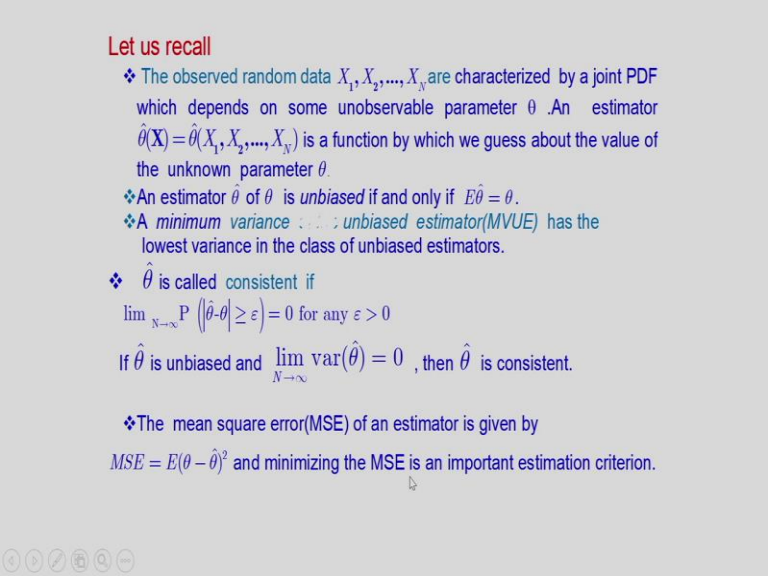
Statistical Signal Processing
Prof. Prabin Kumar Bora
Department of Electronics & Electrical Engineering
Indian Institute of Technology - Guwahati

Lecture 9
Estimation Theory 2: MVUE and Cramer Rao Lower bound

(Refer Slide Time: 00:50)

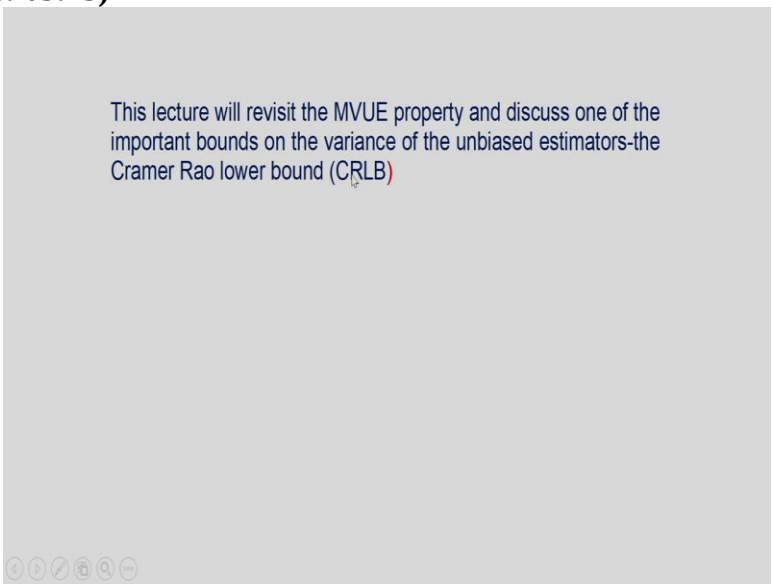


Hello students, welcome to lecture 10 on Estimation theory 2 it will cover MVUE and Cramer Rao lower bound. Let us recall the observed random data X1, X2 up to XN are characterized by a joint PDF with depends on some unobserved parameter theta and estimator theta hat X is a function by which we guessed about the value of the unknown parameter theta. So, corresponding to suppose random data we have a PDF, which is characterized by a parameter theta. And estimator theta hat X is a function by which we guessed about the value of the unknown parameter theta.

An estimator theta hat of parameter theta is unbiased if E theta hat = theta. So, this is the requirement for an unbiased estimator. A minimum variance unbiased estimator MVUE has the lowest variance in the class of unbiased estimators. So this is 1 of the important estimator that is minimum various unbiased estimator MVUE and theta hat that is estimator theta hat is called consistent if this probability of m arbitrary there is summation goes down to 0 as N tends to infinity.

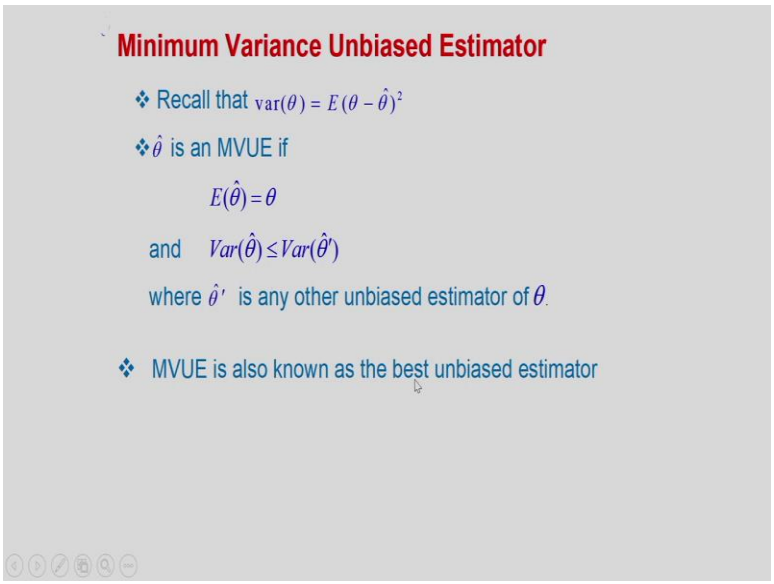
So, this is the definition of consistent estimator. Eventually as we have more and more data, theta as will grow more and more close to theta that is the interpretation of consistent estimator. If theta had this unbiased and limit of variance of theta hat = 0 then theta is consistent that is a simple test for consistent estimator when theta is unbiased. The mean square error MSE of an estimator is given by this MSE = E of theta – theta hat whole square and minimizing the MSE is an important estimation criteria.

(Refer Slide Time: 03:13)



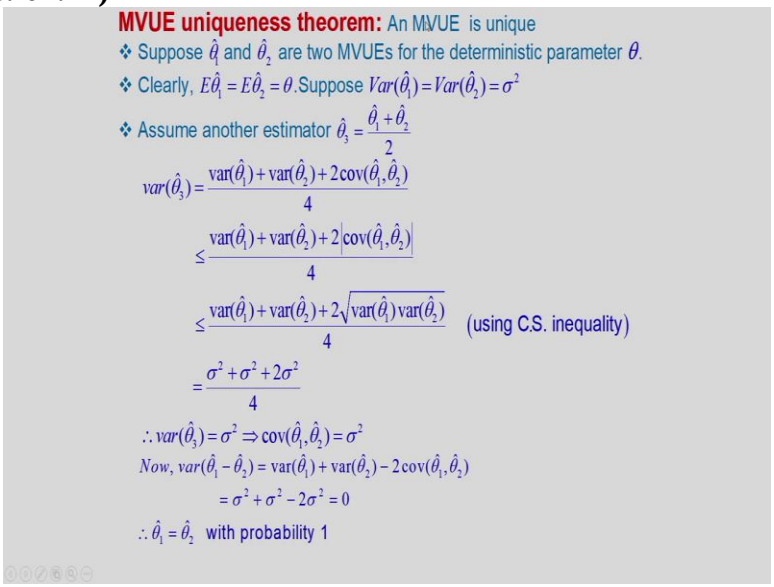
This lecture will revisit the MVUE property and discuss 1 of the important bounds on the variance of the unbiased estimators the Cramer Rao lower bond it is abbreviated as CRLB.

(Refer Slide Time: 03:33)



Minimum Variance unbiased estimator that is MVUE recall that variance of theta as it define it as E of theta - theta hat whole squared. We can write it as theta hat - theta, but it does not matter. I mean both where we can write theta hat is an MVUE if E of theta hat = theta and variance of theta hat is less than equal to variants of theta hat depth where theta hat depth is any other unbiased estimator of theta MVUE also known as the best unbiased estimator because its variance is lowest. Therefore it is also known as the best unbiased estimator.

(Refer Slide Time: 04:24)



We will prove 1 important theorem and MVUE is unique. Suppose theta 1 hat and theta 2 hat are 2 MVUE for the deterministic parameter theta clearly E of theta 1 hat = E of theta 2 hat = theta because of unbiased. Suppose variance of theta 1 hat = variance of theta 2 hat that = sigma

squared because both have the same minimum variance. So Let that minimum variance the sigma squared assume another estimator theta 3 hat which is given by theta 1 hat + theta 2 hat divided by 2, then variance of theta 3 hat will be = variance of theta 1 hat + variance of theta 2 hat + 2 times covariance of theta 1 hat theta 2 hat whole term divided by 4.

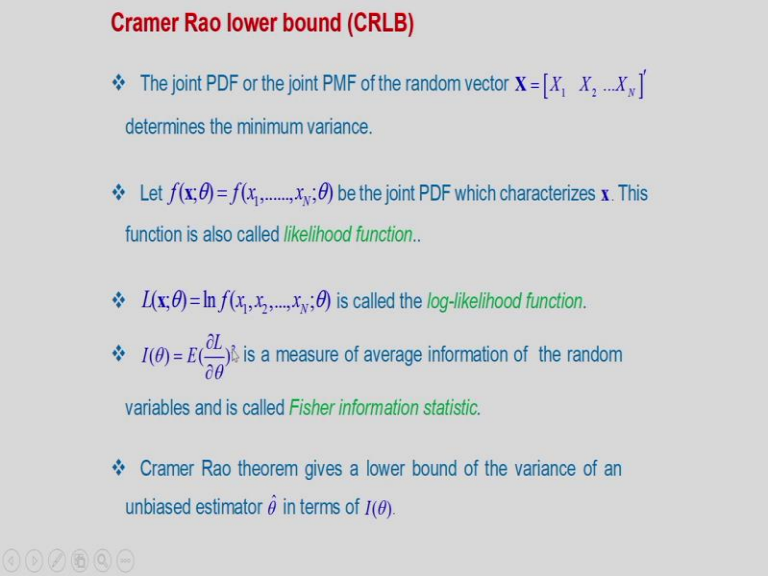
Now, this expression is less than equal to variance of theta 1 hat + variance of theta 2 hat + 2 times mod of covariance of theta 1 hat theta 2 hat divided by 4 because in a number is less than equal to its mod value. Now we will apply the Cauchy-Schwarz inequality, then our expression this expression will be less than equal to variance of theta 1 hat variance of theta 2 hat + 2 times square root of variance of theta 1 hat into variance of theta 2 hat.

Because the magnitude of debt covariance of theta 1 hat theta 2 hat is less than equal to square root of variance of theta 1 hat into variance of theta 2 hat this expression is now = sigma squared + sigma squared + this is sigma squared therefore + 2 sigma squared divided by 4, which = sigma squared. Therefore, what we have established at variance of theta 3 hat is less than equal to sigma squared, but it cannot be less than sigma squared.

Because minimum variance is sigma squared, therefore variance of theta 3 hat is also = sigma squared. And if I substitute this value in this expression, then we will get that covariance of theta 1 hat theta 2 hat is also = sigma squared. So, if there is another estimated theta 3 hat, which = theta 1 hat + theta 2 hat divided by 2, then its variance is sigma squared. And this source that covariance of theta 1 hat theta 2 hat must be = sigma squared.

And now, let us examine these expression variance of theta 1 hat – theta 2 hat this will be = variance of theta 1 hat + variance of theta 2 hat - 2 times covariance of theta 1 hat theta 2 hat. Now, subsidiary the value will get this is sigma squared, this is sigma squared and this is sigma squared. Therefore this expression will be = 0 that is variance of theta 1 hat - theta 2 hat = 0. This implies that theta 1 hat must be = theta 2 hat with probability 1. Therefore an MVUE is unique, minimum variance unbiased estimator is always unique. The uniqueness of MVUE makes it the most desirable estimator.

(Refer Slide Time: 08:15)



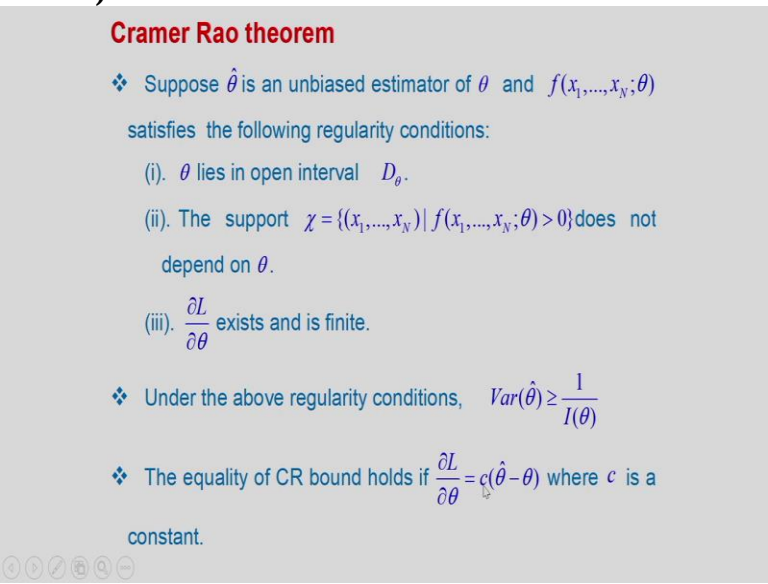
Now we will discuss Cramer Rao lower bound CRLB the joint PDF or the joint PMF of the random vector X that is X1 X2 up to XN transpose determines the minimum variance the joint PDF will only determine the minimum variance of that hat is a function of this random this random variables. Let f x theta that = small f PDF X1 X2 up to XN; theta that is a it is a function of theta be the joint PDF which characterizes the random data vector x.

This function is also called likelihood function. So, this is the likelihood function, which is a function of theta and we denoted by this is the likelihood function. Now, if we take the logarithm capital L x theta that is the log of this likelihood function is called a log likelihood function. Now, we define a term known as Fisher information statistics I theta is the expected value of del L del theta whole square.

So, we take the partial derivative of L with respect to theta and then take square then they variance value of that quantity is a measure of average information and it is called Fisher information statistic like any information, whether it is also positive and we can solve that it is an increasing function Cramer Rao theorem gives a lower bound of the variance of an unbiased estimator theta hat in terms of I theta.

So, the lower and the lower bound in the variance will be given in terms of I theta. So, we have discussed what is I theta that is the information Fisher information statistics that is E of del L del theta squared.

(Refer Slide Time: 10:38)



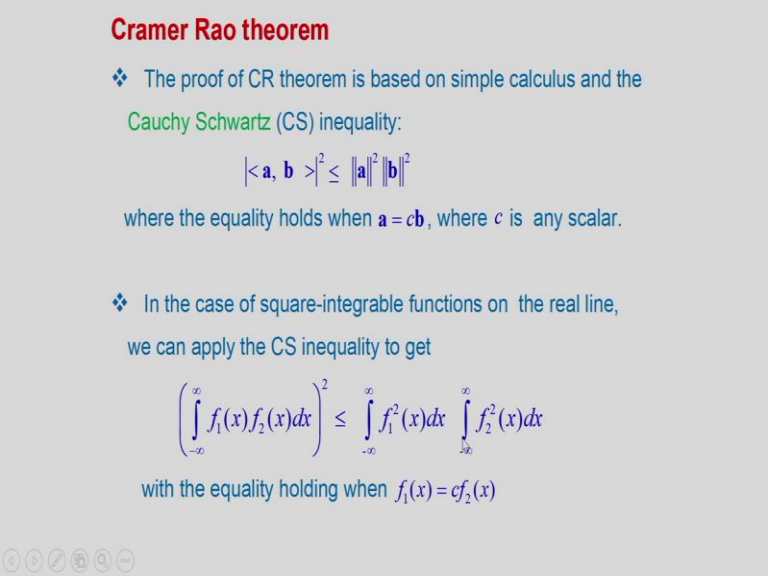
Suppose, theta hat is an unbiased estimator of theta and f x1 x2 up to xN theta satisfies the following regularity condition likelihood function satisfies the following regularity conditions. That is theta lies in open interval D theta there is an interval D theta on which theta is defined. The support chi, this is defined by x1, x2 up to xN such that this likelihood function is greater than 0 wherever likelihood function is 0, that support does not depend on theta.

So essentially we will be using a integration that limited integration does not depend on theta del L del theta partial derivative of L with respect to theta exist and is finite so these 3 are the regularity conditions under the above regularity conditions, variants of theta hat is greater than equal to 1 by I theta where I theta is the Fisher information statistic the equality of CR bound, this is the Cramer Rao bound the equality of CR bound holds if del L del theta = scalar multiple of theta hat - theta where c is a constant scalar it may be a function of theta, but it does not involve theta hat.

So, these regularity conditions are purely general it is not very restrictive and under this regularity conditions Cramer Rao theorem is variance of theta hat is greater than equal to 1 by I

theta the equality holds when the partial derivative of log likelihood function can be written in this policy into theta hat - theta.

(Refer Slide Time: 12:51)

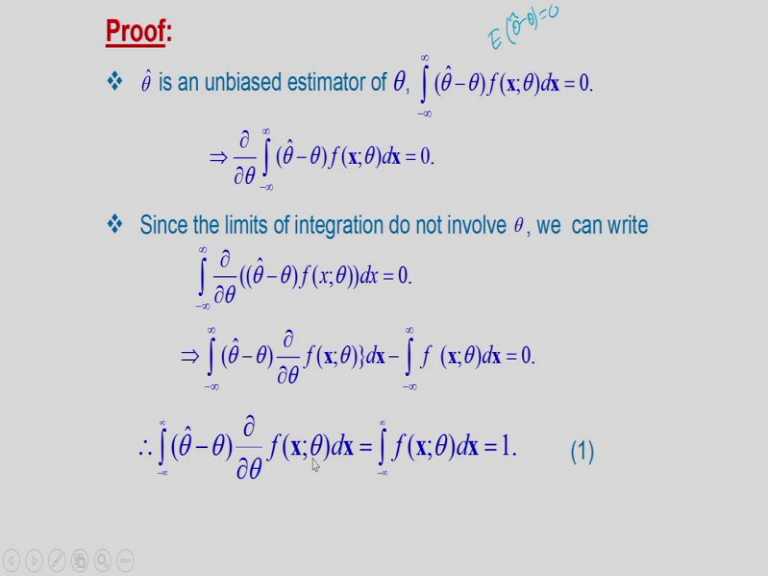


We will try to prove this theorem. The proof of CR theorem is based on simple calculus and the Cauchy Schwartz inequality that is a very important inequality, we have discussed about this earlier and this inequality is given by in norm product this is in norm product magnitude of the in norm product square is less than equal to norm of a squared into norm of b squared, where equality holds when a = c times b where c is any scalar.

So, this is the statement of Cauchy Schwartz inequality in the case of square integrable functions on the real line, we can apply this CS inequality Cauchy Schwartz inequality to get this relationship that is integration from - infinity to infinity f 1 x into f 2 x dx whole squared. This was relation = integration - infinity to infinity f 1 squared x dx integration - infinity to infinity f 2 squared x dx. So, this is the Cauchy Schwartz inequality in terms of 2 functions.

So, here this in norm product is defined in terms of this integration and norm is defined by this integration. So, that way this is the result we get and this result will be using for deriving the processor inequality also in the case of 2 functions, the equality that is, this quantity will be equal to product of this only when 1 function is a scalar multiple of the other function.

(Refer Slide Time: 14:48)

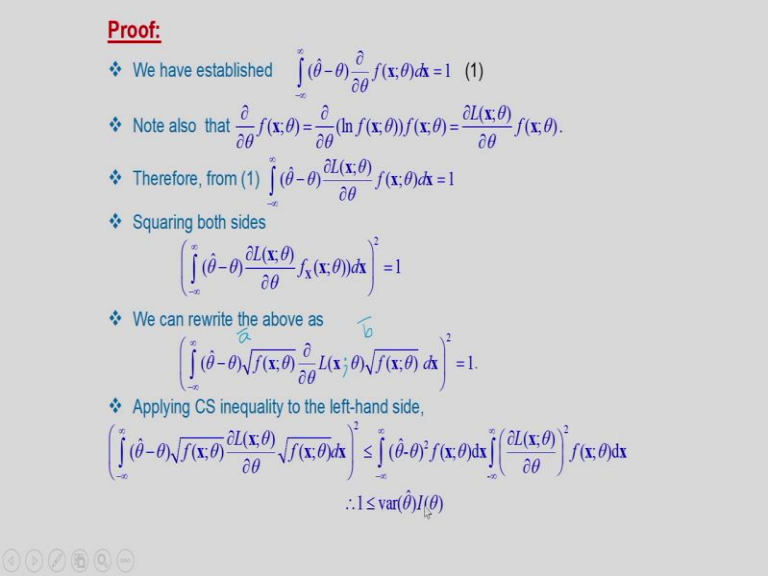


Now will prove this theorem theta hat is an unbiased estimator of theta. This expression actually this is E of theta hat - theta must be = 0 so that if we write in terms of the definition. So we will get integration - infinity to infinity theta hat - theta into f of x theta dx = 0, this dx is a vector so this integral actually this is a multiple integral. So we have supposed and variables x1, x2 up to xN then this is an N holds integration.

Now, since the limits of integration do not involve theta, we can take this partial derivative inside here it was outside the integration now we can take it inside del theta of theta hat - theta into f x theta dx, that must be = 0. This x is a factorial now since this is a product of 2 functions, we can apply the product rule and we will get integration - infinity to infinity theta hat - theta into partial derivative of f x theta with respect to theta into dx.

Now theta hat - theta will have a partial derivative - 1 that way - integration - infinity to infinity, f of x theta dx must be = 0. So this is thus applying differentiation we get this. So we will take this quantity right hand side. So that way it will become positive and then also I know that this integration = 1. Therefore, this quantity integration theta hat - theta into del del theta f of x theta dx that must be = 1 this result is very important for us.

(Refer Slide Time: 17:06)



Now, we have established that integration - infinity to infinity theta hat - theta del del theta f of x theta dx that = 1 this is supposed to call this expression as 1 note that this partial derivative del del theta f x theta that is equal to we can write in terms of log likelihood function like this log of del del theta of f of x theta into f x theta. So, this will be same as this quantity because 1 by f x theta will come and that and this will get cancelled.

So, that way this expression can be rewritten as this so that del theta of f x theta is same as del L x theta del theta into f x. So this is 1 important derivation. Therefore, this expression 1 can read as now integration - infinity to infinity theta hat - theta into del L x theta del theta into f x theta dx = 1. So, this is the modified form of this equation our aim is to apply the Cauchy Schwartz inequality to the left side of this expression.

So, that way we Cauchy Schwartz this expression if I squared then also this will be = 1. Now, we have to write it as a product of 2 functions. And since this quantity is greater than equal to 0. We can take this square root and that way we can write it in terms of 2 function 1 is theta hat - theta into root over f x theta and other 1 is that partial derivative of L x theta into square root of; f x theta dx, so that way we have 2 functions.

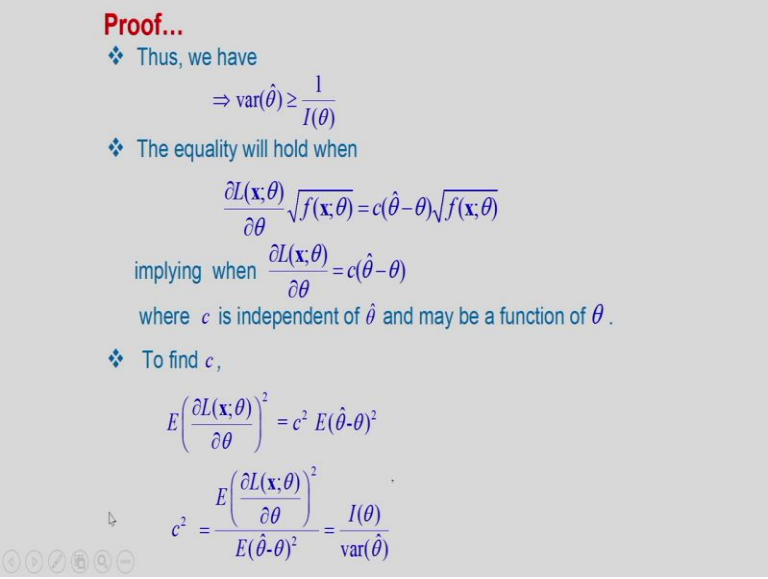
So this expression is rewritten as integration - infinity to infinity integration from - infinity to infinity of theta hat - theta in to root over f x theta into del del theta of L x theta into root over f x

theta dx whole squared = 1. So, we have a known to product up to functions suppose if I call this as a, this is supposed this = a vector and this is b vector, this part is b vector, this part is a vector and this part is b vector, then we can apply this CS inequality.

So, therefore, this is the inner product of 2 functions squared. So, that way this squared must be less than equal to if I consider the individual function norm of individual function, this will be integration - infinity to infinity theta hat - theta whole squared, this is theta hat – theta whole squared and this is square root of f x theta therefore, squaring will get f of x theta dx. So, this is part and first norm square and norm of the second function square will be integration from - infinity to infinity del L del theta whole square into f x theta dx.

So, by definition, this is the variance of theta hat and this part is the Fisher information statistics, so, variance of theta hat into I theta so this because this expression is = 1, therefore, 1 must be less than equal to variance of theta hat into I theta. So, from this now we will get that variance of theta hat is greater than = 1 by I theta.

(Refer Slide Time: 21:42)



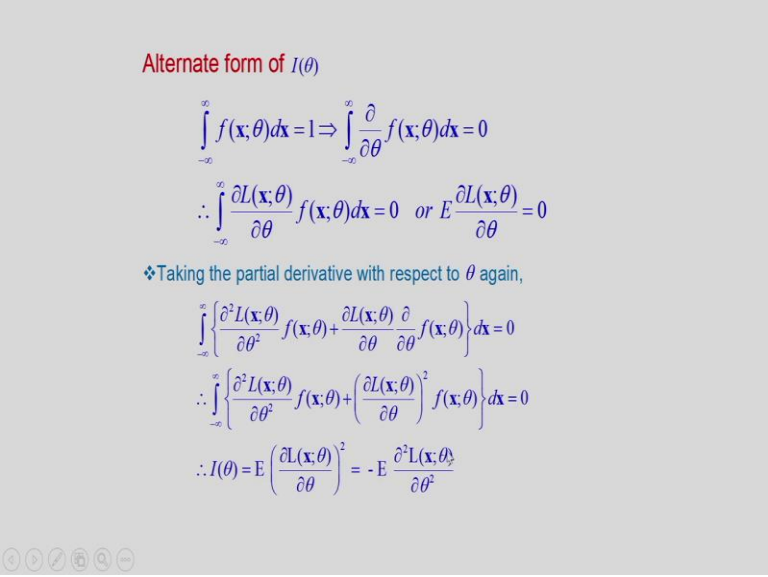
So does we have variance of theta hat is greater than = 1 by I theta the equality will hold when 1 of the function that is suppose pulse function del L x theta del theta into root over f x theta that must be multiple of the other functions 3 times theta hat - theta into root over f x theta that is the condition for equality indicates of Cauchy Schwartz inequality. So, since this quantity, actually

we want this to be greater than 0 in the region of support, that is the regularity condition therefore, we can cancel this to quantity.

So what we will get this equality will hold when del L del theta = c times theta hat - theta. This is the condition for equality where c is independent of theta hat it is a scalar actually. And may be a function of theta c maybe a function of theta but c does not involve x or c does not involve to the hat. Now to find c if we take this square of this and they take the expectation then E of del L del theta whole square will be = c square times E of theta hat - theta whole square.

So, that way c square will be equal to this divided by this E of del L del theta whole square divided by E of theta hat - theta whole square. This is the variance and this is the important statistics. So, that way I theta divided by variance of theta hat, so, c square = I theta divided by variance of theta. So, this scalar c is given by this.

(Refer Slide Time: 23:42)



Now, we know what is I theta but there is a better expression algebraically for I theta that we can derive now, that is integration of f dx = 1, then PDF likelihood function degraded to dx must be = 1. This implies that partial derivative of this quantity must be = 0. Again, under the regularity conditions, we can take this partial derivative inside therefore del del theta of f x theta dx. Integrating from - infinity to + infinity must be = 0.

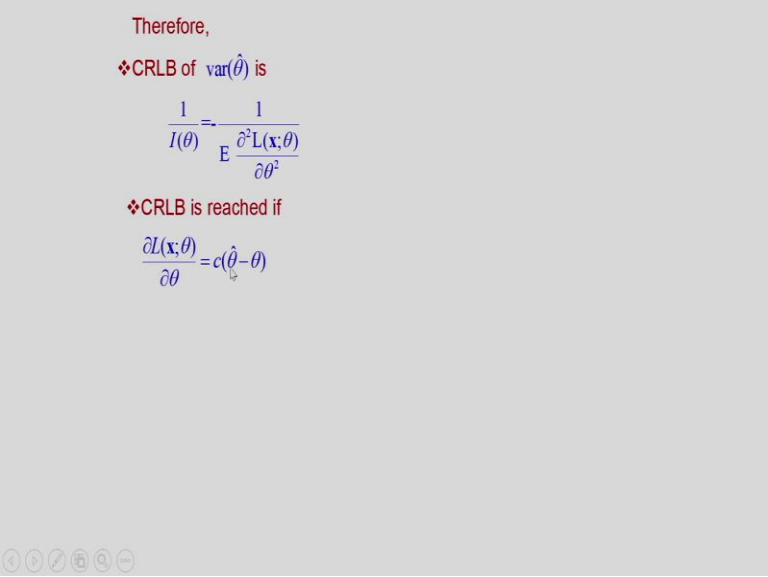
Now this we can write as this quantity, now we can write in terms of log likelihood function. So that way del so integration - infinity to + infinity, del L del theta into f dx that must be = 0. This is the expression we get from here, just violating this expression in terms of log likelihood function as we did earlier. So, now, what does this expression implies that expected value of the partial derivative of log likelihood function is always = 0 this is also 1 important result.

But we are interested to the second partial derivative therefore, will again take the partial derivative of this expression that way will get, because now it is a product of 2 functions. So, if we take the partial derivative with respect to theta again and of course them that this partial derivatives exist. So, second order partial derivative of L with respect to theta into f x theta that is a part term plus again this term is into partial derivative of this so a known this again we will be writing in terms of log likelihood function that way we will get this square here into f x theta dx.

So, from this expression now, if I consider this is nothing, but the expected value of this quantity and if I consider this is the expected value of this quantity. So, that way what I get is that because I know that expected value of this quantity this quantity is the information so, therefore, I theta = E of this quantity and that must be equal to if I take this quantity and said it will be negative, so that way - E of del square L del theta square.

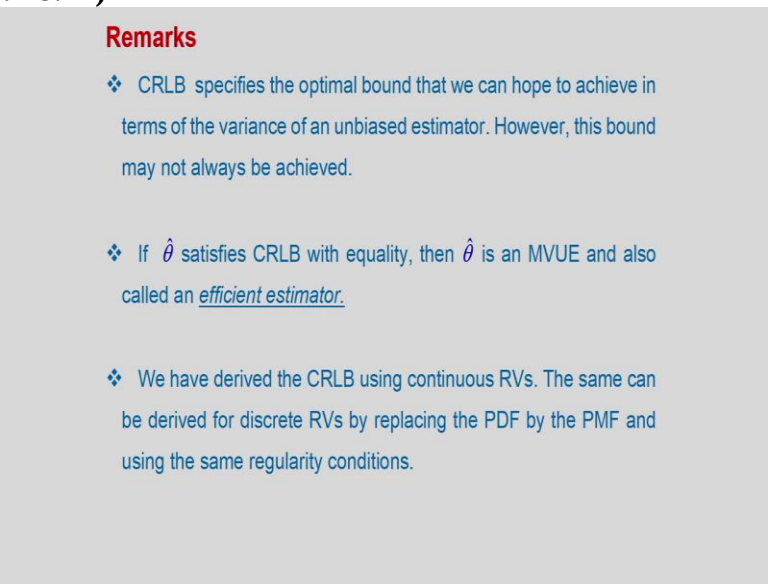
So, - E of del square L del theta square so that way this information statistic I theta can be expressed in terms of the second order partial derivative that is the whereas value of the second order partial derivative with a negative side and compared to taking the square of the partial derivative and then expected value this expected value is easier to compute.

(Refer Slide Time: 27:28)



So, therefore, you can write Cramer Rao lower bound for variants of theta = 1 by I theta that we have already established that = –1 by E of del square L del theta square. So, that way we can simplify this expression and also we have a establish is that CRLB is released if this part of that partial derivative del L del theta = c times theta hat - theta. So, this is the Cramer Rao lower bound given by this expression and this CRLB with this, if this likelihood function satisfy this relationship.

(Refer Slide Time: 28:11)

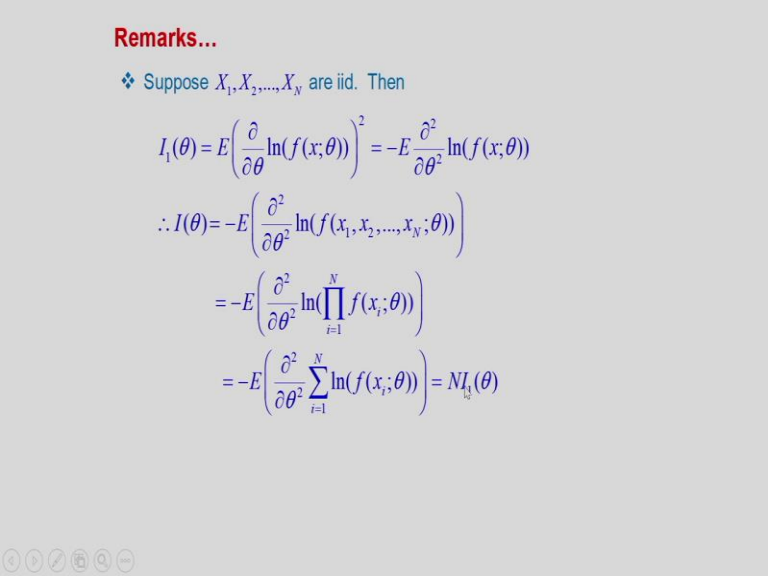


Let us make some important remarks CRLB specifies the optimal bound that we can hope to achieve in terms of the variance of an unbiased estimator, it is an optimal bound. However, this bond may not always be achieved that is a now we may not release this bound if theta hat

satisfies CRLB suppose we have an estimator with reset this CRLB then theta hat as is an MVUE because it is variance will be minimum we cannot have variance lower than that so, theta hat in that case is an MVUE and also called an efficient estimator. That we have defined earlier we have derived CRLB we using continuous random variables.

So, because we derived assume in joint PDF, but the same can be derived for discrete random variables by replacing the PDF by PMF. And using this same regularity conditions so that way, CRLB is more general than what we derive here. It is true for both continuous and discrete random variables.

(Refer Slide Time: 29:33)



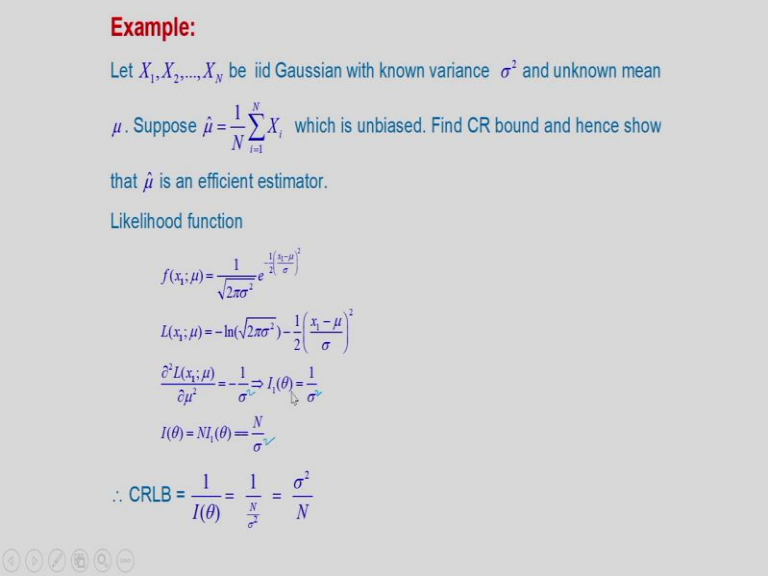
Further suppose X1 X2 up to XN are iid independent an identically distributed then suppose if I consider only 1 random variable I 1 theta = E of by definition, partial derivative of likelihood function whole square and that is also = - E of del 2 del theta square of log likelihood function. So that we have already established this therefore, if there is only 1 variable here I 1 theta I can write log this.

Now I theta is a is log of satisfies we have to take the del 2 del theta square of log likelihood function of N random variables. So, now this joint PDF we can write in terms of marginal PDF like this like product of N marginal PDF. So, that way this will be del 2 del theta square of log of product of N marginal PDF f x I theta. And this product now we can because it is logarithm of

the product we can write it as this term of logarithm so that way del square del theta Square of summation i going from 1 to n log likelihood functions ln f xi theta.

Now this we can take inside this. So that way and all are identical all random variables are identical therefore they will have the same in summation that is I 1 theta therefore, it will be and there are N such as variables therefore it will be N times I 1 of theta so that way if X sides are iid independent and identically distributed, then and the pisser information statistic part is N random variables equal to N times the information statistic per single random variable. So, that is the result for iid this otherwise, if only independent then we have to add all this all in permission we have to it that is the additivity property of information.

(Refer Slide Time: 32:08)



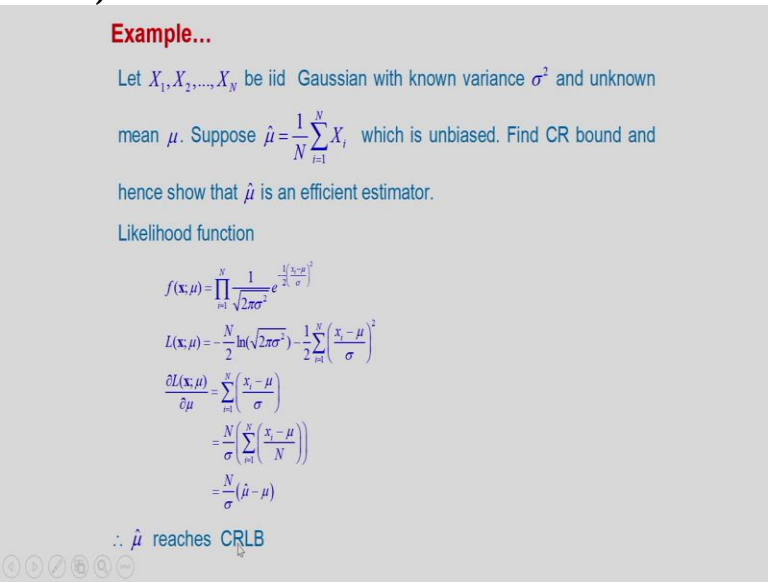
We will give 1 example suppose X1 X2 XN are iid Gaussian with known variance sigma squared and unknown mean mu. Suppose mu hat = this summation Xi, i going from 1 to N divided by N this is the sample mean which is unbiased to have already proved. Find CR bound enhance so, that mu hat is an efficient estimator. So, f is the likelihood function log likelihood function for single variable it will be like this we are taking the logarithm therefore, it will be - half this because it here, 1 by is there and then – half of x1 - mu by sigma whole squared now we have to take this second partial derivative.

So, only this term will involve x mu therefore, we take the partial derivative and then again derivative will get this = - 1 by sigma. Therefore, second order partial derivative del 2 L del mu

square will be = - 1 by sigma square because this is this expression only involved mu if you take twice derivative then it will be so, mu will not be there, so, it will be simply 1 divided by sigma squared.

Therefore, I 1 theta if I consider only 1 random variable x1 then it information will be = 1 by sigma squared and since exercise are iid therefore, I theta will be N times I 1 theta that is equal N by sigma square therefore CRLB Cramer Rao lower bond = 1 by I theta that = 1 by N by sigma squared, that is sigma square by N. So, this is the Cramer Rao lower bound.

(Refer Slide Time: 34:14)

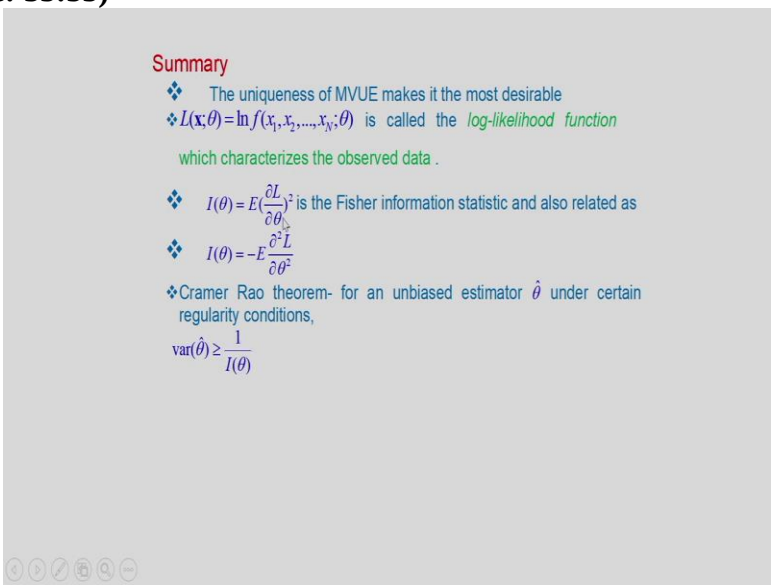


Same example, we are considering so, now, again we consider this likelihood function, this is given by this and if we take the log likelihood function is given by this and now, let us take the first derivative first partial derivative. So, this quantity will be equal to and that is del L del mu will be equal to summation because we will take the personal derivative of this 2 and 2 will get cancel so, that was summation x i - mu by sigma i going from 1 to N that is the partial derivative of the log likelihood function.

So, now here sigma is there this sigma I will take out and I will write N here and divided by N here. So, that this expression can be simplified as N by sigma into mu hat - mu. So, we have seen that del L del mu that is the first partial derivative of log likelihood function is a product of 1 constant term and by sigma into mu hat - mu. Therefore, according to the equality condition of Cramer Rao theorem mu hat will receive reaches CRLB. So, because this partial derivative is for

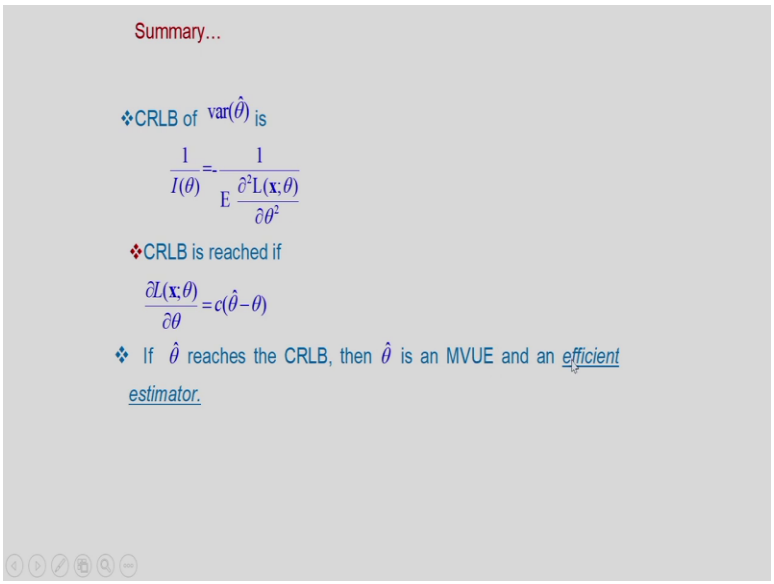
log of this specter mu hat - mu and then scalar therefore mu hat reaches CRLB and hence mu hat is an efficient estimator.

(Refer Slide Time: 35:55)



Let us summarize the lecture the uniqueness of MVUE makes it the most desirable estimator. So, that uniqueness, we establish L x theta that = log of the likelihood function is called the log likelihood function mischaracterizes the observed data I theta that is the Fisher information statistic that = E of del L del theta squared and also related as I theta also we expressed theta = – E of del square L del theta square Cramer Rao theorem that also we established for an unbiased estimator theta hat under certain regularity conditions, variance of theta hat is greater than equal to 1 by I theta is given by this Fisher Information statistics.

(Refer Slide Time: 37:00)



Thus CRLB Cramer Rao lower bound of variance of theta hat is 1 by I theta that is the bound and that = - 1 by expected value of del 2 L del theta square CRLB is released if this partial derivative of log likelihood function is a product of c and theta hat - theta where is a constant. If theta hat reaches the CRLB then theta hat is an MVUE so that is very important. So, our MVUE and if theta as reaches Cramer Rao lower bound then it is an MVUE and therefore, it will be an efficient estimator. Thank you.