Statistical Inference

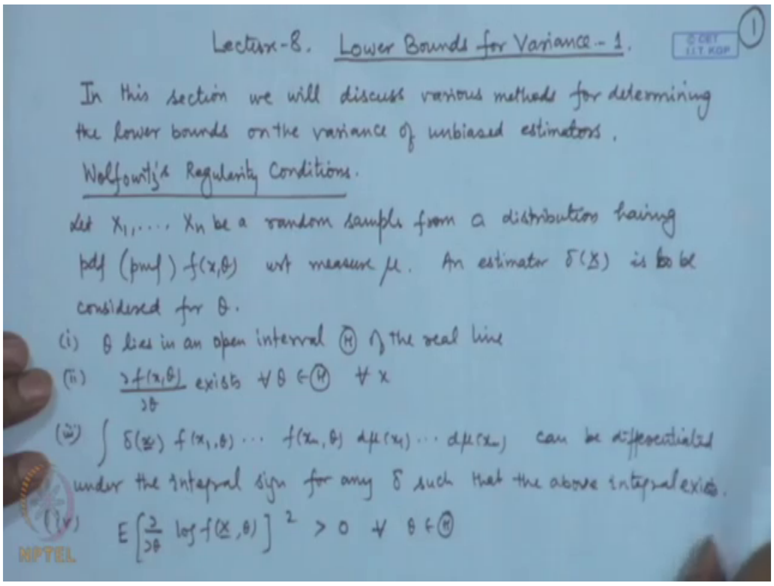
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 15
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 unbiasness 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 consider to be more stable in some sense.

Now therefore, we need to have an estimate of that 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: 01:26)



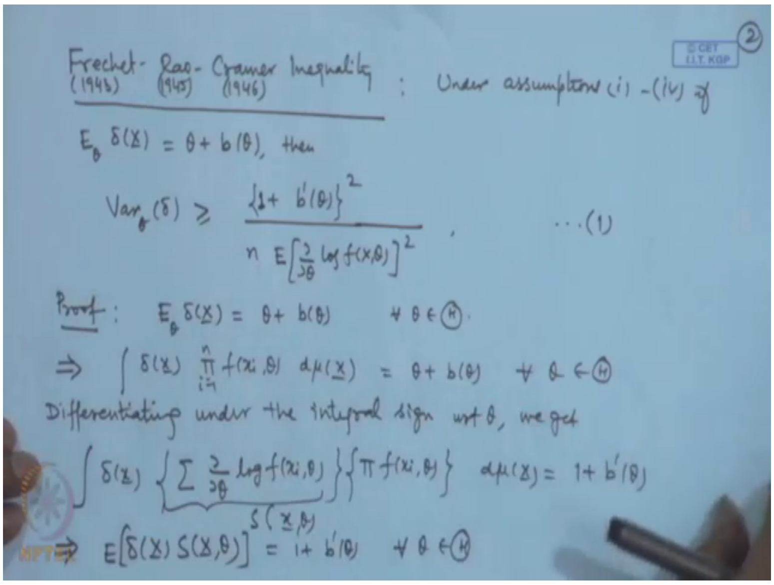
So, in this section, we will discuss various methods for determining the lower bounds on the variance of unbiased estimators. As we have seen in the case of maximum likelihood estimation in the last results that I gave that variance asymptotic variance of the maximum likelihood estimator was 1 by the 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 naught, that means, the Fisher information measure.

The question comes that whether similar result we can give for finite samples. Now, this is the precisely the question that was posed to Indian statisticians 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 Fisher in 1943, by Cramer in 1946, therefore, it is now popularly called Fisher, Rao, Cramer inequality.

Now, once again in order to prove this, we need certain regularity conditions they are known by the name Wolfowitz's wits regularity conditions named after the statistician Jacob Wolfowitz. So, as before we have a random sample be a random sample from a distribution having say PDF. And of course, it could be pmf f x, theta with respect to say measure mean. 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 exist, and of course for all x are for almost all x. The integral I used a more general notation, because if it is discrete this will be replaced by summation, I have written here d mu, so that takes care of both the cases. So, this is a n fold integral or summation. This can be differentiated under the integral sign for any delta such that this is an integral function that means this integral exists that means, for any integral 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 information measure.

(Refer Slide Time: 06:46)



Under these conditions we have the following inequality I will call it fresh Fisher Rao Cramer inequality. Rao Cramer inequality, because Fisher's is 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.

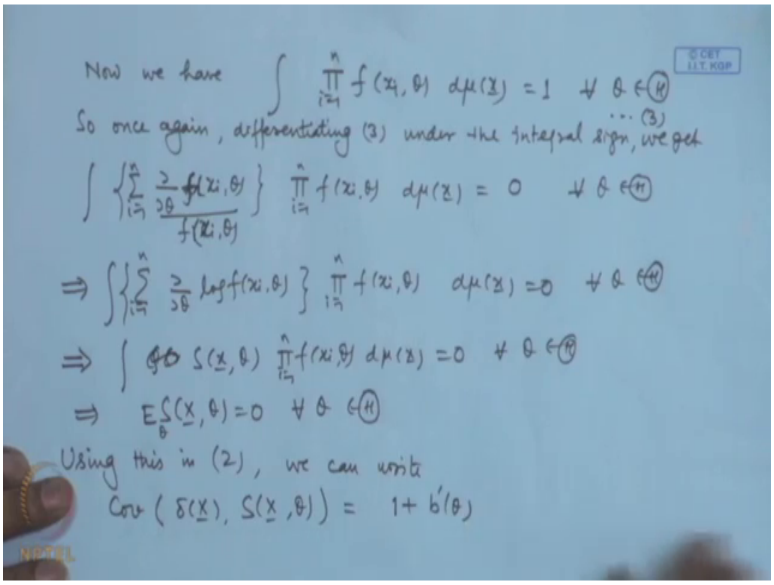
Firstly, let us look at the proof of this. So, what we are doing is that for an estimator delta, we are providing the lower bound for the variance. This right hand side you can see it is not dependent upon the twice of the estimator that we have chosen, that means, any estimator of any unbiased estimator of theta plus b theta will have the minimum variance which will be greater than or equal to this, because this is the lower bound. So, it may be attend or it may not be attend. 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 lebesgue 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 the n fold 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 this is 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 divine distribution of x 1, x 2, x n and d mu x id mu x 1 d mu x 2 direction. 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 exists 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 2 here.

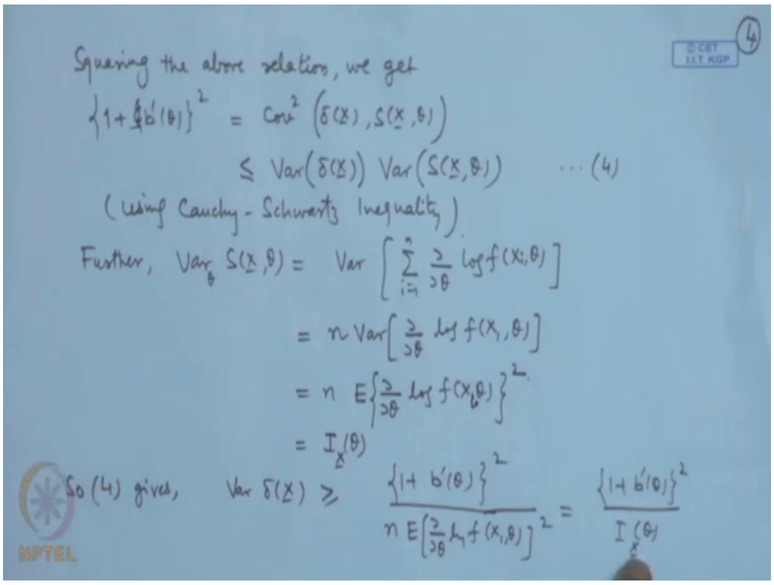
(Refer Slide Time: 11:43)



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 3 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 you 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 equal to the covariance term. So, we can say that using this in 2, we can write that covariance between delta x and S x, theta is equal to 1 plus b prime theta. Now, this relation we square it.

(Refer Slide Time: 15:03)



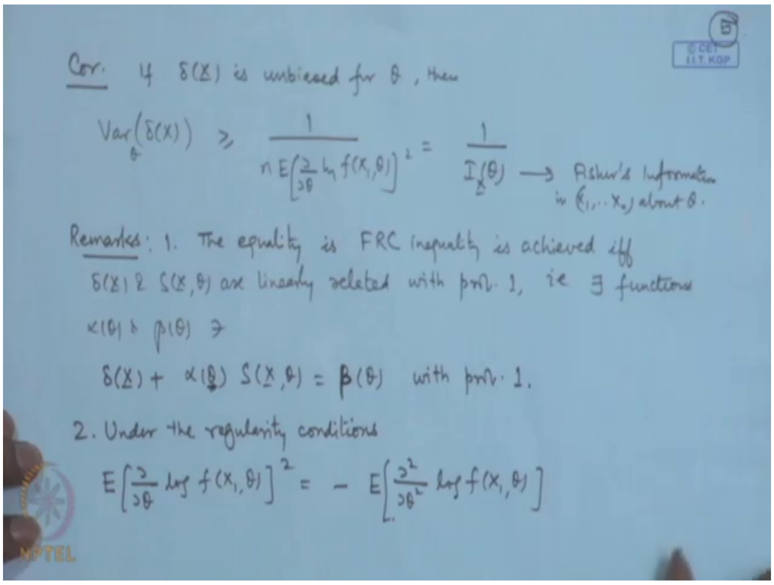
Squaring the above relation we get 1 plus b prime theta square is equal to covariance square delta x S x, theta. Now, covariance square this is less than or equal to the variance of delta into variance of S x, theta if we use Cauchy-Schwarz inequality. So, this is less than or equal to variance of delta 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 x 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's 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 by 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 4. Here we are having variance delta x greater than or equal to 1 plus v 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 Fisher, 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 here 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: 19:35)



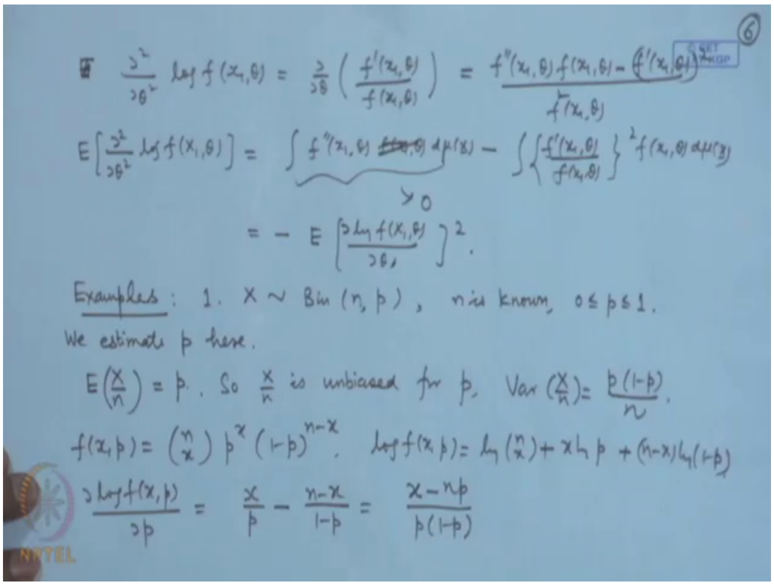
So, we have the following case. As a corollary I write, if delta x is unbiased for theta, then variance of delta x is greater than or equal to 1 by n times expectation del by del theta log of f x 1 theta whole square that is 1 by I theta, this term as I defined Fisher's information in x 1, x 2, x n about 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. Equality is true when delta and s are that means, they are linearly related you 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 there are linear functions with probability 1.

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 their 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: 23:23)



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 cancelled 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 dx.

Now, this term is 0, because of the assumption because integral f x 1 theta d mu x is equal to one. 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 where 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.

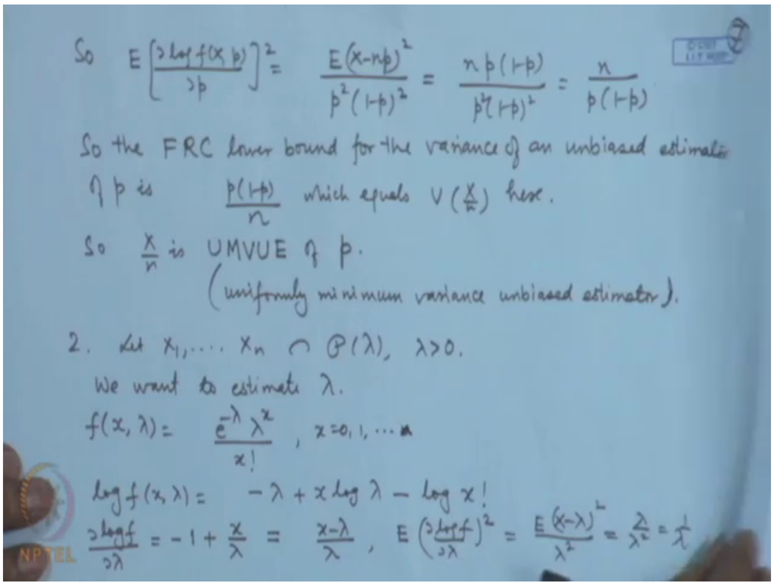
So, there cannot be an either and by the estimator which will have the variance is smaller than this form. So, this is one nice way of proving that a given estimator is minimum variance unbiased estimator. However, in the case when it is not attained, then it is difficult to prove the minimum variance 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, and p takes 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 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 m 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 it and then take the expectation, so that gives us expectation.

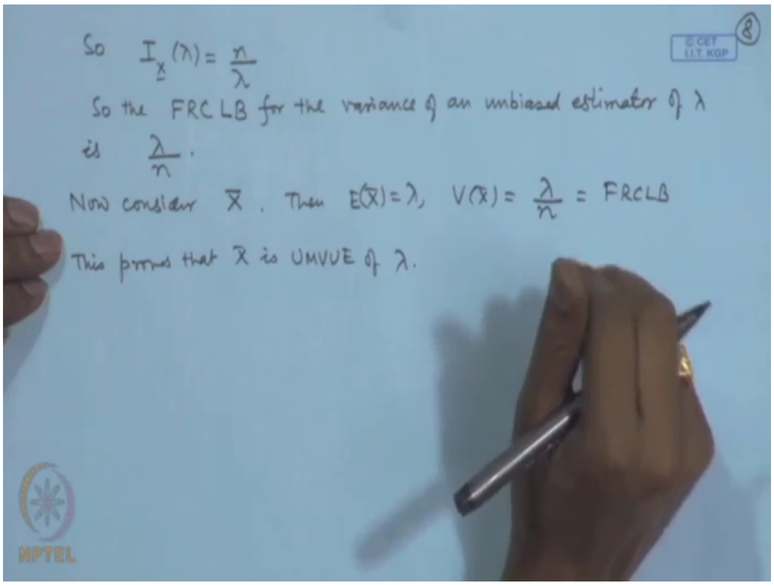
(Refer Slide Time: 28:36)



Log f x p by del p whole square that is equal to expectation of x minus n p square divided by p square and 2 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. So, you can see here the method is quite useful in actually proving that a given estimator is UMVUE naught.

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.

(Refer Slide Time: 32:19)



That gives us 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 estimators. For example, S square I had given x 1 plus x 2 by 2 I had considered each x i's 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.

That is all for today's lecture.