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

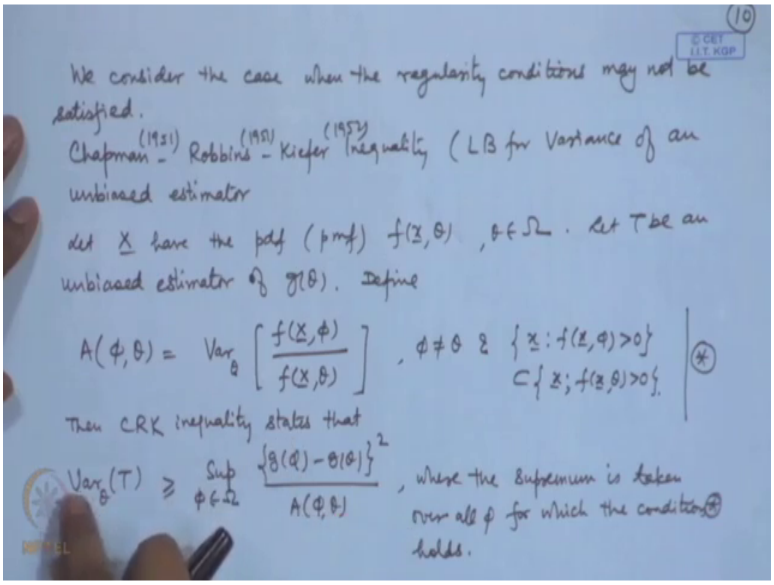
Lecture – 21
Lower Bounds for Variance – VII

In the previous lecture we have discussed the lower bound for the variance of an unbiased estimator when certain regularity conditions are satisfied. The first one assumed first order derivatives. And therefore, we had the Frechet Rao Cremer lower bound and when we assume higher order derivatives existing then we had Bhattacharya's lower bound for the variance. We have seen that Bhattacharya's lower bound is a sharper lower bound.

However, it is not very frequently used because the calculations involved to calculate the Bhattacharya's lower bound are quite involved and higher order movements are frequently used. And therefore, it becomes difficult to use that. Now there are certain densities for example, uniform distribution, exponential distribution with a location parameter Pareto distribution etcetera where the regulatory conditions are not satisfied.

In fact, you can notice that many of these densities are the ones where the range of the variable and the parameter is mixed up for example, in the uniform distribution x lies between 0 to theta. If you consider say exponential distribution then x is greater than theta.

(Refer Slide Time: 01:43)



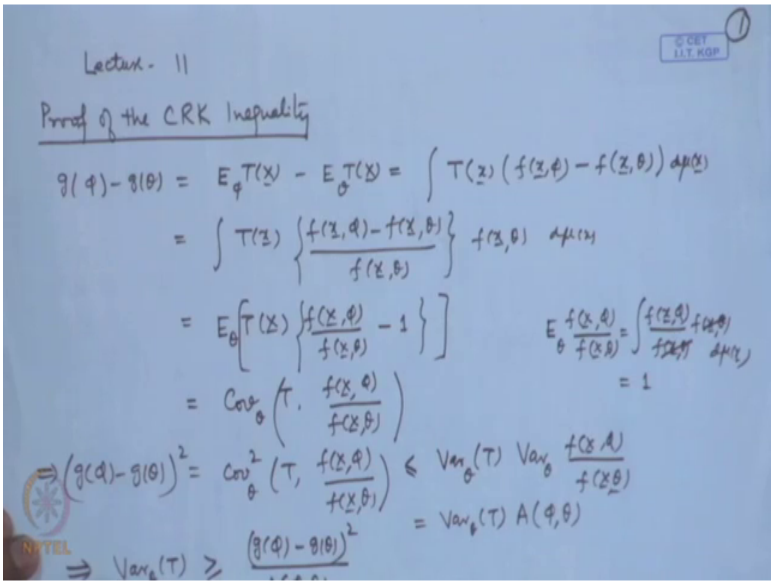
Now, in these cases I mentioned yesterday that we have another inequality that is called Chapman Robbins Kiefer inequality and let me repeat the statement once again.

So, as usual we have a probability density or probability mass function denoted by f X theta, where theta belongs to omega. Now, consider any unbiased estimator of the parametric function g theta, we defined the ratio of the densities f X phi by f X theta at two parameter points phi and theta. Now this ratio should be well defined. That means the set of values where the numerator is positive and the set of values where the numerator is positive.

So, the numerator should be positive more often. So, we have this that the set of x such that f X phi is greater than 0 is a subset of the set of values x for which f x theta is positive. Now for this ratio we consider the variance when the two density is f X theta and we denoted by A phi theta. Then the Chapman Robbins Kiefer inequality says that variance of unbiased estimator T will be greater than or equal to supremum value of g phi minus g theta a square divided by A phi theta where the supremum o is taken over phi for which this condition is satisfied.

Let us look at the proof of this now.

(Refer Slide Time: 03:12)



Let us write g phi minus g theta. Now this is equal to expectation of T X at phi minus expectation of T X at theta. So, that is equal to; now we are assuming the density function or the mass function as f X theta. So, if I make use of the generalized Lebesgue integral then this can be written as T x f x phi. So, let me use multi observations that is x 1 x 2 x n.

So, we are denoting it by x minus f x theta d mu x. Now this one we write as integral T x f x phi minus f x theta divided by f x theta into f x theta. So, if you look at this expression here we have the density and then there is a function here. So, this can be considered as expectation of T X into f X phi by f X theta minus 1. Now this is the expectation when the true density is theta, because here the density function that has been taken is f x theta.

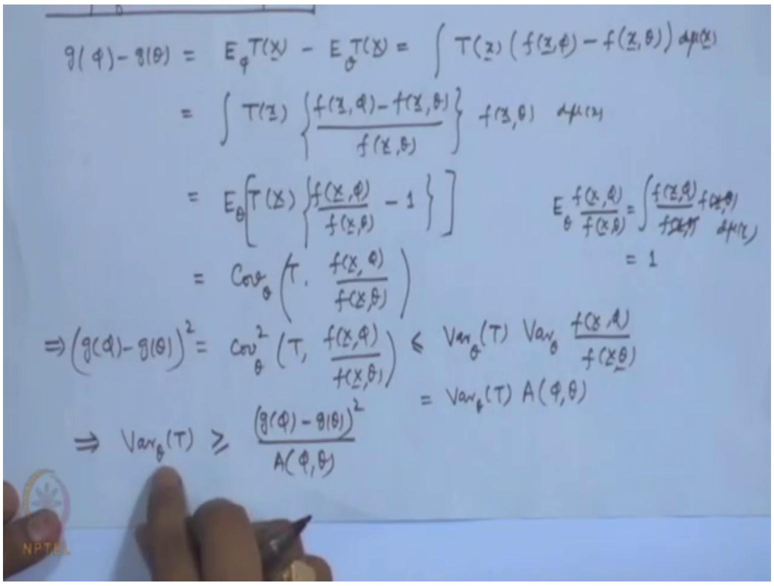
So this, we can write as, now again observe something for example, expectation of f X phi by f X theta, what it is? With respect to theta, that is equal to integral f x phi by f x theta into f x theta d mu x, now this cancels out. So, this becomes integral of the density, this is equal to 1; that means, expectation of this term is equal to 0. Now if I have expectation of product of two expressions and expectation of one of them is 0 then this is nothing, but the covariance between T and f X phi by f X theta.

Therefore we can say that g phi minus g theta a square that is equal to covariance a square of T and f X phi by f X theta at this point I apply the Cauchy Schwarz inequality.

So, covariance square is less than or equal to variance of T into variance of f X phi by f X theta, remember the notation here, variance of f X phi by f X theta had denoted by A phi theta. So, this is equal to variance of theta into A phi theta. So, what we are getting? g phi minus g theta a square is less than or equal to variance T into A phi theta.

So, we can write variance of T is greater than or equal to g phi minus g theta a square divided by A phi theta.

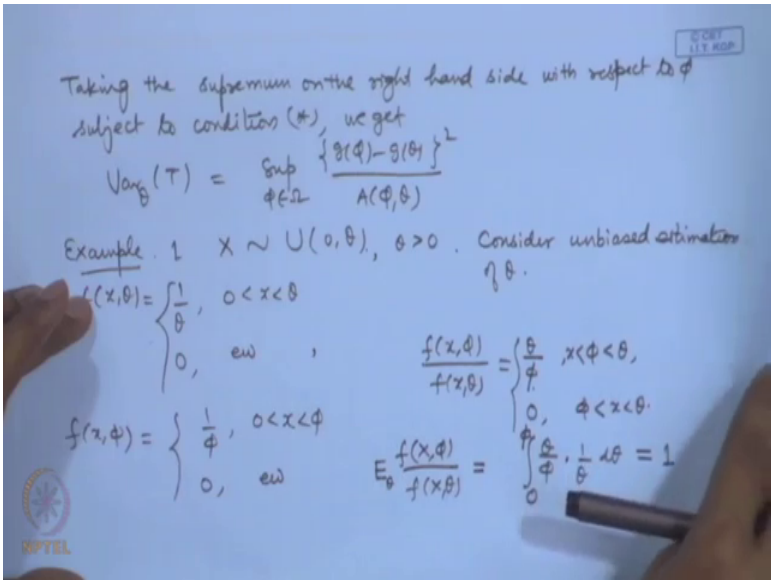
(Refer Slide Time: 06:59)



Now, the left hand side is free from phi the left hand side is dependent only on theta and the right hand side is dependent upon phi and theta both. So, on the right hand side if I take expectation, the maximum over all phi then also this inequality will be true.

Now, when I say supremum over all phi or maximum over all phi then what are the phi's? The phis are the ones which satisfy this condition a star.

(Refer Slide Time: 07:33)



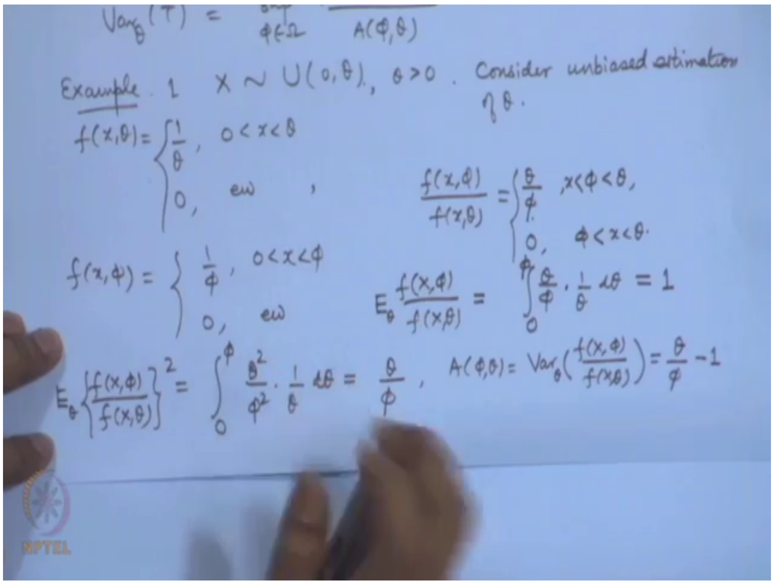
So, we have then this that taking the supremum on the right hand side with respect to phi, subject to condition a star we get variance of T greater than or equal to supremum of phi. And let me write here phi satisfy a belonging to omega g phi minus g theta whole square by A phi theta. So, we have proved the Chapman Robbins Kiefer inequality which we call in abbreviated form as CRK inequality. Let me give example of application of CRK inequality when the regularity conditions are not satisfied. So, let us take say x following uniform distribution on the interval 0 to theta. So, we consider say unbiased estimation of theta.

Now, here we know that the density function is of the form 1 by theta 0 less than x less than theta and it is equal to 0 elsewhere. If I write at another parameter point say f x phi then it is equal to 1 by phi 0 less than x less than phi and 0 elsewhere. So, if we consider the ratio f x phi by f x theta then that will be equal to 1 by phi divided by 1 by theta in this region.

That means it will become theta by phi when we are having phi less than theta and x is less than phi and if phi is less than x less than theta then this will become 0. Now the case when both are 0 we are not considering that thing in. In fact, we can consider the ratio to be 0 by default or by convention in that case, because there this ratio will not be defined there.

So, now once we have the expression for this we can calculate the expectation and the variance of this term. So, for example, expectation of f X phi by f X theta when theta is the distribution so you are getting it as equal to theta by phi integral. Now you have to consider the range of x from 0 to phi here and the density is 1 by theta, because although the density is 1 by theta, but the range of x cannot be 0 to theta, because phi is less than theta here and x is less than phi. So, the range is only this. So, here theta cancels out and you get this value simply as 1.

(Refer Slide Time: 11:14)



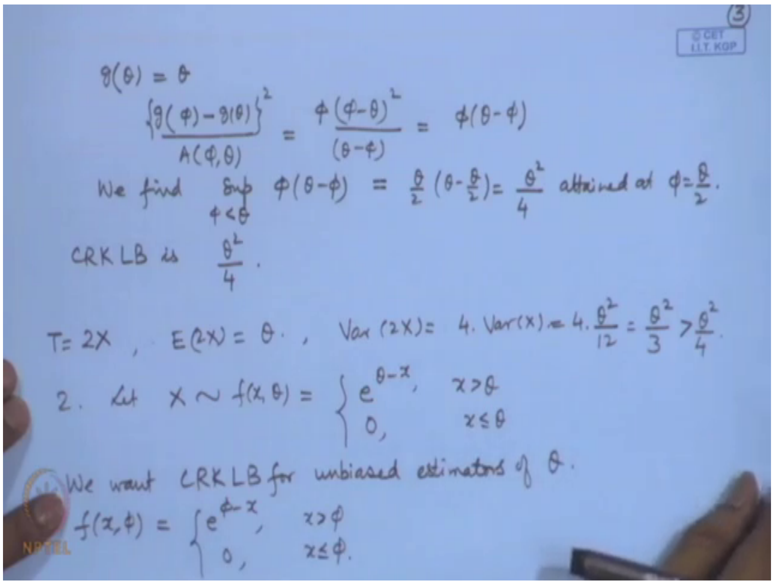
Similarly, if I consider expectation of f x phi by f x theta whole square then this will become 0 to phi theta a square by phi square 1 by theta d theta. So, this is simply theta by phi; that means, a phi theta that is the variance of f x phi by f x theta that will be equal to theta by phi minus 1, this is the variance when the true distribution has been assumed to be theta.

Let us revisit the calculations, we are writing down the distribution at two parameter points theta and phi and then I write down the ratio f x phi by f x theta. Now notice here there is one case when both of them are positive, if both of them are positive then the ratio will be theta by phi.

Now, that is going to be true when x is less than phi less than theta and of course, it will also be true for x less than theta less than phi, but in that case then we have to also take up that the density in the denominator may become 0. So, we will not take that case, it is equal to 0 when x is between phi and theta. Therefore, when you consider the expectation it is theta by phi over this region only that is 0 to phi and when we integrate we get 1.

In a likewise manner the expectation of f X phi by f X theta is square can be calculated and we get the term as theta square by phi square 1 by theta integral of this quantity from 0 to phi, with respect to. So, this is not with respect to theta it is with respect to x here. So, this value turns out to be simply theta by phi, and therefore the variance is expectation of a square minus expectation whole square that is theta by phi minus 1. Now let us consider the CRK inequality.

(Refer Slide Time: 13:49)



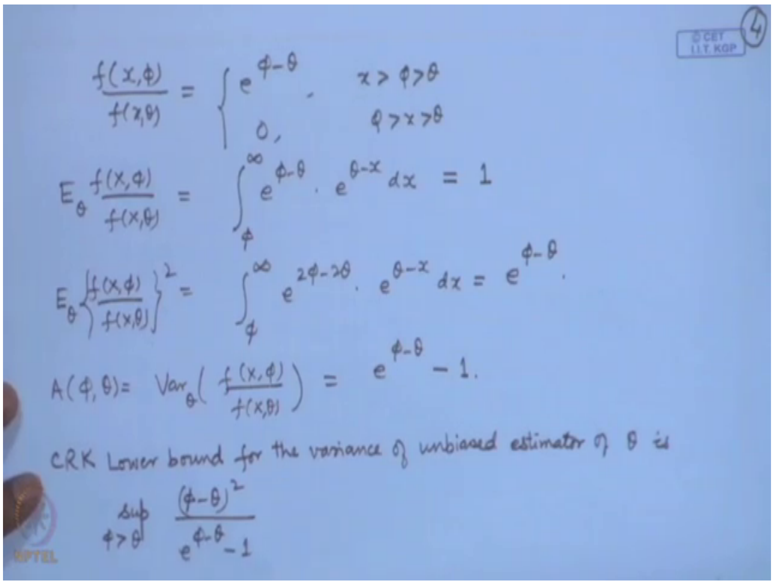
So, for CRK inequality we need g theta g phi. So, here g theta is theta itself. So, if we consider the term g phi minus g theta whole square divided by a phi theta, then that is equal to phi minus theta square divided by theta minus phi into phi. Now in this theta minus phi term will cancel out. So, you get phi into theta minus phi. Now in order to find out the supremum with respect to phi such that the condition is star is satisfied, we should have phi less than or equal to theta. So, we find supremum of this quantity such that phi is less than theta.

So, now this is a simple function here, if you differentiate we will get theta minus 2 phi, and that if you put equal to 0 you will get phi is equal to theta by 2. So, that is equal to theta by 2 into theta minus theta by 2 that is equal to theta square by 4 this is attained at phi is equal to theta by 2. Therefore, CRK lower bound is theta square by 4. So, we have seen here that even if the FRCLB is not available that is Frechet Rao Cramer Lower Bound is not available we can find out lower bound for the variance of an unbiased estimator.

In the case of uniform distribution for example, we know for example, 2 X we can consider then expectation of 2 X is equal to theta. So, this is an unbiased estimator what is variance of 2 X, variance of 2 X is equal to 4 times variance of x that is equal to 4 times theta square by 12 that is theta square by 3. Of course, you can see that this is greater than theta square by 4, we can actually show later on that 2 X is minimum variance unbiased estimator in this problem. We can show it directly also and we will later on use a concept of sufficiency and completeness from there also we will show this thing.

Let us consider another example of non regular distribution, say exponential distribution with a location parameter e to the power theta minus x where x is greater than theta it is 0, for x less than or equal to theta. So, here we want the CRK lower bound for unbiased estimator of theta. So, let us consider f x phi here, f x phi will becoming e to the power phi minus x for x greater than phi and 0 for x less than or equal to phi

(Refer Slide Time: 17:36)



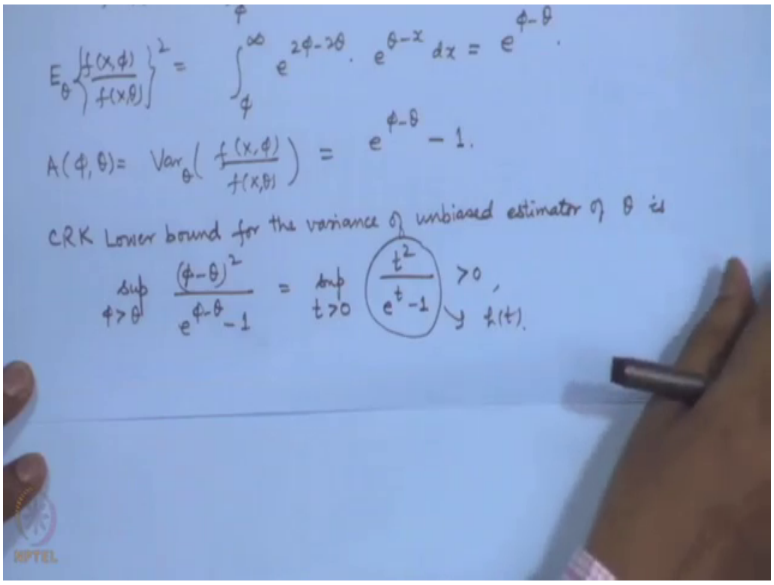
So, once again we consider the ratio f x phi by f x theta; consider the ratio f x phi by f x theta that will be equal to.

Now, e to the power phi minus x divided by e to the power theta minus x; so e to the power minus x will cancel out. And we are left with the term e to the power phi minus theta for x greater than phi greater than theta. And it is equal to 0 for phi less than x greater than x greater than theta, we are not considering the case phi less than theta here because in that case there will be a place where you will have 0 in the denominator. So, we are not considering that case here.

So, expectation of f x phi divided by f x theta when theta is the true parameter value it is equal to e to the power phi minus theta e to the power theta minus x d x from phi to infinity. That is equal to, now if you look at this theta cancels out you get density e to the power phi minus x from phi to infinity. So, the value of integral will be equal to 1; similarly, if we consider the expectation of f x phi by f x theta a square that is equal to integral phi to infinity e to the power twice phi minus twice theta.

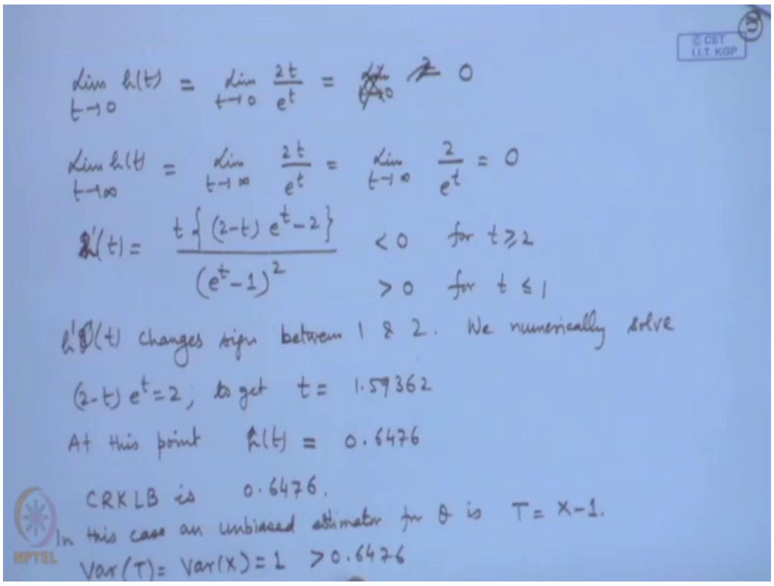
E to the power theta minus x d x that is equal to e the power phi minus theta so A phi theta that is variance of f X phi by f X theta that is equal to e to the power phi minus theta minus 1. Therefore, the Chapman Robbins Kiefer lower bound for the variance of unbiased estimator of theta is supremum of phi minus theta square divided by e to the power phi minus theta minus 1, where phi is greater than theta.

(Refer Slide Time: 20:15)



Now, if phi is greater than theta basically it means we can consider it as a problem supremum say t greater than 0, t square by e to the power t minus 1, because phi minus theta is positive. So, I can replace it by t, now you can notice that this is a positive function we can also notice here that.

(Refer Slide Time: 20:56)



Let me call this as say h t, then you can notice here that limit of h t as t tends to 0 that is, now if you look at this term here this is 0 by 0 form as t tends to 0.

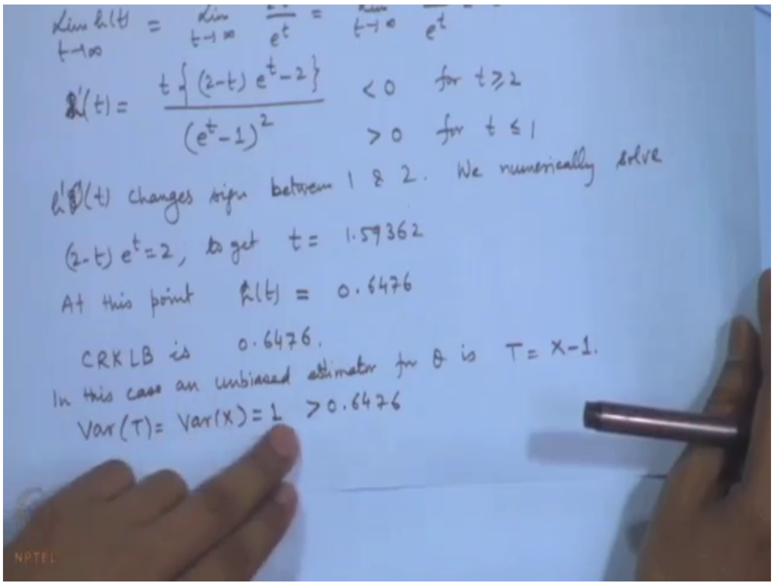
So, we can apply L Hospital's rule. So, you will get limit 2 T by e to the power t s t tends to 0 which is again 0 by 0 form. So, we can further take 2 by e to the power, now this is not 0 by 0 form this is actually 0. Similarly if I consider limit of h t s t tends to infinity that is equal to limit as t tends to infinity to t by e to the power t. That is equal to limit as t tends to infinity of 2 by e to the power t that is equal to 0.

So, as t tends to 0 or t tends to infinity, the function h t tends to this function ht tends to 0. Now let us consider the derivative g prime t that is equal to t times 2 minus t e to the power t minus 2 divided by e to the power t minus 1 square. This is less than 0 for t greater than or equal to 2 and it is greater than 0 for t less than or equal to 1. Actually we can show that g prime t has a change of sign between 1 and 2.

So, you can numerically solve this equal to 0. So, we numerically solve this 2 minus t e to the power t is equal to 2 to get t as approximately 1.59362, at this point h t function sorry this is I was writing h. So, this is will be h prime t and this will also be h prime t. So, h t value will be equal to 0.6476 that is CRK lower bound is 0.6476.

Let us consider say unbiased estimator here. In this case an unbiased estimator for theta is in the exponential distribution; if I take the mean here mean of this distribution is 1 plus theta that is expectation X is equal to 1 plus theta. Therefore, expectation of x minus 1 will be equal to theta. So, an unbiased estimator will be equal to X minus 1, what is variance of this? That is variance of X that is equal to again same 1 it is of course, bigger than the CRK lower bound here.

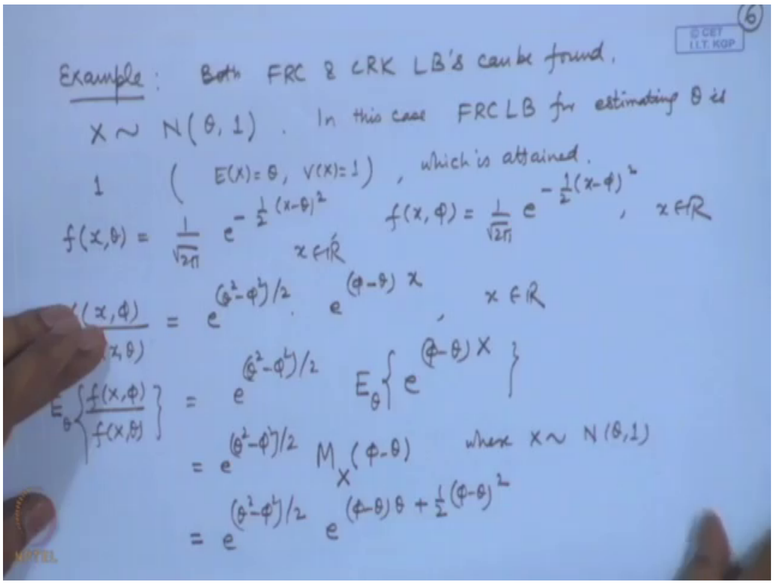
(Refer Slide Time: 24:41)



So, here we are able to obtain a non trivial lower bound for the variance of an unbiased estimator and in this problem we are showing that it is not attend here. In fact, we can show that x minus 1 is minimum variance unbiased estimator by a direct argument, that we will take up little later.

Now, in these two examples that I have given here the regularity conditions which are mentioned in the Frechet Rao Cramer lower bound or the Bhattacharya lower bound they were not satisfied. Now there is an interesting question that if those conditions are satisfied and we find FRC lower bound as well as CRK lower bound, then which one will be sharper? The answer is interesting here I will show it through one example.

(Refer Slide Time: 25:30)



Here both FRC and CRK lower bounds can be found.

Let me take a simple case normal distribution with mean theta and variance unity, suppose we have an observation X from this distribution. In general we have calculated that if X follows normal mu sigma square, the FRC lower bound was sigma a square by N. Now if sigma a square I have taken to be 1 then it will become 1 by N, now that is when we have N observations X 1 X 2 X N here we have only one observation. So, it will become simply 1. So, in this case FRC lower bound for estimating theta is 1 and of course, you had expectation X is equal to theta and variance of X is equal 1.

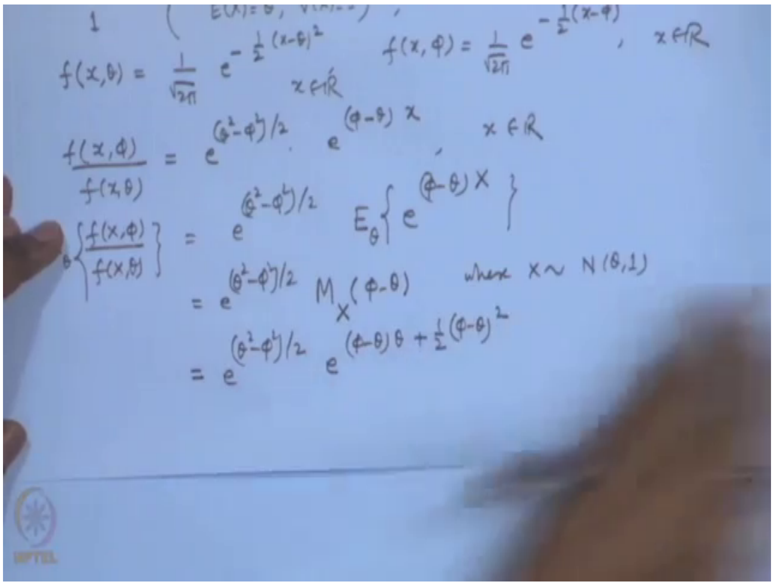
So, it is attained, let us calculate the CRK lower bound here. So, if you want to calculate the CRK lower bound we need to write down the density 1 by root 2 pi e to the power minus half x minus theta a square. We also write this density at another point x phi 1 by root 2 pi e to the power minus 1 by 2 x minus phi a square, notice here that these are defined for all x, x is on the real line here also x is on the real line. So, there is no problem in taking the ratio for all the real values.

So, when I write down the ratio here e to the power minus x square by 1 term cancels out and I will be left with e to the power theta a square minus phi square by 2 into e to the power phi minus theta x. This is valid for all x therefore, when I calculate the expectation when the true density is theta this is equal to expectation of theta square minus phi square by 2.

Expectation of e to the power phi minus theta into X, now this is when the density of X is normal theta 1, now you look at this expression carefully. It is of the form expectation of e to the power t X that is the moment generating function of the normal theta 1 distribution. Now we know that if I have a normal mu sigma a square distribution then the moment generating function at the point t that is given by e to the power mu t plus half sigma a square t square.

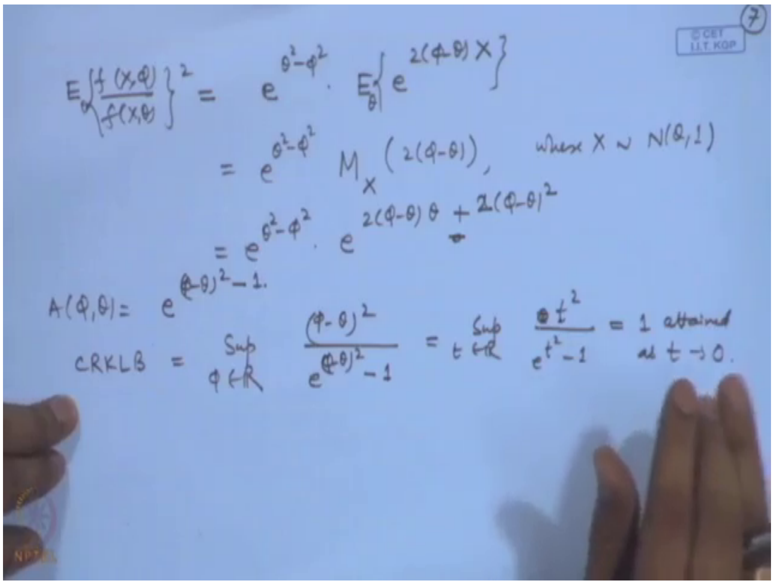
So, in that one we substitute t is equal to phi minus theta and sigma square is equal to 1 and mu is equal to theta. So, this is nothing, but e to the power theta a square minus phi square by 2 into the moment generating function of x at the point phi minus 2 where x follows normal theta 1. So, this value turns out to be e to the power theta square minus phi square by 2 and e to the power phi minus theta into theta plus half phi minus theta whole square.

(Refer Slide Time: 29:57)



In a similar way we can calculate the expectation of expectation of f X phi by f X theta whole square.

(Refer Slide Time: 30:03)



So, this will become equal to expectation of this square, now if I square a I get here e to the power theta square minus phi square which is a constant term. So, it will come out of the expectation sign and then I will get expectation of e to the power twice phi minus theta x. So, this is equal to e to the power theta a square minus phi square into expectation of e to the power twice phi minus theta into x. So, this is nothing, but again of the form of the moment generating function of x at the point twice phi minus theta. So, this is equal to moment generating function of x at the point twice phi minus theta, where X is a normal theta 1 random variable.

So, we substitute in the formula for the moment generating function and we get it as e to the power twice phi minus theta into theta plus twice phi minus theta whole square. So, naturally now the variance that is A phi theta term is equal to e to the power. So, this term minus a square of this term if I square rate this I get e to the power theta square minus phi square which is the same term here. Similarly here I have e to the power twice phi minus theta into theta and here if I square it I get e to the power phi minus theta theta twice.

So, these terms can be taken out and if you take it out what you get here, twice phi theta minus twice theta a square plus theta a square that cancels out minus phi square and if you look at this term here, here I can take come phi minus theta whole square out. So, phi square will come here which will cancell with this and then you get plus theta a square which will again cancel plus theta a square minus twice theta a square plus theta a square. So, all of these terms get cancelled out, you get minus twice phi theta plus twice phi theta.

So, you are left with only e to the power phi minus theta a square minus 1. Now the CRK lower bound is equal to supremum of phi, supremum over phi phi minus theta a square divided by e to the power phi minus theta a square minus 1 this you can simply write something like t. So, it is equal to supremum e to the power t square divided by e to the power t square minus 1 where t is an. Now the analysis of maximization of this is simple in fact, this is a positive term and you can easily show that the maximum is attained at t is equal to 0.

Now, at t is equal to 0 this is having 0 by 0 form. So, you take the limit, this is attained as t tends to 0. Now you notice here in this particular problem the Frechet Rao Cramer lower bound was 1, the variance of the unbiased estimator X was 1 and the Chapman Robbins Kiefer lower bound is also equal to 1. So, in general we cannot say that CRK bound is worse because it does not take care of the regularity conditions. So, in this particular case for example, we get exactly the same.