Statistical Inference

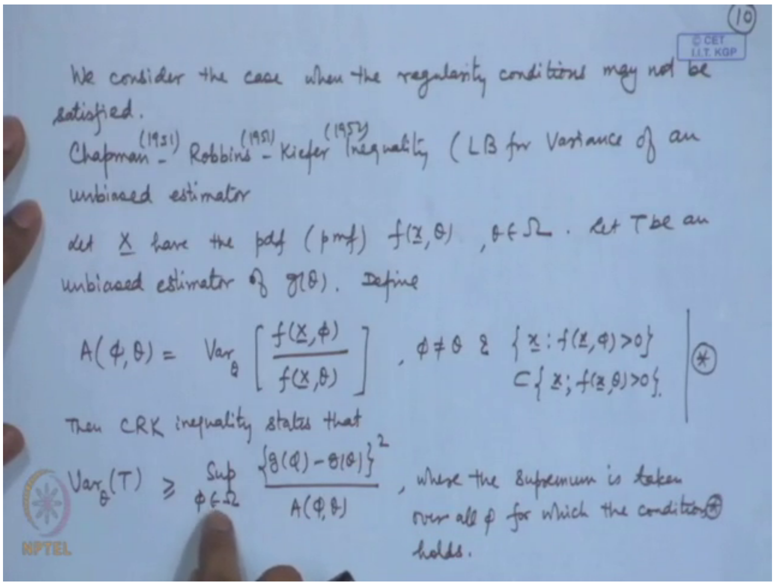
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture No. # 11
Lower Bounds for Variance – IV

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 1 assumed first order derivatives and therefore, we had the Frechet-Rao-Cramer lower bound and when we assume higher order derivatives existing then we had Bhattacharya's lower bound for the variance. We have seen the 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. Very higher order moments 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 regularity 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.

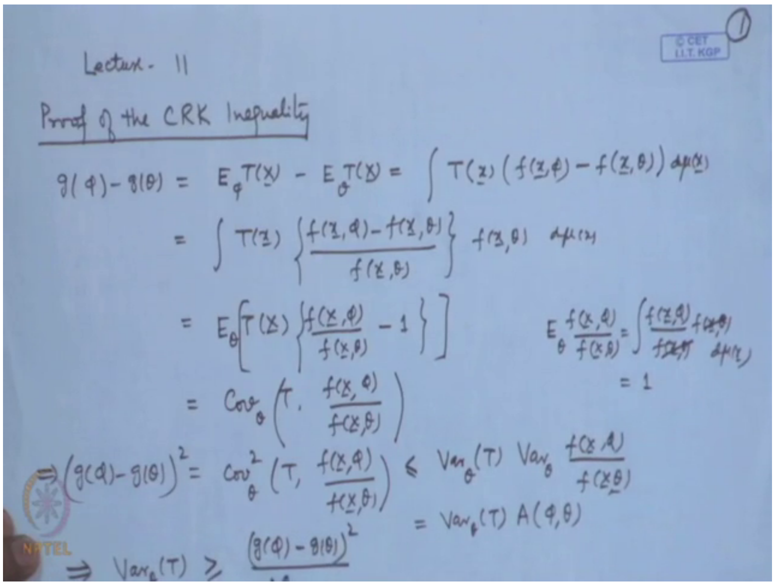
(Refer Slide Time: 01:41)



Now in these cases I mentioned yesterday that we have another inequality that is called Chapman-Robbins-Kiefer Inequality. Let me repeat the statement once again. So, as usual we have a probability density or a 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 define the ratio of the densities f X phi by f X theta at 2 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, that the numerator should be positive more often.

So, we have this that the set of x such that the 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 2 densities f x theta and we denote it 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 square divided by a phi theta, where the supremum o is taken over phi for which this condition is satisfied.

(Refer Slide Time: 03:11)



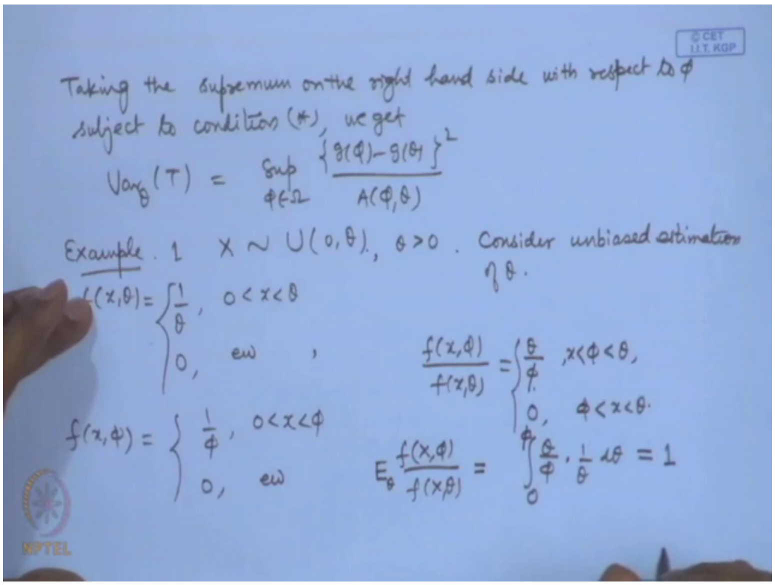
Let us look at the proof of this now. Let us write g phi minus g theta. Now, this is equal to expectation of T X at phi minus expectation of T X at theta. So, that is equal to. Now, we are assuming the density function or the mass function as f x theta. So, if I make use of the generalized Lebesgue-Stieltjes Integral then this can be written as T x f x phi. So, let me use multi observation 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 densities 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 2 expressions and expectation of 1 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 square that is equal to covariance 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 I have 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 is 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 square divided by a phi theta.

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 phi's are the ones which satisfy this condition star.

(Refer Slide Time: 07:32)

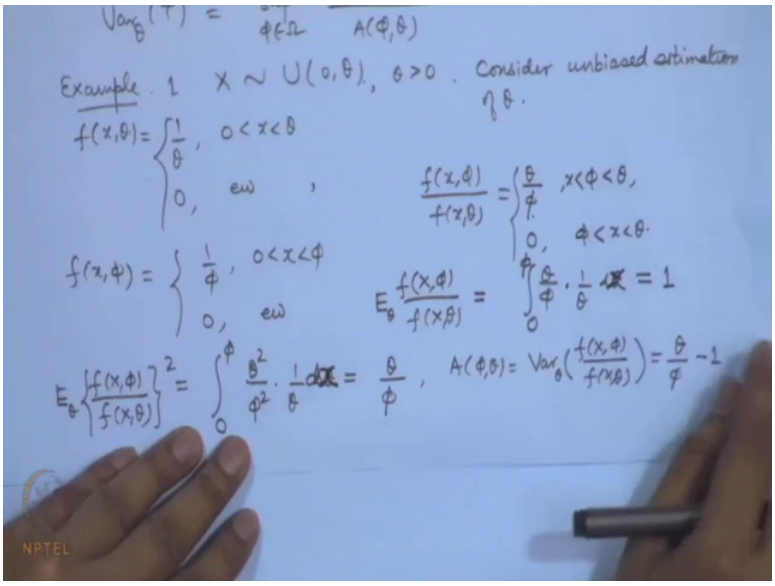


So, we have then this that taking the supremum on the right hand side with respect to phi subject to condition star, we get variance of T greater than or equal to supremum of phi and let me write here phi satisfy belonging to omega g phi minus g theta whole square by a phi theta. So, we have proved the 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 uniformed 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 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 the 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: 07:32)

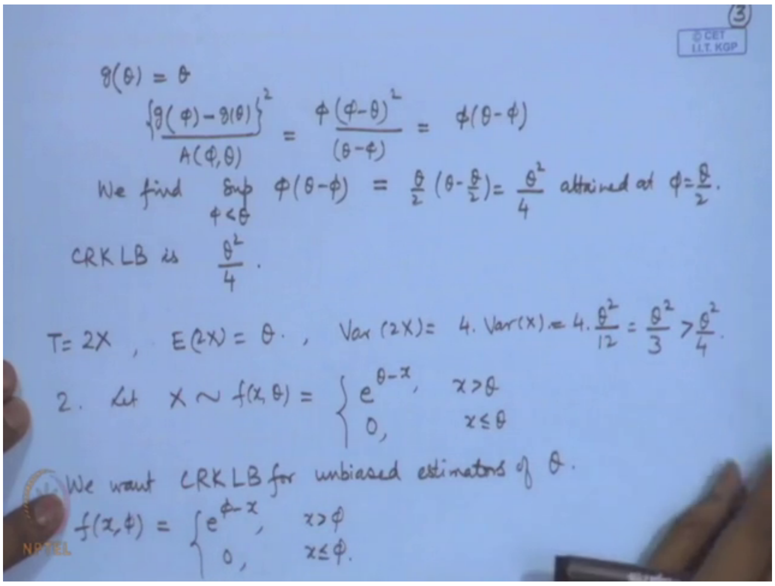


Similarly, if I consider expectation of f X phi by f X theta whole square then this will become 0 to phi theta square by phi square 1 by theta d theta. So, this is simply theta by phi. That means, a phi theta that is the variance of f X phi by f X theta that will be equal to theta by phi minus 1. This 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 2 parameter points theta and phi and then I write down the ratio f x phi by f x theta.

Now, notice here there is 1 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 we 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 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 square minus expectation whole square that is theta by phi minus 1.

(Refer Slide Time: 13:48)



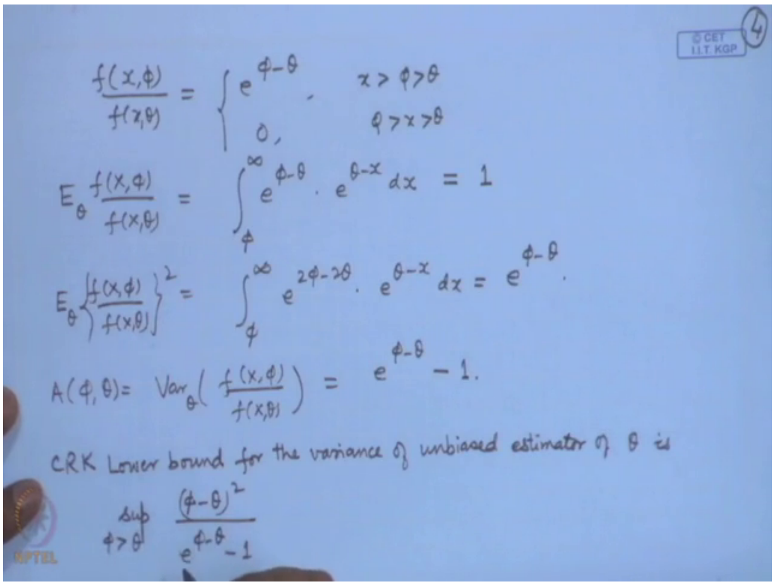
Now, let us consider the CRK inequality. So, for CRK inequality we need g theta g phi. So, here g theta is theta itself. So, if we consider the term g phi minus g theta whole square divided by a phi theta then that is equal to phi minus theta square divided by theta minus phi into phi. Now, in this this 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 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 you 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 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 become 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: 17:34)

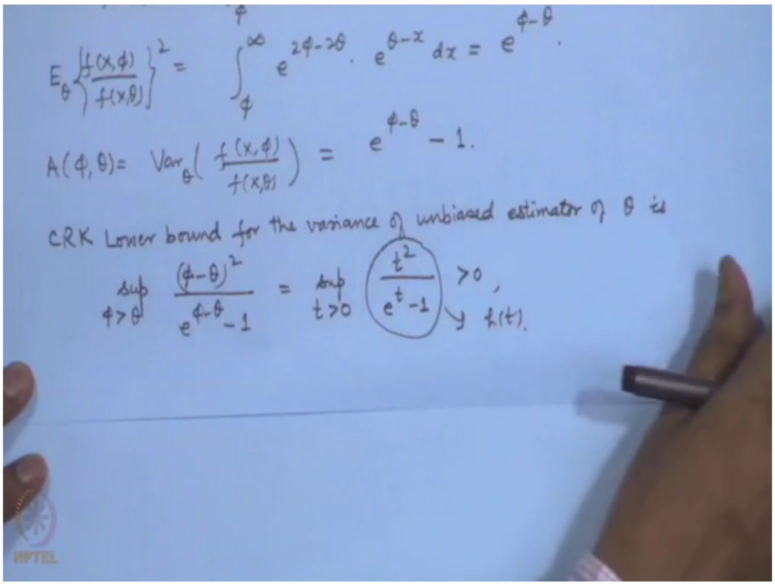


So, once again we consider the ratio f x phi by f x theta. Consider the ratio f x phi by f x theta. That will be equal to now e to the power phi minus x divided by e to the power theta minus x. So, e to the power minus x will cancel out and we are left with the term e to the power phi minus theta for 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 a true parameter value it is equal to e to the power phi minus theta e to the power theta minus x dx 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 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 dx that is equal to e to the power phi minus theta.

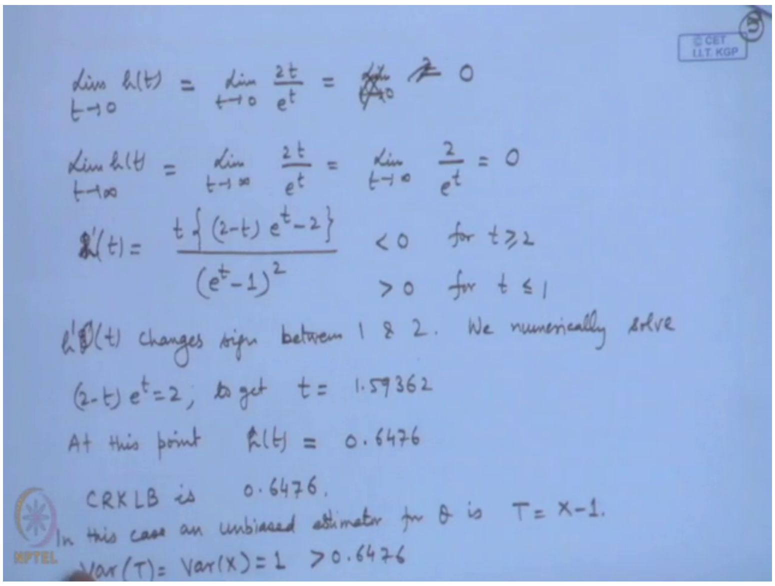
So, a phi theta that is variance of f X phi by f X theta that is equal to e to the power phi minus theta minus 1. Therefore, the Chapman-Robbins-Kiefer lower bound for the variance of unbiased estimator of theta is supremum of phi minus theta square divided by e to the power phi minus theta minus 1 where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

(Refer Slide Time: 17:34)



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 e minus 1 because phi minus theta is positive. So, I can replace it by t here. Now, you can notice that this is a positive function. We can also notice here that let me call this as say h t.

(Refer Slide Time: 20:54)



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 rule. So, we will get limit 2 t by e to the power t as 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 as t tends to infinity that is equal to limit as t tends to infinity 2 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 h t tends to 0.

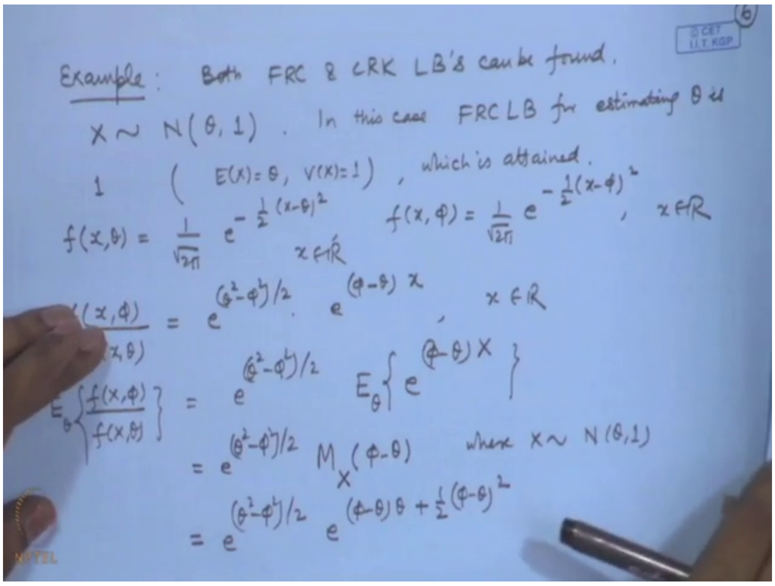
Now, let us consider the derivative g prime t that is equal to t times 2 minus t e to the power t minus 2 divided by e to the power t minus 1 square. This is less than 0 for t greater than or equal to 2 and it is greater than 0 for t less than or equal to 1. Actually, we can show that g prime t has a 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 I was writing h. So, this 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 and 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.

So, here we are able to obtain a nontrivial lower bound for the variance of an unbiased estimator and in this problem we are showing that it is not attained 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 2 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 1 will be sharper? The answer is interesting here

(Refer Slide Time: 25:29)



I will show it through 1 example. 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 square by n. Now, if sigma 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 1 observation. So, it will become simply 1.

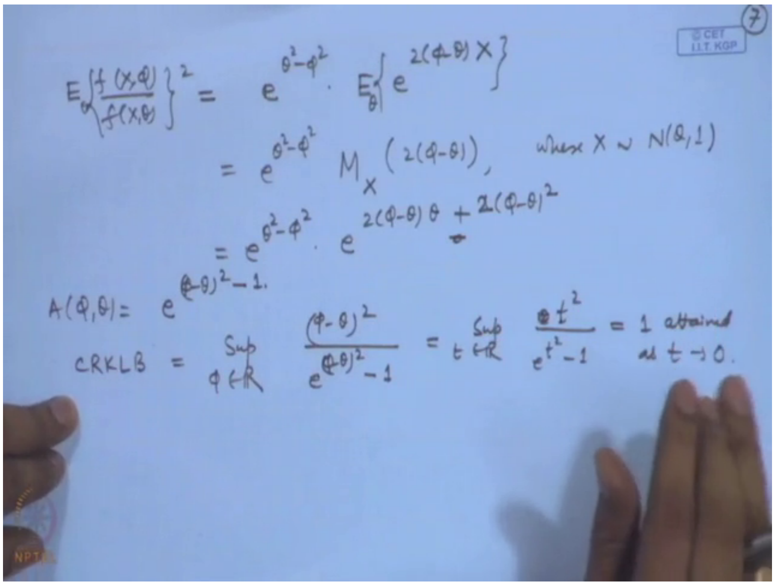
So, in this case FRC lower bound for estimating theta is 1 and of course, you had expectation x is equal to theta and variance of x is equal to 1. So, it is attained. Let us calculate the CRK lower bound here. So, if you want to calculate the CRK lower bound we need to write down the density 1 by root 2 pi e to the power minus half x minus theta 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 pi 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 2 term cancels out and I will be left with e to the power theta 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 square distribution then the moment generating function at the point t that is given by e to the power mu t plus half sigma square t square. So, in that 1 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 square minus phi square by 2 into the moment generating function of x at the point phi minus theta, 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: 30:01)



In a similar way we can calculate the expectation of expectation of f X phi by f X theta whole square. So, this will become equal to expectation of this square. Now, if I square rate it 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 the I will get expectation of e to the power twice phi minus theta X. So, this is equal to e to the power theta 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.

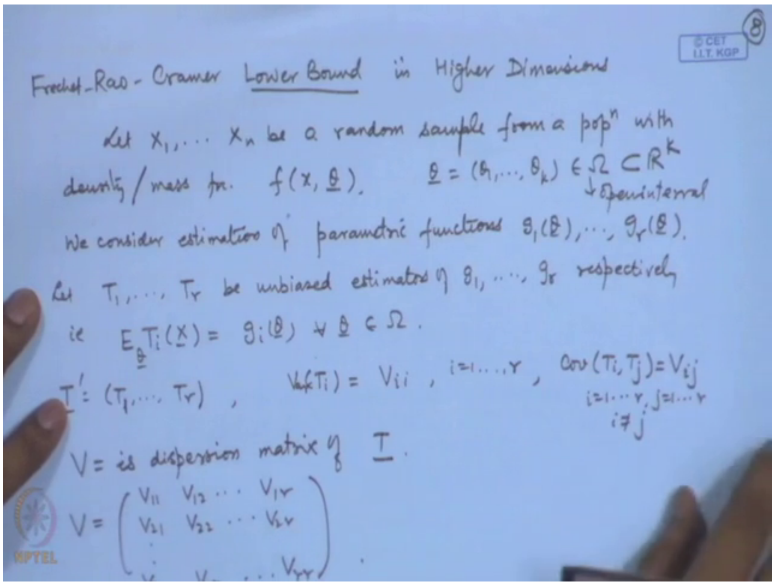
So, naturally now the variance that is a phi theta term is equal to e to the power. So, this term minus square of this term. If I square rate this i get e to the power theta square minus phi square which is the same term here. Similarly, here I have e to the power twice phi minus theta into theta and here if I square rate 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 square plus theta 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 cancel with this and then you get plus theta square which will again cancel plus theta square minus twice theta square plus theta square. So, all of these terms get cancel out, you get minus twice phi theta and plus twice phi theta. So, you are left with only e to the power phi minus theta square minus 1. Now, the CRK lower bound is equal to supremum of phi supremum over phi phi minus theta square divided by e to the power phi minus theta 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 a.

Now, the analysis of maximization of this is simple. In fact, this is a positive term and we 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. Now, we move to another generalization of the Rao-Cramer lower bound that is the case of several parameters. The lower bounds that I have discussed so far here we are assuming or we are calculating the derivatives with respect to 1 parameter that is theta in the problem and of course, you may consider a function of theta for the estimation problem but my density function itself may be a function of say a k dimensional parameter say theta 1, theta 2, theta k. Now, we consider this generalization here.

(Refer Slide Time: 34:54)

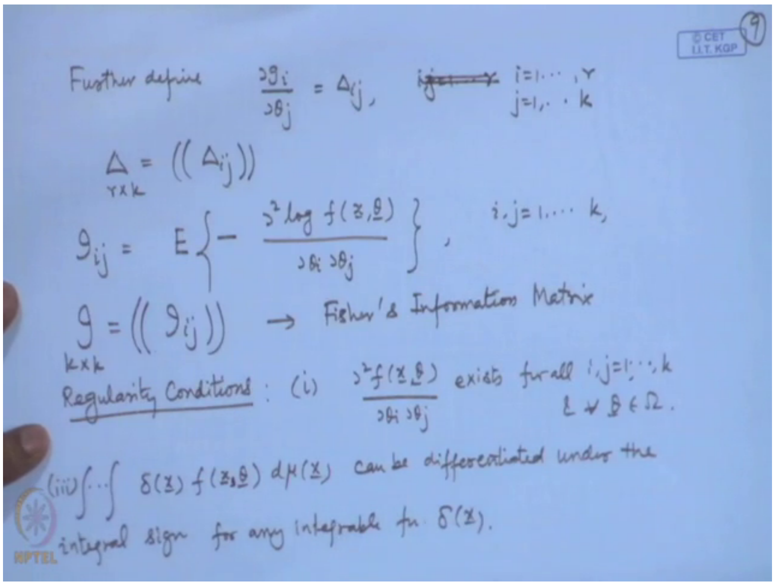


So, Rao-Cramer let me put Frechet Rao-Cramer lower bound. Now, it is not necessarily just a lower bound actually we will call it inequality in higher dimensions. So, let us consider say X 1, X 2, X n be a random sample from a population with. Now, once again we may have a density or mass function f x theta. Now, in the case of 1 dimension we have assumed that theta lies in an open interval and the real line. If we are considering k dimensional parameter here theta is equal to theta 1, theta 2, theta k belonging to omega then this is a subset of k dimensional euclidean space.

But we have to make an assumption that we may consider an open interval in r k. So, what is the meaning of open interval? It can be a ball or a open disk. So, omega is open interval in k dimensional euclidean space and we are considering parametric functions say g 1, g 2, g r etcetera. We consider estimation of parametric functions say g 1 theta, g 2 theta, g r theta. Now, let us consider say T 1, T 2, T r be unbiased estimators of g 1 g 2 g r respectively. That is expectation of T i is equal to g i theta. What we do we define a variance covariance matrix for T 1, T 2, T r. Let us call T as T 1, T 2, T r vector let us define variance of T i as V i i. That is variance for i is equal to 1 to r.

We also define covariance between say T i and T j as V i j for i is equal to 1 to r, j is equal to 1 to r, i not equal to j. So, V is the dispersion matrix of T that is the terms of V are V 1 1, V 1 2, V 1 r, V 2 1, V 2 2, V 2 r and so on. V r 1, V r 2, V r r. Let us make certain regularity assumptions here. Also, we give some notation here.

(Refer Slide Time: 39:18)



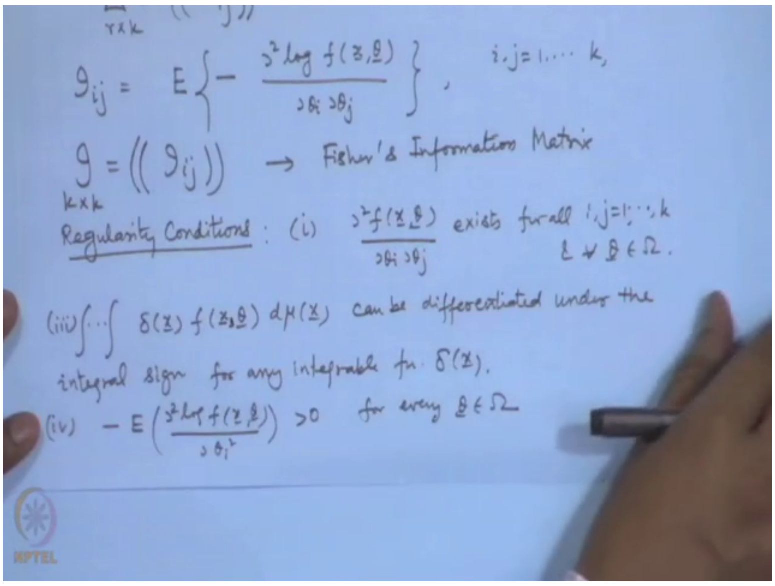
We define say further define delta g i by delta theta j as the terms delta i j for i and j equal to 1 to r. Now, you see here we are considering theta to be k dimensional and g 1 g 2 r r parametric functions are there. So, when I write del g i by del theta j this i will be from 1 to r and j will be from 1 to k. That means I am considering all partial derivatives of g i functions with respect to each of theta 1, theta 2, theta k and delta is the matrix of delta i j that means, it is an r by k matrix. Let us also define a term i j that is equal to expectation of minus del 2 log of f x theta divided by del theta i del theta j.

Once again these are for all i j 1 to k and when i is equal to j this will become second order derivative with respect to theta i. In other cases it is it rated second order partial derivative once with respect to theta i and another time respect to theta j. Once again we are making certain regularity assumptions like second order differentiability like in the Frechet-Rao-Cramer lower bound for one dimensional parameter in that case the order will not make a difference. Whether we write del theta i del theta j or we write del theta j del theta i both will be same under the regularity conditions. i is the matrix of i j's. So, this is a k by k matrix this is called Fisher's information matrix.

Notice, in the case of 1 dimension we have written e to the power expectation of minus del 2 log f x theta by del theta 2 or expectation of del by del theta log f x theta whole square both the quantities were same and I would define it as the Fisher's information. So, now when we have a multidimensional parameter we are defining Fisher's information matrix. Then let us make the regularity assumptions, regularity conditions as in the case of 1 dimensional. We have already made the assumption that the parameter space is an open interval in k dimensional euclidean space. Then we have to make the assumption about the existence of the partial derivatives. So, del 2 f by del theta i del theta j exists for all i j equal to 1 to k and for all theta.

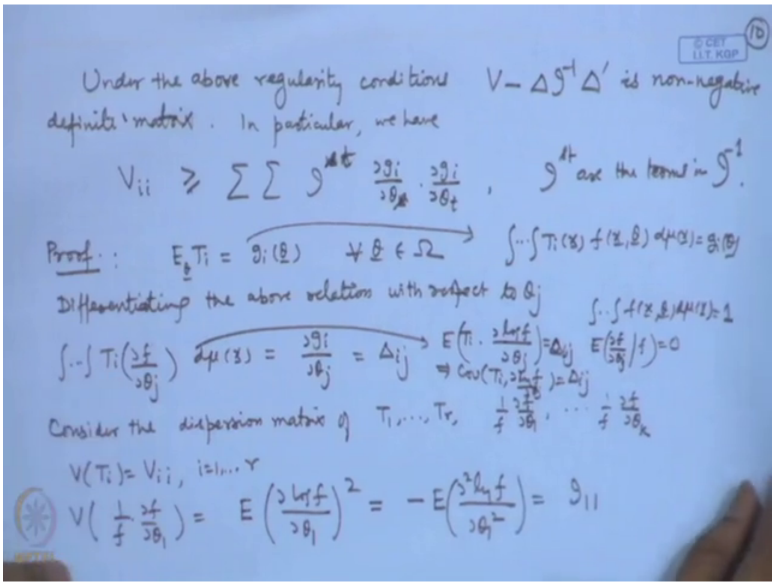
We have to also make the assumption about the differentiability under the integral sign that is del x, delta x f x. So, let me write the joint density as f x theta d mu x can be differentiated. So, this is an iterated enfold integral this can be differentiated under the integral sign for any intergrable function delta.

(Refer Slide Time: 39:18)



We also assume that expectation of del 2 log f x theta by del theta i 2. This is positive for every theta belonging to omega. Basically, the purpose is to have this Fisher's information matrix as a invertible matrix.

(Refer Slide Time: 44:45)



Under these regularity conditions, under the above regularity conditions variance of t. In fact, we can write V minus delta i inverse delta prime is non-negative definite matrix. In the case of 1 dimension we had the term to be non-negative. Here we are saying it is because here we are dealing with the matrix notation this becomes a nonnegative definite matrix. However, for a non-negative definite matrix we know that the diagonal elements are also non-negative now the diagonal elements of this will be of what form in particular if I write only for the diagonal elements, we can write that variance of T i that is for estimation of g i theta this is greater than or equal to double summation i m n del g i by del theta m del g i by let me not take m n let me put here say s T s del theta t.

Where this i s T are the terms in i inverse matrix. So, this Fisher's information matrix i which I have taken if you take the inverse of that s T f element of that I am denoting by i s t. So, this is the lower bound for the variance of unbiased estimator of the i-eth function. Let us look at the proof of this, let us consider expectation of T i is equal to g i theta. Now you differentiate this is true for all theta you differentiate this with respect to say theta j, differentiating the above relation with respect to theta j. So, how will you differentiate actually this relation you can write as T i f x theta d mu x is equal to g i theta.

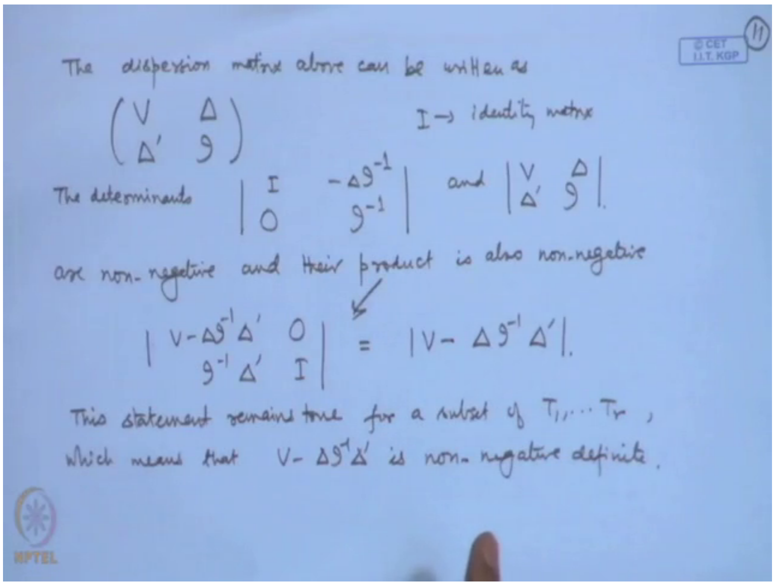
So, if you differentiate this this term will be differentiated because this term does not involve theta. So, we get it as equal to T i del f by del theta j into d mu x is equal to del g i by del theta j that is the term which I define as delta i j and we can also consider. So, this is delta i j also consider the variance covariance are the dispersion matrix of T 1, T 2, T r and 1 by f del f by del theta 1 and so on. 1 by f del f by del theta k. If we consider this r plus k by r plus k dimensional dispersion matrix what kind of terms will occur here.

We will have the variance of T 1 that is V 1 1, variance of T 2 that is V 2 2, variance of T r that is V r r, the variance of 1 by f del f by del theta 1. Now, we have already seen what kind term this will be. Actually, if we consider this here integral of f x theta d mu x that is equal to 1 because this is the density function. If I differentiate this with respect to any theta I will get 0 that term will give me expectation of del f by del theta j divided by f equal to 0 this will be true for all j's. That means, variance of 1 by f del f by del theta 1 it will be equal to expectation of del log f by del theta 1 square or it is equal to minus of expectation of minus del 2 f by del theta 1 square. Let me write this.

So, variance of T i's are V i i for i is equal to 1 to r. Let us consider say variance of 1 by f del f by del theta 1 that is equal to expectation of del log f by del theta 1 square that is equal to minus expectation del 2 log f by del theta 1 2 that is equal to i 1 1 term. Why? Because if I define i i j as expectation of minus del 2 log f by del theta i del theta j here if I take i is equal to j then I get exactly this term. So, this is i 1 1.

So, therefore, variance of 1 by f del f by del theta k etcetera that will be i k k. Now, there will be correlation co covariance term. So, covariance between T 1 T 2 that is V 1 2 and so on. So, these terms will be coming. Now, what other type of terms will come. We will get the covariance between T 1 and 1 by f del f by del theta 1. You look at this relation that we have derived here. Here we are getting expectation of T i into 1 by f del f by del theta j into f. So, this term is reducing to expectation of T i into del log f by del theta j equal to 0 that is giving that covariance T i into del log f by del theta j is equal to not 0 it is equal to delta i j, equal to delta i j. So, the covariance terms between these will give me again delta i j terms.

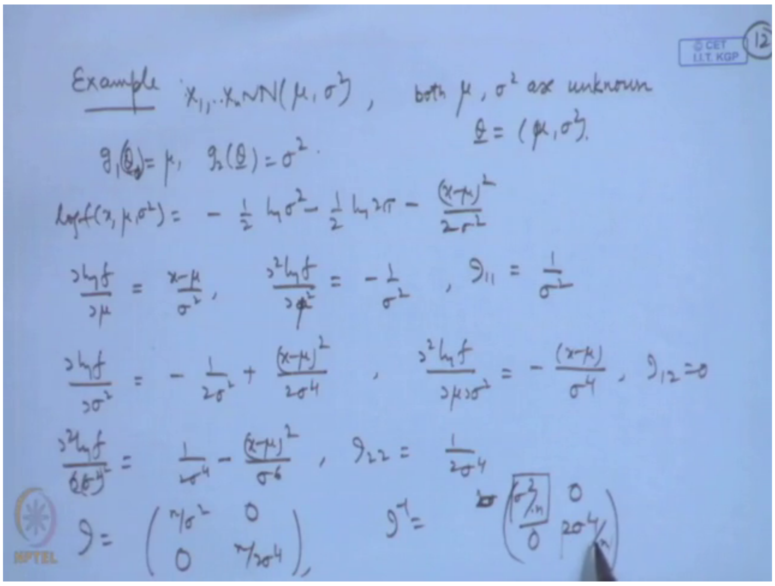
(Refer Slide Time: 52:49)



So, we are getting the the dispersion matrix above can be written as V delta delta prime I. Now, if we consider here the determinants here I denotes identity matrix. So, minus delta I inverse null matrix and I inverse this is information matrix inverse of that and if we consider say V delta delta prime I, these are non-negative and their product is also non-negative. What is the product? Product is this product is V minus delta I inverse delta prime null i inverse delta prime I. That is V minus delta i inverse delta prime.

Now, this is a dispersion matrix therefore, its determinant must be non-negative. Now, the same thing will be true if I take any subset of T 1, T 2, T r and here also any subset of this therefore, for any dimension this determinant will be non-negative. That means, this matrix is non-negative definite this statement remains true for a subset of T 1, T 2, T r which means that V minus delta I inverse delta prime is non-negative definite and if you consider the diagonal elements of this then that would lead to the this statement that is the generalized Rao-Cramer inequality for the k dimensional parameter.

(Refer Slide Time: 55:48)

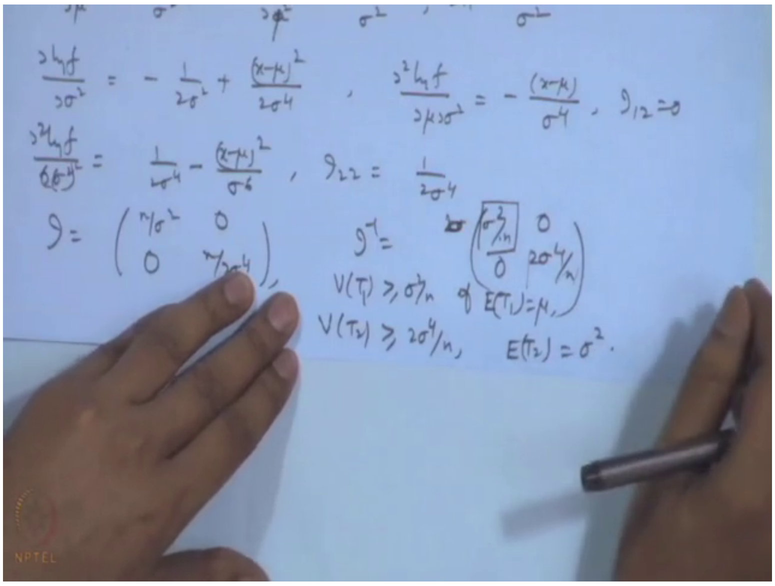


Let me end this lecture by an example let us consider say normal mu sigma square. So, we have a sample X 1, X 2, X n from normal mu sigma square distribution. Here both mu and sigma square are unknown. That means, theta is equal to mu sigma square here. So, the problem is to find out the Rao-Cramer inequality for the unbiased estimator of mu and sigma square. So, I am considering g 1 as mu and g 2 theta as sigma square. So, we consider here the density function log of f will be equal to minus 1 by 2 log sigma square minus 1 by 2 log 2 pi minus x minus mu square by 2 sigma square.

If we consider del log f by del mu that is x minus mu by sigma square del 2 log f by del sigma mu 2 that will be equal to minus 1 by sigma square. So, I 1 1 term is simply minus of this expectation that is 1 by sigma square. Similarly, if I consider del log f by del sigma square i get it as minus 1 by 2 sigma square plus x minus mu square by 2 sigma to the power 4, del 2 log f by del mu del sigma square that will be equal to minus x minus mu by sigma to the power 4, if I take expectation of this it will become 0.

So, i 1 2 is 0 similarly del 2 log f by del sigma 2 that will be equal to 1 by 2 sigma to the power 4 minus x minus mu square by sigma to the power 6. So, that gives us i 2 2 as equal to 1 by 2 sigma to the power 4. So, i matrix simply becomes n by sigma square 0 0 n by 2 sigma to the power 4. So, i inverse is equal to 2 sigma sigma square by n sigma sigma 2 sigma to the power 4 by n 0 0. So, of diagonal is here is 0.

(Refer Slide Time: 55:48)



So, variance of an unbiased estimator of mu will be greater than or equal to sigma square by n, the variance of an unbiased estimator of sigma square will be greater than or equal to 2 sigma to the power 4 by n. So, variance of T 1 will be greater than or equal to sigma square by n if expectation of T 1 is mu and variance of T 2 will be greater than or equal to 2 sigma to the power 4 by n if expectation of T 2 is equal to sigma square.

We can also develop this rao-cramer inequality in the higher dimensions for various practical examples like a bivariate normal distribution where we have 5 parameters mu 1, mu 2, rho sigma 1 square, sigma 2 square etcetera So, we have considered in detail 1 method for finding out the minimum variance unbiased estimator and this method is not only useful for finding out the minimum variance unbiased estimator, it is also used in other applications of decision theory such as proving admissibility or minimaxity of estimators also. In the next lectures we will take up another concept that is of sufficiency.