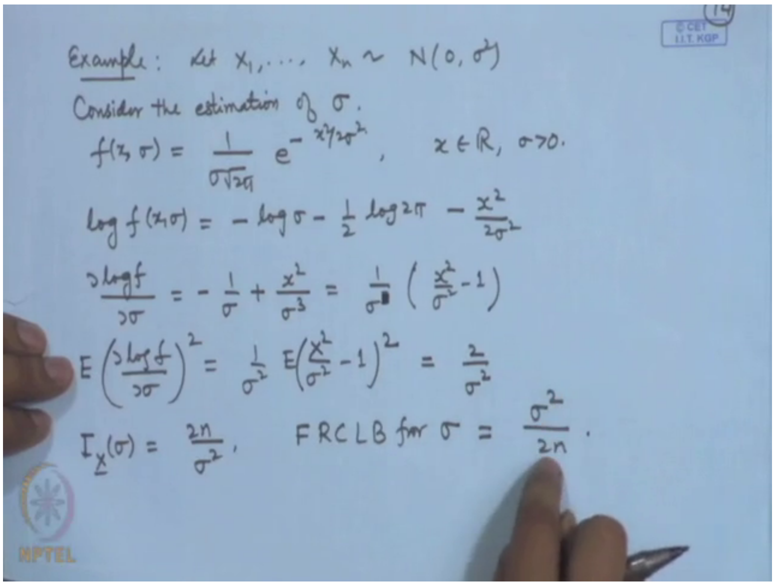
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

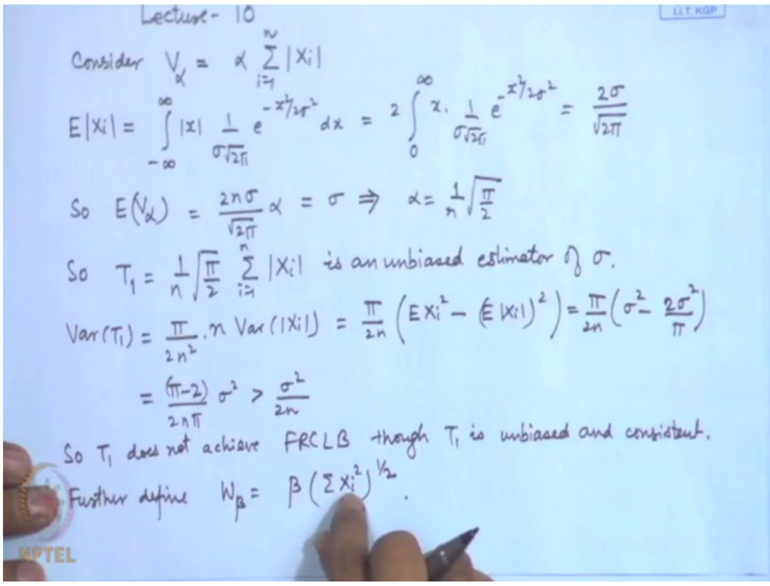
Lecture No. # 10
Lower Bounds for Variance – III

(Refer Slide Time: 00:25)



In the last class, I have discussed one example. Let me continue with that example. 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 for estimation of sigma is sigma square by 2 n. Now, I will propose two estimators for the estimation of sigma, 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 two examples.

(Refer Slide Time: 01:16)



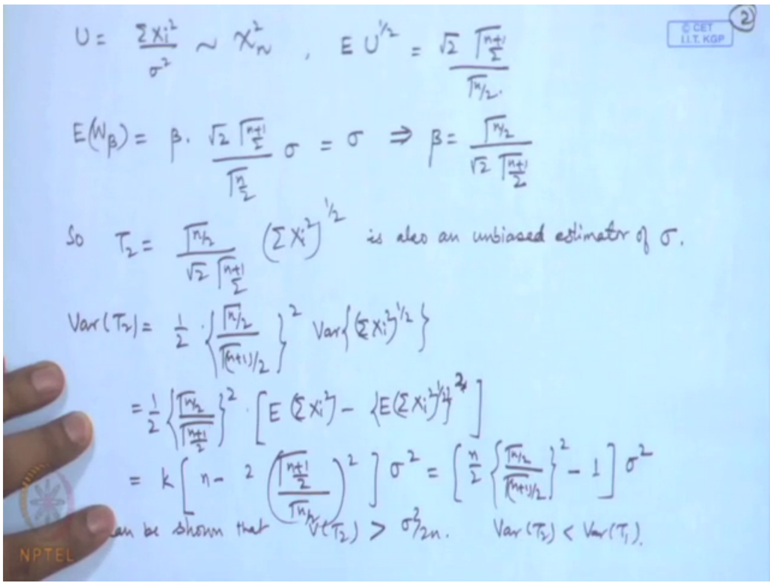
So, let me take the first example consider an estimator of the form say V alpha is equal to alpha into sigma modulus of X i is equal to 1 to n. Now, if you consider say expectation of modulus X i that is equal to integral from minus infinity to infinity, modulus X 1 by sigma root 2 pi e to the power minus x square by 2 sigma square dx. Now, this is an even function. So, this will become two 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 minus by 2 sigma square is e to the power minus x square by 2 sigma square into X by sigma square. So, 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. So, let us look at 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, but 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 I will adjust this term 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. Now, if we want to evaluate the expectation of this we can consider if X i is follow normal 0 sigma square then X i by sigma that will follow normal 0 1. So, the sum of the squares of standard normal variables when they are independent is a chi square random variable.

(Refer Slide Time: 06:26)

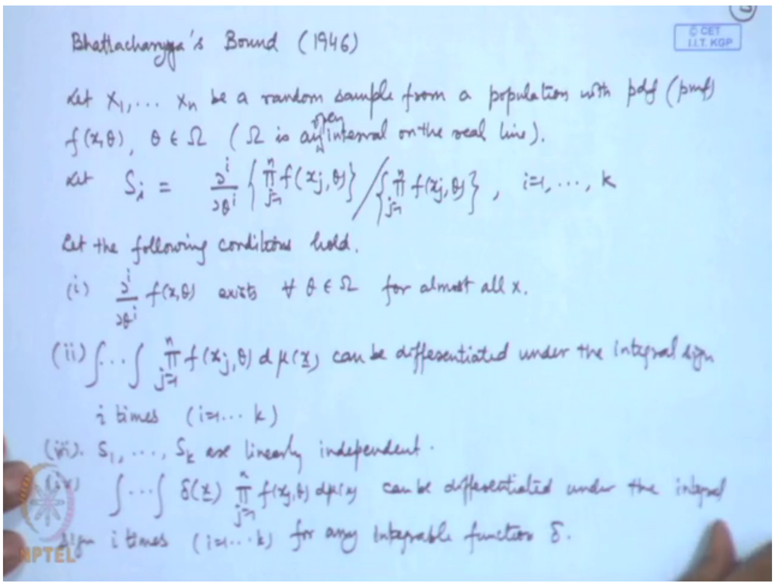


So, we get here that U is equal to sigma X i square by sigma square, this follows chi square distribution on n degrees of freedom. Now, if I have a chi square then expectation of U will become root 2 gamma n plus 1 by 2 by gamma n by 2 therefore, expectation of W beta that turns out to be beta times 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 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 2 whole 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 by T 1.

In fact, one can show that this also goes to 0 as it tends to infinity. 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 sharpen it or whether, we can extend 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.

(Refer Slide Time: 10:29)



The first half this is known as Bhattacharyya bound. So, this was proposed by A Bhattacharyya in 1946. Now, in the Frechet 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.

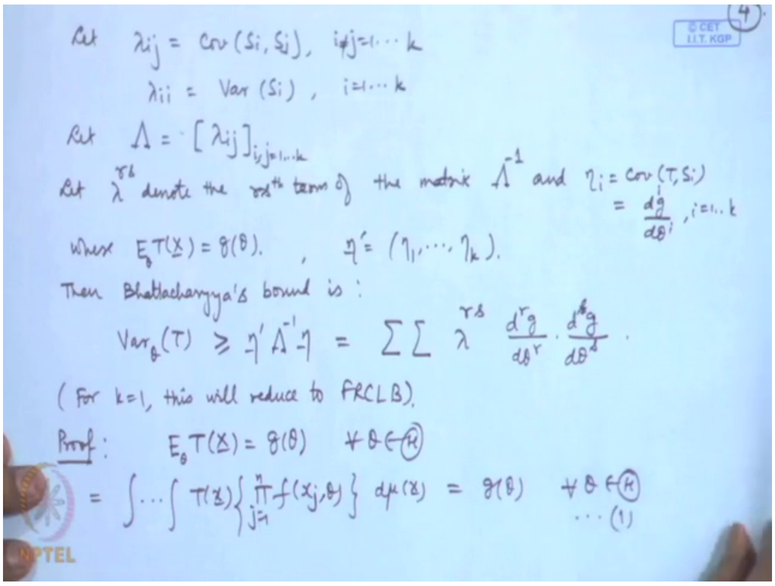
So, once again as in the Rao Cramer lower bound let us consider the regulatory conditions in the same way. So, we have a random sample let X 1, X 2, X n be a random sample from a population. Now, again it may have a probability density function or probability mass function say, f x theta, theta belonging to omega, where omega is a, is an interval on the real line. Let us define S i to be i-th 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 to and. So, on suppose I am assuming up to order k. Let the following conditions hold. So, you had already assumed that the parameter space is an interval in the real line, let us consider open interval, let us assume that the i-th 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 multifold integral and this is a generalized integral; that means, it takes care of the discreet 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 it means, that none of this 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 would exist for any integrable function delta.

(Refer Slide Time: 15:20)



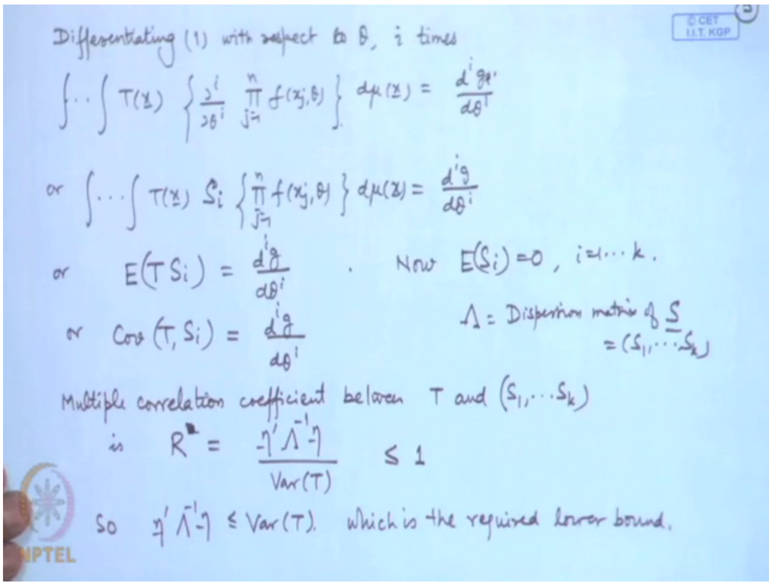
Let us define, let us define say lambda i j to be the covariance between S i, S j for i j equal to 1 to k. So, if i is not equal to j then this will be covariance and lambda ii is variance of S i for i is equal to 1 to k and let lambda be the matrix of lambda i j for i j equal to 1 to k. Let us denote by lambda say r s denote the r s-th 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 d g, d i g by d theta i. Now, what is p here p is an unbiased estimator of g theta. So, let us look at the problem here once again, we have a probability mass function are probability density function f x theta, we have a random sample X 1, X 2, X n from here.

The parameter space omega is 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 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 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 S. So, if I consider the i-th derivative it will give me expectation of T S i since, expectation of S i is 0 this becomes covariance and i denote it 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, d r g by d theta r d, r d s g 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, these statements are true for all theta.

(Refer Slide Time: 20:52)

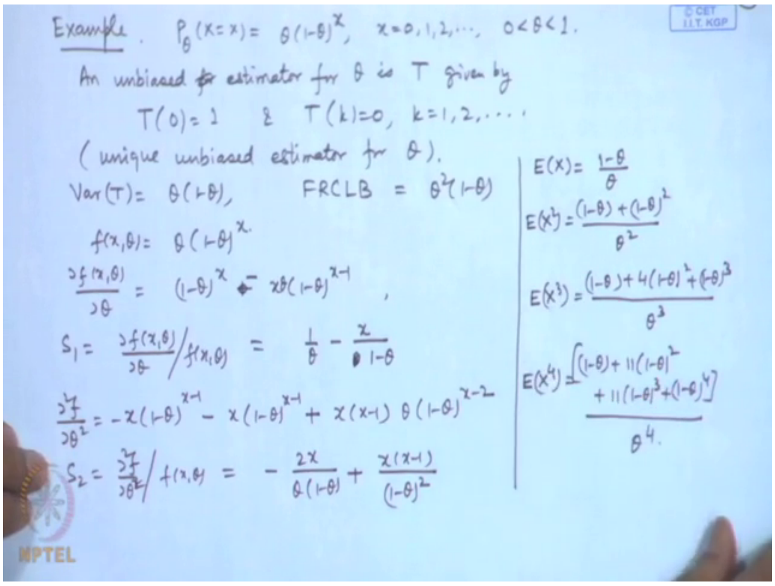


Now, this relationship we differentiate let me call it 1. Differentiating 1 with respect to theta i times. So, I will get integral T x del i by del theta i product f of x j theta j is equal to 1 to n, d mu x is equal to on the right hand side we had g. So, d i g by d theta i. Now, this term we can consider as T x del i by del is this. I divide it by the i divide it by product of f x j theta. If I divide it 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 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 equal to 0.

Similarly, if I differentiate it twice and again divide 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 d i g by d theta. 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 Bhattacharya’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 Frechet Rao Cramer lower bound, we had used that the correlation is less than or equal to 1 and here we are using. In fact, correlation square is less than or equal to 1. Here we are using multiple correlation square is less than or equal to 1.

(Refer Slide Time: 25:58)

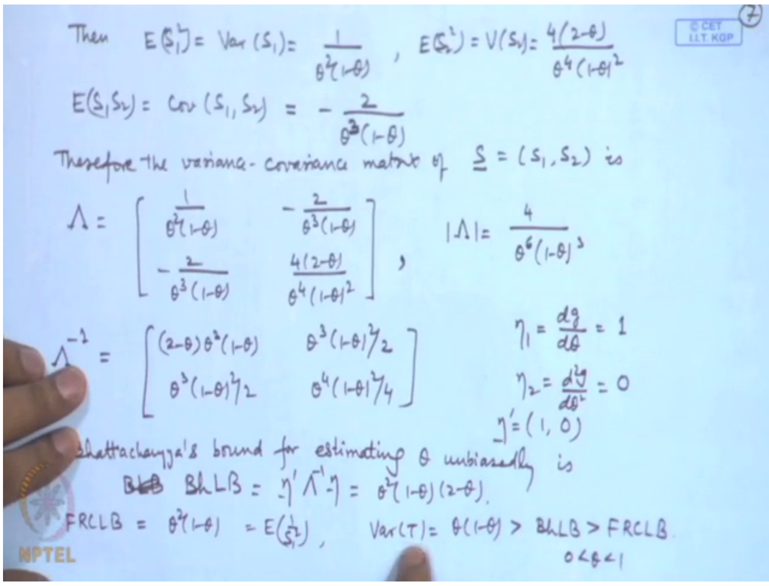


Let me explain through an example here, we consider our example of the geometric distribution that is P theta x is equal to x is equal to theta into 1 minus theta to the power x for x is equal to 0, 1, 2 and so on. In fact, for this problem you already shown that an unbiased estimator for unbiased estimator, for theta is T given by that T 0 is 1 and T k is equal to 0 for k equal to 1, 2 and so on. In fact, this is the only unbiased estimator unique unbiased estimator and we have already seen that variance of T is theta into 1 minus theta and the FRC lower bound was theta square into 1 minus theta. Now, let us apply Bhattacharyya's bound here. So, let us calculate here f x theta is equal to theta into 1 minus to the power x.

So, del f x theta by del theta that is equal to 1 minus theta to the power x plus x into 1 minus theta to the power x minus 1 with a minus sign. So, S 1 is del f x theta by del theta divided by f x theta that will be equal to 1 by theta minus x by 1 minus theta to the power x minus 1 divided by this term. So, we will get it as theta into 1 minus theta to the power x. So, I am sorry this is theta here. So, I will get here x by 1 minus theta similarly, if we consider say second derivative here del 2 f by del theta 2, we get here x into 1 minus theta to the power x minus 1 minus x into 1 minus theta to the power x minus 1 minus plus x into x minus 1 theta into 1 minus theta to the power x minus 2 these two terms can be combined. So, S 2 that is del 2 f by del theta 2 by f x theta that becomes minus 2 x by theta into 1 minus theta plus x into x minus 1, 1 minus theta square.

Now, for this geometric distribution if i want to calculate variance covariance matrix of S 1 then I need various expectations. So, let us see. In fact, I will need expectation of X, expectation of X square and here I need expectation X, expectation X square and expectation X cube and expectation X to the power 4 also. So, let us see for this geometric distribution you will have expectation X is equal to 1 minus theta by theta expectation of X square that is equal to 1 minus theta plus 1 minus theta square divided by theta square. Expectation of X cube is equal to 1 minus theta plus 4 into 1 minus theta square plus 1 minus theta cube divided by theta cube expectation of X to the power 4 that is equal to 1 minus theta plus 11, 1 minus theta square plus 11, 1 minus theta cube plus 1 minus theta to the power 4 divided by theta to the power four. So, if we use these expectations.

(Refer Slide Time: 30:42)

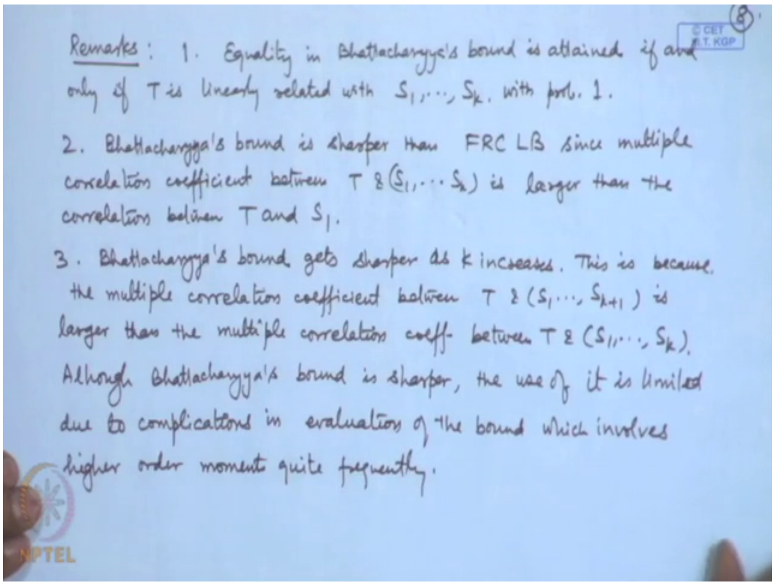


We can easily write down expectation of S square that is variance of S 1 as 1 by theta square into 1 minus theta expectation of S 2 square that is variance of S 2 that is equal to 4 into 2 minus theta divided by theta to the power 4 into 1 minus theta square. We also need the covariance between S 1, S 2 that is expectation of S 1, S 2 because expectation as 1 and expectation as 2 or 0 this is equal to minus 2 divided by theta cube into 1 minus theta therefore, the variance covariance matrix of S is equal to S 1. So, here we are going only up to second stage that is lambda 1 by theta square into 1 minus theta minus 2 by theta cube into 1 minus theta and 4 into 2 minus theta divided by theta to the power 4 into 1 minus theta square.

Now, the inverse of this can be written easily if you look at the determinant of this it is 4 divided by theta to the power 6 into 1 minus theta cube and the inverse is then, simply obtained as 2 minus theta, theta square into 1 minus theta. Theta cube 1 minus theta square by 2 theta cube into 1 minus theta square by 2 theta to the power 4 into 1 minus theta square by 4, we also look at what is eta, eta 1 is d g by d theta that is 1 eta 2 will become d 2 g by d theta 2 that is equal to 0. So, your eta vector is 1, 0. So, Bhattacharyya's bound for estimating theta unbiased is I will call it BLB Bhattacharyya lower bound or say B h L B that is equal to eta prime lambda inverse eta.

Since eta is 1, 0. So, you will get actually the first term that is theta square into 1 minus theta into 2 minus theta. What was FRC? Lower bound here that was theta square into 1 minus theta that is expectation of 1 by S 1 square here and what is variance of T variance of the unbiased estimator T that was theta into 1 minus theta. So, it is greater than Bhattacharyya lower bound and that is greater than FRC lower bound for theta lying between 0 and 1. Now, what you observe here is that although this unique unbiased estimator T. So, therefore, it is the best unbiased estimator it does not achieve the Bhattacharyya lower bound, but Bhattacharyya lower bound is sharper than the FRC lower bound. So, in that sense this is an improvement over the FRC lower bound although, we are making an assumption about the differentiation of the density function a higher number of times.

(Refer Slide Time: 35:12)



So, let me give you a few comments here about Bhattacharyya's bound. Equality in Bhattacharyya's bound is attained if and only if T is linearly related with S 1, S 2, S k. Now, why is this because actually we are using that the multiple correlation coefficient is less than or equal to 1. So, multiple correlation coefficients is equal to 1 provided the dependent variable and independent variables are completely linearly related. So, that is the condition here because we are considering multiple correlations between T and S here. So, they must be linearly related with probability 1.

Then we have observed that Bhattacharyya's bound is sharper than the Rao Cramer lower bound why because the Bhattacharyya bound is using multiple correlation coefficients between T and S 1, S 2, S k and Frechet Rao Cramer lower bound has only the correlation between T and S 1. So, certainly this multiple correlation coefficient will be higher than that.

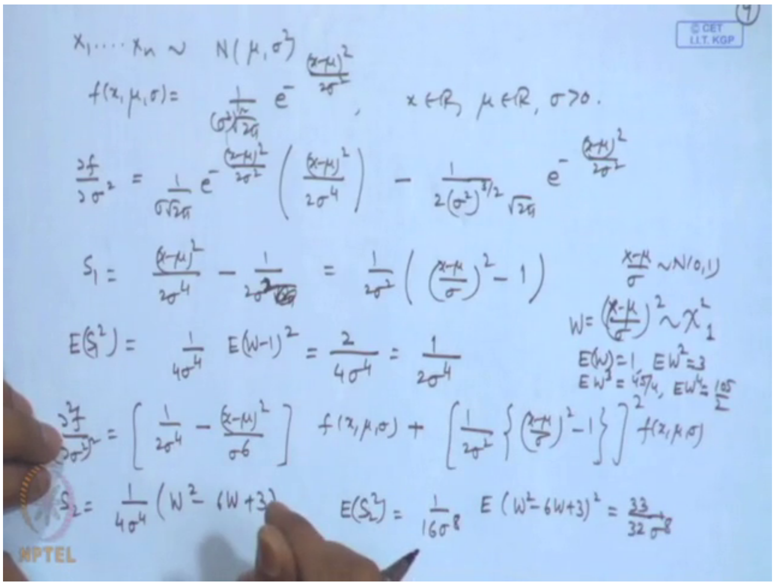
So, we can say in general that Bhattacharyya bound is sharper than FRC lower bound since, multiple correlation coefficient between T and S 1, S 2, S k is larger than the correlation between T and S 1 another thing that you observe here I had considered derivative up to order k suppose, I consider order up to k plus 1 in that case the inequality will be dependent upon the multiple correlation between T and S 1, S 2, S k plus 1.

Now, if you increase the number of variables the multiple correlation coefficient increases; that means, the Bhattacharyya bounds gets sharper and sharper as k increases. So, we can say that Bhattacharyya's bound gets sharper than, sharper as k increases this is because the multiple correlation coefficients between T and S 1, S 2, S k plus 1 is larger than the multiple correlation coefficients between T and S 1, S 2, S k. Now, you can see the historical development the Frechet Rao Cramer bound was obtained in 1943, 44, 45 and it was dependent upon one derivative or first order derivative; however, this Bhattacharyya's bound which was developed immediately after that it is sharper. It in uses higher order derivatives. Now, theoretically speaking this should be used more often; however, it is not very popular or you can say not frequently used.

The main reason is that the calculations become very, very complicated, if we use higher order derivatives I have shown the example of second order here. So, if we are using the second order we are actually, making use of the expectation X to the power 4 that is the fourth order moment. Now, if you consider distributions like normal distributions etcetera where already x square comes. So, if you consider the second order derivative you will get power 4. Now, if you take the variance of that you will get expectation of X to the power 8 kind of term and therefore, if I go to third order or fourth order the number of terms will be formidable and therefore, even though you get sharpness the method of Bhattacharyya bound has not been used much far for finding out the lower bounds for the variance of unbiased estimators.

I will just consider one example here. Let us take say normal distribution and I will show that how the calculations become complicated, although Bhattacharyya's bound is sharper, the use of it is limited due to complications in evaluation of the bound which involves higher order moments quite frequently.

(Refer Slide Time: 42:16)



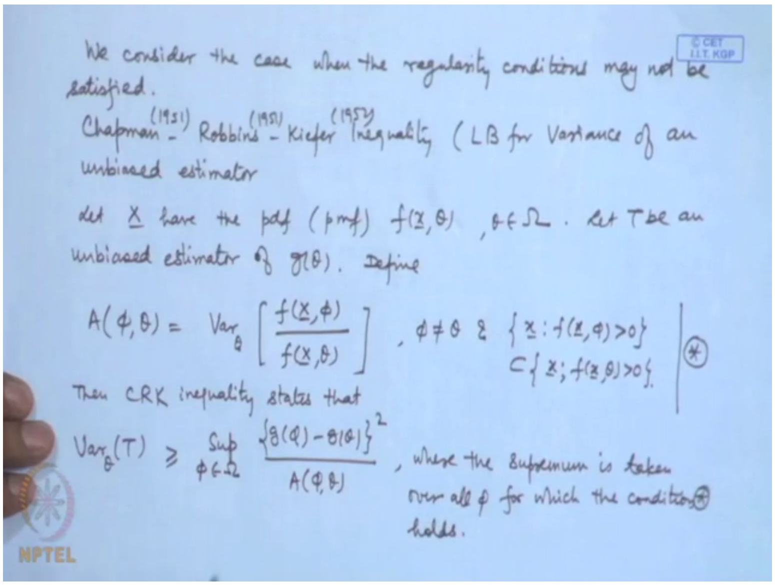
Let me give an example of this say X 1, X 2, X n follow normal mu sigma square. So, here the density function is 1 by sigma root 2 pi e to the power minus x minus mu square by 2 sigma square. We are considering sigma here. So, the derivative of this with respect to sigma del f over del sigma square. So, that will involve derivative of this that will be e to the power minus x minus mu whole square by 2 sigma square and of course, 1 by sigma root 2 pi and derivative of this term that is x minus mu whole square by 2 sigma to the power 4. Now, we consider derivative of this now, this term we will consider as sigma square to the power half. So, the derivative of that will become minus 1 by 2 sigma square to the power 3 by 2 and then we have root 2 pi e to the power minus x minus mu square by 2 sigma square.

Now, you can see this is S 1 term itself will be equal to x minus mu square by 2 sigma to the power 4 minus 1 by 2 sigma cube root 2 pi. Now, this term of course, will cancel out. So, you will get sigma square here that is 1 by 2 sigma square x minus mu by sigma whole square minus 1. Now, if I want to calculate expectation of S 1 square that will involve fourth order moment here of course, you may take help of the calculation that x minus mu by sigma that follows normal 0 1. So, x minus mu by sigma whole square let me call it w that follows chi square on 1 degree of freedom.

So, expectation of S 1 square can be written as 1 by 4 sigma to the power 4 expectation of w minus 1 whole square. So, if w is chi square 1 expectation of w is 1. So, this is variance term. So, that becomes 2 by sigma to the power 4 that is 1 by 2 sigma to the power 4. Now, if we calculate S 2. S 2 will involve the second derivative here. So, if we consider the second derivative of this density multiplied by this term you have to differentiate and then the differentiate the density also. So, you will get the terms like this del 2 f by del sigma square, square that is equal to 1 by 2 sigma to the power 4 minus x minus mu square by sigma to the power 6 into the density plus 1 by 2 sigma square x minus mu by sigma whole square minus 1 whole square into the density. So, your S 2 then turns out to be you can write using this w term as follows 1 by 4 sigma to the power 4 w square minus 6 w plus 3.

Naturally, you can see that expectation of S 2 square will involve expectation of w to the power 4 and these terms you can see here expectation of w is 1 expectation of w square is 3 expectation of w q that turns out to be forty 5 by 4 expectation of w to the power 4 turns out to be 100 5 by 2. So, you can calculate expectation of S 2 square as 1 by 16 sigma to the power 8, expectation of w square minus 6 w plus 3 whole square which is 33 by 32 sigma to the power 8. So, you can see here the terms become complicated increasingly, as we increase the order of derivatives in the Bhattacharyya's bound. Here we have considered only second order if we take third order and. So, on then it will be very ,very cumbersome calculations. So, therefore, the use of Bhattacharyya's bound is restricted .Now, I mentioned about two other things one is that case of multi parameter situation, what happens to the lower bounds in that case and another is that what if the lower bounds are not there, sorry if the regulatory conditions are not satisfied then what happens to the lower bounds.

(Refer Slide Time: 48:08)



So, we consider the case when the regularity conditions may not be satisfied. So, we have the. So, called Chapman Robbins and Kiefer inequality or lower bound for variance of an unbiased estimator. So, this is developed by D G Chapman Robbins and Kiefer. So, let x have the probability density function or probability mass function f x I am already writing for example, here where theta is belonging to omega, let T be an unbiased estimator of g theta and define, the term like a phi theta this is defined to be variance under the two distribution f x theta of f x phi divided by f x theta.

That means, I am considering the joint distribution at the parameter 0.5 and the joint distribution at the point theta, let us consider the ratio and the variance of this is considered when the true distribution is f x theta; obviously, when we write this ratio we should have certain conditions for example, I should not have the case when f x theta is 0 and f x phi is non 0 because then this will give me an infinite term; that means, the set of values for which the density function f x phi is positive should be a subset of the set of points for which f x theta is positive.

So, we should say here phi not equal to theta and the set x such that, f x phi is positive is a subset of the set such that f x theta is positive. Now, then CRK that is Chapman Robbins Kiefer inequality states that, variance of T is greater than or equal to supremum of g phi minus g theta whole square divided by a phi theta. Now, this supremum is considered over all phi belonging to omega, let me call this condition as star. Where the supremum is taken over all phi for which the condition star holds.

So, this Chapman Robbins Kiefer inequality this gives the lower bound for the variance of an unbiased estimator of a parametric function g theta, but we have not placed any condition on the density function like in the case of Rao Cramer or Bhattacharyya's bound. We have placed conditions on the existence of the derivatives existence of the derivatives of the integrals etcetera here there is no such condition. The proof of this we will be considering in the following lecture and you will again see that, the proof is dependent upon the variance, covariance, inequality or you can say Cauchy Schwarz inequality that is the correlation coefficient is less than or equal to 1.

So, in the next lecture we will be proving this CRK inequality.