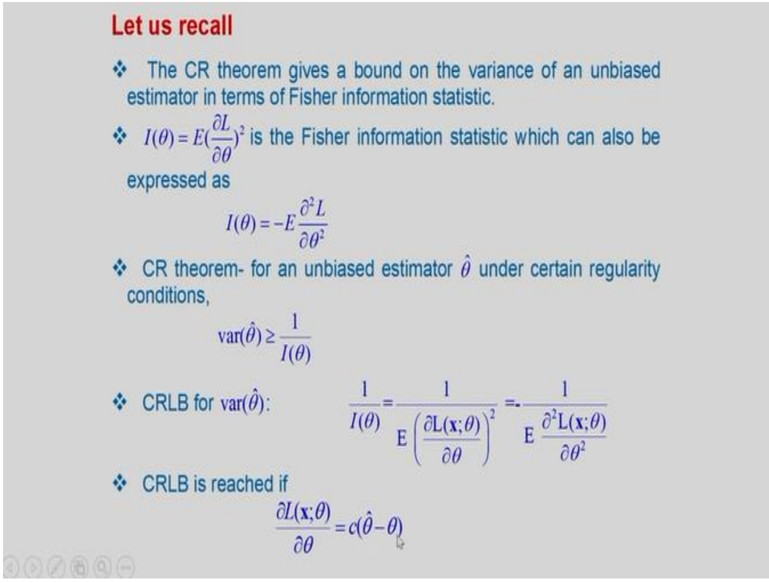
Statistical Signal Processing

Prof. Prabin Kumar Bora
Department of Electronics & Electrical Engineering
Indian Institute of Technology, Guwahati

Lecture No 10
Estimation Theory 3 Cramer Rao Lower Bound II

Hello students, welcome to lecture 10, Cramer Rao Lower Bound two.

(Refer Slide Time: 00:42)

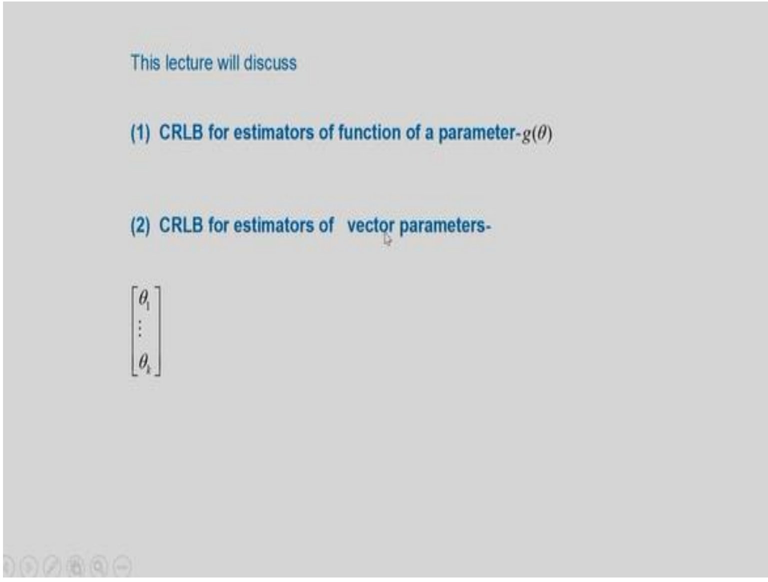


Let us recall, the Cramer Rao theorem gives a bound on the variance of an unbiased estimator in terms of Fisher information statistics. Now what is a Fisher information statistic? I theta that is expected value of del L del theta Squad is the Fisher information statistic, which can also be expressed as I theta equal to minus of expected value of del L, del theta squared.

CR theorem, or Cramer Rao theorem for an unbiased estimator theta head under certain regularity conditions, variance of theta head is greater than equal to one by I theta. So this is the statement of Cramer Rao theorem. Therefore, Cramer Rao lower bound theorem CRLB for variance of theta head is equal to one by I theta, that is equal to one by expected value of del L del theta whole squared and this can be altered it in a one by minus expected value of del2L, del theta squared.

And one important point is CRLB is reached if the partial derivative of log likelihood function that is del L del theta can be expressive to theta head minus theta, versus theta can be a function of theta.

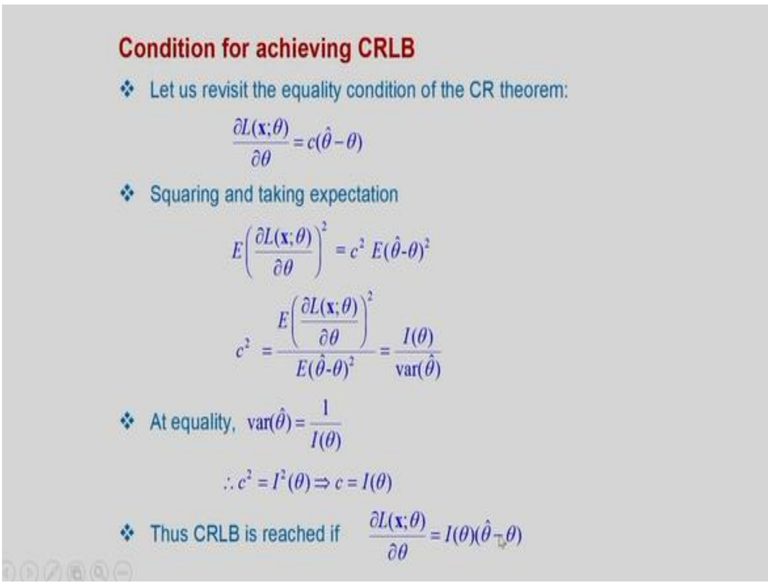
(Refer Slide Time: 02:16)



In this lecture we will discuss Cramer Rao lower bound theorem CRLB for estimator of function of a parameter g theta earlier, we discussed the estimator theta but now we will discuss what will be the CRLB for an unbiased estimator for g theta. Second thing will be discusses CRLB for estimator of vector parameters.

So, parameters are supposed multiple parameters are there is that theta one up to theta k which are emerged as vector. So what will be there will be CRLB for unbiased estimators for this vector parameter.

(Refer Slide Time: 02:52)



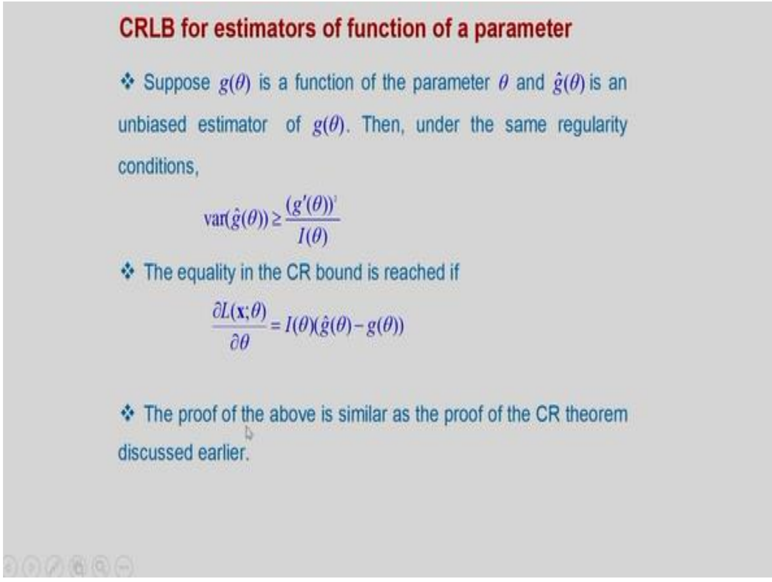
Let us revisit the condition for achieving CRLB, we know that for CRLB we have the equality condition, given by the del L del theta is equal to c times theta head minus theta.

Squaring and taking expectation we get the sum of del L del theta whole squad is equal to c squad times theta head minus theta whole squared. So we will squared is the expectation of both sides.

So that way we get this from this we get, c squared is equal to E of del L del theta Squared, that is I theta divided by up to that theta head minus theta squared. That is called million of theta head. So this is the expression for C squared, which we derived in the last lecture. Now, at the equality, we know that variance of theta head is equal to one by I theta. So if I put variance of theta is equal to one by I theta.

Then, I will get c squared is equal to I square theta from this c square is equal to variance of I theta and will interpret that is equal to one by ethicality, that the condition, Therefore CRLB is reached if del L del theta equal to I theta into theta head minus theta.

(Refer Slide Time: 04:28)

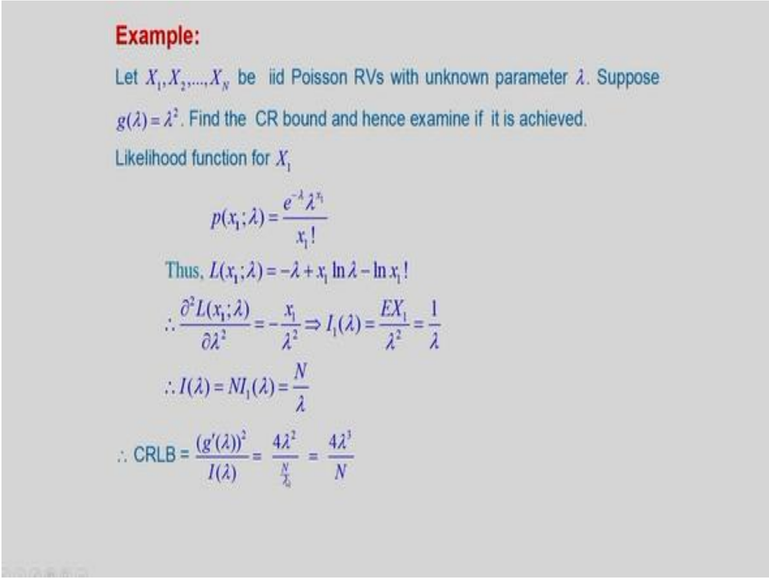


We will discuss about CRLB for the estimators of function of a parameter. So suppose g theta is a function of the parameter theta and the theta is an unbiased estimator of g theta. Then under the same regularity conditions, a variant of the head theta is greater than and equal to the g dash theta whole square divided by I theta, that it is the first order derivative with respect to theta.

So this is the Cramer Rao theorem for the function of a parameter and the equality in CR bound for this inequality, the equalty will be reached if partial derivative of that is L x, theta that is partial derivative of the likelihood function is equal to I theta multiplied by the g head theta minus g theta. So you can factorize del L, del theta to form I theta into g head theta minus g theta.

So this is the condition for reaching the CR bound, the proof of the above is similar to the proof of CR theorem, we have discussed earlier.

(Refer Slide Time: 05:51)

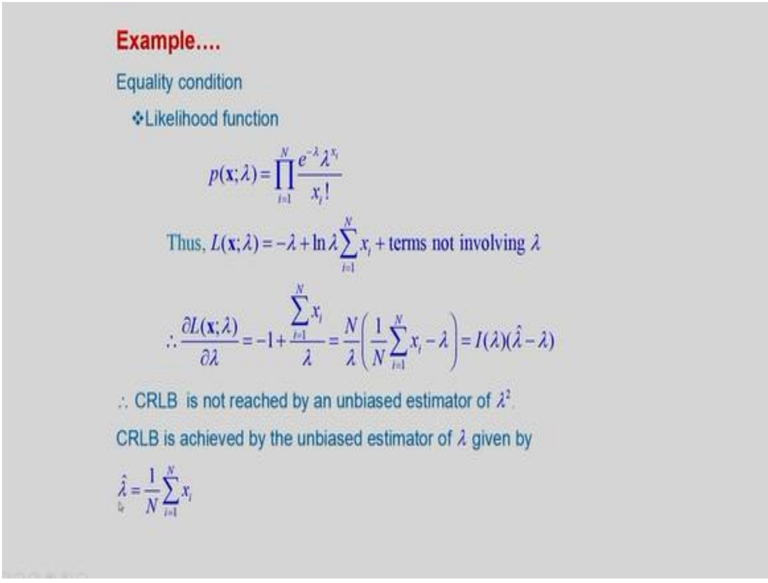


We will consider one example; Let X1, X2, up to Xn be iid poisson random variables with unknown parameter lambda. Suppose the g lambda is equal to lambda square. So find the CR bound for the unbiased estimator of the g lambda. And hence, examine if it is achieved. Now exercise the iid we will find out the likelihood function of x1, that is equal to PMF, Px1 is the function of lambda that is equal to e to the power minus lambda into lambda to the power x1 divided by x1 factor.

Thus the likelihood function of x1will be equal to will take the logaritham, therefore it will be minus lambda plus x one in the log of lambda minus log of x and factor in. And if we take the second order partial derivative with respect to lambda will get delta squared is equal to minus x and y lambda square. So, for information, we have to take the expected value, therefore x1 of lambda will be E of X1 divided by lambda squared, that is equal to one by lambda.

Since there are n iid random variable, therefore I lambda will be equal to N times I one lambda that is equal to n by lambda. Therefore, CRLB is equal to the g dash lambda squared divided by I lambda that is equal to 4 lambda square. Because if we take the derivative this will be two lambda square, will be 4 lambda squared divided by N lambda, and that is equal

to 2 lambda cube by N, and this is a serial way for any unbiased estimator of the lambda.
(Refer Slide Time: 07:59)

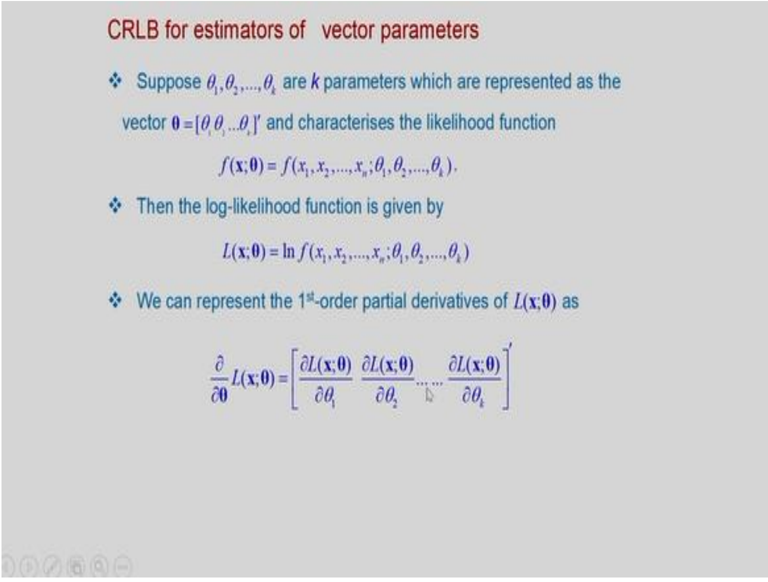


Now let us examine the equality condition, the likelihood function, we are again relating this P of X vector lambda, has a function of lambda is equal to product of e to the power minus lambda, lambda to the power xi divided by factor xi, i going from one to N, and this is the join pyramid or likelihood function in terms of lambda, and taking the logarithm L x; lambda will be equal to minus lambda plus log of summation x i, either in from one to N plus terms, not involving lambda.

Therefore, del L, del lambda will be equal to minus one derivative of this is minus one plus summation xi, i going from one to N, and derivative of log N is one by lambda. So, that this is del L, del lambda is equal to minus one plus summation xi, i going from one to N divided by lambda, and this can be simplified as N by lambda into summation xi, i going from one to N by N minus lambda, and this we can write as, I lambda into lambda head minus lambda here it is lambda not lambda square.

Therefore, we conclude that CRLB is not reached by an unbiased estimator of lambda square. It is risk by an unbiased estimator of lambda, which is given by lambda head is equal to one by N times summation xi, i going from one to N.

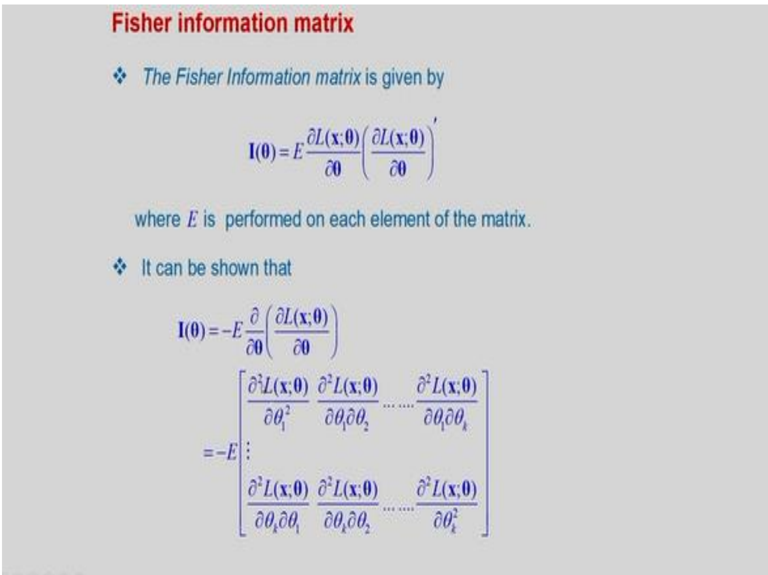
(Refer Slide Time: 09:48)



Now we will be considering CRLB for estimator of vector parameters, suppose theta1, theta2 up to theta k, are k parameters which are represented as a vector, this is a K dimensional vector theta given by theta1, theta 2 up to theta k transpose this is a column vector. This vector characterizes the likelihood function and likelihood function is given by F x theta that is f of x1,x2 up to xn and parameter started theta1,theta2 up to theta k.

Then the log-likelihood function is given by we will take a logarithm so this is the log-likelyhood function. Now because it is a function of several variables or we can get the partial derivative vector. So, we can take the del L del theta 1, del L del theta 2 like that so one column we can take this column is the vector representation for the partial derivatives of del L del theta.

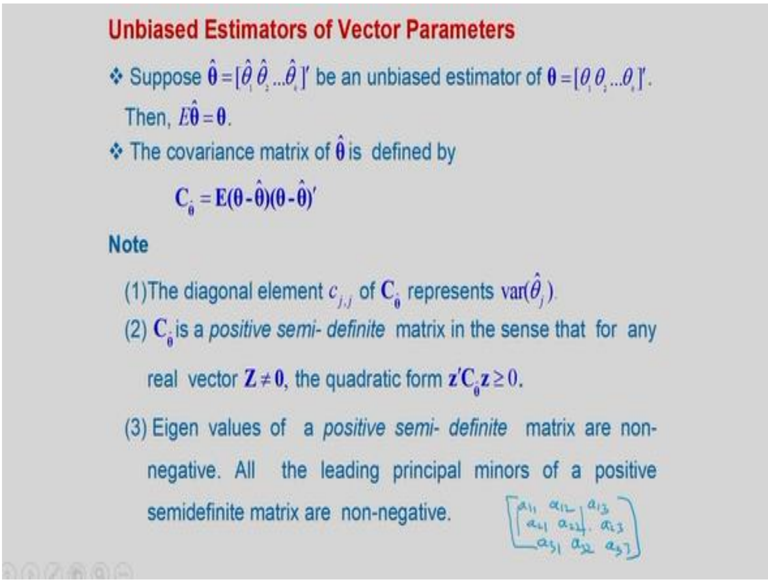
(Refer Slide Time: 10:56)



Fisher information matrix, the Fisher information matrix is given by I theta matrix. That is given by E of del L del theta, this is a vector, del L, del theta transpose this product will be a matrix and then we take the expectation of the individual element of matrix. So E is performed on its element of the matrix. So this is the Fisher information matrix and it can be also saw that I theta is equal to minus expectation of del del theta of the del L del theta vector.

So that way this will become a matrix now and that matrix is given where minus expectation of del 2 L del theta l square, del 2 L del theta l del theta 2, del 2 L, del theta 1, del theta k like that last row will be del 2 L del theta k del theta l, del 2 L del theta k del theta 2, up to del 2 L, del theta k square. So this is the Fisher information matrix.

(Refer Slide Time: 12:15)



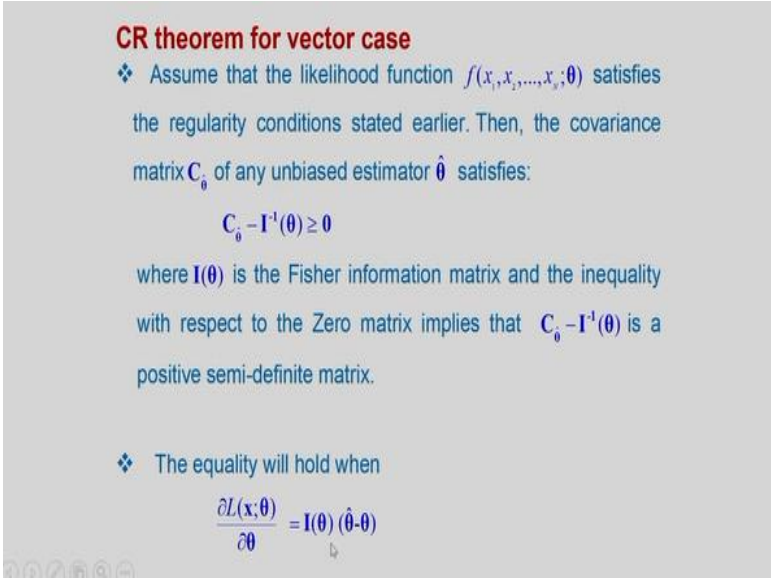
Now will be considering suppose, unbiased estimators for vector parameters. That is our concern, suppose theta head is equal to theta 1 head, theta 2 head up to theta k head transpose, this is a vector be unbiased estimator of theta, theta equal to theta1, theta2, up to theta k transpose. So, then E theta head equal to theta that is unbiased estimators and the covariance matrix of theta head is defined by C theta head that is equal to E of theta minus theta head in to theta minus theta head transpose this is the definition of covariance matrix.

Now, we examine some interesting properties of C, this covariance matrix, the diagonal element C j,j of C theta head represents the variance of theta head j. So, the j th diagonal element will be the variance of theta head. Similarly, C theta head is a positive semi definite matrix that is very important, this covariance matrix is a positive semi definite matrix, how do I define positive semi definite matrix for any real k dimensional vector Z not equal to zero.

The quadratic form this quantity z transpose C theta head into z will be greater than equal to zero, this is the definition of positive semi definiteness and if it is greater than zero, then it will be positive definite. Now, test for positive definiteness is Eigen values of positive semi definite matrix are non negative. So, if we get all Eigen values greater than equal to zero then it will be positive definite matrix and all the leading principle minors of positive semi definite matrix are non negative, this is a test for simple case.

Suppose I have some matrix like this a11, a12, a13 suppose a21, a22, a23, a31, a32, a33. So this is the matrix. Now we have to consider all leading principal minor that this we have to consider on the left hand side, we have to consider minor parts minor will be a11 this will be greater than equal to zero second minor will be determinant of this, this, and this would be also greater than equal to zero. And third, leading principle minority determinant, that's also greater than equal to zero. So this is the test for positive semi definiteness.

(Refer Slide Time: 15:11)

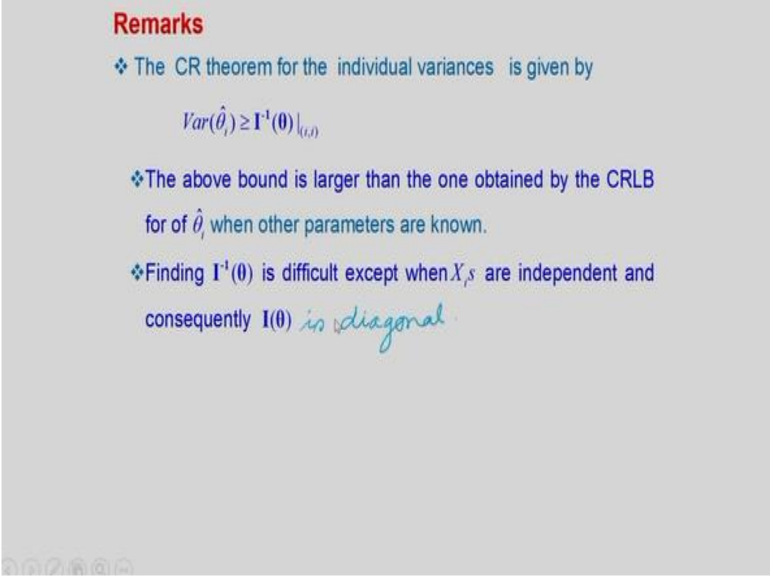


Now, we can state this CR theorem for vector case, assume that the likelihood function, this is the likelihood function satisfies the regularity conditions stated earlier. Then, the covariance matrix C theta head of any unbiased estimator theta head that is the covariance matrix satisfy this relationship, covariance matrix minus I inverse of theta, theta is the fisher information matrix that will be greater than equal to zero.

So C of theta head minus I inverse of theta, will we always greater than equal to zero matrix. Where I theta is the Fisher information matrix and the inequality now this inequality the matrix in inequality, this inequality, with respect to zero matrix implies that this quantity theta head minus I inverse theta is a positive semi definite matrix. So you have to test whether this matrix C theta head minus I inverse is a positive semi definite matrix.

This is the condition for Cramer Rao Lower Bound. So this C of theta head by the inverse of theta is a positive semi definite matrix. Now again the eqality will be coming in the power vector the del L del theta will be equal to I theta of theta head minus theta. So this is the condition to test the weather this one bound will be reached.

(Refer Slide Time: 16:52)

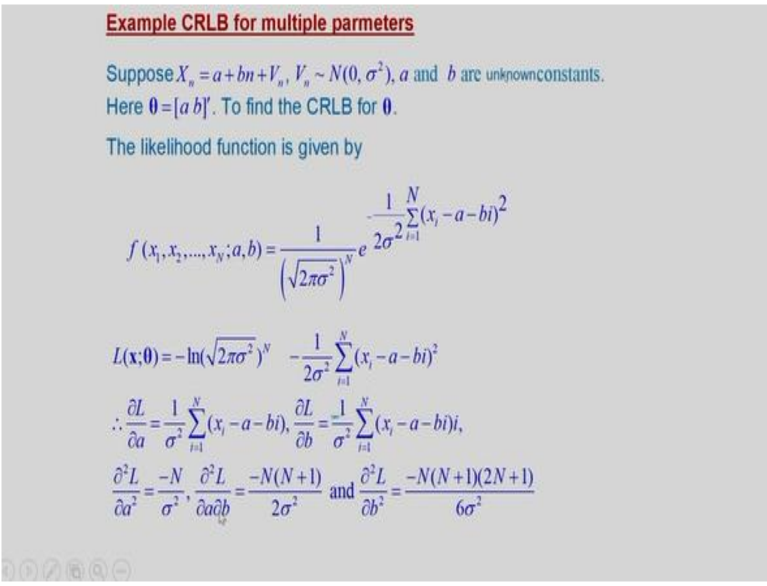


We will make some remarks. Now, what will be the individual suppose we have found this bound in terms of C of theta head, but what will be the real bound for individual element of C theta head, so that the individual variance will be given by variance theta i head will be greater than equal to I diangonal inverse of theta. That is I diagonal elements of I inverse theta. This is the variance of theta i.

The bound is larger than the one obtained by CRLB for of theta one head when other parameters are known. For example we can find out individually, the bound on variants of theta i head. Assuming that other parameters are known, in that case we will get a lower bond. So, this bound what do we get through this matrix from these matrix, and that is larger than what we get, individually.

That is larger than the one obtained by CRLB for theta i head when other parameters are known. We will illustrate this concept by finding I inverse theta is difficult except when XiS are independent. In that case I theta is diagonal matrix, so finding out I theta inverse is easier.

(Refer Slide Time: 18:25)

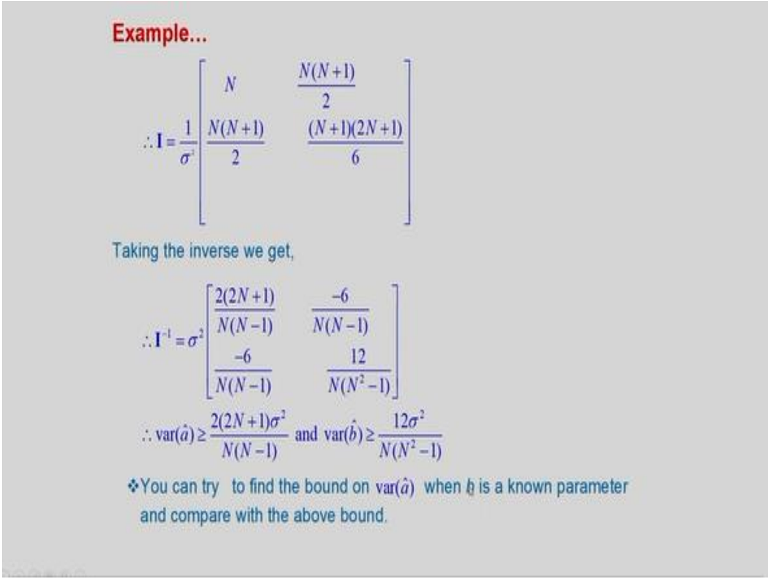


We will consider an example of CRLB for multiple parameters. Suppose Xn that is equal to a+bn+Vn, this is a constant and it is a trend here bn plus Vn. Where Vn is a normal Gaussian noise and noise sequences are independent, a and b are unknown parameters. Here, parameter vector is equal to a, b transpose. Now we have to find out the CRLB for theta, so in this case because these variances are independent.

Therefore, we can write the join PDF that is the likelihood function that will be given by this expression 1 by root 2 pai sigma square to the power N into e to the power minus half of 1 by 2 sigma square in to summation xi minus a minus bi whole squared i going from 1 to N this is the N dimensional Gaussian and we can take the logarithm to get the log likelihood function given by this.

This is this expression for log likelihood function and if I take the partial derivative with respect to a will get this quantity. Similarly, del L del b will be given by this, because bi is there. So, we will have I here and similarly, we can take the second order derivative now, suppose it del L by del a is given thought we can find out del L del a squared, we can find out del2L del a del b, del 2 L del b square.

(Refer Slide Time: 20:16)

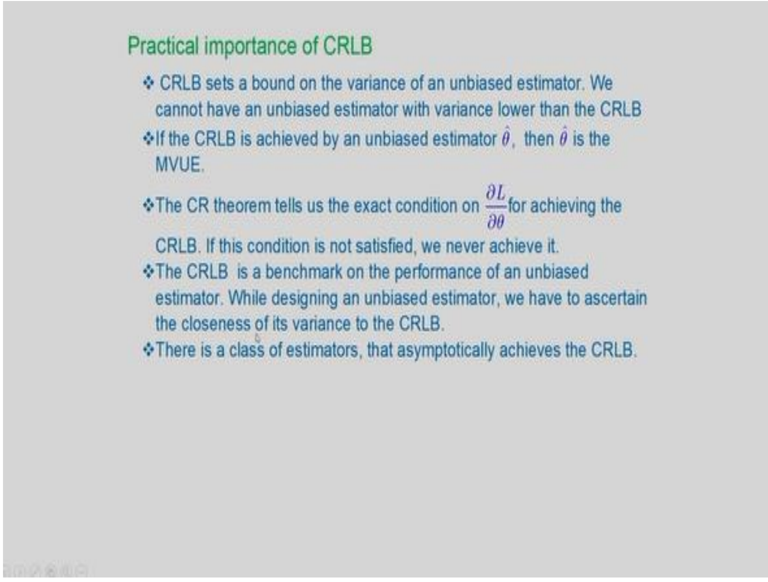


So, these are the fellows therefore I will be equal to one by sigma square times the matrix, first element will be N, and then also second element will be N into N +1 divided by 2. Similarly, second row, first element is N into n+1 divided by 2 and N +1 into twoN+1 divided by 6, this is the information matrix. Taking the inverse, which is a two by two matrix we can easily take the inverse.

The inverse, I inverses is equal to given by sigma square times, or sentiment is 2 into 2N+1 divided by N into N-1, is minus 6 by N into N-1, similarly this element and this element is same. This last element is 12 by N into N squared minus 1. Now, this term is related to the variance of parts parameter, it is variance of a head. Similarly this part is related to the variance of the second parameter that is variants of b head is greater than equal to 2 into 2N+1 time sigma squared by N into N-1.

And variance of b head is greater than equal to 12 sigma squared by N into N squared minus 1. You can try to find the bound on variance of a head when b is a known parameter and compare with the above bound, we have found a bound, when both a and b are unknown, but suppose b is known then try to find out the bound on the variance of a head and compare this bound.

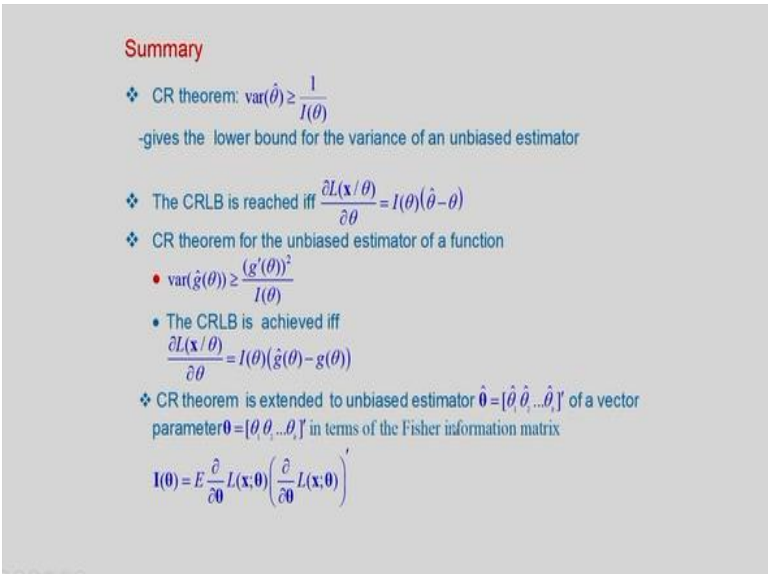
(Refer Slide Time: 22:07)



Practical importance of CRLB, Cramer Rao Lower Bound sets a bound on the variance of an unbiased estimator. We cannot have an unbiased estimator with a variance lower than the CRLB and if CRLB is achieved by an unbiased estimator theta head, then theta head is the MVUE. So that is an important aspect because MVUE will be the CRLB. This CR theorem came tells us the exact condition on del L by del theta for achieving the CRLB.

If this condition is not satisfied, we never achieve it. This CRLB is a benchmark on the performance of an unbiased estimator. While designing an unbiased estimator, we have to ascertain the closeness of its variance to the CRLB. There is a class of estimators that asymptotically achieves the CRLB.

(Refer Slide Time: 23:15)

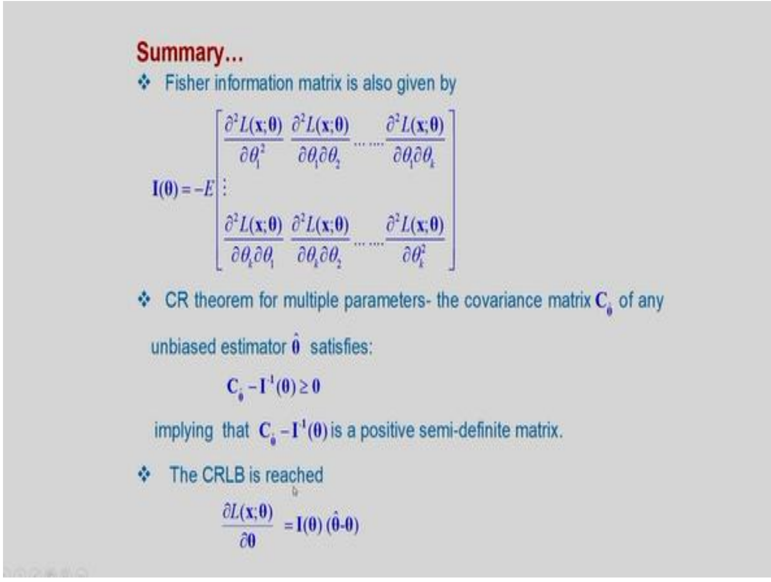


Let us summarize the lecture; first Cramer Rao theorem variance of theta head is greater than equal to one by I theta. It gives the lower bound for the variance of an unbiased estimato. Theta head is an unbiased estimator of theta. This CRLB is reached iff del L del theta equal to I theta into theta head minus theta. Cramer Rao theorem unbiased estimator of a function is given by variants of g head theta is greater than equal to g dash theta whole squared divided by I theta.

Where g dash theta is the derivative of g theta with respect to I theta. This CRLB is achieved iff the del L del theta is equal to I theta into g head theta that estimator minus g theta. This is the condition for receiving the CRLB by g head theta. CR theorem is extended to an unbiased estimator theta head of a vector parameter theta, which is given by theta 1, theta 2, up to theta k transpose and it is given in terms of Fisher information matrix.

It is given in terms of fisher information matrix I theta, this is a matrix. This is given by del L del theta this is a vector into del L del theta transpose. So this matrix is the Fisher information matrix.

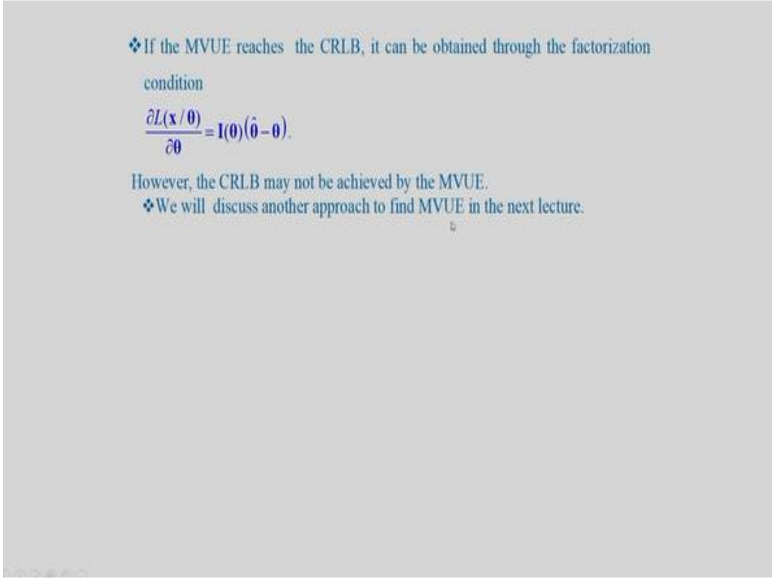
(Refer Slide Time: 24:52)



Fisher information matrix is also given by this expression, I theta matrix is equal to minus expectation of del 2 L del theta l square, del 2 L del theta l del theta 2, del 2 L, del theta 1, del theta k and going this way, last row will be del 2 L del theta k del theta l, del 2 L del theta k del theta 2, up to del 2 L, del theta k square.

Cramer Rao theorem for multiple parameters, the covariance matrix C theta head of any unbiased estimator theta head satisfies this is covariance matrix, minus inverse of information matrix will always greater than equal to zero matrix. This implying that covariance matrix minus inverse of information matrix, is a positive semi definite matrix. And similarly the CRLB is reached if del L del theta, this is a vector equal to I theta of theta head minus theta vector.

(Refer Slide Time: 26:08)



If the MVUE reaches the CRLB, it can be obtained through the factorization condition del L del theta is equal to I theta into theta head minus theta. However, the CRLB may not be achieved by the MVUE. We will discuss another approach to find MVUE in the next lecture. Thank you.