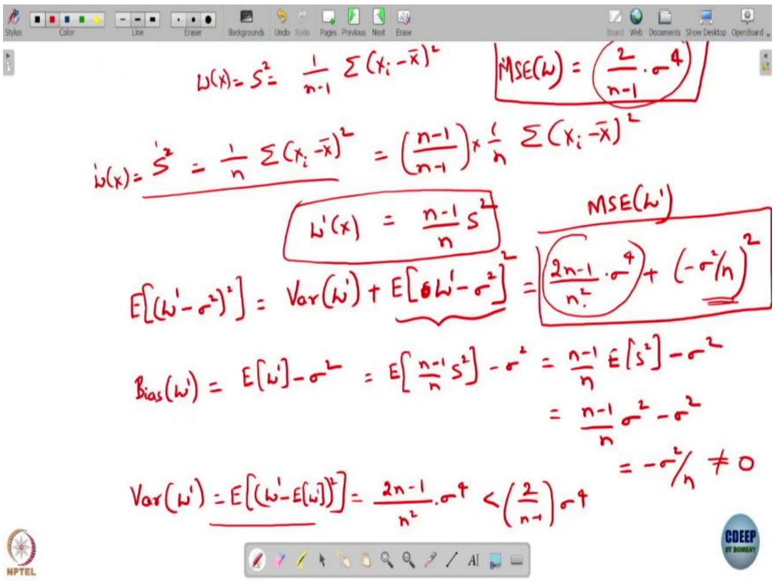
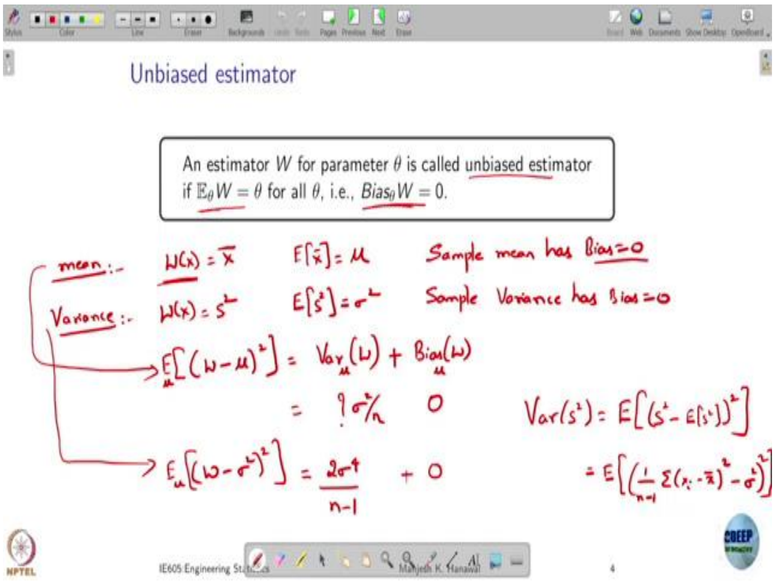
Engineering Statistics
Professor Manjesh Hanawal
Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay
Lecture 40
Evaluating Estimator, Cramer Rao Bound, Fisher Information (Contd.)

(Refer Slide Time: 0:15)





Now what is the variance of W prime? You can again compute variance of W prime, you have to expand it. Let us try to see if there is something simpler here. I do not think, maybe I will just write it from here. Let us see quickly, I do not have it here. The variance of this is expected value of W prime minus sigma square whole square, this is by definition.

Am I correct in this? What is wrong in this? Is this correct variance of W prime is this? What is wrong? Sigma, so instead of this I cannot write sigma square here, right? I have to write expected value of W prime here and this square is what is the variance of W prime and you can actually compute this, I am simply going to write it again. 2 n minus 1 divided by n square times sigma 4.

And now if you want to compare it with our previous thing, this was 2 sigma 4 n minus 1 and now what we have is 2 n minus 1 by n square sigma 4. You can simplify this as 2 n minus 1 divided by sigma 4. So, only thing you need to verify is 2 n minus 1 divided by n square, I can upper bound it by 2 to the power n minus 1. So, now what?

I have this quantity here, now 2 n minus 1 by n square into sigma square into sigma 4 and plus minus sigma square by n. So, by the way we have made one more mistake, right? We wanted to take square of this that is by definition. So, we should be taking square of this. So, now let us compare it with what we had for expected value of W. Expected value of W are simply 2 to the power n minus 1 sigma 4.

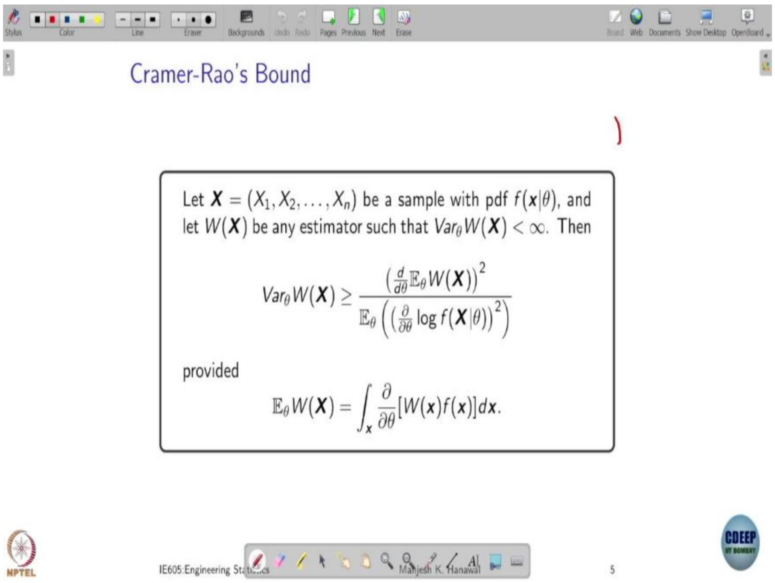
Now let us compare one expression we have for this and this is for W prime. We just argued that this quantity here is going to be smaller than this quantity. So, the first term is going to be smaller than this quantity but there is an another quantity that has come here because of this biasness. Can I say that the mean squared error of W prime is going to be less than mean squared error of W? By the way this is not, this is basically this is like variance, sorry this is like MSE of your W and this guy what I have here is MSE of your W prime.

Now can we say anything like whether MSE of W prime is going to be smaller than MSE of W? We know that the first term is going to be smaller but there is also second term here which is adding positively to it. So, it is unclear, right? Whether W prime is going to have a better mean squared error or W is going to have a mean squared error.

So, it is not necessary that if you have an unbiased estimator it is going to also give you smaller mean squared error. It may happen that, here W prime which is biased here which may end up to be smaller than it may have a smaller mean squared error than your unbiased estimator.

So, anybody has any doubt in this example that we did? Is this you are able to comprehend what I have written here? So, is the motivation clear? I may have an unbiased estimator, this bias is going to be 0 fine but when it comes to its mean squared error, it may not be the lowest one. I may have another bias estimator which may give me smaller mean squared error. And I want to estimate my, I want to evaluate my estimator with respect to the mean squared error part.

(Refer Slide Time: 6:39)

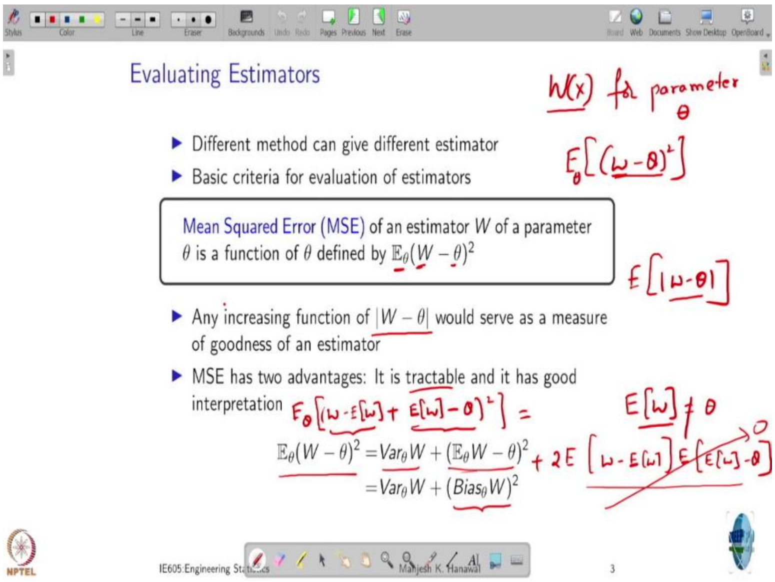


So, then the question comes what is the smallest mean square error I can get. And that is where the Cramer's Rao Bound comes into picture which kinds of provides me some bound on the variance of my estimators. So, let me give you two, let us say I give you W and W prime, I said that W is unbiased but W prime has a smaller mean squared error which one you choose? You like unbiased, I mean W you want. Anybody here who feels W prime should be better?

W prime should be better why not W? That depends, right? So, what is W being unbiased like on an average it is giving the right value. Unbiased is not saying that but what unbiased is saying on an average I will be closer to the point.

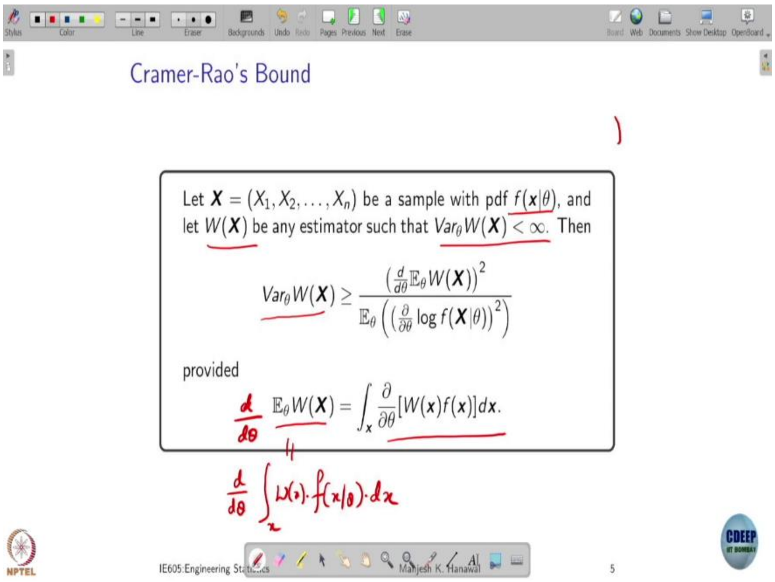
Now let me ask you another question. Let us say W and W prime both are unbiased, which one you are going to choose? Or let us say W and W prime both unbiased and W prime has a smaller mean squared error, which one you are going to choose? W Prime, if that is the criteria, once I said both of them are unbiased, so their bias term is going to be 0.

(Refer Slide Time: 8:34)



And in that case what matters to be is only the variance of the estimator because the second term is anyway 0, then I should be worried about only which of these estimator is going to be the smaller.

(Refer Slide Time: 8:49)



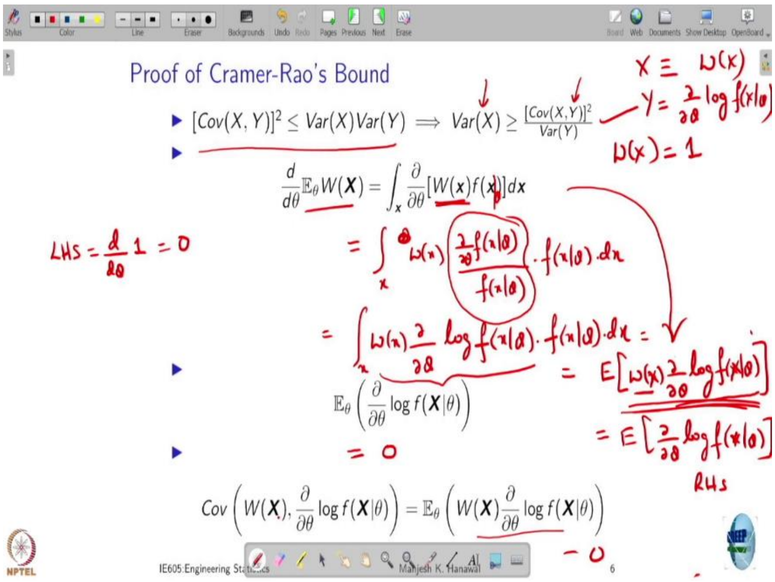
Now that is where the Cramer's bound is coming and it is saying that, what is the smallest variance one we can expect. So, Cramer Rao Bound says that suppose if we have a samples coming from a distribution PDF parameterized by theta and W(X) is any estimator, right now we are not saying is there a biased or unbiased.

Then it is saying that and we are assuming that its variance is going to be finite. If that is the case, the variance of that estimator has to be at least this much. It is not that I am going to achieve an arbitrarily small variance. If it is an unbiased, I mean whatever the estimator here there is some lower bond which is going to be governed by this. And this holds, I mean this bound comes under certain condition, one condition is the expected value of W, I should be, there is a typo here this should be d by d theta here.

When you look into the derivative of the expected value of W of X as a function of theta, this should be like this. So, what is this basically condition telling, can anybody look, can anybody see what this condition is asking for. Let us write the left hand side here, what is this left hand side? If I have to expand this is d by d theta of integration of W of x, f of x given theta d of x over x, right? The left hand that is the expected value.

Now what is the difference between right hand side and the left hand side? The derivative, I have taken inside. So, basically saying I will be able to interchange my integration and derivative. So, is it possible that you can interchange this integration and differentiation? Not always but that is why this condition. Whenever this is possible, whenever your PDF function satisfies that then we are saying this.

(Refer Slide Time: 11:20)



Let us have a quick discussion about how this Cramer Bound comes out. Actually, this Cramer Rao Bound is an intelligent application of bound on your covariance function or bound on exploiting the other property of your correlation function or correlation coefficient. So, by the way what is the correlation coefficient? Correlation coefficient of two random variables x and y, what it is? Covariance between x and y divided by the standard deviation of x and y are square root of the variance of x and variance of y.

And we know that this ratio has to be always minus 1 and n. So, if I have to take the square, I can get this condition. And now I am going to, this is going to give me one way to get a lower bound on the variance term. So, now by manipulating this variance of x's covariance of x y whole square divided by variance of y. Everybody agree with this, now what Cramer Rao Bound's is doing is appropriately define your x and y random variables to get this bound. So, let me see if we can quickly argue this. We agree with this.

Now let us start with this function, d by d theta of this quantity. This is our assumption, so what we will now do is x d by anyway this W of x is a function of x but x I think I should have written x given theta here, f of x given theta is a function of theta. So, what I will do is W of x d of d of x, f of x given theta. And then what I will also do is, to this I am going to add and multiply x given theta. Everybody agree but I have done nothing. I have just doing a algebraic manipulation.

Now this quantity here, this whole quantity here can I write it as log of this quantity W of x, log of, sorry d by d theta of log of f of x given theta, f of x given theta still remains. So, now I am going to look into this quantity, now this is nothing but expectation of W of x, Wx log sorry d by d theta of log y, I do not have space here.

So, this quantity is nothing but here expected value of W of x, sorry capital X, d del del theta, log of f of x given theta and this is, see notice that X and this X both are capital because now they are random variables. So, this is one random variable here and this entire thing I have taken it as another random variable.

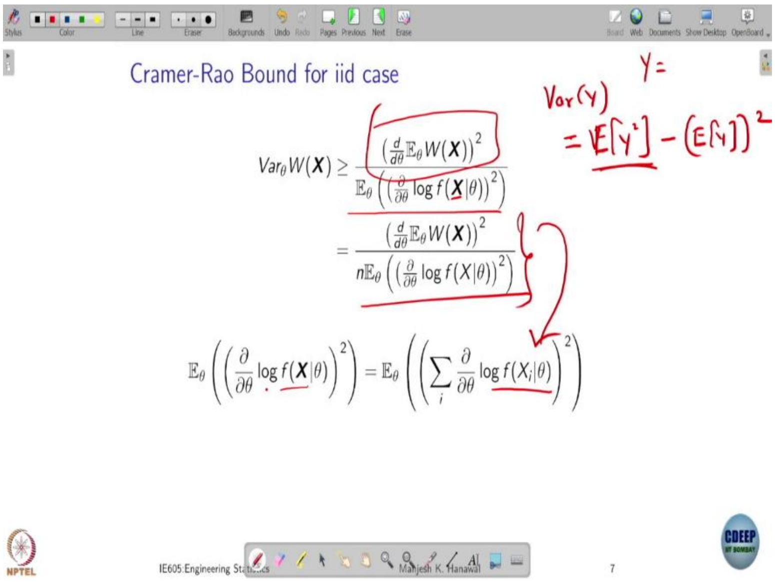
Now everybody agree with this relation? What we basically did this quantity d by del theta we are able to show this quantity. Now this should hold for any W(X), we did not make any assumption on what of W(X) is and what now we will assume is, I am going to assume a estimator which is constant all the time, W of x equals to 1.

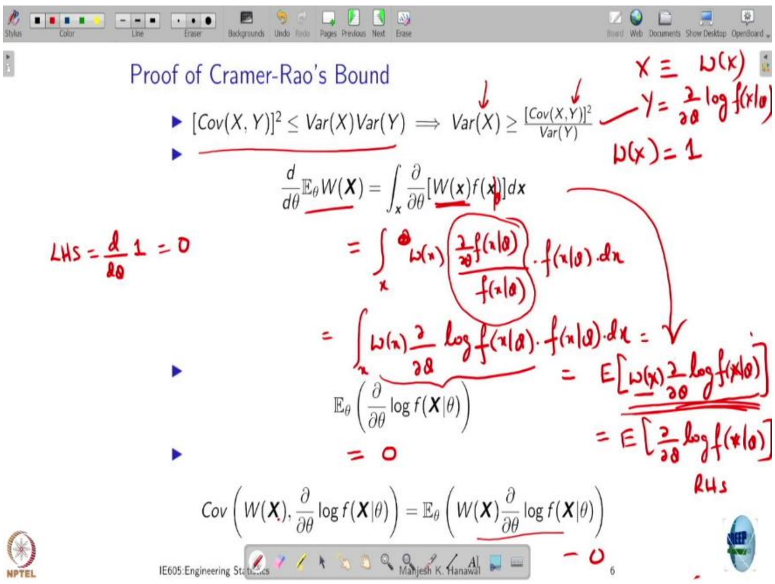
So, this relation should hold for even W of x equals to 1. Now let us see for this W of x, my LHS is going to be, what is this LHS is going to be expectation of 1. It is going to be 1 and what is d by d theta of 1? 0 because 1 does not depend on theta, so in d by d theta of this is 1 which is 0.

And what is on the right hand side? The right hand side is expectation of W X, I have set it to be 1 it is del upon del theta log of f of x given theta. This is my RHS, so what I showed this quantity here now is equals to 0, everybody agree? This is my right hand quantity and left hand quantity, they are equal irrespective what is your W this is the case.

Now this quantity here, I can interpret as this is a correlation between W of X and this other random variable. Can I interpret like that? And now the covariance between W of X and this quantity is nothing but expected value of their product minus expected value of W X and expected value of this quantity which is already shown to be 0. So, the other quantity is 0, so only that is why this quantity is going to be this much.

(Refer Slide Time: 17:52)





So, now that is it. We plug in all these quantities here, so what we did is, now let us go back here. Sorry, variance X, I have taken it to be X, I have calling W of X to be X and I am going to be calling it as d upon del theta log of f x given theta. This covariance of X comma Y, that is W(X) and this quantity I have just demonstrated it to be this quantity and what we want to say is, we actually said that this quantity, this quantity here is nothing but d by d theta expected of W of X.

We just argued that this quantity is nothing but d by d theta expected, so I am going to replace this quantity the numerator here by this quantity. Everybody agree with the numerator?

Now the denominator, denominator is what del upon del theta log of f of x given theta. But what I actually have in the denominator here is variance of Y, what is variance of Y? I know that variance of Y is expectation of Y square minus expectation of Y whole square. But now Y is we already said, right? Y is this quantity and we have just argued that its expected value is 0. So, what matters is only expectation of Y square and have just put it here.

So, that is how Cramer Rao Bounds work and it so happens that if my samples X are i.i.d. samples, then one can simplify this instead of taking the vector, we can take a task n times this quantity. That is a simplification.

So, with this we have this a lower bound. So, I have just skipped this step but this step here why n comes? It follows from the simplification, just go and work out that because this Xi's are i.i.d. I should be able to write it as a product of individual terms because of the log, I get the summation and then I have squared here when you take a square and expand, you will get expectation of these terms whose value is 0. And because of the summation adding or n you will get the n term here.