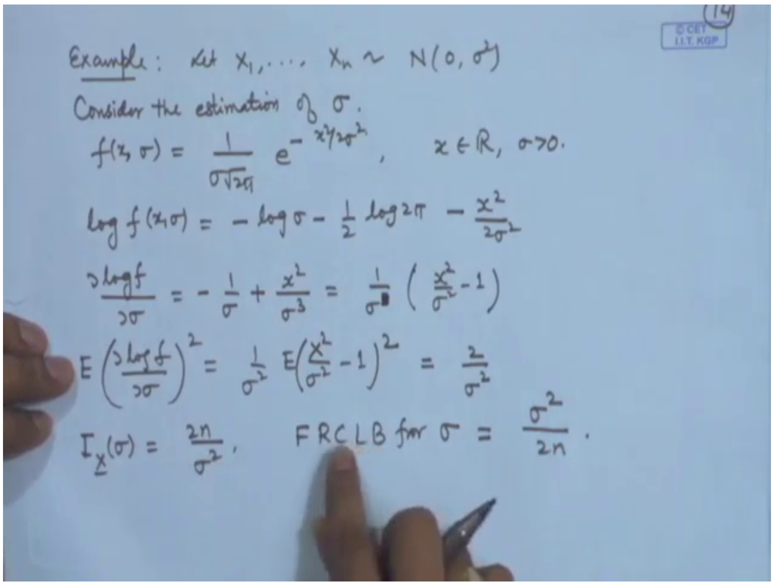
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 19
Lower Bounds for Variance – V

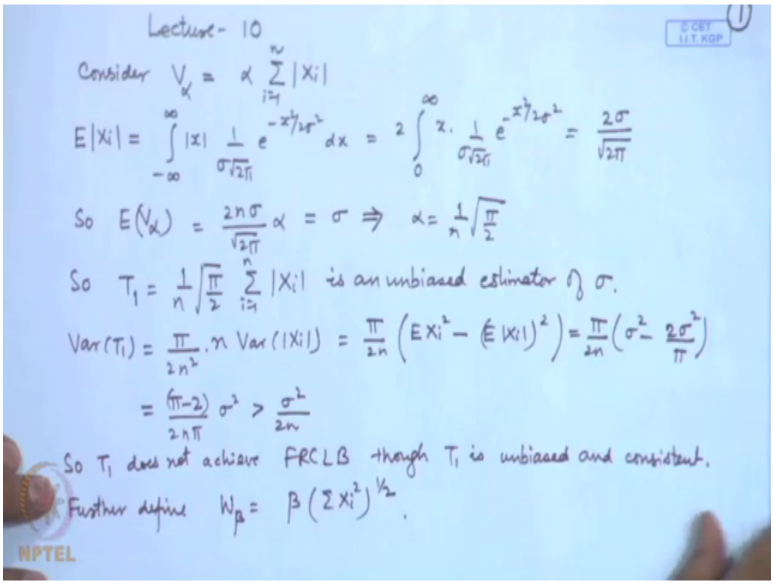
In the last class I have discussed one example, let me continue with that example.

(Refer Slide Time: 00:26)



We have a random sample from normal 0 sigma square distribution. And we are considering the estimation of sigma in place of sigma square. So, what I showed in the last class is that the Rao Cramer lower bound lower bound for estimation of sigma is sigma square by 2 n. Now, I will propose two estimators for the estimation of sigma and we will consider their variances and then we will see whether the FRC lower bound for them is attained or not. In fact, we have seen that for sigma square it is attained; now sigma is not a linear function of sigma square. Therefore, this bound may not be will not be attained. however, we will consider 2 examples.

(Refer Slide Time: 01:18)



So, let me take the first example. Consider an estimator of the form say V alpha is equal to alpha into sigma modulus of X i, i is equal to 1 to n. Now, if you consider say expectation of modulus x i, that is equal to integral from minus infinity to infinity. Modulus x 1 by sigma root 2 pi e to the power minus x square by 2 sigma square d x.

Now this is an even function so this will become 2 times 0 to infinity x 1 by sigma root 2 pi e to the power minus x square by 2 sigma square. Now this can be easily evaluated because derivative of e to the power minus x square by 2 sigma square is e to the power minus x square by 2 sigma square into x by sigma square. So, if you evaluate this integral this turns out to be simply 2 sigma divided by root 2 pi.

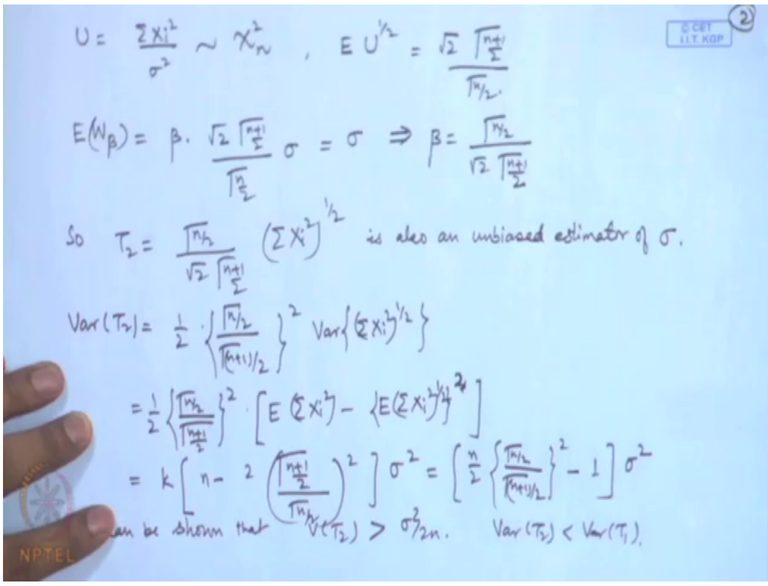
So, if we consider expectation of V alpha that will be equal to twice n sigma by root 2 pi alpha. Now, if we want that V alpha be an unbiased estimator of sigma then we substitute this equal to sigma; that gives the value of alpha is equal to 1 by n root pi by 2. So, what we are getting is that let me call this estimator as T 1, by substituting alpha is equal to this value.

That is 1 by n root pi by 2 sigma modulus of X i this is unbiased estimator of sigma. Let us look at the variance of T 1. So, what is variance of T 1? Variance of T 1 will become pi by 2 n square into n times variance of modulus X i; now this becomes pi by 2 n. Now variance of X i is expectation, modulus X i square that is expectation of X i square and minus expectation of modulus X i whole square.

Now, since we have considered here the normal 0 sigma square. So, expectation of X i square is nothing the variance that is sigma square. So, this value is equal to sigma square. And expectation of modulus X i we have just now calculated. So, if we substitute the square of that I get 2 sigma square by pi. So, this can be written as pi minus 2 by. So, what I am doing is a larger system pi minus 2 by pi and then 2 n is there. So, 2 n pi sigma square. Now, this can be shown that this is bigger than sigma square by 2 n..

Similarly, so we can say that T 1 does not achieve FRC lower bound. And if you look at the estimator here see the variance is certain term divided by n. So, as n tends to infinity this goes to 0 and it is unbiased. so, this is unbiased T 1 is unbiased as well as consistent. Let me define another estimator here let me call it say W beta that is equal to beta times sigma X i square to the power half.

(Refer Slide Time: 06:04)



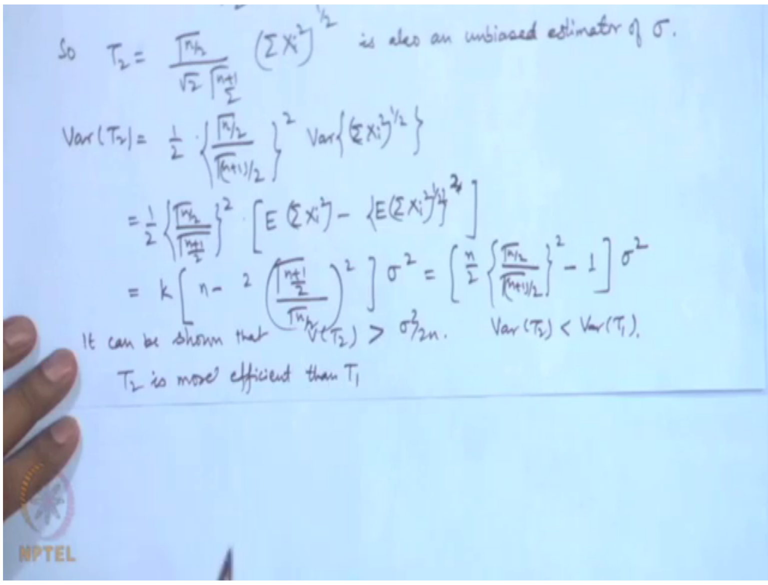
Now, if you want to evaluate the expectation of this we can consider. If X i’s follow normal 0 sigma square then X i by sigma that will follow normal 0 1. So, the sum of squares of a standard normal variables when they are independent is a chi square random variable. So, we get here that U is equal to sigma X i square by sigma square this follows chi square distribution on n degrees of freedom.

Now if I have a chi square then expectation of U will become root 2 gamma n plus 1 by 2 by gamma n by 2. Therefore, expectation of W beta that turns out to be beta times root 2 and here we will get gamma n plus 1 by 2 by gamma n by 2 sigma. Once again if I want this to be unbiased then I equate it to sigma ; that means, beta should be equal to gamma n by 2 divided by root 2 gamma n plus 1 by 2..

So, T 2 is equal to gamma n by 2 by root 2 gamma n plus 1 by 2 sigma X i square to the power half this is also an unbiased estimator of sigma. Let us look at what is variance of T 2; variance of T 2 is half gamma n by 2 divided by gamma n plus 1 by 2 whole square into variance of sigma X i square to the power half. Now, variance of sigma X i square to the power half. That is expectation of sigma X i square minus expectation of sigma X i square whole square;. sigma X i square to the power half whole square.

Now these terms we have already calculated so that becomes let me call it this some constant n minus twice gamma n plus 1 by 2 by gamma n by 2 whole square sigma square. That we can write after simplification as n by 2 gamma n by 2 divided by gamma n plus 1 by 2whole square minus 1 sigma square. It can be shown that variance of T 2 is greater than sigma square by 2 n. And variance of T 2 is less than variance of T 1. In fact, one can show that this also goes to 0 as n tends to infinity.

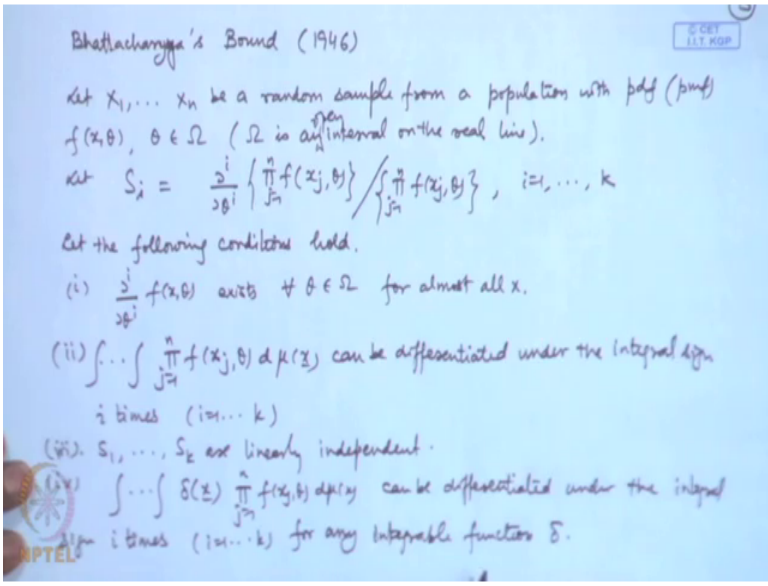
(Refer Slide Time: 09:29)



So, T 2 is more efficient than T 1. Now we have discussed in detail one lower bound for the variance of an unbiased estimator and this lower bound takes into account one derivative of the log of the density function.

Now, naturally there is a question whether one can further sharpen it or whether we can extended to multi parameter case or whether if the regulatory conditions are not satisfied; then this will be true or not. Fortunately in all the directions the extensions of this result have been done. So, let me discuss this here the first of this is known as Bhattacharyya bound.

(Refer Slide Time: 10:30)



So, this was proposed by a Bhattacharyya in 1946. Now, in the fisher Rao Cramer lower bound we had considered first order derivative and of course, second order derivatives condition was assumed. However, in the Bhattacharyya bound higher order derivatives are used. And therefore, we have to make the assumptions accordingly. So, once again as in the Rao Cramer lower bound let us consider the regularity conditions in the same way.

So, we have a random sample let X 1 X 2 X n be a random sample from a population. Now, again it may have a probability density function or probability mass function say f x theta; theta belonging to omega, where omega is a is an interval on the real line.

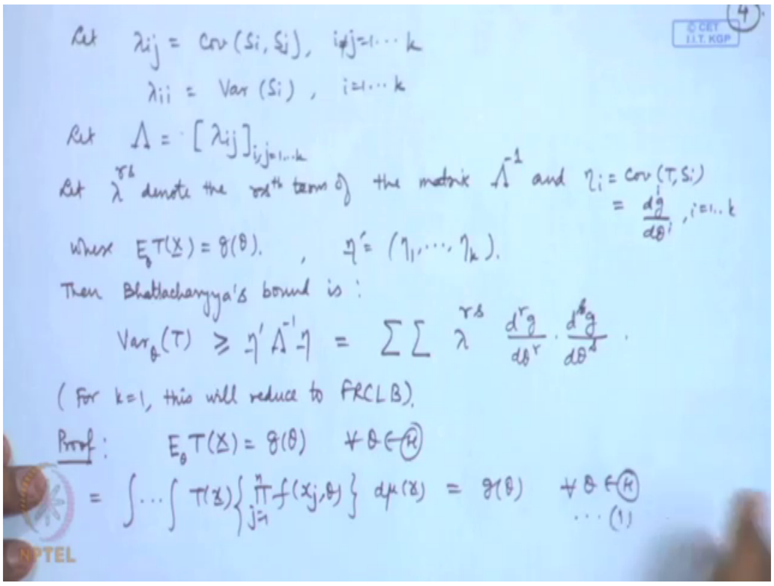
Let us define S i to be ith order derivative of the joint distribution divided by the joint distribution. You compare it with the Rao Cramer lower bound in the Rao Cramer lower bound we had first order derivative here. Now, I am defining higher order derivatives also. Because in the first order derivative it will become del by del theta of the density divided by the density that is del log of that. But here it is higher order here.

So, i is equal to 1 2 and so on suppose I am assuming up to order k let the following conditions hold. So, we have already assumed that the parameter spaces and interval in the real line let us consider open interval. Let us assume that the ith order derivative of the density exists for all theta for almost all x; by almost all x means that the set where this is not existing will have probability 0.

The density function once again I am writing this is a multifold integral and this is a generalized integral; that means, it takes care of the discrete case also in that case this will be summation. This can be differentiated under the integral sign at least i times. Let us define S i as this term. Then we assume that S 1, S 2, S k are linearly independent.

By linearly independent means that none of their can be expressed as function as linear combination of the others. And this manifold integral this also can be differentiated under the integral sign i times; i is equal to 1 to k for any integrable; that means, this should exist for any integrable function delta.

(Refer Slide Time: 15:23)



Let us define; let us define say lambda i j to be the covariance between S i S j for i j equal to 1 to k. So, if i is not equal to j then this will be covariance and lambda i i is variance of S i for i is equal to 1 to k. And let lambda be the matrix of lambda i j S for i j equal to 1 to k. Let us denote by lambda say r s denote the rsth term of the matrix lambda inverse.

Then and also we can write here let us write eta i vector to be covariance of T S i that is equal to dg d i g by d theta i. Now what is T here? T is an unbiased estimator of g theta. So, let us look at the problem here once again. We have a probability mass function or probability density function f x theta we have a random sample X 1, X 2, X n here. The parameter space omega is on it has an open interval on the real line. We define the derivatives of the joint density divided by the density as S i.

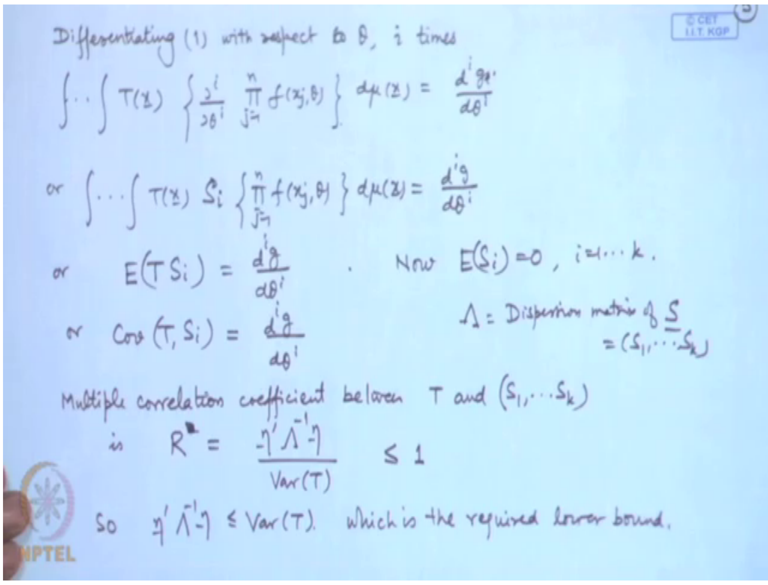
And then we have certain conditions because for the existence of this we should have the derivatives existing. Then we should also have and this should be true for i is equal to 1 to k then this integral we should be able to differentiate under the integral sign. Then the terms S 1, S 2, S k should be linearly independent. And for any integrable function delta x we should be able to once again differentiate this integral delta x product of f x j theta d mu x.

Further we define certain quantities; let us call this lambda to be the variance covariance matrix of a S 1, S 2, S k and we consider lambda inverse. And the terms of lambda inverse we denote by lambda r s. I am defining some additional things let T be an unbiased estimator of g theta. So, I am considering in general estimation problem for any parametric function say g theta. So, T is an unbiased estimator let us consider the derivative of expectation T X.

So, if I consider the ith derivative it will give me expectation of T S i. Since expectation of S i is 0 this becomes covariance and i denoted by eta i for i is equal to 1 to k. And let us denote eta vector to be eta 1 eta 2 eta k. Then Bhattacharyya's bound is that variance of theta variance T is greater than or equal to eta prime lambda inverse eta; which is nothing, but lambda r s dr g by d theta r dr dsg by d theta s.

You can see that if I had considered k equal to 1 then this will reduce to the FRC lower bound for k equal to 1. This will reduce to FRC lower bound. Let us look at the proof of this. So, expectation of T X is equal to g theta which we can write as integral T x the joint distribution of X 1, X 2 X n d mu X is equal to g theta. So, this is the statements are true for all theta. Now, this relationship we differentiate let me call it 1.

(Refer Slide Time: 20:53)



Differentiating 1 with respect to theta i times. So, i will get integral T x del i by del theta i product f of x j theta j is equal to 1 to n d mu x is equal to on the right hand side we had g so d i g by d theta i. Now this term we can consider as T X del i by del theta in this I divided by the I divided by product of f x j theta.

If I divided by this term this becomes nothing, but S i. And then I can express it as S i into product f x j theta j is equal to 1 to n d mu x is equal to d i g by d theta i. This is nothing, but expectation of T into S i. Now since we are assuming that the density can be differentiated under the integral sign. Therefore, if we differentiate this relationship this is equal to 1. So, if i differentiate this i will get expectation of S 1 is equal to 0. Similarly if I differentiate it twice and again divided by that I will get expectation of S 2 is equal to 0.

So, what we are getting now expectation of S i is 0 for i is equal to 1 to k. So, this relation is an equivalent to covariance between T and S i is equal to dig by d theta i. So, now, let us consider the multiple correlation coefficient between T and S 1, S 2, S k. That is equal to let me use a notation say capital R square that is equal to eta prime lambda inverse eta divided by variance of T.

Because lambda was the dispersion matrix of S that is S is equal to S 1, S 2, S k. So, if I apply the formula for the multiple correlation coefficient I get eta prime inverse lambda inverse eta divided by variance of T. Now this is less than or equal to 1 because multiple correlation coefficient lies between 0 and 1. Now let me write R not R square so I get eta prime lambda inverse eta less than or equal to variance of T.

Now this is nothing, but the Bhattacharyya's bound variance theta T is greater than or equal to let me call it a star. And if I expand these terms then I get this. So, you notice here that in the fresher Rao Cramer lower bound we have used that the correlation is less than or equal to 1. And here we are using in fact, correlation is square is less than or equal to 1. Here we are using multiple correlation is square is less than or equal to 1.