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

Lecture – 17
Lower Bounds for Variance – III

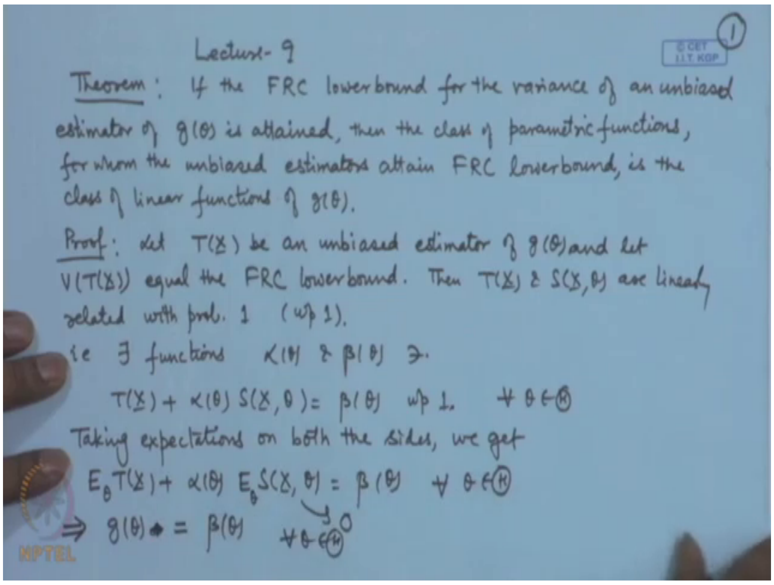
In the previous lecture I explained the method of finding out a Lower Bound for the Variance of an unbiased estimator for a given parametric function. As I mentioned it was derived independently by three statisticians Frechet Rao and Cramer. And therefore, we have named it as a Frechet Rao Cramer lower bound that is FRC lower bound for the variance of an unbiased estimator.

We have seen that there are cases where we can find out an estimator for which this lower bound is attained, there are also cases where it is not attained. We gave a condition under which an unbiased estimator will attain this lower bound. The condition was in the terms that it should be linearly related with a function S X theta with probability 1. This method as I explained, this method of lower bounds is very very useful from 2 points of view.

One is that given any estimator we can compare its variance with the lower bound and therefore, we know that how far we are from the actual and; that means, what could be the best possible way minimum variance and where are we; that means, where is our estimator is standing in its relative position. And second thing is that if we are able to obtain an estimator for which it is equal to the lower bound then certainly it is the minimum variance unbiased estimator that is among the unbiased estimator it will be certainly the best.

So, from this point of view this method of lower bounds is extremely useful. We have seen that FRC lower bound as I call it is dependent upon certain regularity conditions that is when the density or the mass function under consideration satisfy certain conditions then only this lower bound is valid. We also see in this: what are the parametric functions for which this lower bound is attained. So, let me give it in the form of a theorem.

(Refer Slide Time: 02:25)



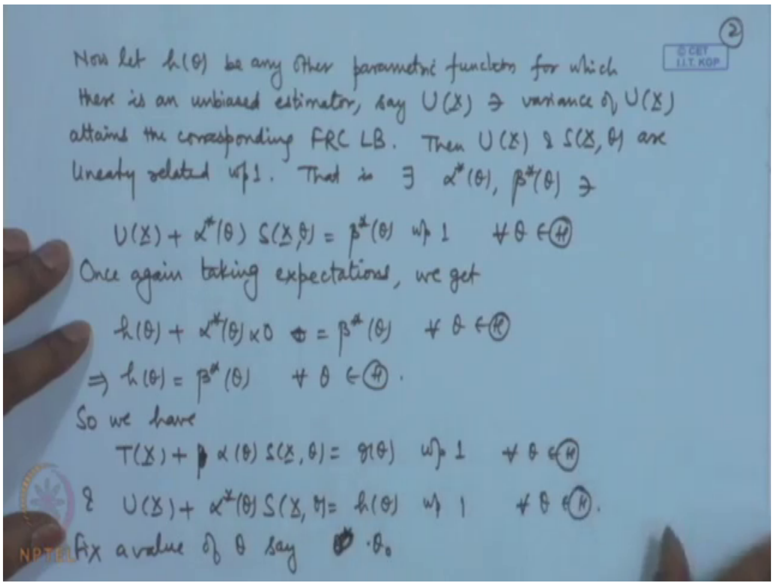
So, we have a random sample X 1 X 2 X n and we know that the F RC lower bound for the variance of an unbiased estimator of say g theta is attained; then what are the parametric functions apart from g theta for which this will be attained. Then there answer is that they are actually the linear functions of g theta, then the class of parametric functions for whom the unbiased estimators attain this FRC lower bound; then this class is the class of linear functions of g theta.

Like I said what is the unbiased estimator for which the lower bound will be attained, that should be a linear function of S X theta with probability 1. Now what are the parametric functions for which it will be attained then they should be simply the linear functions of g theta that is the statement of this theorem; let me prove this theorem here. So, let us consider say T X let T X be an unbiased estimator of g theta and let variance of T X equal the FRC lower bound.

Then certainly we know that T X and S X theta they are linearly related with probability 1 we will use this with probability 1 as an abbreviation here. So that means there exist functions say alpha theta and beta theta such that say T X plus alpha theta S X theta is equal to say beta theta with probability 1; this should be true for all theta ok. Now in this relation let us take expectation on both the sides.

So, expectation of T X plus alpha theta expectation of S X theta is equal to beta theta for all theta. Since this statement is true for all; that means, for random variable X here it is true with probability 1. Therefore, it is possible to take the expectations basically expectation means either we have taken summations or we have taken the integrals or a mixture of the 2. Therefore, we will get expectation of this equal to beta theta. Now p is unbiased estimator for g theta; that means g theta now expectation of S X theta that is 0. Therefore, this is simply giving you beta theta because this is equal to 0. So, in this relationship beta theta has turned out to be g theta here.

(Refer Slide Time: 07:30)

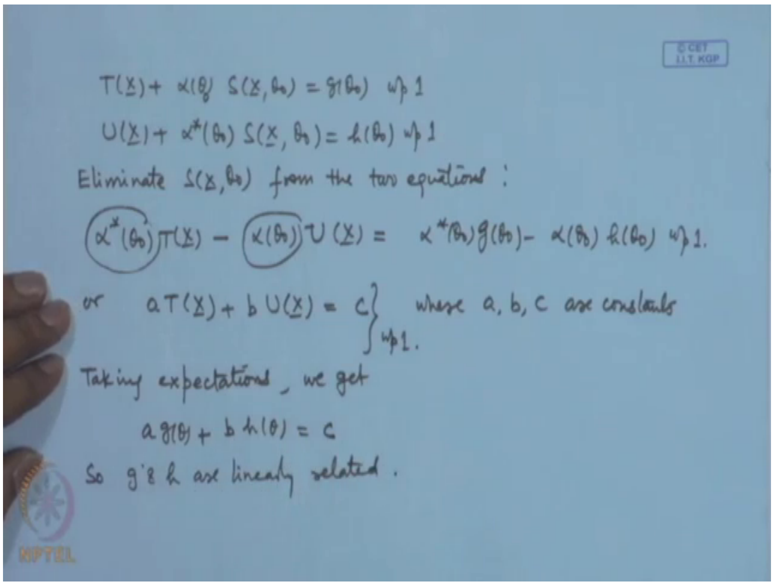


Now let us consider now, let h theta be any other parametric function for which there exist an unbiased estimator for which the lower bound is attained ok. So, for which there is an unbiased estimator say U X such that variance of U X attains the corresponding FRC lower bound. We have seen that even if we change the parametric function the lower bound is changed, but the condition for attaining the lower bound remains the same.

Therefore, so, there will exist that U X and S X theta are again linearly related with probability 1; that means, we can say that there exist say functions alpha star theta and beta star theta such that U X plus alpha star theta into S X theta is equal to beta star theta with probability 1 for all theta. Once again since this statement is true with probability 1 we can take expectations. So, if we take expectations we get expectation of U X will be equal to h theta plus alpha star theta into expectation of S X theta is 0 is equal to beta star theta. So, we are getting h theta is equal to beta star theta.

So, if we look at these two equations now T X plus alpha theta S X theta that will be equal to g theta and U X plus alpha star theta S X theta is equal to h theta. So, we have T X plus alpha theta S X theta is equal to g theta with probability 1 for all theta and U X plus alpha star theta S X theta is equal to h theta with probability 1 for all theta belonging to theta. If this relationship is true for all theta we can fix a value of theta fix a value of theta say theta star or let me put theta naught because already. So, many stars are there.

(Refer Slide Time: 11:17)

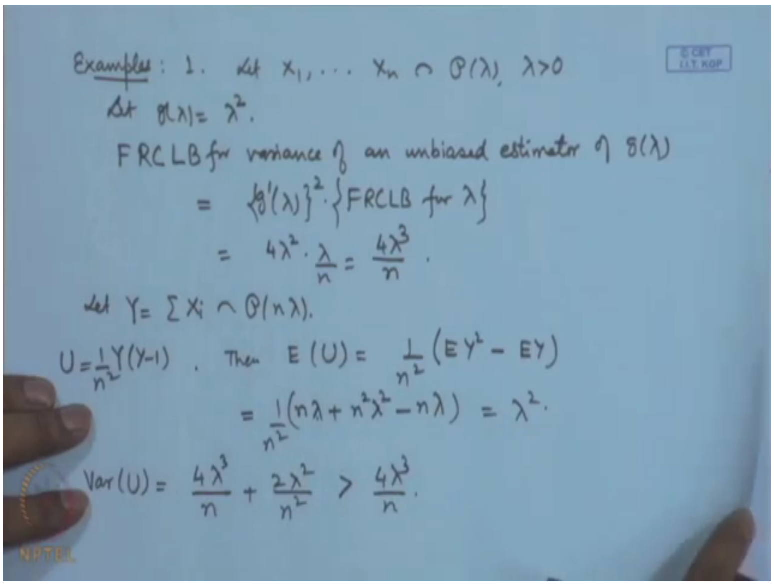


So, in that case we can write the relationship as T X plus alpha theta naught S X theta naught is equal to g theta naught with probability 1 and U X plus alpha star theta naught S X theta naught is equal to h theta naught a with probability 1. That means, what I have done is that these two relations I have written for a fixed value of theta that is theta naught. Now in both of these equations S X theta naught is appearing so, I can eliminate that. So, eliminate S X theta naught from the two equations that is in the first equation multiply by alpha star theta naught in the second equation multiply by alpha theta naught and then subtract.

So, we get alpha star theta naught T X minus alpha theta naught U X is equal to alpha star theta naught g theta naught minus alpha theta naught h theta naught with probability 1. Now, once again you can take the expectation because what is happening here is that this coefficient is a fixed number this coefficient is a fixed number and right hand side is also a fixed number. So, we can say that a times say T X plus say b times U X is equal to c where a b c are constants and this statement is true with probability 1.

So, we can again take expectations if we take expectations we get a times g theta that is expectation of T X plus b times h theta is equal to c. Now you look at the significance of this I started with a function g for which the FRC lower bound was attained, I assumed h theta to be any other parametric function for which the lower bound is attained and now we are getting that such g and h will be related using linear relationship here. So, g and h are linearly related therefore; all functions for which the FRC lower bound will be attained, they will be linear functions of g. Now, in yesterdays lecture I have given examples in some examples the lower bound was attained let us take one such example say a Poisson distribution.

(Refer Slide Time: 14:48)



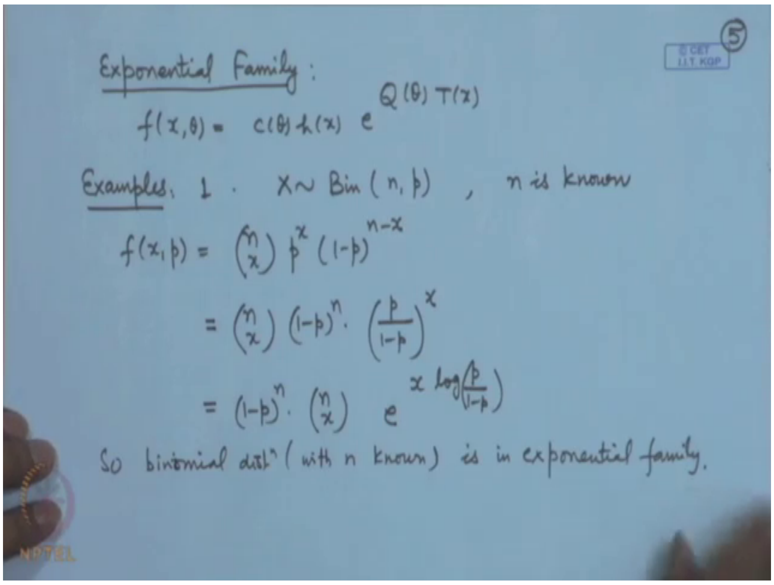
So, we had X 1 X 2 X n following Poisson lambda where lambda is positive; we have seen that X bar was unbiased for lambda and variance of X bar was lambda by n which was also the FCR lower bound for unbiased estimator of lambda. So, if I consider say lambda square let g lambda be equal to lambda square, in that case what you will get; the FRC lower bound for variance of an unbiased estimator of g lambda now that will be equal to g prime lambda square into the FRCLB for lambda.

So, this will become 2 lambda square that is 4 lambda square and this is lambda by n. So, it is equal to 4 lambda cube by n. Now, let us consider say Y is equal to sigma X i of course, this will follow Poisson n lambda. And you can look at Y into Y minus 1 by n square let me call it to be say U, then expectation of U it is equal to 1 by n square expectation of Y square minus expectation of Y that is equal to. Now this will become equal to n lambda plus n square lambda square minus n lambda expectation of Y square is n lambda plus n square lambda square because, if we consider Poisson distribution with parameter lambda the second moment is lambda square plus lambda and expectation Y is equal to n lambda.

So, this divided by n square so, that is equal to lambda square. But if we consider say variance of U that will be equal to this can be calculated easily that will turn out to be because this will involve expectation of U square minus expectation of U whole square. Now, expectation of U is lambda square and expectation of U whole square will involve expectation of Y to the power 4, expectation of Y cube and expectation of Y square which is available all the expressions are there for the Poisson distribution. After simplification you get it as 4 lambda cube by n plus twice lambda square by n square.

Now, you can easily see that this is bigger than 4 lambda cube by n. It is understood that this statement should be true because lambda square is not a linear function of lambda here. We have already shown that for lambda the variance of the unbiased estimator attains the lower bound. Therefore, all other functions for which it will be attain they will be of the form a lambda plus b and this is lambda square. So, certainly this cannot be attained. Later on we will show that actually this is minimum variance unbiased estimator using another method.

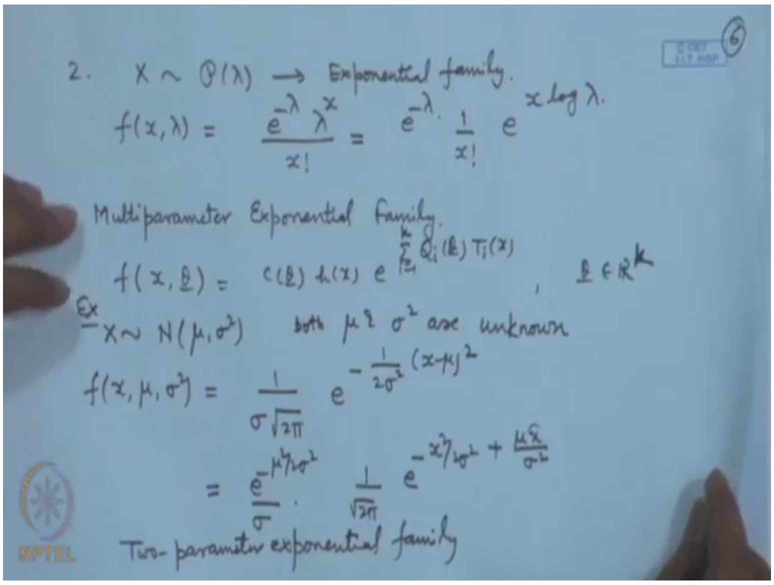
(Refer Slide Time: 18:33)



Let us consider general form of an distribution in the exponential family. So, let us consider a density in the exponential family. What is exponential family? The density is of the form c theta h x e to the power Q theta T x. Now if we have a distribution of this form it is said to be a distribution in the exponential family; we can see examples here say x follows binomial n p, where n is known.

Then the form of the distribution is n c x p to the power x 1 minus p to the power n minus x this we can write as n c x 1 minus p to the power n p by 1 minus p to the power x this we write as 1 minus p to the power n n c x e to the power x log p by 1 minus p. So, if you compare it with this form here you have a function of the parameter that is c theta here theta is p h x is n c x here e to the power Q theta T x. So, here Q theta is a function here log p by 1 minus p and x is the term T x so, this is a distribution. So, binomial distribution; binomial distribution with n known is in exponential family. Let us take some more popular examples in the statistics.

(Refer Slide Time: 21:05)



Let us consider say x following Poisson lambda distribution. The form of the probability mass function is f x lambda it is equal to e to the power minus lambda lambda to the power x by x factorial this we express as e to the power minus lambda 1 by x factorial e to the power x log lambda.

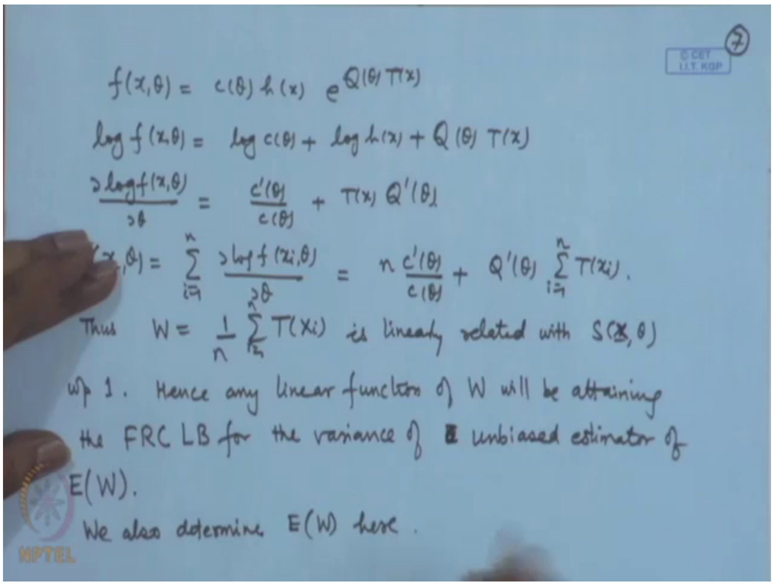
Once again if we compare it with this particular form you can see here e to the power minus lambda is a function of lambda 1 by x factorial is a function of x. So, you can call it h x x can be written as T x and Q theta it is log lambda here. So, you can easily see that this is also a distribution in exponential family we can actually also consider this we can consider as a 1 parameter exponential family we may also consider multi parameter exponential family here parameter could be multi parameter here. So, here we write c theta h x e to the power sigma Q i theta T i x i is equal to 1 to k. So, theta could be say p dimensional and we may have this particular form here.

So, this is actually called see if we have the same dimension here k then this is called a k parameter exponential family. Let us consider say x following normal mu sigma square here both mu and sigma square are unknown f x mu. Sigma square we can write as 1 by sigma root 2 pi e to the power minus 1 by 2 sigma square x minus mu whole square this we express in the following fashion.

If we expand this term you get a term mu square. So, you get minus mu square by 2 sigma square and there is 1 by sigma here 1 by root 2 pi you have e to the power minus x square by 2 sigma square plus mu x by sigma square. Now, this is a function of parameters here. So, this can be considered as a c theta function this constant 1 by root 2 pi can be considered as a function of x alone and then you have you can write here T 1 x is equal to x square and Q 1 theta is equal to minus 1 by 2 sigma square. Similarly here T 2 x can be taken to be x and q 2 theta can be considered to be mu by sigma square.

So, this is a distribution in 2 parameter exponential family most of the standard distributions in statistics that we use for example, gamma distribution with r known and lambda unknown that is a distribution and exponential family; if we consider a negative exponential distribution with the scale parameter that is also in the exponential family. So, there are various distributions which are actually in the exponential family. Now, exponential families have some important feature and in particular with respect to the FRC lower bound.

(Refer Slide Time: 25:37)



So, let us consider in the context of the lower bound. So, if we are writing 1 parameter exponential family as let us take log of this that is equal to log of c theta plus log of h x plus Q theta T x. If we consider the derivative of this with respect to theta we get c prime theta by c theta plus T x into Q prime theta. Now, if you remember your S x function it is nothing, but sigma del log f x i theta by del theta for i is equal to 1 to n. So, this becomes simply n times c prime theta by c theta plus q prime theta sigma T x i.

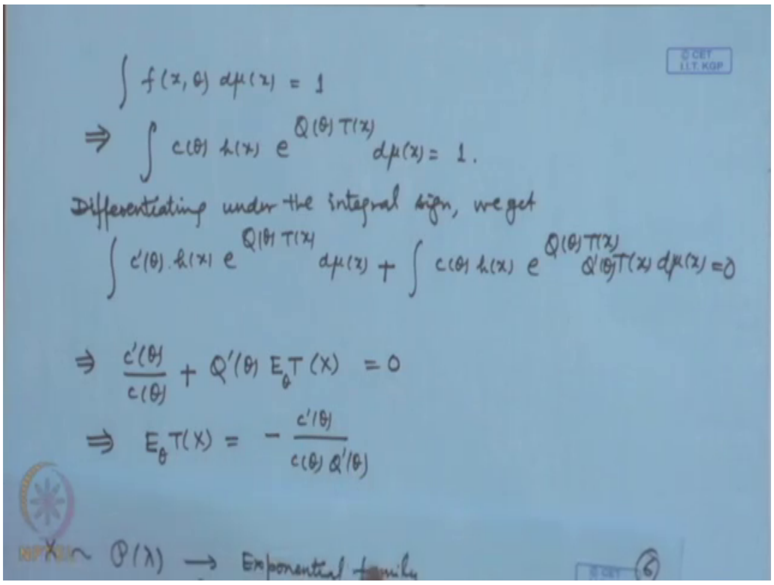
Now, you see here this is constant as far as variable is concerned. So, this is actually a linear function of sigma T x i. So, S x theta is a linear function of sigma T x i. So, in the distributions which are in the exponential family the variables or you can say the estimators which are linear functions of sigma T x i. The variances of them will be attaining the lower bound for the estimation of the expectation of this. So, what we are saying is let us call it say W that is 1 by n sigma T x i.

So, this is linearly related with S x theta with probability 1. Hence any linear function of W will be attaining the FRC lower bound for the variance of expectation for the variance of unbiased estimator of expectation w. We can also see that what will be this expectation in general see in this particular case see we discussed some examples like a Poisson distribution. Now, in this Poisson distribution if you see c theta is e to the power minus lambda its derivative will also be equal to e to the power minus lambda. So, you will get minus n here and q is log lambda. So, q prime will become 1 by lambda.

So, you are getting minus n plus lambda and this will become sigma x i. So, when we say v v is equal to x bar and this W is equal to x bar here. So, x bar is attaining the FRC lower bound for expectation of x bar that is lambda. So, we already proved this statement I am just once again demonstrating that if the distribution is in the exponential family then all the linear functions of 1 by n sigma T x i they will have variance equal to the FRC lower bound.

So, this is a remarkable thing whenever we are having distributions and the exponential family there will be certain parameters for which the lower bound will certainly be attained. Now, let me also obtain the expression for this what is expectation of w? So, let us also we also determine expectation of W here.

(Refer Slide Time: 30:25)



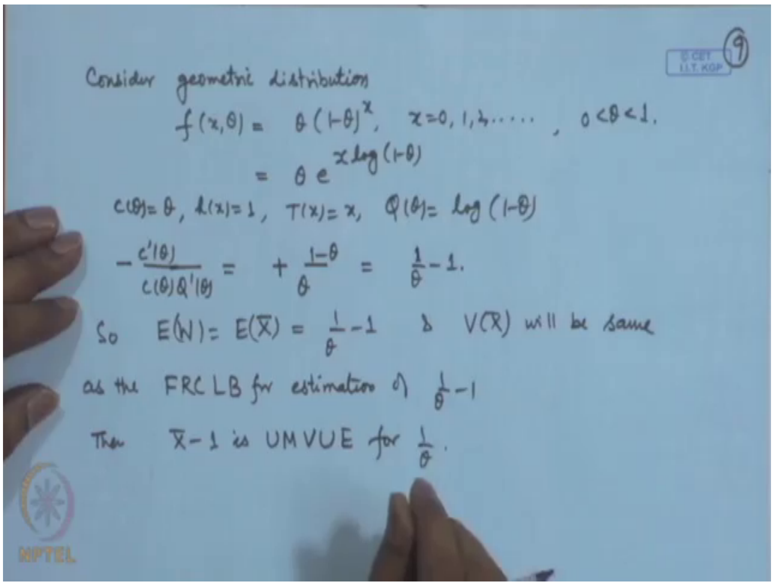
So, that will be let us consider the integral or the summation of the density function or the mass function will be equal to 1. So, I had written a general (Refer Time: 30:38) still just integral meaning that it covers the discrete and continuous cases both. So, c theta h x e to the power Q theta T x is equal to 1.

Now, we make certain assumptions here like differentiation under the integral sign will be assumed because in the Rao, Cramer lower bound itself we made certain assumptions certain regularity assumptions. So, that assumption should be true here also. So, if we assume that then we can differentiate under the integral sign. So, we will get here there are 2 terms which involve theta. So, if we take the first one we will get c prime theta h x e to the power Q theta T x d mu x. And if you differentiate the second term you will get c theta h x e to the power Q theta T x into T x d mu x and of course, Q prime theta will also come this is equal to g or the right hand side is 1 so, the derivative is going to be 0.

Now, this term we can write as divided by c theta multiplied by c theta then that will be integral of the density once again. So, that will become equal to 0. If you look at the second term this density is as such then you are getting this term as additional term. So, Q prime theta expectation of T X this is equal to 0; that means, what we are saying that expectation of theta expectation of T X is actually equal to minus c prime theta by c theta into Q prime theta. Consider for example, that case of Poisson distribution in the case of Poisson distribution c was e to the power minus lambda. So, c prime theta by c theta will become equal to minus 1 that is minus minus becomes plus Q prime theta that will become 1 by lambda. So, if you put it in the denominator you will get lambda here.

So, in the case of Poisson distribution this will become lambda sigma of T X i by n was x bar. So, the statement is that x bar will attain FRC lower bound for the estimation of lambda. So, that statement we verified directly now if we are having the distribution the exponential family this will be always true.

(Refer Slide Time: 33:51)



Let me take one more application here; consider say geometric distribution yesterday we have seen here the form of the distribution we have see taken theta into 1 minus theta to the power x for x is equal to 0 1 2 and so on. So, we can write this is equal to theta e to the power x log 1 minus theta. So, here c theta is equal to theta if we compare with the distribution in the exponential family h x is 1 T x is equal to x and Q theta is equal to log of 1 minus theta.

So, naturally minus c prime theta by c theta q prime theta that is going to be equal to minus 1 c theta is theta Q prime theta will become equal to minus 1 by 1 minus theta. So, it is equal to 1 by theta minus 1; that means, and here x T x is equal to x so, W is equal to x bar. So, expectation of W that is equal to expectation of x bar is equal to 1 by theta minus 1 and variance of x bar will be attaining the Rao Cramer lower bound it will be same as the FRC lower bound for estimation of 1 by theta minus 1. Now if we consider a linear function of it so, 1 by theta also we can consider. So, we can say that X bar minus 1 is minimum variance unbiased estimator for 1 by theta. So, this statement is also true.