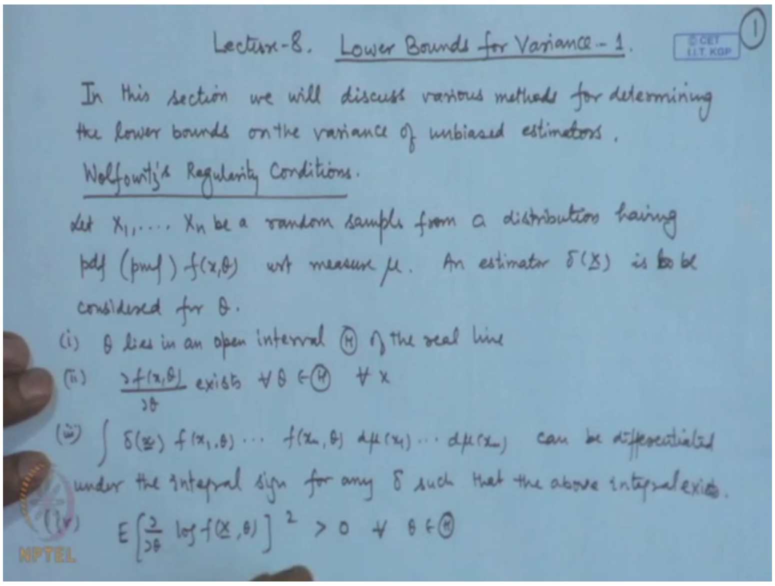
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture No. # 08
Lower Bounds for Variance – I

So, now we will take up another topic, that is for the lower bounds for the variance. Now, what is this concept earlier, we have seen that unbiasedness is a desirable property or desirable criteria to use an estimator; however, we have also seen the example that in a given problem there can be several unbiased estimators. Now, if there are several unbiased estimators which one to choose.

Then we can decide some additional criteria such as variance. The one which has smaller variance will be considered to be more stable in some sense. Now therefore, we need to have an estimate of what could be the variance or what could be the minimum variance. So, this gives the idea or you can say this led to the development of methods for finding out lower bounds for the variance of an unbiased estimator.

(Refer Slide Time: 01:25)



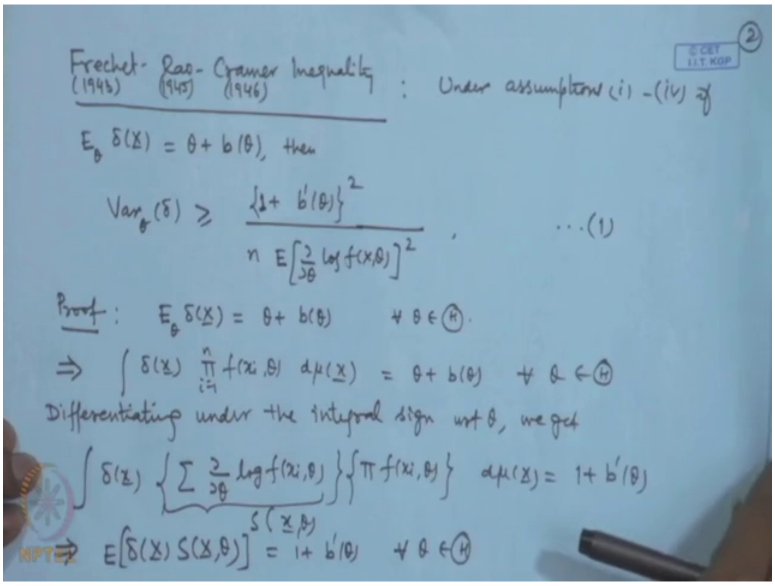
So, in this section, we will discuss various methods for determining the lower bounds on the variance of unbiased estimators. As we have seen in the case of maximum likelihood estimation the last results that I gave, that variance asymptotic variance of the maximum likelihood estimator was one by the information now, this is asymptotic variance. So, if the maximum likelihood estimator is the best in some sense, then its variance will not be below 1 by I theta and; that means, the fisher's information.

The question comes that whether similar result we can give for finite samples. Now, this is precisely the question that was post to indian statistician c r rao in his class in 1943 at Indian statistical institute and he started working out for finite samples and it led to the famous lower bound by rao; however, at the same time the result was also proved by Frechet in 1943, by Cramer in 1946 therefore, it is now popularly called Frechet rao Cramer in equality.

Now, once again in order to prove this, we need certain regularity conditions they are known by the name Wolfowitz regularity conditions named after the statistician Jacob Wolfowitz. So, as before we have a random sample be a random sample from a distribution having p d f and of course, it could be p m f, f x theta with respect to say measure mu. So, we assume the usual conditions for the existence of a density function or the mass function etcetera. Now, an estimator delta X is to be considered for the parameter theta. We make the assumptions that theta lies in an open interval of the real line. The derivative of the density or the mass function exists and of course, for all X or for almost all X.

The integral I have used a more general notation, because if you this discrete this it will be replaced by summation I have written here d mu. So, that takes care of both the cases. So, this is a enfold integral or summation, this can be differentiated under the integral sign for any delta, such that this is an integrable function; that means, this integral exists; that means, for any integrable function its expectation should be or its integral should be differentiable. So, that the above integral exists. This is positive for all theta, once again this is related to the fisher's information measure.

(Refer Slide Time: 06:45)



Under these conditions, we have the following inequality. I will call it Frechet-rao-Cramer inequality, because Frechet's paper appeared in 1943, Rao's paper appeared in 1945, Cramer's paper appeared in 1946. So, they all seem to have done it independently. Under assumptions 1 to 4, if expectation of delta X is equal to theta plus b theta, then variance of delta is greater than or equal to 1 plus, b prime theta whole square divided by n times expectation del by del theta log of f X theta whole square.

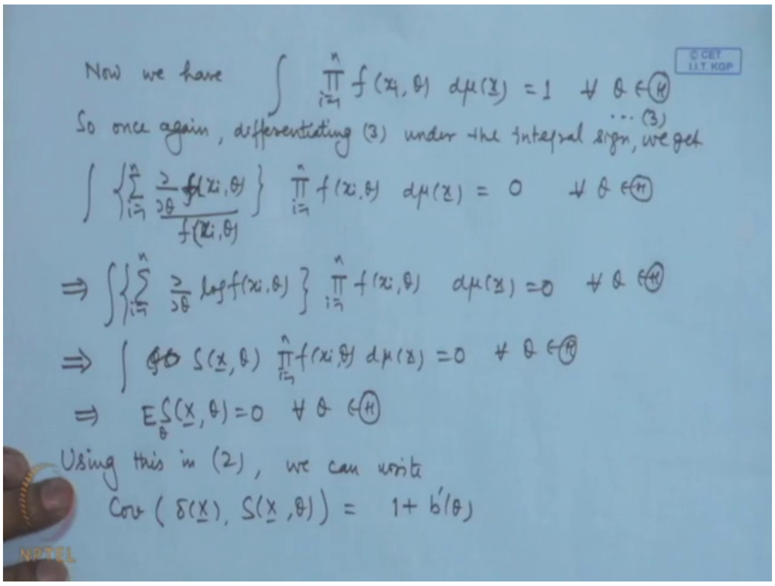
Firstly, let us look at the proof of this. So, what we are doing is that for an estimator delta we are providing the lower bound for the variance, this right hand side you can see it is not dependent upon the choice of the estimator that we have chosen; that means, any estimator of any unbiased estimator of theta plus b theta, we will have the minimum variance which will be greater than or equal to this, because this is the lower bound. So, it may be attained or it may not be attained let us look at the proof of this result first of all.

So, expectation of delta X is equal to theta plus b theta. Now, this is of course, true these statements are true for all theta. Now, we are assuming that we can differentiate under the integral sign. So, this is delta product f of X i theta, d mu X now, this denotes d mu X 1, d mu, X 2 d mu, x n this is equal to theta plus b theta for all theta differentiating under the integral sign. Let me again emphasize that this integral is a generalized label still just integral; that means, if we are dealing with the discrete distributions then this will be replaced by the summation. So, this is delta X 1, X 2, X n product of f of X I, theta d mu X 1, d mu X 2, d mu X n.

So, this is a enfold integral. So, if you differentiate with respect to theta, we will get delta X now, derivative of the product that you can easily write as sigma del by del theta log of f X i theta multiplied by product f X i theta, d mu X that is equal to 1 plus b prime theta. Now, we use some notation this term I call say s X theta, then I am getting delta X into s X theta into the joint distribution of X 1, X 2, X n and d mu X i, d mu X 1, d mu X 2, d mu X n. So, this we can write as, expectation of delta X into S X theta it is equal to 1 plus b prime theta.

Now, what we can see that this term if we look at this, we have made the assumption here that for any function delta for which this integral exist this can be differentiated. So, if we look at this particular term that is S X theta, then expectation of S X theta can also be differentiated under the integral sign, if we look at that then this is going to be 0 let us see this let me give this two here.

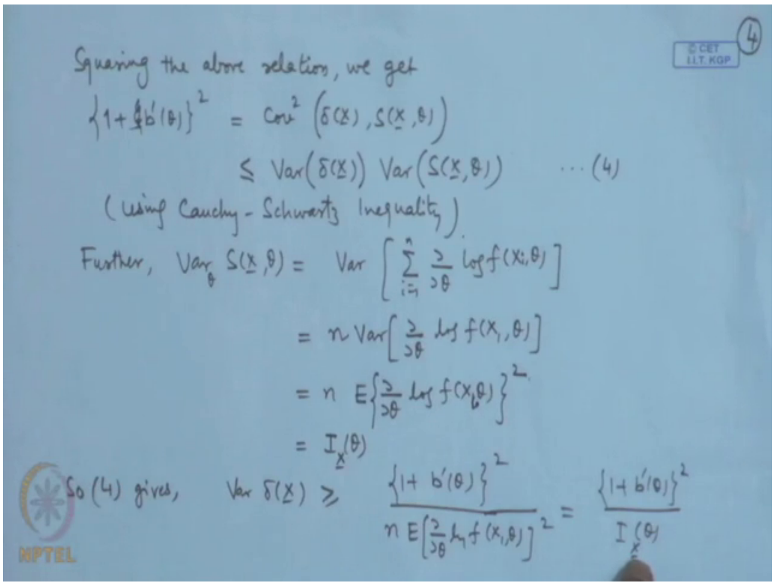
(Refer Slide Time: 11:43)



Now, we have the integral of the distribution of X 1, X 2, X n, equal to 1, by the property of the distribution that the integral or the summation should be equal to 1 over the whole range. So, once again if we differentiate, let me call it relation three. Under the integral sign, we get sigma del by del theta f I, f X i theta into product of f X i theta, see if we differentiate one particular term then other will be there. So, we can keep that also and then divide by that. So, this becomes sigma del by del theta f X i theta by f X i theta product f X i theta, d mu X is equal to 0 now, this term I can write as del by del theta log of f X i theta. Now, compare this here we defined S X theta to be sigma del by del theta log of f X i theta and this is the term.

So, what we have got here, we have got integral of S X, S x theta product f X i theta, d mu X is equal to 0; that means, expectation of S X theta is 0, if expectation of a random variable is 0, then expectation of that random variable multiplied by another will be equal to the co-variance term. So, we can say that using this in two, we can write that co-variance between delta X and S X theta is equal to 1 plus b prime theta.

(Refer Slide Time: 14:59)



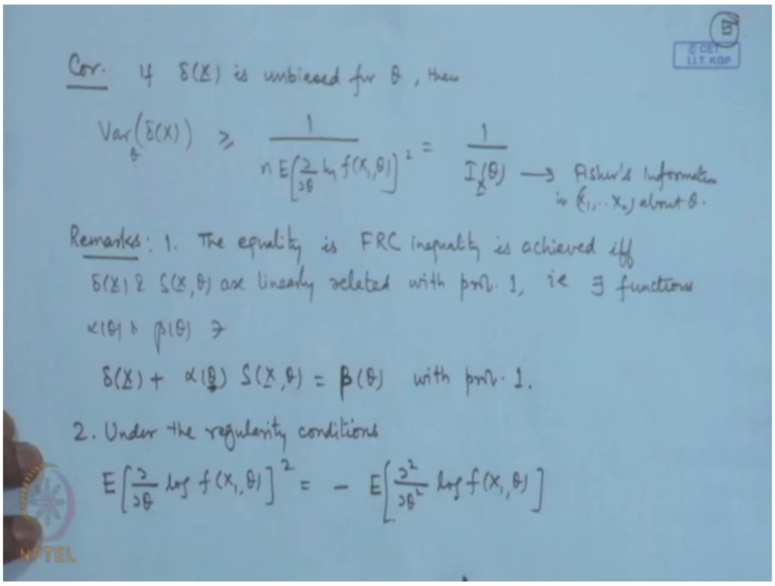
Now, this relation we squire rate. Squaring the above relation, we get 1 plus b prime theta square is equal to co-variance square delta X, S X theta. Now, co-variance square this is less than or equal to the variance of delta into variance of S X theta, if we use Cauchy-Schwarz inequality. So, this is less than or equal to variance of delta X into variance of S X theta this is true in general, let me say it here using Cauchy-Schwarz inequality. Now, once again since expectation of S X theta is 0 variance is nothing, but expectation of S square or we can also say that, variance of S X theta now, that is equal to variance of sigma del by del theta log of f X i theta. Now, this is variance of a sum.

Now, each term in the sum involves each X i, X i, are independent and identically distributed random variables. So, this becomes nothing, but the n times we can say variance of del by del theta log of say f X 1 theta since, expectation of del by del theta log f X theta is 0 this is nothing, but expectation of del by del theta log f x theta square. So, this is equal to n times expectation del by del theta, log of f X theta square. So, if we are using the notation I theta for this term then this is nothing, but the fisher's information. In the sample we can say fisher's information contained in the full sample.

So, this we can then write four, here we are having variance delta X greater than or equal to 1 plus b prime theta whole square divided by this and that term is this. Variance of delta X greater than or equal to 1 plus b prime theta see this will be whole square here, divided by n times expectation del by del theta log of f x 1 theta whole square, which we can also write as 1 plus b prime theta square by I theta in the sample, this means the random sample is X 1, X 2, X n. So, this is exactly the statement of the Cauchy-Schwarz of the Frechet-rao-Cramer inequality.

Now, we can look at the various ramifications of this. First of all in the assumption we have taken, the delta estimator to have expectation theta plus b theta, suppose our parameter of interest is theta and delta is an unbiased estimator then b theta will be 0, if b theta is 0 then this term will vanish. So, the lower bound will come as simply 1 by the information or 1 by n times expectation del by del theta log of f X theta.

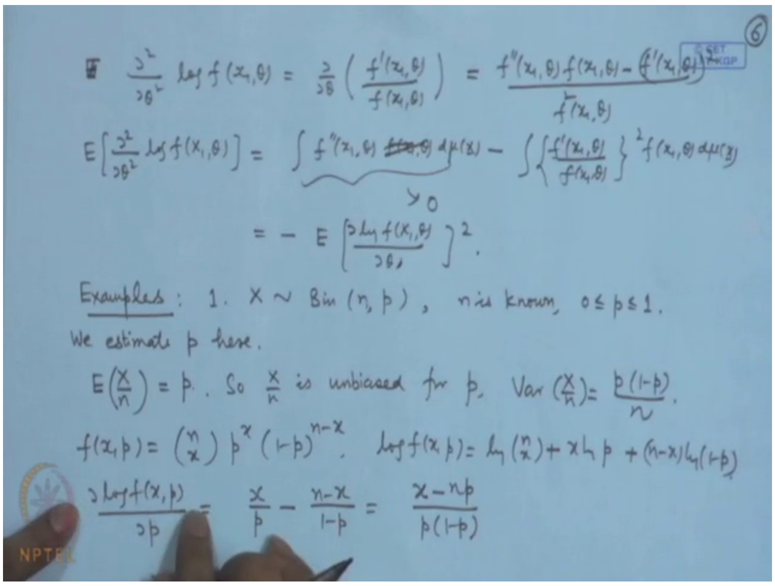
(Refer Slide Time: 19:34)



So, we have the following case as a corollary I write, if delta X is unbiased for theta then variance of delta X is greater than or equal to 1 by n times expectation del by del theta log of f x 1 theta whole square that is 1 by I theta. This term as I have defined fisher's information in X 1, X 2, X n, about theta another point that let us see the Rao-Cramer inequality, that we have proved the proof used Cauchy-Schwarz inequality. Now, Cauchy-Schwarz inequality has a condition for the equality, also when is that true the equality is true when delta and S are; that means, they are linearly related we can say that S is a linear function of delta or delta is a linear function of S since, here the random variables are involved.

We have to say that the they are linear functions with probability one. So, we can say as a remark the equality in FRC inequality is achieved, if and only if delta X and S X theta are linearly related with probability 1, that is there exist functions say alpha theta and say beta theta such that, we can say delta X plus alpha theta S X theta is equal to say beta theta with probability 1. Now, another point I have been using that expectation of del by del theta log f x theta square and earlier, I wrote this also as minus expectation del 2 by del theta square log f x theta. Now, that is true provided the regularity conditions are satisfied. So, let me prove that also here. Under the regularity conditions, under the regularity conditions expectation of del by del theta log of f X theta square is equal to minus expectation del 2 by del theta 2 log of f X theta.

(Refer Slide Time: 23:21)



So, let us look at the proof of this. Expectation of see, we have to consider the second derivative here. So, let us write this del 2 by del theta square log of f X 1 theta that is equal to del by del theta of first derivative. Now, the first derivative is nothing, but f prime by f. So, if you differentiate this you will get second derivative, here multiplied by f minus derivative of this and this. So, that becomes square divided by f X 1 theta square. So, if we consider expectation of this, that is equal to integral of f double prime X 1 theta f X 1 theta d.

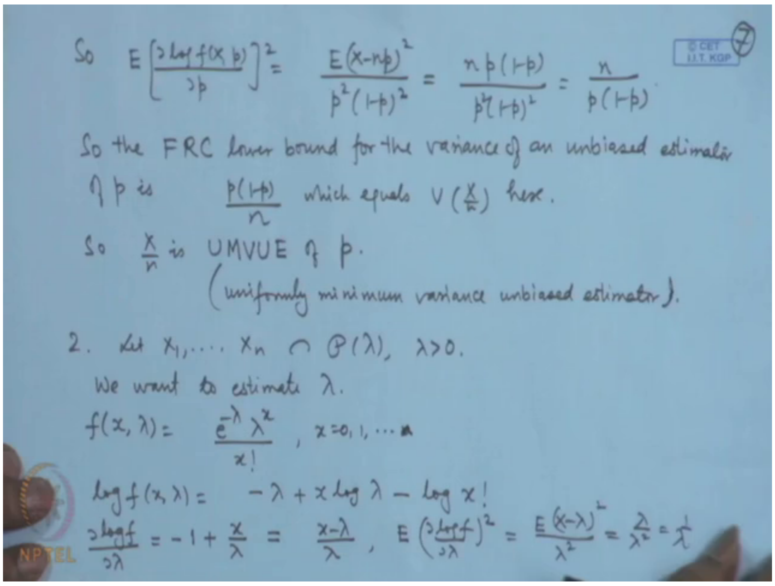
So, this will be canceled out, because when we multiply by f X 1 theta f X 1 theta and f X 1 theta square that will cancel out minus, second term will become f prime X 1 theta by f X 1 theta whole square f X 1 theta d mu X. Now, this term is 0 because of the assumption, because integral f X 1 theta d mu X is equal to 1. So, you differentiate under the integral sign. So, this becomes 0. So, this is nothing, but minus expectation of del log f X 1 theta by del theta whole square. So, these are two alternative ways of evaluating this fisher's information measure. Now, let me give examples of the situations we had the lower bound is attained and also the examples where the lower bound is not attained. Certainly whenever, the lower bound will be attained the unbiased estimator will become minimum variance unbiased estimator, because it is attaining the lower bound.

So, there cannot be another unbiased estimator which will have the variance smaller than this bound. So, this is one nice way of proving that a given estimator is minimum variance unbiased estimator; however, in the case when it is not attained, then it is difficult to proof the minimum variance unbiased estimator using this approach, for that we will take up another case or another approach here. So, let me start with the some of the standard distributions, let us consider say binomial distribution with parameters n and p, where n is known. So, the parameter is actually p and p x any value between 0 and 1.

So, we have to consider the estimation of p here. Now, easily you can see that X by n is an unbiased estimator of p, X by n is unbiased for p and also let us look at what is variance of X by n variance of this is simply p into 1 minus p by n. Now, let us look at the lower bound, here if it is unbiased then the lower bound is simply equal to 1 by the information measure. So, here we can calculate this the density function is n c x, p to the power x into 1 minus p to the power n minus x. So, we take log of this that is equal to log of n c x plus X log p plus n minus x log 1 minus p.

So, derivative of this with respect to p will give x by p minus, n minus x by 1 minus p, which we can write as x minus n p divided by p into 1 minus p. So, in order to apply the lower bound, we calculate the information and the information term is equal to n times expectation del by del theta log f x 1 theta square since, in this case we have only one observation. So, n will not be there we simply calculate this. So, we have already evaluated the derivative del log f x p by del p now, we square rate it and then take the expectation.

(Refer Slide Time: 28:34)

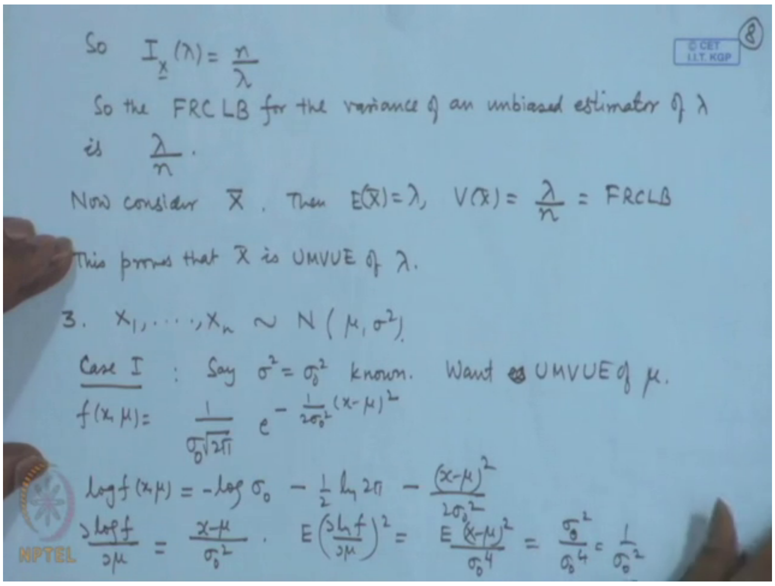


So, that gives us expectation del log f x p by del p whole square that is equal to expectation of X minus n p square divided by p square into 1 minus p square. Now, this is nothing, but the variance of X that is n p into 1 minus p in a binomial distribution. So, you get it as n by p into 1 minus p. So, the FRC lower bound for the variance of an unbiased estimator of p is p into 1 minus p by n. Now, in this particular case you observe, here variance of X by n was equal to p into 1 minus p by n which equals variance of X by n here.

So, X by n is uniformly minimum variance unbiased estimator of p. So, that is uniformly minimum variance unbiased estimator of p. So, you can see here the method is quite useful in actually proving that a given estimator is UMVUE are not. Now, let us take say poisson example. So, suppose we have a random sample from poisson distribution with the parameter lambda. So, naturally we want to estimate lambda now, let us consider the density function e to the power minus lambda, lambda to the power x by x factorial, log of f that is equal to minus lambda plus X log of lambda minus log of x factorial. So, if we consider the derivative of this with respect to lambda, then we get minus 1 plus x by lambda, that we can write as x minus lambda by lambda.

So, expectation of del log f by del lambda square that will be equal to expectation of x minus lambda square by lambda square. Now, in the poisson distribution case expectation of x is lambda therefore, this is nothing, but the variance and this is also lambda. So, this is lambda by lambda square that is equal to 1 by lambda that gives us.

(Refer Slide Time: 32:18)

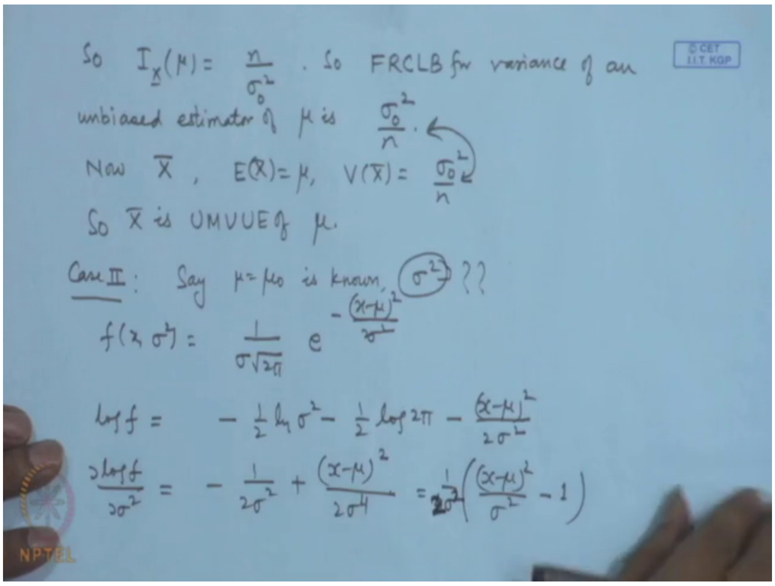


So, you get here the information as n by lambda. So, the FRC lower bound for the variance of an unbiased estimator of lambda is lambda by n. Now, consider say X bar, then expectation of X bar is lambda, what is variance of X bar? Variance is equal to lambda by n, which is equal to this FRC lower bound, this proves that X bar is UMVUE of lambda. In this particular case in the poisson example, I had given several unbiased estimator for s square I had given X 1 plus X 2 by 2 I had considered each X i is also unbiased for lambda, but you can see that among all of them X bar will be preferred, because this is the uniformly minimum variance unbiased estimator.

Let us take another popular example, that is the normal distribution. So, let us take say X 1, X 2, X n, following normal mu sigma square. Now, as before we will consider different cases sigma square is equal to sigma naught square, that is known. In that case we want estimate of say UMVUE of mu. So, if we write down the distribution here 1 by sigma root 2 pi. So, here it will become sigma naught e to the power minus 1 by 2 sigma naught square x minus mu whole square. So, log of f is equal to minus log of sigma naught minus half log 2 pi minus x, minus mu square by 2 sigma naught square.

So, if we consider derivative of this with respect to mu, we get simply x minus mu by sigma naught square. So, expectation of del log f by del mu whole square, that is equal to expectation x minus mu square by sigma naught to the power 4. Once again in the normal distribution this is reducing to the variance term, that is expectation of x minus mu square is variance, that is sigma naught square by sigma naught to the power 4 that is 1 by sigma naught square.

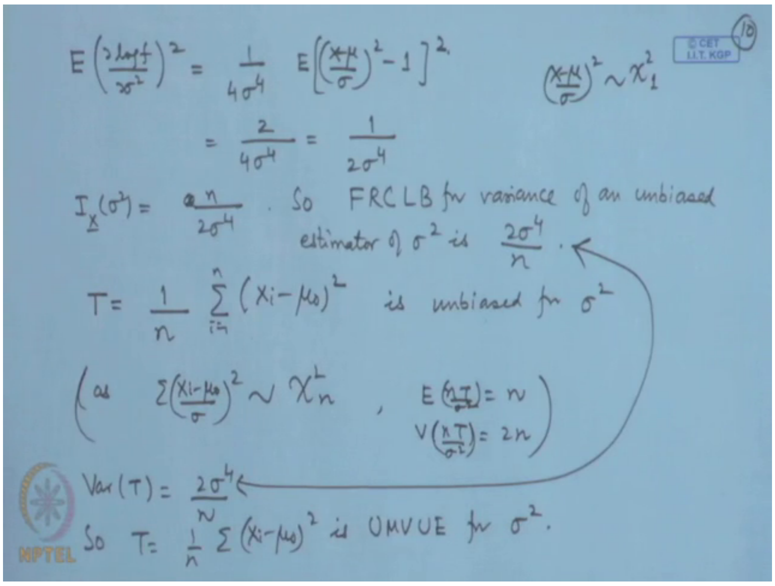
(Refer Slide Time: 35:44)



So, information contained in this will be n by sigma naught square in the sample. So, the Frechet-Rao-Cramer lower bound for variance of an unbiased estimator of mu is sigma naught square by n. Now, if you consider say X bar then expectation of X bar is mu and what is variance of X bar? That is sigma naught square by n that is equal to this value. So, X bar is UMVUE of mu, let us take another case when say mu is known and we want to estimate say mu is equal to mu naught is known and we want sigma squares estimator.

So, here the density function will be written as a function of sigma square 1 by sigma root 2 pi, e to the power minus x minus mu square by 2 sigma square. So, log of f becomes minus half log sigma square minus half log 2 pi minus x minus mu square by 2 sigma square. So, differentiation of this with respect to sigma square gives minus 1 by 2 sigma square plus x minus mu square by 2 sigma to the power 4, which I can write as X minus mu square by sigma square minus 1, 1 by 4 sigma 1 by 2 sigma to the power 1 by 2 sigma square.

(Refer Slide Time: 38:30)



So, if we consider expectation of del log f by del sigma square whole square that is equal to 1 by 4 sigma to the power 4 expectation of X minus mu by sigma whole square minus 1 whole square.

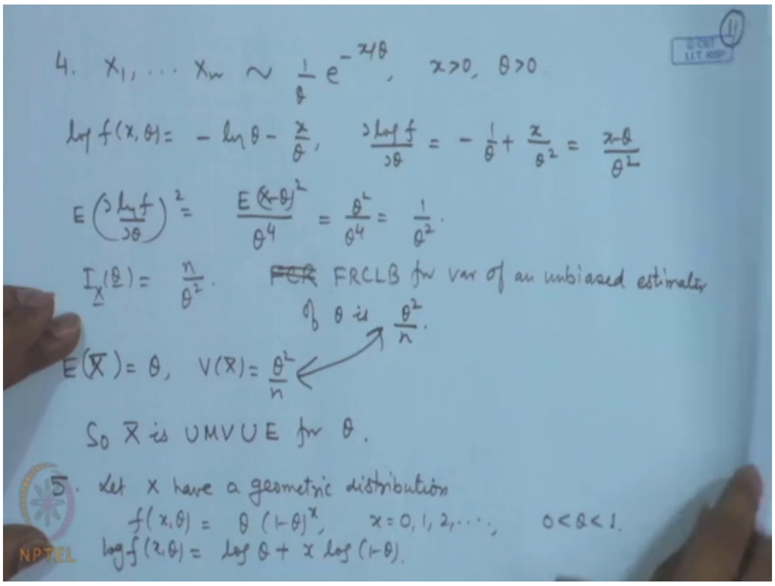
Once again you look at this, X minus mu by sigma is a standard normal variable X minus mu by sigma square it will follow chi square 1. So, expectation of this is equal to 1 and therefore, this term reduces to the variance. So, variance is twice the degrees of freedom that is equal to 2 by 4 sigma to the power 4. So, you get 1 by 2 sigma to the power 4. So, the Fisher's information in this problem will be 2 n by 2 sigma to the power 4. So, the Frechet-Rao-Cramer lower bound for variance of an unbiased estimator of sigma square is 2 sigma to the power 4 by n.

Now, in this case let us consider see the maximum likelihood estimator for example, or the method of moments estimator. So, that would be for example, 1 by n sigma X i minus mu naught square. So, this is now you can see here X i minus mu naught by sigma that will follow standard normal. So, some of squares will be chi square n. So, expectation of that is n. So, this divided by n will have expectation 1. So, if you multiply by sigma square, we will get sigma square. So, this is unbiased for sigma square, because we can see here as sigma X i minus mu naught by sigma whole square that follows chi square on n.

So, expectation of n T is equal to n and variance of n T sorry this divided by sigma square and this divided by sigma square that will be equal to 2 n. So, we will get variance of T as equal to twice sigma to the power 4 by n, because this will go here and n square will come to below. So, we will get 2 sigma to the power 4 by n, which is same as this value once again here. So, T is equal to 1 by n sigma X i minus mu naught square it is the minimum variance unbiased estimator for.

Obviously, you can see here that if mu was not known, then you could not have used this estimator. So, this solution is specific to this problem, that is when we are dealing with one parameter case mu naught is known to us.

(Refer Slide Time: 42:03)



Let us consider say a random sample from exponential distribution with mean say theta. Now, in this case the density function is this. So, log of the density function is minus log of theta minus x by theta. So, the derivative with respect to theta will be minus 1 by theta plus x by theta square that is x minus theta by theta square.

So, expectation of del log f by del theta whole square, that is expectation of x minus theta square by theta to the power 4. In the exponential distribution with mean theta variance is equal to theta square. So, this term becomes theta square by theta to the power 4, that is equal to 1 by theta square. So, the information in the sample about theta is n by theta square and the lower bound Frechet-Rao-C ramer lower bound for variance of an unbiased estimator of theta is theta square by n.

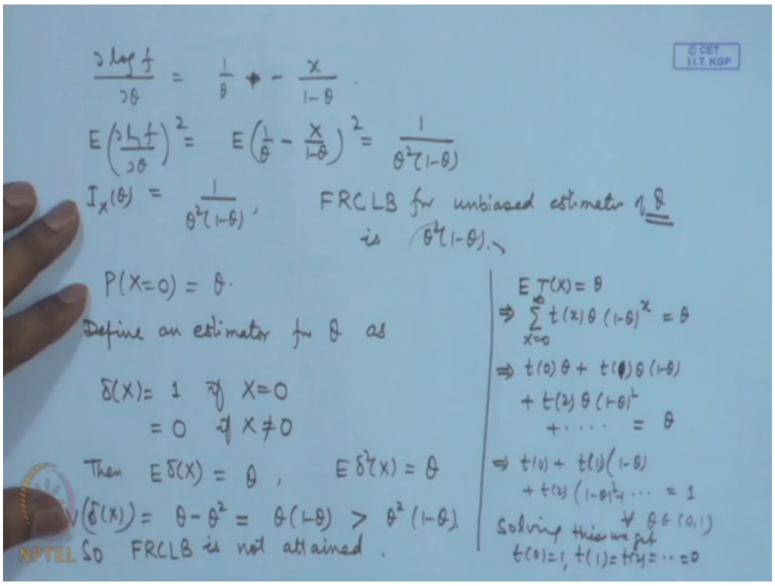
If we consider X bar then expectation of X bar is equal to theta and variance of X bar is equal to theta square by n. So, this will prove that X bar is minimum variance unbiased estimator for theta in the case of negative exponential distribution, it is not necessary that the lower bound is always attained. In fact, if you see carefully in each of these problems we have calculated the derivative here. So, s function you can see here for example, s would have become here n by minus n by theta plus sigma X i by theta that is n X bar.

So, this is linearly related with X bar and therefore, X bar must attain the variance lower bound for its expectation. If you see the previous problem for the estimation of sigma square, here del log f by del sigma square is this function. So, if we look at s function S X i sigma square that would have become minus n by 2 sigma square, sigma X i minus mu square by something which is a linear function of sigma X i minus mu whole square and therefore, it is natural that sigma X i minus mu naught whole square by n will attain the lower bound here.

So, if you see the estimation of the poisson distribution case, here the derivative is equal to minus 1 plus x by lambda. So, if you look at s function it would have become minus n plus sigma X i by lambda, which is again linearly related with X bar therefore, X bar must the lower bound for the variance of its unbiased estimation. So, in all these problems it is naturally coming let me take another example, where it may not be natural and therefore, the lower bound may not be attained.

Let us consider say, let x have a geometric distribution and we consider the following form theta into 1 minus theta to the power x, where theta is any number between 0 and 1. So, here the problem is of estimation of theta. So, let us look at log of f x theta, that is equal to log of theta plus x times log of 1 minus theta.

(Refer Slide Time: 46:55)



So, if we consider del log f by del theta we get 1 by theta plus with a minus sign x by 1 minus theta here. So, if we look at the expressions here expectation of del log f we can use the moment structure of the geometric distribution. If we use that this is equal to expectation of 1 by theta minus X by 1 minus theta whole square. So, after simplification this turns out to be 1 by theta square into 1 minus theta. So, since I have taken only one observation here the information will remain the same and the lower bound for unbiased estimator of theta is theta square into 1 minus theta. Now, here theta is not the mean actually, if you look at the mean of this distribution that will be 1 minus theta by theta.

So, X will attain the lower bound for that for the variance of an unbiased estimator for 1 minus theta by theta, but suppose we are considering estimation of theta, if we are estimating theta here, then it will not be attained. So, you can see here, what is the interpretation of say theta here, theta is the probability of X is equal to 0, because if in the probability mass function we put X equal to 0 here I get theta. So, if I define an estimator for theta as say delta X is equal to 1 if X is equal to 0 it is equal to 0 if X is not equal to 0; that means, if X equal to 1 2 and so on.

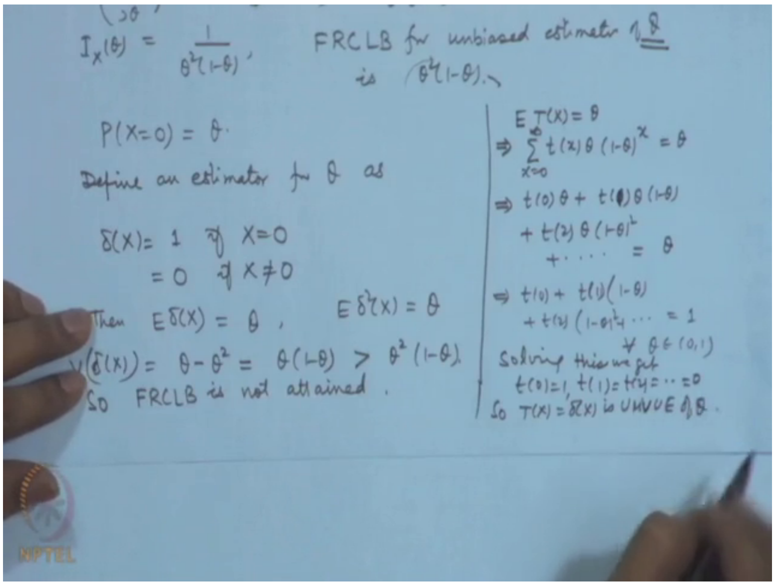
Then expectation of delta X will be equal to 1 into probability X is equal to 0 plus 0 into probability X not equal to 0; that means, it will be simply equal to theta and what is expectation of say delta square X, that will also be theta. So, variance of delta X that will be equal to theta minus theta square that is equal to theta into 1 minus theta. Now, here if you compare with this lower bound, here lower bound is theta square into 1 minus theta and theta is any number between 0 and 1. So, this one will be naturally bigger than this. So, the lower bound is not attained. So, we do not know whether delta is minimum variance unbiased estimator here, we may try another approach here.

Let us consider expectation of t x is equal to theta, if we consider this then we will get sigma t x into theta into 1 minus theta to the power x is equal to theta as x varies from 0 to infinity, that will give me t 0 into theta plus t 0 into theta into 1 minus theta plus sorry plus t 1 plus t 2 into theta into 1 minus theta square and so on. is equal to theta. Now, you look at this, what we are getting that coefficient of theta here if you see. So, this you can cancel out actually, t 0 plus t 1 into 1 minus theta plus t 2 into 1 minus theta square and so on is equal to 1 this is true for all theta belonging to the interval 0 to 1.

Now, if you see this carefully what is the solution, see if you look at the coefficient of say theta here theta will have coefficient t 1 see for example, if I look at the coefficient of the constant term, constant term is t 0 plus t 1 plus t 2 and. So, on that should be equal to 1, if you take coefficient of theta then you get minus t 1 minus 2 t 2, then in the next 1 also minus 3 t 3 and so on, that should be equal to 0 then if you look at the coefficient of theta square you will get t 2 then here in the second one, it will become 3 t 3 and. So, on.

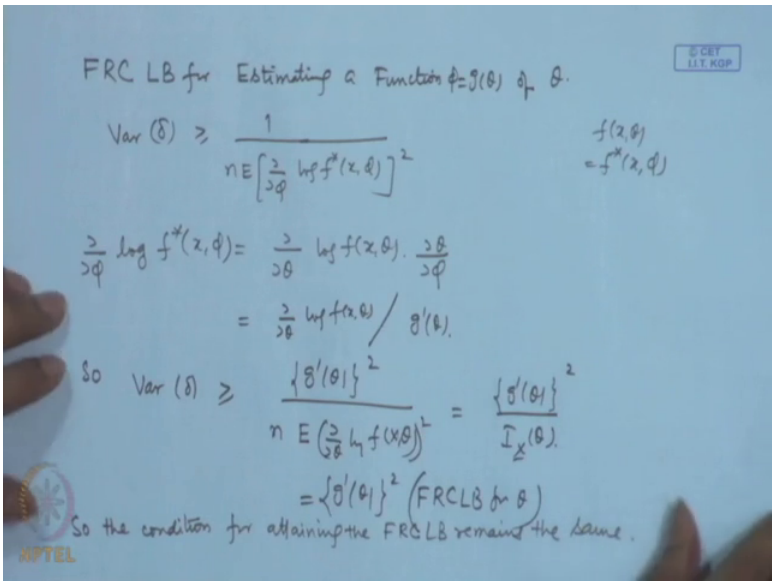
So, if you solve this solving this, we get t 0 is equal to 1 t 1, t 2 and so on is equal to 0 which is nothing, but this t function then it is becoming same as this.

(Refer Slide Time: 46:55)



So, we have proved otherwise, that t x that is equal to delta x is UMVUE, because this is the only unbiased estimator, which we obtained through solving the equation itself; however, using the method of lower bounds we are not able to prove this result here. Now, many times we may not be interested directly in the theta itself, we will be interested in some function say g theta of theta. In that case what we can do is we can modify this lower bound formula like.

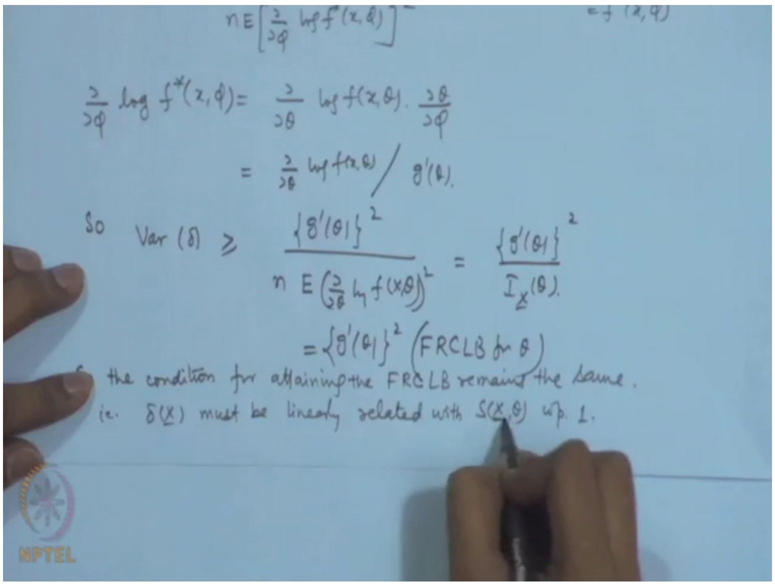
(Refer Slide Time: 53:27)



So, FRC lower bound for estimating a function g theta of theta. So, let me call it phi. So, we will write variance of delta greater than or equal to 1 by n times expectation del by del phi log of f star x phi, because f x theta density now I am writing as f star x phi, because we have substituted theta by g inverse phi in whatever form we are able to do that. So, if you look at this derivative here del by del phi, log of f star x phi, you can apply the chain rule you can write it as del by del theta log of f x theta into del theta by del phi, this you can write as del by del theta log of f x theta divided by g prime theta. So, if you substitute this function here, we get variance of delta greater than or equal to g prime theta square divided by n times expectation of del by del theta log of f x theta whole square, that is equal to g prime theta whole square by the information in the sample about theta.

That means, if we have the lower bound for variance of an unbiased estimator of theta then from there we can derive for any other function what we have to do, we have to multiply by the lower bound by g prime theta square. So, this we can say it is equal to g prime theta square into the Frechet-Rao-Cramer lower bound for theta. So, this new formula can be obtained. Moreover, the condition for obtaining the lower bound for attaining the lower bound that will remain the same, because the condition is coming only from the Cauchy-Schwarz inequality, which was dependent upon the estimator being linearly related with s x theta. Now, the g theta function does not affect that thing. So, the condition for the condition for attaining the f remains the same.

(Refer Slide Time: 53:27)



That is your delta x must be linearly related with S X theta with probability one. Tomorrow's class, we will be considering further properties and further ramifications of this lower bound, as well as we will see some extensions there can be 2 types of extensions, one is the extension to the higher dimension; that means, if in place of 1 dimensional parameter I have several dimensional parameter, then what will be the form of the Row-Cramer inequality. Similarly, here we have used first order derivative in the lower bound now, if we consider second and higher order derivatives then the level of the inequality can be changed. So, they are generalization into another direction.

Another thing is that whenever we are considering differentiation in some sense, we are taking the limits suppose, we do not take the limits in place of that we write the difference for example, we are saying derivative. So, we are writing down the value of the function at two points theta and theta plus delta say. So, we consider the difference there and then look at the inequality, that inequality will be called the equality without the regularity condition, because when we are having regularity conditions then we are considering the derivative and other things, but if that is not satisfied then what. So, we will have another extension in that direction.

So, in the next lecture, we will be considering extensions to these things and then further applications of this that is all for today's lecture.