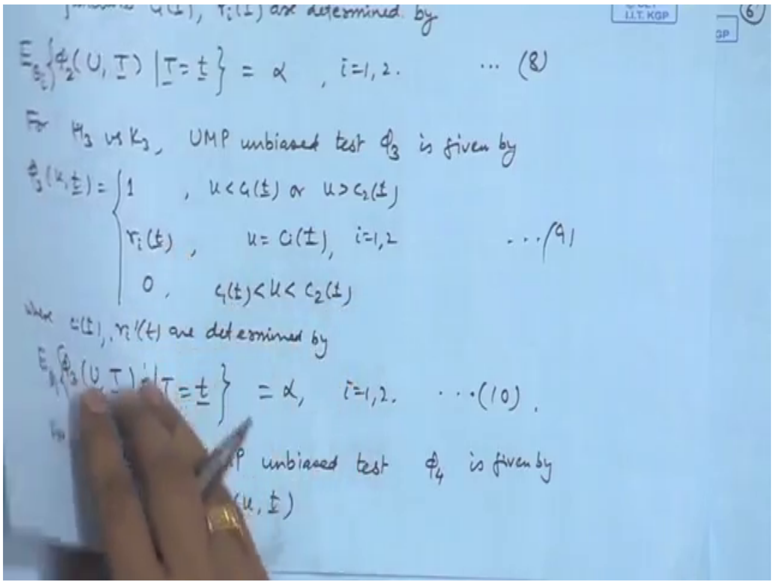


(Refer Slide Time: 30:41)



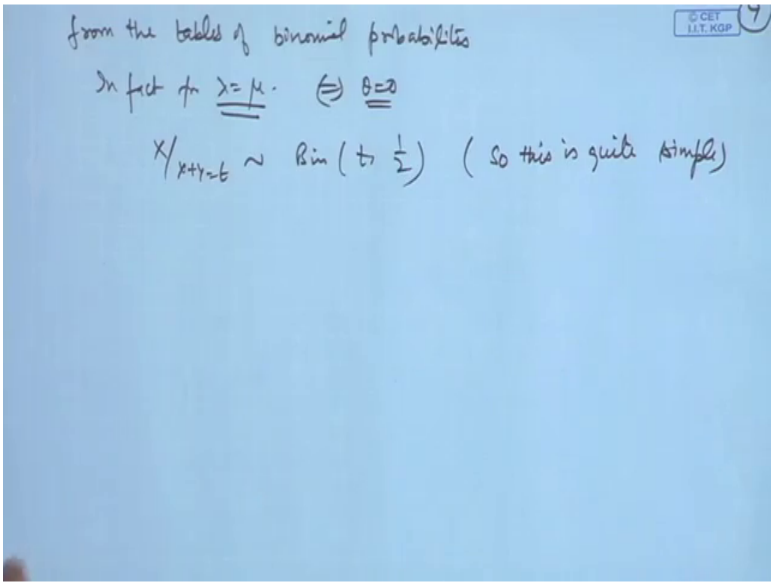
Therefore, in the test function if I am considering c naught and gamma naught here from this condition or in phi 4 function if we are considering these coefficients that is c 1, c 2s; gamma 1, gamma 2 etcetera through these conditions they all can be determined from the tables of the binomial distribution.

(Refer Slide Time: 30:47)



So, the terms c 1, c naught t gamma naught t; c 1 t, c 2 t gamma 1 t and gamma 2 t in test criteria for H 1 versus K 1 and H 4 versus K 4 can be determined from the tables of binomial probabilities.

(Refer Slide Time: 31:43)



In fact, when we are considering the point lambda is equal to mu; then that case that is theta is equal to 0 then X given X plus Y is equal to t this is simply following binomial t half. So, this is simply it can be calculated directly without even seeing the tables here; so this is quite simple.

I have shown you two applications of the UMP unbiased test, where we are considering two different populations. Like in the case of binomial we had two proportions p 1 and p 2; in Poisson arrival we have two rates lambda and mu and we were able to compare. We will further show the applications to comparing the means etcetera for normal populations and variances; also testing for the means and variances in normal population.

So, these are applications I will be; so I will be covering it in the next lecture.