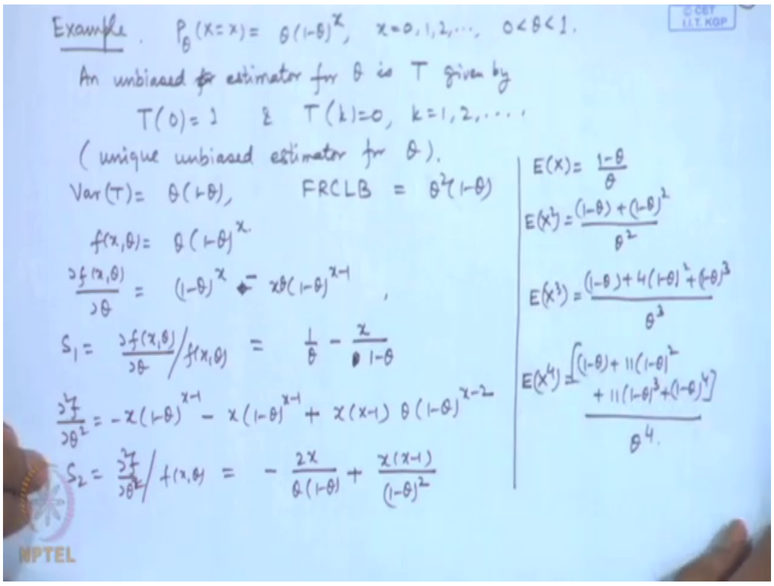
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 20
Lower Bounds For Variance – VI

Let me explain through an example here.

(Refer Slide Time: 00:31)



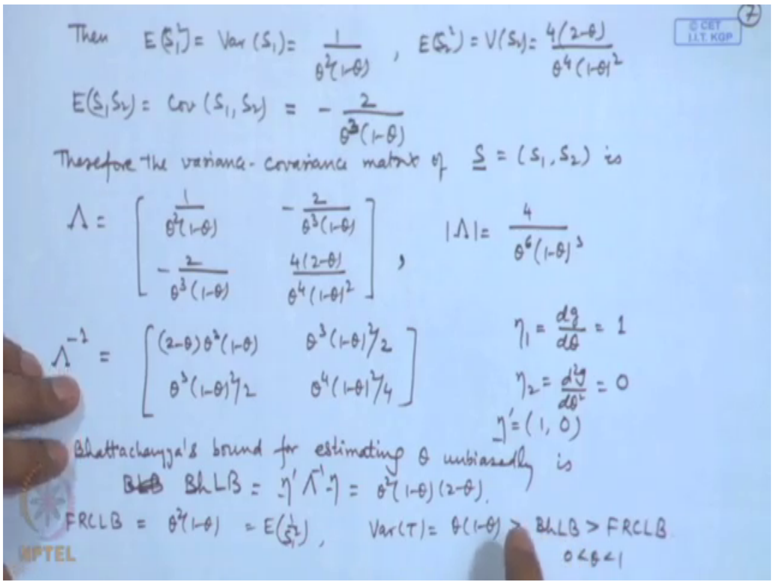
We consider our example of the geometric distribution that is P theta X is equal to x is equal to theta into 1 minus theta to the power x for x is equal to 0, 1, 2 and so on. In fact, for this problem we have already shown that an unbiased estimator for unbiased estimator for theta is T given by that 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 a unique unbiased estimator.

And we have already seen that variance of T is theta into 1 minus theta and the FRC lower bound was equal to 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 theta 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 2 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 if 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 4. (Refer Slide Time: 05:19)



So, if we use these expectations we can easily write down expectation of S square that is variance of S 1 as 1 by theta S 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.

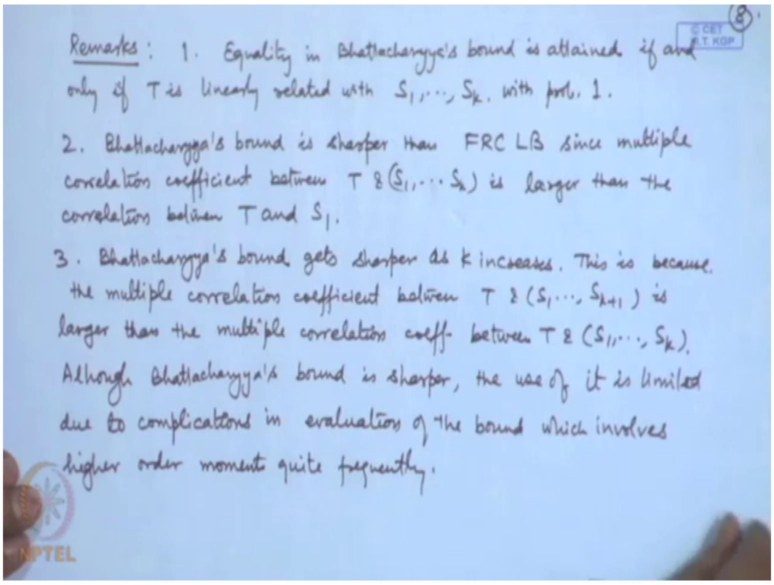
We also need the covariance between S 1 S 2 that is expectation of S 1 S 2, because expectation S 1 and expectation S 2 are 0, this is equal to minus 2 divided by theta cube into 1 minus theta. Therefore, the variance covariance matrix of S is equal to S 1. So, here we are going only up to second stage, 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 dg 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 unbiasedly is I will call it BLB Bhattacharyya lower bound or say Bh LB 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 layer. 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 based 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: 09:49)



So, let me give 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 Sk. Now why is this? Because actually, we are using that the multiple correlation coefficient is less than or equal to 1. So, multiple correlation coefficient is equal to 1 provided the dependent variable and the independent variables are completely linearly related.

So, that is the condition here because, we are considering the multiple correlation 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 Crammer lower bound, why? Because the Bhattacharyya bound is using multiple correlation coefficient between t and S 1 S 2 Sk and Fresher 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's 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 have to consider derivative up to order k suppose, I consider up to order 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.

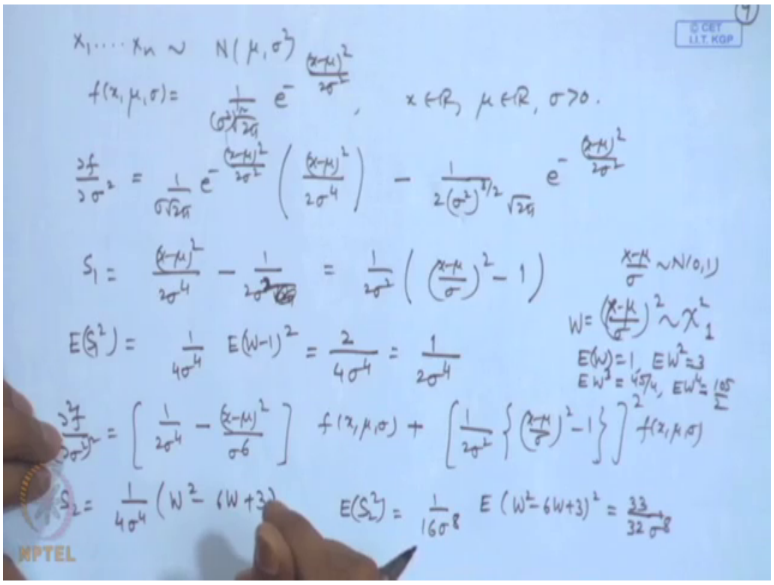
So, we can say that Bhattacharyya's bound gets sharper than sharper as k increases, this is because the multiple correlation coefficient between T and S 1 S 2 S k plus 1 is larger than the multiple correlation coefficient between T and S 1 S 2 S k.

Now, you can see the historical development the Fresher Rao Creamer lower bound was obtained in 1943, 44, 45 and it was dependent upon one derivative or first order derivative; however, this Bhattacharyya bond, 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 distribution 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 and the method of Bhattacharyya bound has not been used much 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, let me give an example of this.

(Refer Slide Time: 16:51)



Say x 1 x 2 xn 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 you have root 2 pi e to the power minus x minus mu square by 2 sigma square.

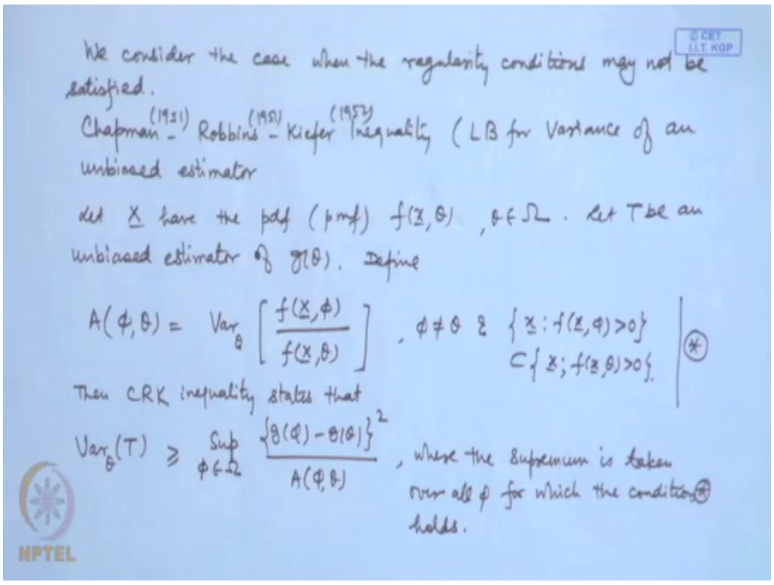
Now, you can see this is S 1 term itself will be equal to x minus mu square by 2 sigma to the power 4 minus 1 by 2 sigma cube root 2 pi. Now this term of course, will cancel out. So, you will get sigma square here that is 1 by 2 sigma square x minus mu by sigma whole square minus 1. Now, if I want to 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 4 sigma to the power 4 that is 1 by 2 sigma to the power 4. Now, if we I calculate S 2 S 2 will involve the second derivative here. So, if we consider the second derivative of this then, this density multiplied by this term, you are 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 we 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 even 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 cube that turns out to be 45 by 4, expectation of W to the power 4 turns out to be 105 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, it will be very very combustion calculations. So, therefore, the use of Bhattacharya bounds is restricted. Now I mentioned about two other things, one is the 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 regularity conditions are not satisfied then what happens to the lower bounds.

(Refer Slide Time: 22:41)



So, we consider the case when the regularity conditions may not be satisfied. So, we have the so called Chapman Robbins and Kiefer inequality are 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 fx, I am already writing for sample here, where theta is belonging to omega, let T be an unbiased estimator of g theta and define a term like A phi theta. This is defined to be variance under the true distribution fx theta of f x phi divided by f x theta; that means, I am considering the joint distribution at the parameter point phi.

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 fx theta. Obviously, when we write this ratio, we should have certain conditions for example, I should not have the case, when fx theta is 0 and fx y is nonzero, because then this will give me an infinite term; that means, the set of values for which the density function fx phi is positive, should be a subset of the set of points for which fx theta is positive. So, we should say here phi not equal to theta and the set x such that fx phi is positive is a subset of the set such that f x theta is positive.

Now, then CRK that is Chapman Robins Kiefer inequality, it 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 the 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 a star holds. So, this Chapman Robins 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.