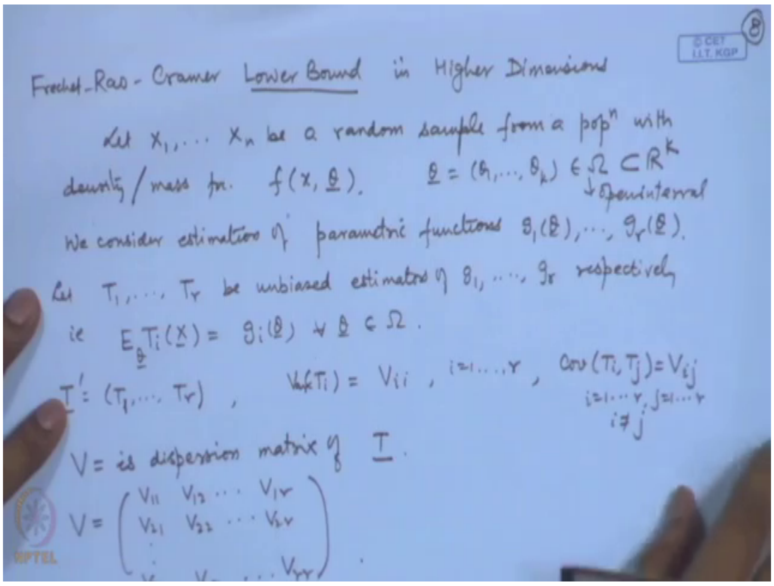
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institutes of Technology, Kharagpur

Lecture – 22
Lower Bounds For Variance – VIII

Now, we move to another generalization of the Rao Cramer Lower Bound; that is the case of several parameters. The lower bound that I have discussed so far here we are resuming or we are calculating the derivatives with respect to 1 parameter that is theta in the problem. And of course, we 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: 01:01)

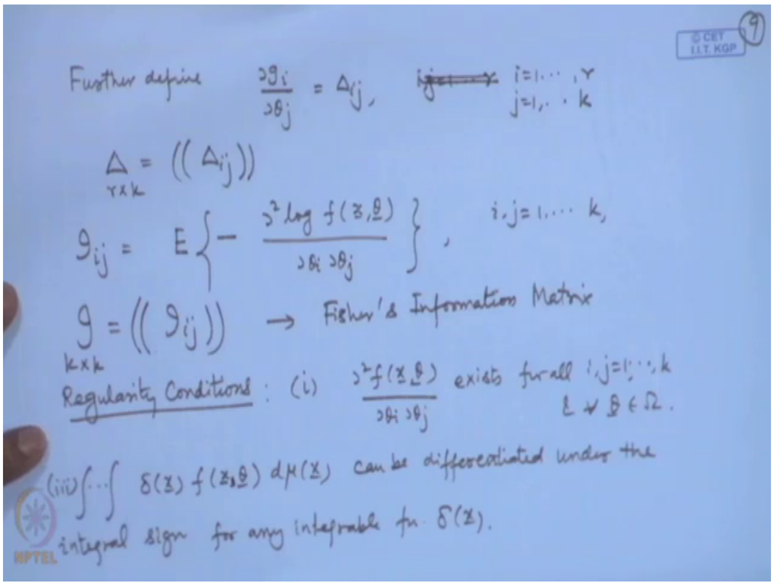


So, Rao Cramer, let me put Frechet Rao Cramer lower bound, now it is not necessary just the lower bound actually we will call it in equality in higher dimensions. So, let us consider say X 1, X 2 X n be a random sample from a population with now, once again we may have a density or mass function f x theta.

Now, in the case of one 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 another assumption that we may consider an open interval in r case. So, what is the meaning of open interval? It can be a ball or a open desk. So, omega is open interval in k dimensional Euclidean space and we are considering parametric functions say g 1 g 2 g r etcetera ok. 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 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: 05:26)



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 right 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 ok.

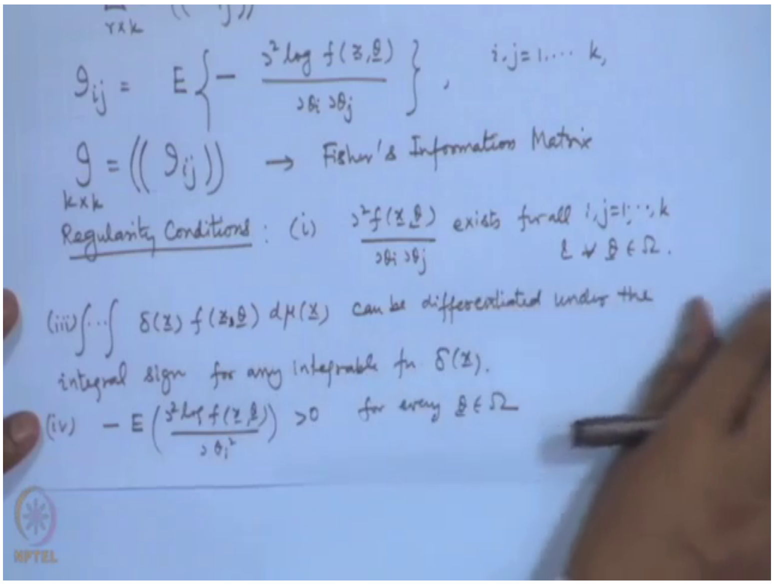
Let us also define a term i ij, that is equal to expectation of minus del 2, log of f x theta divided by del theta i del theta j. Once again these are for all i j 1 to k and when i is equal to j, this will become second order derivative with respect to theta i, in other cases it is it rated second order 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 Fisher or 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 are we right del theta j del theta i, both will be same under the regulatory conditions. I is the matrix of i j, so this is a k by k matrix, this is called Fisher's information matrix. Notice in the case of one dimension, we have written e to the power expectation of minus del to log f x theta by del theta 2 r 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 ok.

So, now when we have a multi-dimensional parameter, we are defining Fisher's information matrix ok. Then let us make the regulatory assumptions, regularity conditions as in the case of one 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.

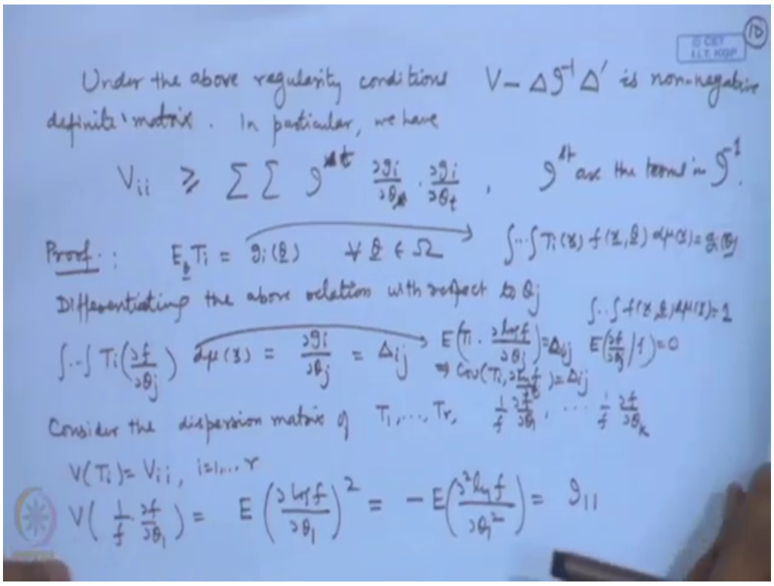
We have to also make the assumption about the differentiability under the integral sign that is del x, delta x f x. So, let me write the joint density was fx theta, d mu x can be differentiated So, this is an treated n fold integral, this can be differentiated under the integral sign for any integrable function delta.

(Refer Slide Time: 10:22)



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 only 1. Basically the purpose is to have this Fisher's information matrix as a invertible matrix.

(Refer Slide Time: 10:53)



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 one 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 non negative 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 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 th 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 do 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 is equal to T i del f by del theta j into d mu x is equal to del gi by del theta j, that is the term which I defined 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.

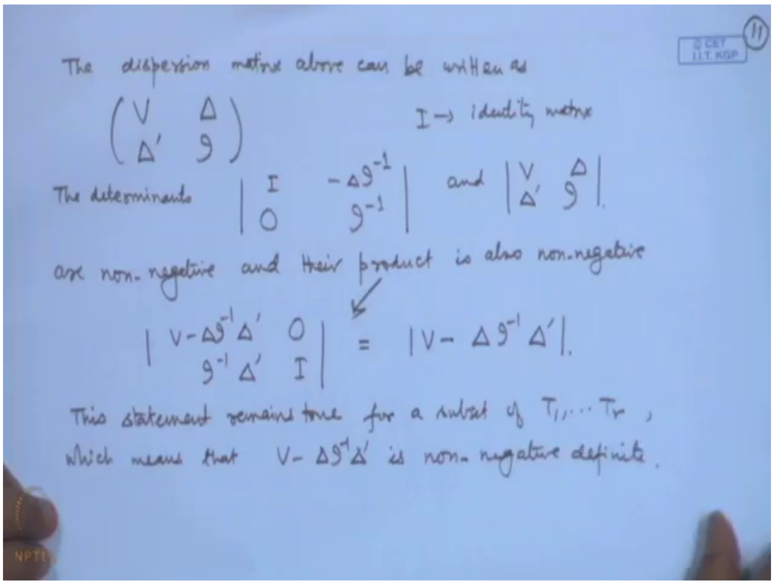
If we consider this r plus k by r plus k dimensional dispersion matrix, what kind of terms will occur here? We will have the variance of T 1, that is V 1 1, variance of T 2 that is V 2 2, variance of T r that is V r r, the variance of 1 by f del f by del theta 1. Now we have already seen what kind of terms this will be. Actually if we consider this here integral of fx 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 one by f del f by del theta 1, it will be equal to expectation of del logo f by del theta 1 is 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 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. Why? Because if I define I i j as expectation of minus del to 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 11. So, therefore, variance of 1 by f del f by theta kth sector that will be i k k.

Now, there will be correlation, co-variance terms. So, covariance between T 1 T 2 that is V 1 V 2 and so on. So, this term 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 ij, equal to delta ij. So, the covariance terms between these will give me again delta i j terms.

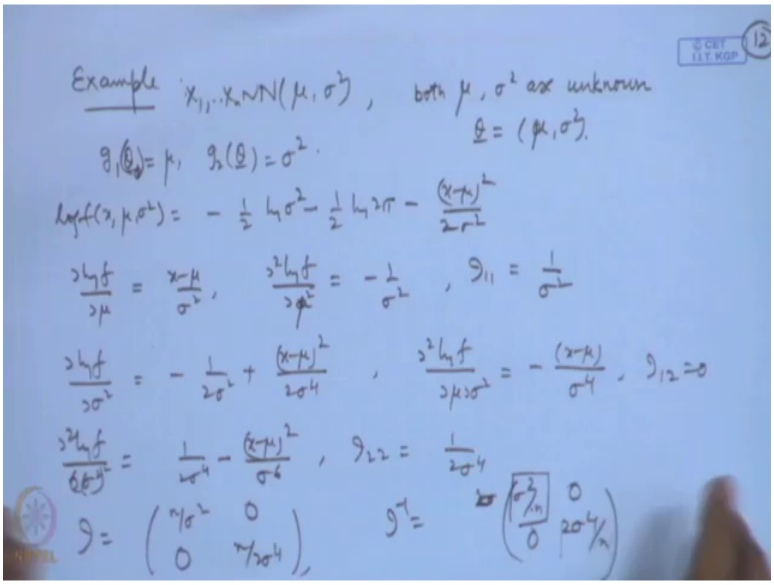
(Refer Slide Time: 18:56)



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 considered 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, you consider the diagonal elements of this then that would lead to this statement, that is the generalized Rao Cramer inequality for the k dimensional parameter.

(Refer Slide Time: 21:55)



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 in equality for the unbiased estimator of mu and sigma square. So, I am considering g 1 as mu and g 2 theta s sigma square.

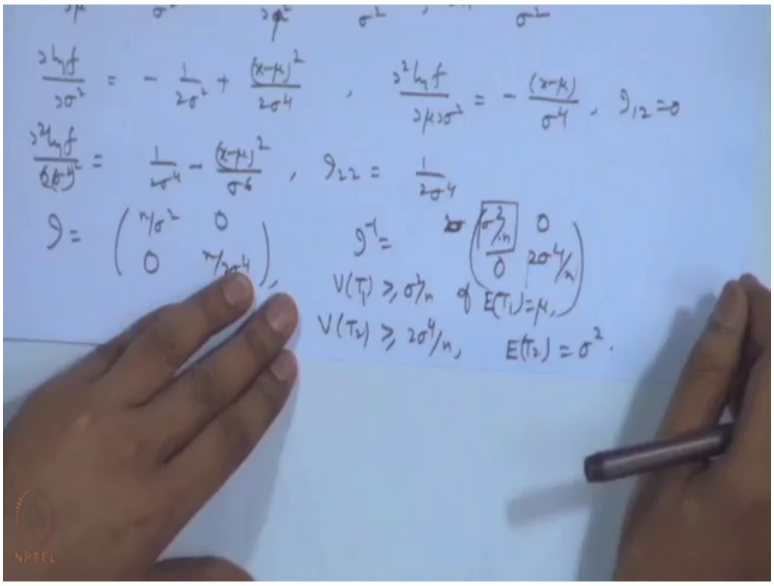
So, we consider here the density function log of f will be equal to minus 1 by 2 log sigma square minus 1 by 2 log 2 pi minus x minus mu square by 2 sigma square. If we consider del log f by del mu, that is x minus mu sigma square del 2 log f by del sigma is mu that will equal to minus 1 by sigma square.

So, I 11 terms is simply minus of this expectation that is 1 by sigma square. Similarly if I considered 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 four del 2 log f by del nu 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 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 is 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 2 sigma to the power 4 by n 0 0.

So, half diagonal here is 0. 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.

(Refer Slide Time: 24:48)



So, variance of T1 will be greater than equal to sigma square by in if expectation of tone is mu and variance of T 2 will be greater than or equal to 2 sigma to the power 4 by n, if you expectation of T 2 is equal to sigma 2. We can also develop this Rao Cramer inequality in the higher dimension for various practical examples like a bivariate normal distribution, where we have 5 parameters, m 1 mu 2 row sigma 1 square sigma 2 square etcetera.

So, we have considered in detail one method for finding out the minimum variance and bi estimators. And, this method is not only useful for finding out the minimum variance and bi estimator; it is also used in other applications of decisions theory; such as proving admissibility or minimaxity of a estimators also. In the next lectures we will take up another concept; that is of sufficiency.