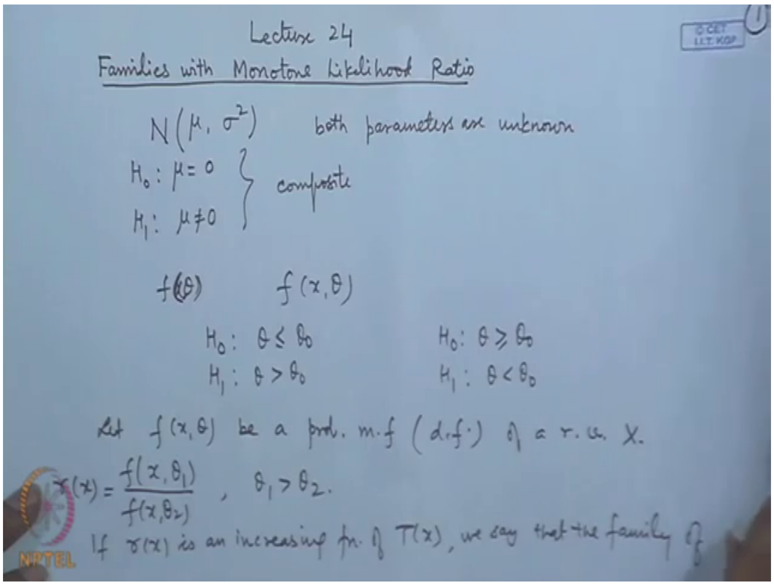
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 37
UMP Tests – I

So, far the testing procedure that we have discussed was based on the Neyman Pearson and fundamental lemma. The main assumption that we made in deriving the test procedure was that, the null hypothesis and the alternative hypothesis both were considered to be simple. And in this case when we fix the probability of type 1 error, then we were able to derive the test which is having the minimum probability of type 2 error or the maximum power and we called it the most powerful test.

However in most of the real life situations, we do not come across the simple hypothesis versus simple hypothesis problems in most of the complex situations we have composite hypothesis.

(Refer Slide Time: 01:07)



As a very simple case we may have the family of distributions as normal mu sigma square distributions, and we may like to even now we may like to test something like whether mu naught mu is equal to 0 or r mu is not equal to 0.

Note here that now this H naught is not a simple hypothesis this is composite because sigma square is unknown. Here we have assumed both parameters to be unknown both parameters are unknown. Therefore, these are now composite both hypotheses are composite hypothesis and therefore, the Neyman Pearson lemma does not help us to give a solution in this particular problem; that means, it does not give a most powerful test.

The simplest composite hypotheses are of this nature that we may have a one parameter family say family of distributions with one parameter theta say f x theta, and we may like to test about say H naught theta less than or equal to theta naught against say theta greater than theta naught or alternatively we may have say H naught, theta greater than or equal to theta naught against h 1 theta less than theta naught. Now let us remember our cases that some examples we considered for the Neyman Pearson lemma.

Where we had considered theta is equal to theta naught against theta is equal to theta 1. I had considered 2 cases; 1 was theta naught less than theta 1 and another was theta naught greater than theta 1. When theta naught was less than theta 1; we got a 1 sided testing region that is the rejection region that is for larger values of x bar we were rejecting H naught.

Now, in that problem in place of theta 1 suppose we replace it by another value theta 2; suppose we replace by another value theta 3 the testing procedure remains the same as long as this s second value in the alternative hypothesis remains larger than theta naught. In a similar way, if we are considering the reverse case theta naught greater than theta 1, then the rejection region was for s smaller values of x bar.

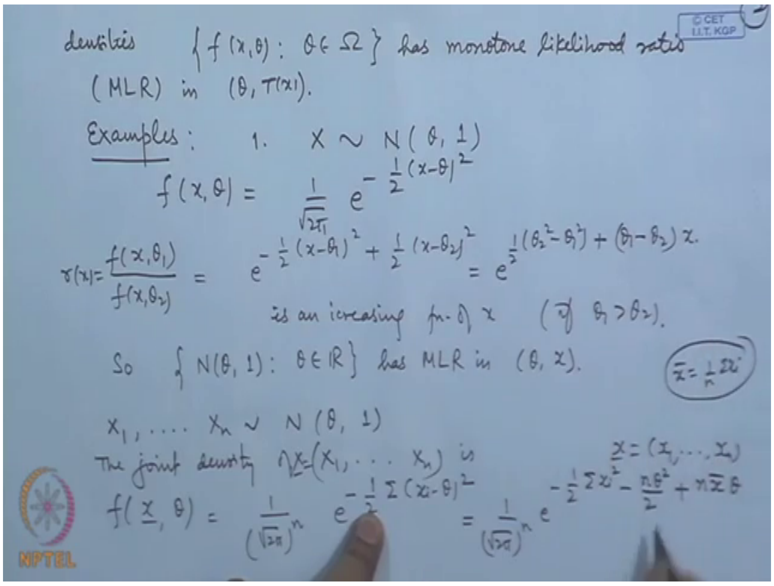
And once again if we replace this alternative hypothesis theta 1 in the same direction; that means, value which is le greater than theta naught or less than theta naught, then the rejection region does not get affected. What does it mean? It means that for those values we are getting the most powerful tests; that means, this normal distribution with 1 parameter the second parameter sigma square was considered to be known has certain property.

Now, in these situations for the changing values, we get the maximum power at each of the values this is called uniformly most powerful test. Now this family of distributions which will satisfy this property; that means, where we will get such test, it is having some particular name it is called the families with monotone likelihood ratio property.

In particular for the one sided testing of hypothesis problems like theta less than or equal to theta naught against theta greater than theta naught or theta greater than or equal to theta naught against theta less than theta naught etcetera for such cases we are actually getting the uniformly most powerful test. The results that are proved they are actually you can say they are extensions of the Neyman Pearson fundamental lemma.

So firstly, let me define this families. So, let f x theta be a probability mass function or density function of a random variable say x. Let us write down the ratio f x theta 1 divided by f x theta 2 let us call this name this ratio let me call it r x and let us take say theta 1 greater than theta 2. If r x is a an increasing function of some variable say T x then we say that the family of densities.

(Refer Slide Time: 05:59)



The word densities means it includes the probability mass functions. So, that is f x theta, theta belonging to the parameter space has monotone likelihood ratio, that we call MLR in theta T x.

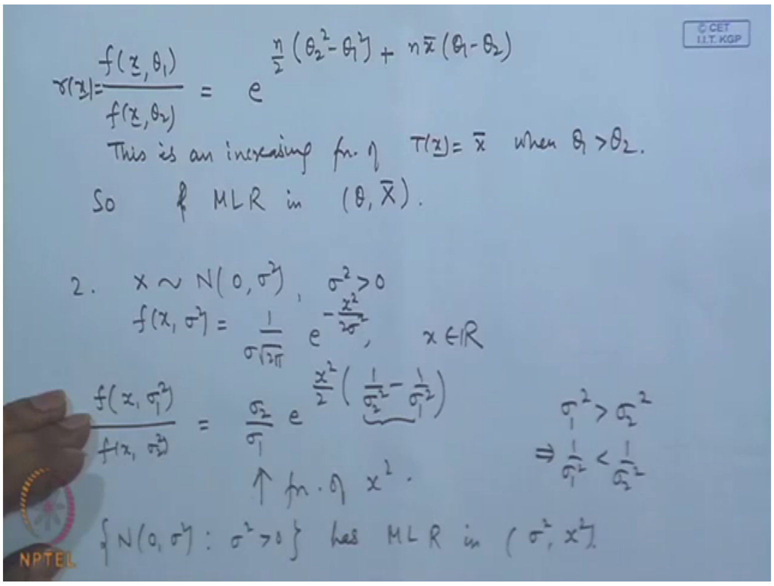
Let me give an example here. Say let us consider say x following a normal distribution with mean theta and known variance 1. Let us write down the distribution f x theta is 1 by root 2 pi e to the power minus half x minus theta square. Let us consider this ratio r x that is f x theta 1 divided by f x theta 2. Now when you write this ratio this gets cancelled out and you have e to the power minus half x minus theta 1 square plus half x minus theta 2 s square, that is equal to e to the power half theta 2 square minus theta 1 square and then you will have plus theta 1 minus theta 2 x.

So, you can look at this, this is an increasing function. If I am taking increasing function of x, if theta 1 is greater than theta 2 because this is constant and if theta 1 is greater than theta 2 e to the power this becomes an increasing function of x. So, this family of distributions normal theta 1, where theta belongs to real line this has monotone likelihood ratio in theta and x.

Now I have written here the distribution of 1 observation, suppose in place of x; I have x 1 x 2 x n suppose I have X 1 X 2 X n. In this case f x theta we have to write the joint distribution of X 1, X 2, X n. So, the joint density of X 1 X 2 X n. So, let me give the notation f x, here x is standing for the values x 1 x 2 x n of capital X 1 capital X 2 capital X n.

So, this becomes 1 by root 2 pi to the power n e to the power minus 1 by 2 sigma x i minus theta square. Let us simplify this we can write it as 1 by root 2 pi to the power n e to the power minus half sigma x i square minus n theta square by 2 plus. Now you have the cross product term twice x i theta with a minus sign and minus minus will become plus and 2 will cancel out. So, you get twice n x bar theta where x bar is the 1 by n sigma x i.

(Refer Slide Time: 10:03)



.So, now you write down this ratio f x theta 1 divided by f x theta 2 that is turning out to be. Now when you write the ratio this constant term will get cancelled out e to the power minus half sigma x i square will get cancelled out. We will be left with e to the power n by 2 theta 2 square minus theta 1 square plus n x bar theta 1 minus theta 2.

Now, this is constant, for theta 1 greater than theta 2 this becomes an increasing function of x bar. So, this ratio and increasing function of T x is equal to x bar when theta 1 is greater than theta 2. So, this family of distributions normal theta 1, when we are having n observations; so we have MLR in theta and x bar we can say.

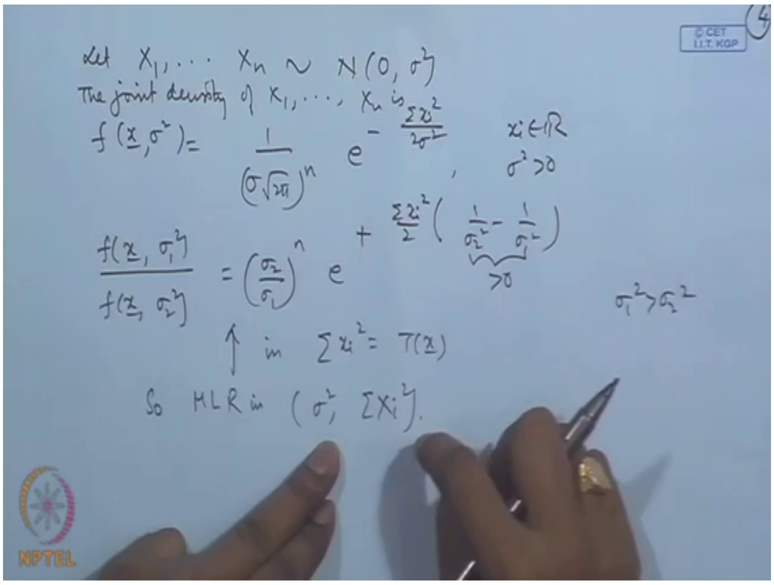
Now, the similar thing we can observe for various distributions let me give a couple of more examples. Here I have considered the normal distribution when the variance is assumed to be known. Now there can be other case where mean may be known and the variance may be unknown let us take that case. Let me again consider say 1 observation and then I will consider N observations generally we are dealing with the sample; so let me take this case.

Here sigma square is a positive parameter, if we consider the density function here it is 1 by sigma root 2 pi e to the power minus x square by 2 sigma square where x is any real number. Therefore, if I consider the ratio f x sigma 1 square divided by f x sigma 2 square. Now this will give me sigma 2 by sigma 1 this 1 by root 2 pi will get cancelled out, e to the power x square by 2, 1 by sigma 2 square minus 1 by sigma 1 square.

Now, let us take say sigma 1 square greater than sigma 2 square; that means, 1 by sigma 1 square is less than 1 by sigma 2 square. So, this term becomes positive and therefore, this is increasing function of x square. So, this family of normal 0 sigma square distributions, this has monotone likelihood ratio in sigma square x square.

Now note here that suppose I take a sample here in place of X; let us take sample say x 1 x 2 x n following normal 0 sigma square and let us write the same thing once again.

(Refer Slide Time: 13:17)

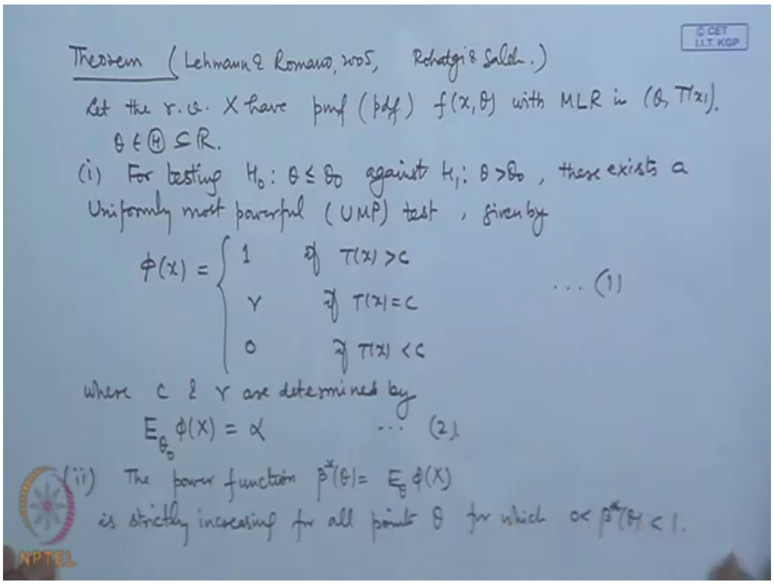


The joint distribution, the joint density of X 1, X 2, X n that will become 1 by sigma root 2 pi to the power n e to the power minus sigma x i square by 2 sigma square where sigma square is positive and each x i is on the real line. So, when we write down the ratio, now this term get cancelled out we will get sigma 2 by sigma 1 to the power n e to the power minus sigma x i square by 2 1 by sigma 2 square minus 1 by sigma 1 square.

So, this I will put plus here. Once again you note here this is positive if sigma 1 square is less than sigma 2 square sorry if sigma 1 square is greater than sigma 2 square, then this term becomes positive. So, this is increasing in sigma x i square, that we will call T x. So, this family has monotone likelihood ratio in sigma square and sigma X i square.

Now, this T x has a special role. When we will derive the uniformly most powerful test you will see that the test will depend upon this itself. So, I will discuss a few more applications a little later let us look at the main result of this section now, that is an application of the monotone likelihood ratio property how the uniformly most powerful test exists.

(Refer Slide Time: 15:09)



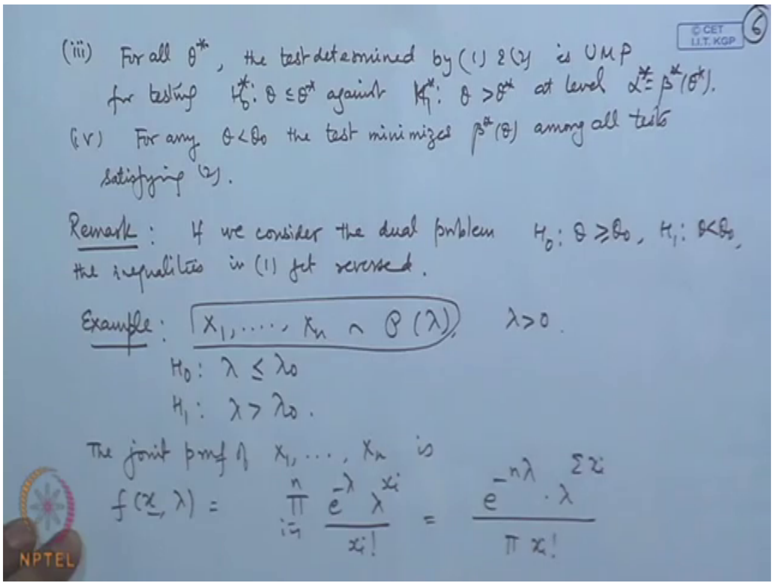
So, I state the theorem. For a proper statement of this theorem you may look at the books of Lehmann and Romano 2005 or you may look at Rohatgi and Saleh books.

The proofs are also given there. So, I am not discussing the proof here. So, let us consider let the random variable X have probability mass function or probability density function say f x theta with monotone likelihood ratio in theta T x. And of course, here theta is a real parameter theta belongs to say theta which is subset of the real line.

So, the result that we are having here is that, for testing H naught theta less than or equal to theta naught against H 1 theta greater than theta naught, there exists a uniformly most powerful test that is UMP test, given by. As before, we will use the phi notation for the test function. So, you reject if T x is greater than C, you reject with probability gamma if T x is equal to C and you accept if T x is less than C; where C and this gamma are determined by expectation of theta naught phi X is equal to alpha, let me call this conditions 1 and 2.

Note here that similarity with Neyman Pearson lemma; in the Neyman Pearson lemma we had written f 1 by f naught greater than k. Now if f 1 by f naught ha is an increasing function of T x, then that region is transformed to T x greater than C. So, it is as I mentioned it is direct extension of the Neyman Pearson fundamental lemma only, the result is coming from there. The power function that is we have used the notation say beta is star theta, that is equal to expectation theta phi x is a strictly increasing for all points theta for which it lies between 0 and 1.

(Refer Slide Time: 19:05)



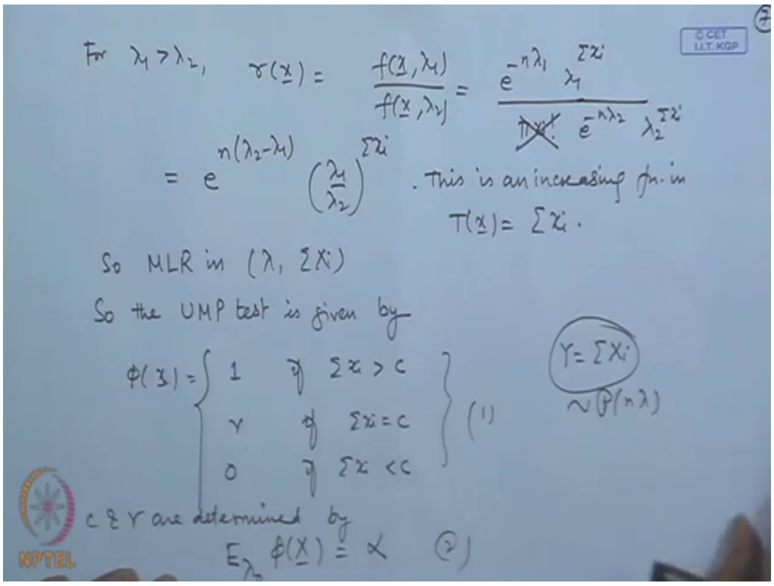
For all theta star the test determined by 1 and 2 is uniformly most powerful for testing H prime theta less than or equal to theta prime against k prime theta, greater than theta prime at level say alpha prime is equal to let me put a star here. This new hypothesis I am calling H naught and h 1 star at alpha star is equal to beta star of theta star. And for any theta less than theta naught, the test minimizes beta star theta among all tests satisfying the condition 2. I will skip the proof here 1 can look at the book of Lehmann for the detailed proof of these statements.

Now, one may note here I have considered theta less than or equal to theta naught against theta greater than theta naught. As I gave the heuristic argument that in the Neyman Pearson lemma and as also I gave the normal distribution example when we were testing for the mean, the rejection region for the larger value of x bar and here it is for the larger value of T x. So, if we reverse like for the null hypothesis region we consider grater and for the null alternative hypothesis we consider less than or equal to then the rejection region will also get the reverse.

So, what I just give it as a comment here. If we consider the say dual problem say H naught theta greater than or equal to theta naught against H 1 theta less than theta naught, the inequalities in 1 get reversed. So, you have the solution in a similar manner there. Let me take an application here.

Say we have a random sample say X 1 X 2 X n from say Poisson lambda distribution. And we consider say hypothesis lambda less than or equal to say lambda naught against say lambda greater than lambda naught. Now let us look at this family of Poisson distributions whether it has a monotone likelihood ratio or not. So, the joint probability mass function of X 1 X 2 X n; So, we write it as f x lambda product i is equal to 1 to n e to the power minus lambda lambda to the power x i by x i factorial. That is equal to e to the power minus n lambda, lambda to the power sigma x i divided by product x i factorial.

(Refer Slide Time: 22:59)



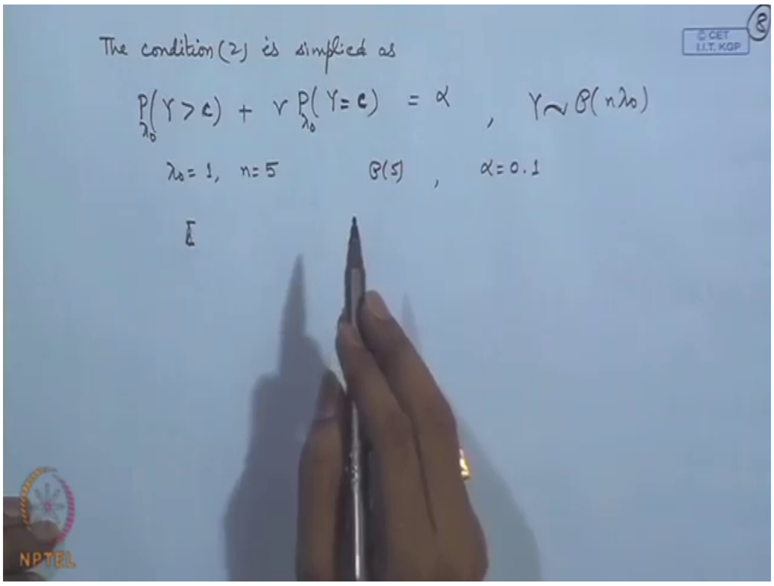
So, if we consider the ratio for lambda 1 greater than lambda 2 let us consider the ratio f x lambda 1 by f x lambda 2. So, it is becoming e to the power minus n lambda 1, lambda 1 to the power sigma x i divided by product x i factorial divided by e to the power minus n lambda 2, lambda 2 to the power sigma x i and this term will get cancelled out.

So, we can write it in a simplified fashion as lambda 2 minus lambda 1 lambda 1 by lambda 2 to the power sigma x i. This lambda 1 by lambda 2 is greater than 1 because lambda 1 is greater than lambda 2 therefore, this will become an increasing function, this is an increasing function in T x is equal to sigma x i.

So, we have monotone likelihood ratio in lambda and sigma x i. So, we can apply the theorem that I gave. If the family has monotone likelihood ratio in theta and T x then for one sided null hypothesis versus 1 sided alternative hypothesis the uniformly most powerful test is obtained here. So, let me write it here. So, the UMP test is given by phi x, this here x means x 1 x 2 x n it is rejecting if sigma x i is greater than C it is rejecting with probability gamma if sigma x i is equal to C it is 0 if sigma x i is less than C.

Now, the sigma X i is actually let me write say it has equal to y, then that will follow Poisson distribution n lambda. Now C and gamma are determined by the condition expectation of lambda naught phi x is equal to alpha. Now this is reducing to let me write it as say 1 and this as 2.

(Refer Slide Time: 25:45)



So, this condition 2 let me simplify the condition 2. So, expectation of lambda naught phi x since this phi x is completely dependent upon sigma x i that is y. So, it is becoming probability of Y greater than c plus gamma times probability of Y is equal to c this is equal to alpha when the true parameter value is lambda naught.

Now another point which I would like to explain here in the case of simple versus simple hypothesis, we had the probability of type 1 error as a single value. But when we have composite hypothesis for the null hypothesis, then the probability of type 1 error is a function. However, this is an increasing function which I mentioned in the statement of the theorem also that, the power function is strictly increasing function. So, the probability of type 1 error is increasing.

So, when you are getting theta is equal to theta naught, then at that point the maximum value is are obtained. So, effectively this condition is actually the size condition, that is expectation of phi x equal to alpha lambda naught this is the maximum probability of type 1 error here, that we are fixing to be equal to alpha. So, the size condition now it is reduced to a condition which is really involving the distribution Poisson n lambda naught. Therefore, from the tables of the Poisson distribution one can calculate this.

Suppose I say lambda naught is equal to 1 and n is equal to say 5 then basically we are looking at the tables of Poisson 5 distribution. Suppose I say alpha is equal to 0.1, then basically what we are seeing here is that what is the point from where. Now this C could be need not be an integer actually we may fix that thing in such a way. If it is an integer then this value may be positive if it is not integer then this may become 0.

Now, you may see from the tables that whether this randomization with probability gamma is required or not. If it is not required then this probability can be taken to be 0 there will be a point where after you will have the probability alpha. In case that is not possible then we suitably choose a value where we a lot of probability and then we may give some value of gamma also. So, now, these can be calculated from the tables of the from the tables of the Poisson distribution.

We can also see a like a binomial distribution, we suppose we have a hyper geometric distribution in all of suppose we have a negative binomial distribution in all of these distributions we may able we are able to find out the uniformly most powerful tests. I will consider this derivation of the tests in the following I will be discussing it in the next lecture.