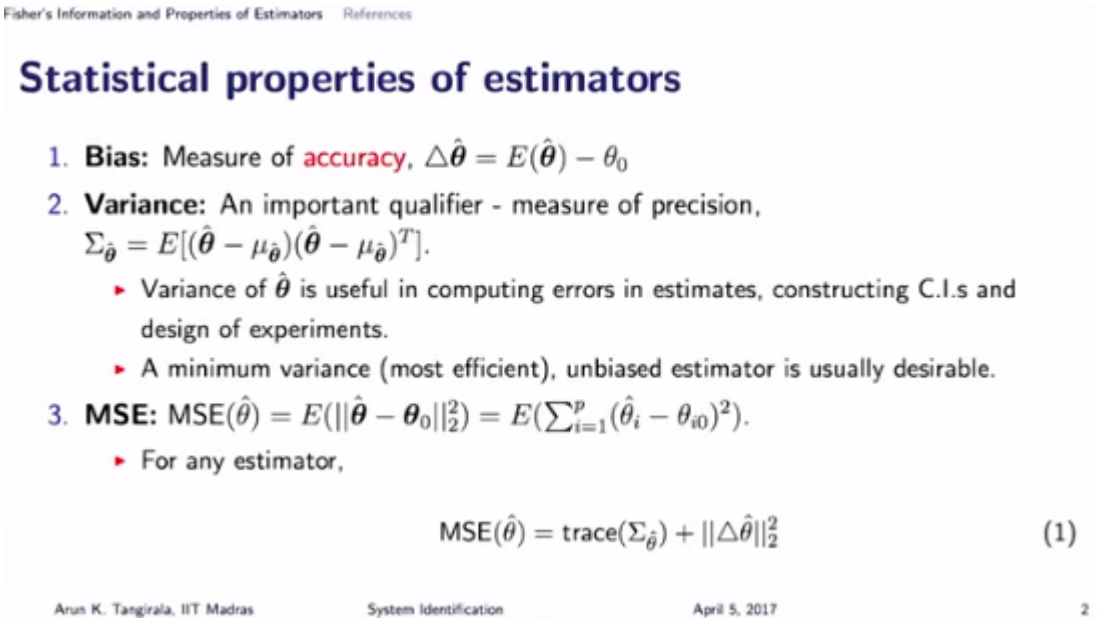
CH5230: System Identification

Fisher's information and properties of estimators

Part 05

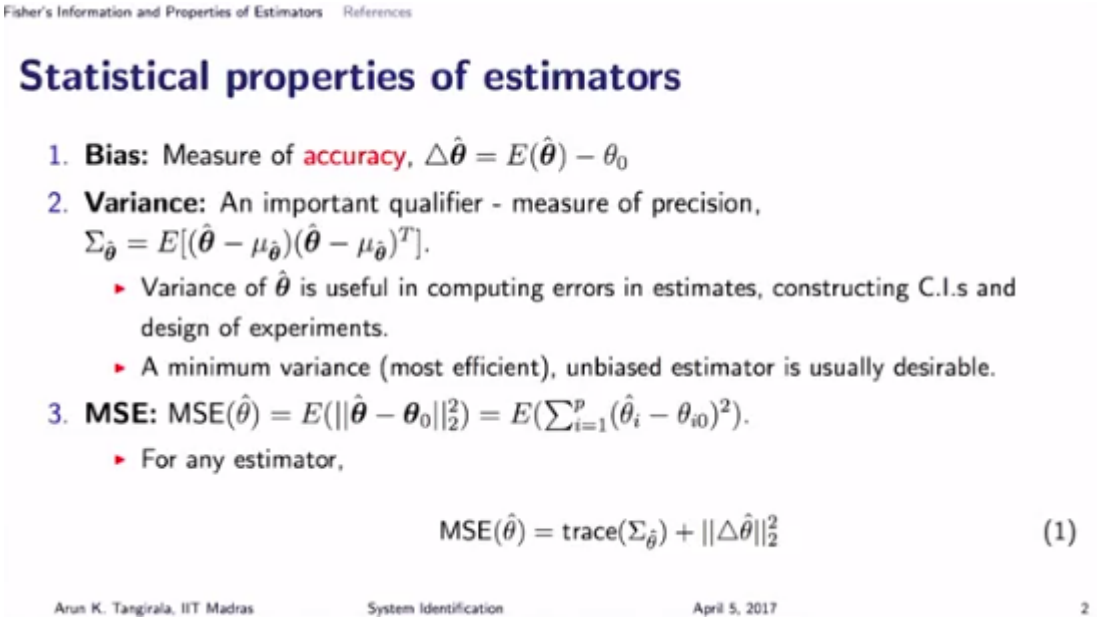
Very good morning. What we shall do today is, as I explained this today, I expect you to go over the video lectures detailing the goodness of estimators. What I'll do today is just summarize some of the important aspects of the properties of estimators. Then we'll go through a couple of examples that probably wouldn't find, at least one example that you wouldn't find in the video lectures. And then conclude our discussion on the properties of estimators. If time permits, we'll get started on least squares or methods of estimation impulse. So just to summarize as I had written on the board yesterday, when it comes to looking at estimators, goodness of estimators, you have two classes of properties, the statistical properties and the asymptomatic properties.

(Refer Slide Time: 1:01)



The statistical properties of estimators tells us how the estimator, how good the estimator is for a finite sample size, right? And it's looking at how the estimator behaves when you change the realization or across the realizations. And in this regard, we have three important properties namely Bias, Variance and Mean Square Error, right? Bias looks at the accuracy of the estimator and on the other hand Variance is concerned with the precision of the estimator. So you should distinguish between these two terms accuracy and precision. Accuracy has got to do with, both are telling you something about the estimates, how close they are to the truth, but in accuracy, you're looking at on the average whether you are hitting the truth, whereas precision, which is a more important property of the estimator, looks at the spread of your estimates, right?

(Refer Slide Time: 2:12)



And you can always recall the line that we drew yesterday, where the estimates are spread around the truth. So the spread is different from the accuracy. You can have a highly accurate estimate or you can say you can have an accurate estimator with that means on the average you can hit the truth, but you can have a very poor spread, right? So your estimates for example, can be all over the place with respect to the truth can be from, you know, minus infinity to plus infinity, just to give you a feel. And on the average, the estimates maybe at the truth that is expectation of theta hat can be equal to theta naught. But the spread may be too much, which is not desirable.

What is important to us is spread, of course, we want an accurate estimate, but what is easier to fix is the bias rather than the precision. Because if I know there is a bias, I can correct for it in my estimate. But if I know there is a problem with the precision, it's very difficult to correct for it. You have to search for another estimator. All right? So I always give this example where you have selection process where a person is being selected for shooting, and there is one candidate who actually shoots all over the place, okay? So that the organizers are also running away from the scene, but on the average he's on the target and there is another shooter who has actually not managed to hit the target on average but he's far away--his average is far away from the target, however his spread that means repeated trials has actually produced very small spread around the target.

So the question is which shooter is preferred? Of course both are having their own problems. So in one case, you have, so let's say in both cases this is the truth, theta naught, in one case you have estimates all over the place. So it could be here and then, you know, somewhere here, here, here, on average it's the truth, even, you know, outside this room. And then you have another situation, where the shooter is actually, the marks of just spread around a small value. So this is the average here expectation of theta hat. The shots fired by the second shooter are only confined within this small region. Now, this is the case of inaccurate, but relatively more precise estimate. We say this is relatively efficient, in fact, much more efficient than the first guy here. So, the question is which one is preferred? You will encounter this situation in estimation as well. You may have one estimator giving you with some biased estimate, but more precise estimate.

And on the other hand, you have this estimator which is accurate but poor precision. The preference would always go for the precision, right? Because precision is a measure of reliability. Given one chance, what is this estimator doing? Given one realization, is it producing an estimate that is highly unreliable or reliable? Because if I repeat the experiment, I may get a completely different estimate that is not, that means I cannot trust a single estimate too much. So in this case, I can easily adjust the bias by either moving the target, okay? Then that is match fixing. Or you actually maybe give corrective lenses to this candidate, so that the bias is corrected, it's easier to correct the bias.

Or you make the shooter look at a different angle and, you know, aim at the target. All of this is feasible but what is very difficult to fix is precision. Because that's an inherent property of the person. So an estimation you should remember, variability of the estimator is given far more importance than the bias. In the ideal estimator that you're looking for is minimum variance, unbiased estimator, MVUE as you would have learnt in the notes. So a minimum variance that is the most efficient unbiased estimator is always the desired estimator, but if I have to sacrifice among variability and bias, I would sacrifice bias. Then you have mean square error, which is not different from variance, the only difference between variance and mean square error is what? What is the difference?

[7:31 inaudible]

Sorry.

Mean square error is.

Measuring, what does it measure? Spread around the truth. True value, right? Whereas variance is looking at a spread around its own average. Now variance is a more practically calculable quantity, because it doesn't require the knowledge of the truth whereas mean square error requires knowledge of the truth. That's why initially when we talk of estimators, we talk of variance rather than MSC, but gradually when one moves to base in estimation, it is possible to look at mean square error. The ideal estimator is the one that minimizes the main square error, right? Because what I want is minimum spread around the truth. If I have an estimator that produces estimates, which are spread at a minimum, within a minimum range of the truth that is what I would like to have. But the question is, is it easy to design such an estimate? It is possible, in fact, Kalman Filter is a minimum mean square error estimator and so on.

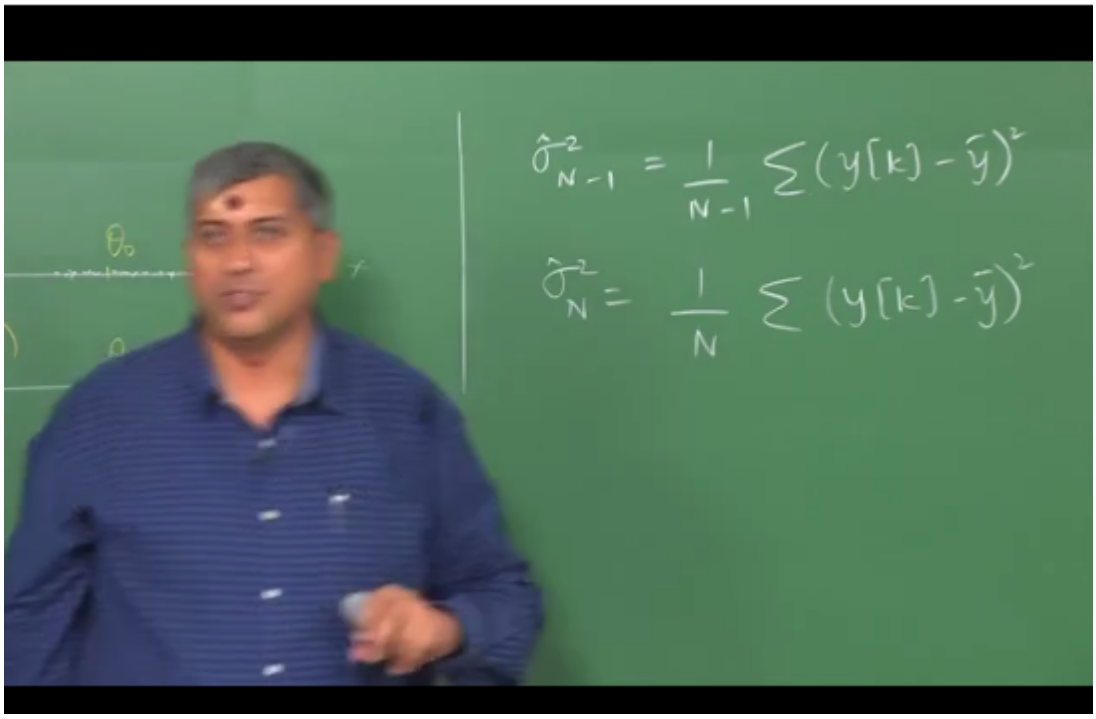
Now the good news is, if I have an unbiased estimator, mean square error and variances are the same. Right? Because expectation of theta hat is theta naught. So if I design a minimum variance unbiased estimator, what am I actually designing? A minimum mean square error estimate. So all minimum variance unbiased estimators are MMSEs. But are all MMSEs is minimum variance unbiased estimators? Can you say that? You can't. Why? Because mean square error is actually a sum of bias square plus variance square. If you're minimizing mean square error, you're actually minimizing bias square plus variance, which need not imply minimum variance unbiased, right? The sum of the squared bias plus variance is what you're minimizing. When you're minimizing mean square error.

That means by minimizing mean square error, I'm somehow trying to strike an optimal trade-off between bias and variance. This is always the classic they say Classical Dichotomy in estimation, where you trade-off bias for variance. Suppose, I want to get lower variability, then I may have to sacrifice on the bias. Ideally, I would like to get the lowest variance and zero bias. But what this minimum mean square error concept tells us is, I can achieve lower variability, then what a minimum variance estimator can give me, provided I'm willing to sacrifice on the bias. So, you have, for example, two standard estimators, right? As you must have also heard in the videos of variance. So,

you have one which is based on the estimator sample variance one which is based on 1 over n minus 1. Right. So you have y[k] minus y bar square and then you have the other estimator, which is this?

We know that this is an unbiased estimator of variance and this isn't, because if this is unbiased this cannot be. But what is the difference between these two apart from the bias? Yes, I have sacrificed on the bias when I worked with this estimator, but you can show that this has a lower variability, then the unbiased one. Okay?

(Refer Slide Time: 11:28)



So, of course, there are other reasons why people use this, but you should remember that between estimators there is always this choice that you have whether you want an unbiased estimator for sure, then you have no choice, you have to stick to that unbiased estimator. But if you have an option of working with a biased estimator then you will choose the one that gives you a minimum variance, lower variance, right. But then what happens is when you're willing to sacrifice the bias, when you're willing to sacrifice the bias, there is no end to how much you want to sacrifice the bias on. When it comes to unbiased estimators, it's okay. But when it comes to sacrificing the bias, you say, it doesn't matter to me, the estimator can have some bias, then what does it mean? I'm willing to tolerate any amount of bias. So by significantly sacrificing on a bias, I can achieve a lower variance. Right? That means there is no end. However, therefore, what we do is we say we want a minimum mean square error estimate. What does minimum mean square error estimate gave me? A healthy trade-off between bias and variance. At some point, see look at MSE, right? MSE, the expression for MSE is, the bias that is delta theta square, if I look at the single parameter, plus sigma square, theta hat. This is MSE. What I was saying earlier is, if you start, if you want to get lower variance then a minimum variance unbiased estimator, then you can start sacrificing on the bias. This is like this. Suppose, I want to fit a model, right. In system identification, we routinely fit a model. Suppose, I'm fitting a model to some data, input-output data, I want low variance of parameter estimates. That means low error. Then what would I do? I would actually choose the simplest model with a single parameter. Whatever data you give me, I will always fit a single parameters model, because I'm so worried about the error. But what

will happen? If I always choose to stick with a one parameter model? Whatever you give me, I'm going to fit only one parameter. I'm going to fit maybe a pure delay model with amplitude, that's it. What happens? What do you expect to happen?

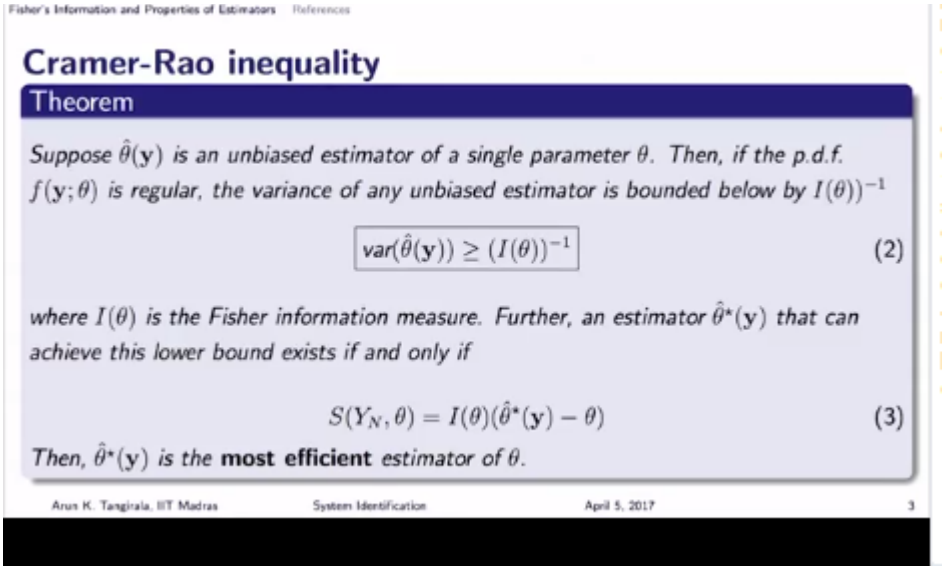
You'll not be able to get [14:26 inaudible].

It will not be able to fit the data. There is no, you know, there's no thinking here. It's very straightforward. Which means, they'll be huge bias in the prediction? That is what we mean by bias there. Here, we're talking of bias of theta hat, but there we are talking of bias of y hat. Whether it's theta hat or y hat, it's almost the same. Because if there is a bias in theta hat, you should expect to see a bias in y hat, so by sacrificing here on the bias in y hat, I have achieved a low error in theta hat. Because if I want to reduce the bias in y hat, what do I have to do? I have to include more parameters in the model. As I include more parameters in the model, what happens, the variance of theta hat starts to increase. The errors in theta hat starts to increase, because as you will, you will also see in least squares estimates, the expression for sigma square theta hat will have the number of parameters in the denominator, which means as I increase the number of, it will be proportional to N minus P, n is a number of observations, P is a number of parameters. As P increases, sigma square also starts to increase. All right.

But what do I achieve? An improvement in y hat. At some point, I would have achieved a very good trade-off that means, now my model explains the data very well. Yes, the parameter estimates have higher error than a simpler model, but my focus is on my hat, right. Beyond that point if I start to include more and more parameters in a bit to reduce the bias, then what happens? The variance starts to shoot up. So if you recall, I talked about this Akaike Information criteria, and BIC and so on. They're all looking at this bias variance trade-off. In some sense, they're looking at minimum mean square error, but not explicitly. So you will always find this kind of, I mean, ideally, you will find this trade-off.

If you were to plot MSE, for example, you will see this. Where p is the number of parameters that you're fitting. So, minimum main square error estimator tries to achieve a healthy trade-off between bias and variance. It doesn't necessarily give you a minimum variance unbiased estimator. If there exists, it may give you. So, that's the difference between these two.

(Refer Slide Time: 17:08)



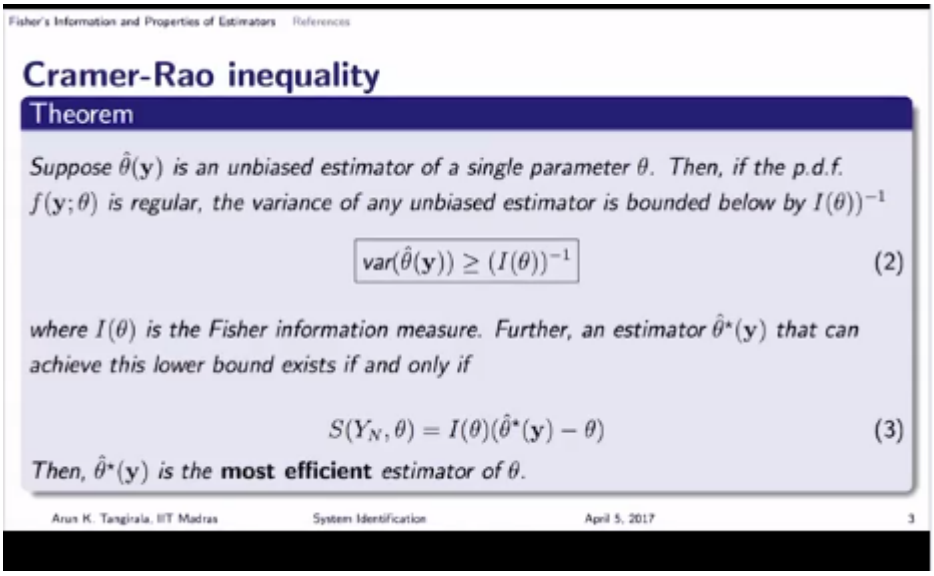
Now, the Cramer-Rao's inequality tells me what is the lowest achievable minimum variance? Oh, sorry, lowest achievable variance among all unbiased estimators. I cannot ask the question what is the minimum variance that you will get among all estimators? Why? Why should I have the unbiased tag here? Why can't I ask the question, what is the minimum variance that I can expect among all estimators? We have discussed so much now until, at least for last 10 minutes we've been discussing something. You should be able to get something from the discussion and answer that question. Why can't they asked what is the minimum variance I can expect among all estimators? Why do I need to restrict to unbiased only?

If that is when variance should be, like when you switch to unbiased, the variance will be, it will be less.

Okay. What happens among, to biased system?

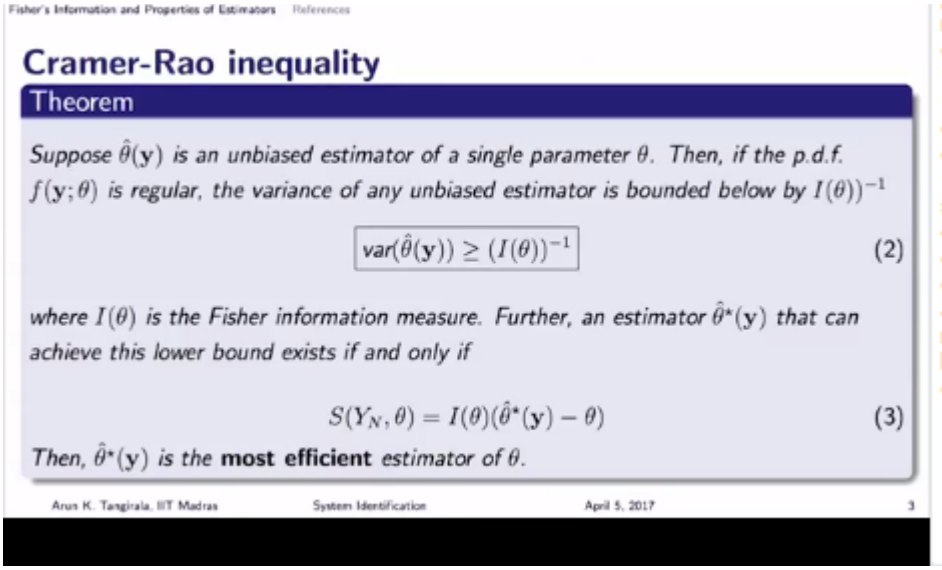
Biased is the case that we have to find out what is the minimum variance because that is the problem of variance, that means when it is unbiased, so the concern is variance so that is why we're checking for [18:45 inaudible].

(Refer Slide Time: 18:46)



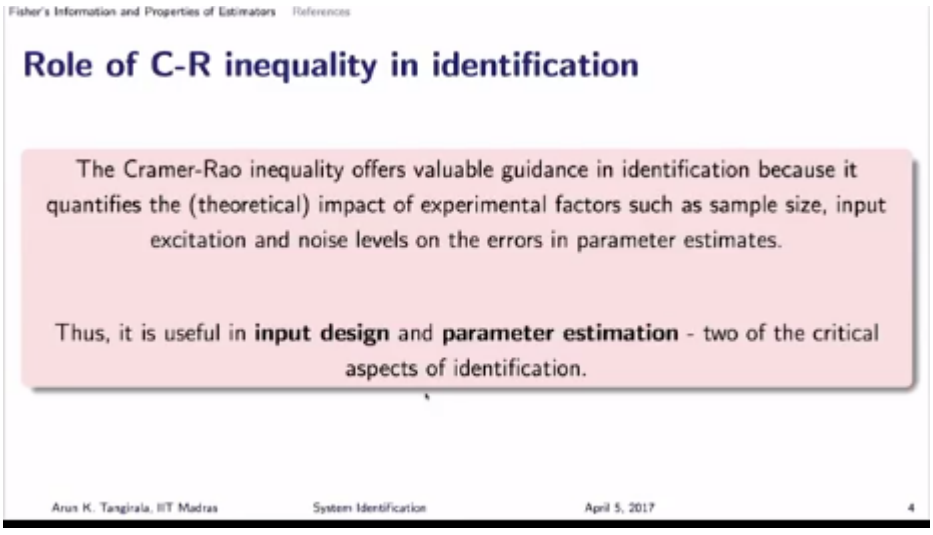
Why can't I check for biased? When it comes to seeking minimum variance among biased estimators, the minimum variance that I can achieve depends on how much bias I'm willing to sacrifice. It's very easy to see. Right. That means, the minimum variance that I can achieve depends on the bias that I'm willing to tolerate. If I say this is the bias, then if I fix the bias, then I can ask what is the minimum variance for a fixed bias. But without fixing the bias, if I want to hunt for a minimum variance estimator, then that's not a well-posed problem, because it depends on the bias. So, you can look at the problem this way, I fixed the bias to zero. I can change this question to saying, for this bias, what is the minimum variance? For a fixed bias, what is the minimum variance, I can ask. You can think of, therefore, this statement as the bias being fixed to zero. So this is also a question of searching for minimum variance estimator, but by fixing the bias to zero. Because if I do not fix the bias to some value, then the choices are infinite, it's not a well posed problem. Right? Because you can, if I don't fix the bias, you can say, look, I found a minimum variance estimator with some bias. I can come up with another estimator which has lower variance than what you have by sacrificing more on the bias. So, there is no comparison. In order to be fair, we have to fix the bias. So, we are fixing the bias to zero. That is how you can look at it as. And the Cramer-Rao's inequality tells me what is the bound on the any estimator with zero bias. There are versions of Cramer-Rao's inequality for non-zero bias as well. We'll not go into that but as long as you understand the nature and the essence of this result, you're okay. What is this result telling me? For a given unbiased, I mean, for a class of unbiased estimators, for the class of unbiased estimators, for a given problem, what is the minimum variance that you should expect to see. That is half of the result. The other half of the result says, what is that estimator which gives me the lower bound, if it exists? Sometimes, it may be just an imagination. It may be a mathematical estimator but not a physically realizable estimator. So, Cramer-Rao's inequality has two parts to it.

(Refer Slide Time: 21:25)



One, it says that the lower bound is the inverse of Fisher's information. That means, if you come to me and say you have an unbiased estimator and you claim that the variance is lower than what the bound says that you have done a mistake. You cannot achieve a bound lower than this among unbiased estimators. If you're willing to sacrifice the bias you can get lower bound in this. That is a different story. And then the second part of the result says, what is that estimator, what is the formula that gives you an estimate with this error or squared error. This variability. Right. So let's look at an example here, in a system identification, the Cramer-Rao's inequality is used in input design and parameter estimation as I mentioned, yesterday, as well.

(Refer Slide Time: 22:15)



And there's, in the video lectures, you will find an example where we derive a minimum variance unbiased estimator for estimating mean. See, here, the approach to designing an estimator is different from what we have learned earlier. When we went through the early example, the first example in estimation, there we went through a systematic procedure where we set up an optimization problem and so on. But we never imposed any constraints on how good the estimator should be across realizations and so on. All we said is, it should minimize some square prediction error. But the approach here is different. The approach is directly attacking the property of the estimator and says, I don't know what you do, but you have to get my minimum variance unbiased estimator.