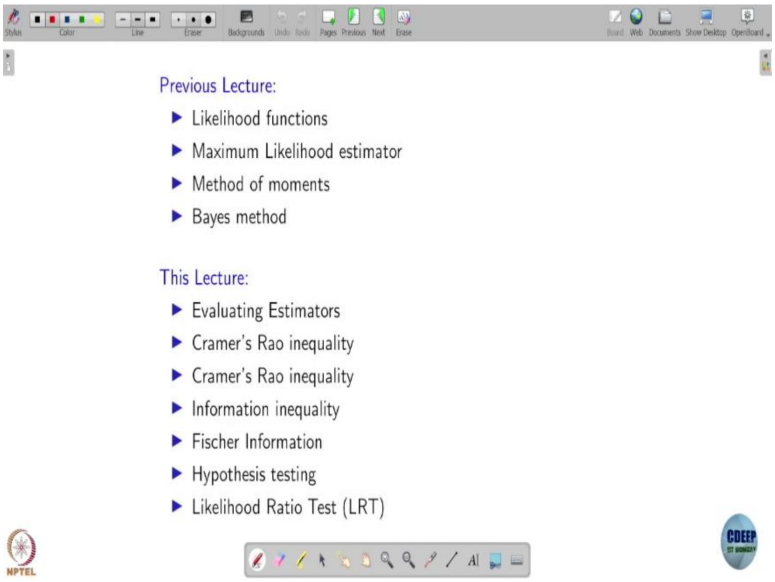
Engineering Statistics
Professor Manjesh Hanawal
Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay
Lecture 39
Evaluating Estimator, Cramer Rao Bound, Fisher Information

(Refer Slide Time: 0:29)



Okay then, let us get started. Any questions so far on what we have discussed? So far, we have discussed likelihood function, maximum likelihood estimator, method of moments, Baye’s methods discussed and there is one more thing called expectation maximization algorithm. I will see that if I have time I will come back to that. Any questions so far on this?

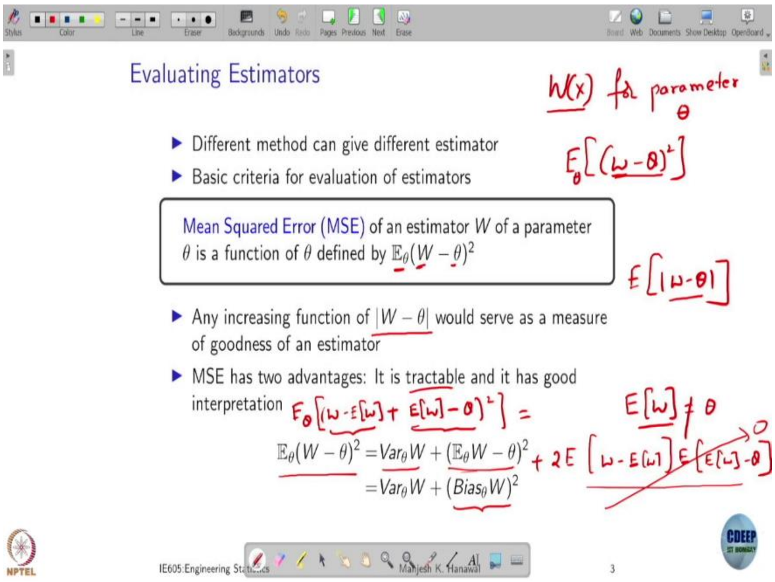
So, we basically looked into how to get the different estimators based on these different principles. One was based on sufficiency principle then based on likelihood principle, then other methods we looked into method of moments and Bayesian methods. And one thing is these are all like a point estimators. What is a point estimator?

You are given a data sample random sample, now you use some statistics on that or you do maximum likelihood estimator on that whatever, you are going to get a value, one value based on your data sample and which you are going to take as an estimate of that underlying parameter. So, from data you convert to one value which you are going to tell, this is an estimate of this underline. That is like a point estimator.

So, later we will see also something called interval estimators. So, there you will say that this is not point, I am not just giving you a point but I am given to give an interval and say that the parameter that you will be interested is going to line this interval.

Now we have seen so many estimators so far based on all this method. Then the question naturally comes to all of us is, which among these estimators we get is a good one. Then there has to be some criteria to evaluate the estimators we obtain. So, we are going to start talking about how to evaluate my estimators. Then we will look into one of the fundamental principles or property of these estimators which is going to be captured through this Cramer's Rao inequality. Then we talk about some terms related to that called as information inequality and Fischer inequality.

(Refer Slide Time: 3:12)



So, what should be the criteria for evaluating your estimators? As we said different methods may give different estimators what is the criteria? One possible criteria is take just the mean squared error. What is the mean squared error is doing? Suppose, let us say W is your estimator and of course this depends on your data samples x, which I have suppressed in this definition. And this is an estimator for parameter theta let us say.

Now I am going to define its mean squared error as simply expected value of difference between this W and the theta. So, theta is a constant, W is a random variable here. It depends on your data, right? So, you are trying to find the expected value of this square difference and we are denoting it with theta here, that is just to denote that these are all with respect to the parameter theta. So, I am interested in knowing how well this W captures my theta I am just looking into this is like, W is a point estimator here.

I am looking into the difference between these two points and since W is in random quantity, I am going to look at the expected value of this difference, square difference. You may ask why you take the square and then take expectation, why do not we simply take this quantity absolute value of the difference? And why do not we take simply this instead of taking the square one. This is also have the same property, it is also penalizing both negative and positive values but this is like penalizing larger by taking the square.

So, from the computational point of view MSE is better and one is that is tractability and as we will see it has also one good interpretation. What is that interpretation? Let us write this quantity here, what we can do is notice that this W is not necessary that this W is going to be unbiased. I am not making any assumption here. This W is just I am starting with an estimator, it need not be an unbiased.

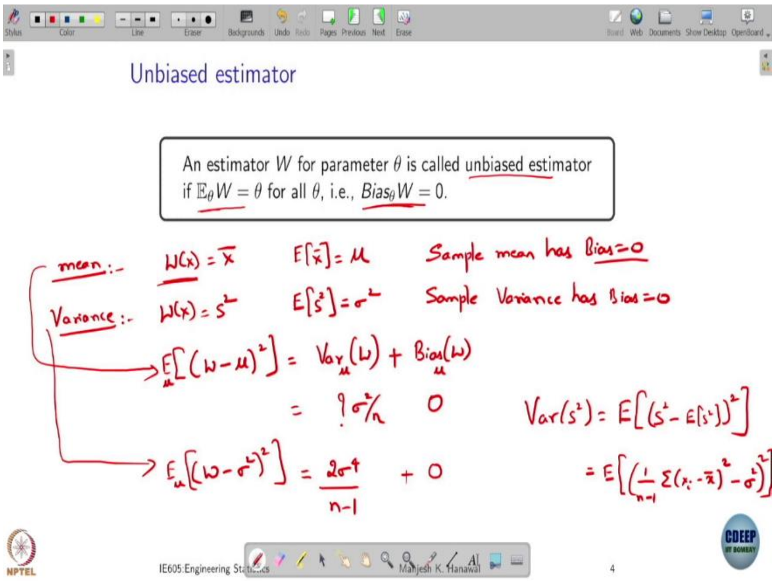
So, if it is an unbiased quantity, what is the expected value of W? If it is an unbiased value this is going to be theta. If not this need not be this. So, whatever is the mean value of the W, let us simply take it as expected value W. What I will do is then expected value, I will do W plus expected value of W minus expected value of W plus sorry minus theta, I will do this and this is same as this quantity and this is same as this quantitative because simply have added and subtracted this quantity.

Now, if you simply square it and expand it, you will see that one term you are going to get is W minus expected value of W squared which is nothing but variance of W and another term you are going to get is expected a value of W minus theta squared, that is the quantity I have written here and in addition there has been another quantity here which is 2 times expected value of W minus expectation of W into expectation of W minus theta.

If I had squared this. So, squaring I have basically grouped them maybe what I should have done is, first I will W and then minus. So, these two quantities and these two quantities I have grouped. The first, when I expand this the first term will give me variance when I expand the second term will give me this quantity and their cross product is this. But what is the value of this cross product, 0 because expectation of W is going to cancel as expectation of W this is going to be 0. So, we will get this.

Now this term here, expected value of W minus theta we are going to call it as biased of my estimator W. Is that fine? So, if W happens to be unbiased estimator, what is its bias is going to be? It is going to be 0. So, in this case if it is not an unbiased estimator, it could be positive but we are going to simply take that the difference of these two quantities we are going to simply define as bias of my estimator W.

(Refer Slide Time: 9:39)



So, I just said that if an estimator W for parameter theta. So, now we are basically defining. Earlier we have already said this definition but we are redefining it. An estimated W for parameter is called unbiased if expectation of W is going to be theta. This should happen for all theta and now that means basically alternately I can define it, its bias is 0. So, we are simply putting it like it is going to be unbiased, if it is bias is going to be 0. And we know what is the bias area.

Now let us look into some estimators we have. So, let us take W of x to be simply x bar. What is x bar? Sample mean we already know this and x bar, we use it used it as an estimator for my mean parameter. Now if I am going to compute this x bar, this is going to be mu. So, if I am going to use this estimator to estimate the mean parameter, what is the bias of your sample mean with respect to the mean parameter, it is going to be 0.

So, sample mean has bias 0. It is same as saying that it is unbiased. Now I am interested in variance term. So, this is for mean. Let us now look into variance. Now my estimator for variance I am going to simply take it as S square.

Now, what is the expected value of s square? Sigma square. So, is sample variance what is the bias of sample variance to estimate the parameter sigma square? That is going to be 0. So, sample variance has bias again 0. So, that is fine.

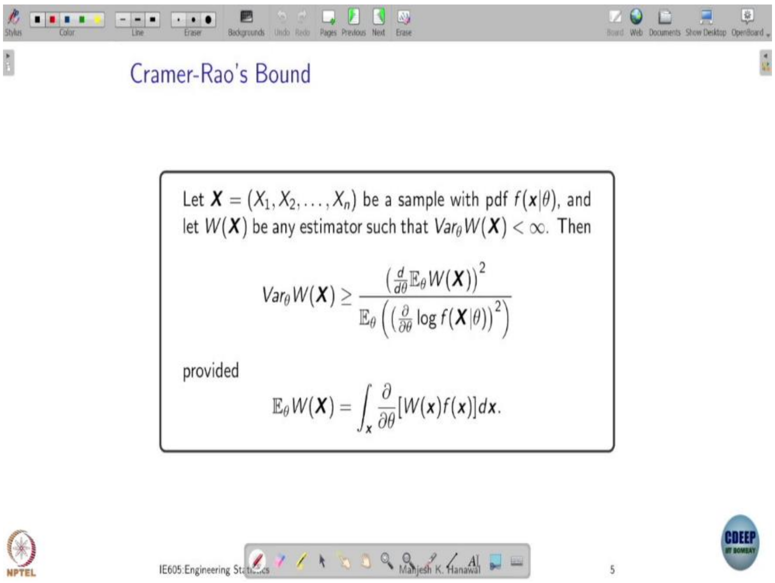
Now let us look into this quantity that does, let us look into for each one of them. The mean squared error that we are interested in. First case, I am again taking this W to be with respect to my mean parameter. I am going to take the mean as mean and we know that this is nothing but variance of W plus bias of W with respect to my parameter mu. These are all with respect to parameter mu I am computing now.

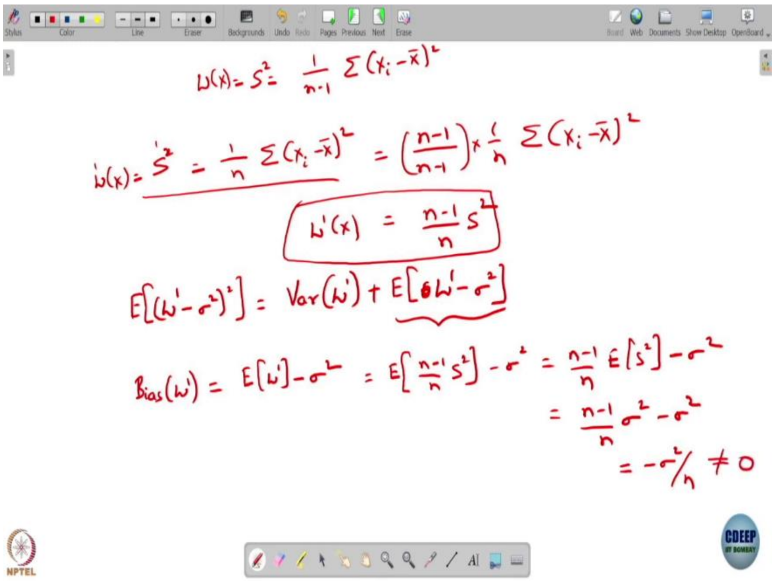
We just notice that if W happens to be a sample mean, this quantity is 0 what about this quantity? Sigma square by n. What is Sigma? These are the variance of the samples. Now let us do the same thing for variance. Now I am taking W to be S square, that is a sample variance. And now I am interested in this quantity. So, first of all what is its bias is going to be, 0. What is its variance though? So, what is going to be the variance of your variance estimator. Have you calculated it before, what is the variance of your sample variance?

I am going to directly write here, this is going to be 2 sigma 4 divided by n minus 1. This is, you just need to do computations. So, how we are going to do this? Variance of s square is nothing but expectation of s square minus expected value of s square whole square. By definition this and you know definition of s square, what is the definition of s square?

We will write it later this part but what is the expected value of s square? We know it is going to be sigma square. Then, all you need to do is 1 upon n minus 1 summation xi minus xi bar whole square minus sigma square whole square. So, you can compute this expression variance of your square and I am just writing it for you this is going to be 2 sigma square divided by n by 4.

(Refer Slide Time: 15:53)





This is for the case of unbiased estimator. But now let us look into the case where, let us look into the case. So, we always took s square to be 1 upon n minus 1 expectation of, we took it as summation xi minus xi bar whole square but we deliberately chose it to be n minus 1 though naturally we felt that it would have been s square, let us put it as s square prime to be 1 by n, summation xi minus xi bar whole square.

So, now let us say that but this could be also an estimator, who is stopping you to take this as an estimator? Let us say this is your one estimator for variance, this could be your another variance let us call this W prime. And I know that this W prime is not unbiased. So, now let us write it, n minus 1 n minus 1, I have just multiplied it and 1 by n and let us write it as xi minus xi bar square. So, this is nothing but if I do this and now if I want to compare it with this quantity, this is going to be n minus 1 by n s square.

My W bar of x is not a square but it is scaled by factor n minus 1 by n. Everybody agree? This is another estimator. Now let us compute its mean squared error. So, what is the mean squared error for W? By our definition this is W prime minus we are still considering it to be an estimator for sigma square. So, this is sigma square whole square, this I know as sorry variance of W prime plus bias. How is bias defined? Bias is s, sorry W prime minus theta, sorry sigma square.

So, now let us compute each of these terms. So, simpler is to compute expected value of W prime minus sigma square. Can you compute it? Now let us take this. So, bias of W prime is this quantity which I am going to write it as simply expected value of W prime minus sigma square.

Everybody agree? And now what is W prime? W prime I expressed in this format which is nothing but n minus 1 into s square minus sigma square but n minus 1 by n is a constant for me, I can pull out this from the expectation, n minus 1 expectation of s square minus sigma square but what is expectation of s square? I know that to be sigma square minus sigma square.

And if I simplify that this quantity what I am going to get? So, n and n cancels, minus by n. And now this is, why this is unbiased because this quantity need not be 0.