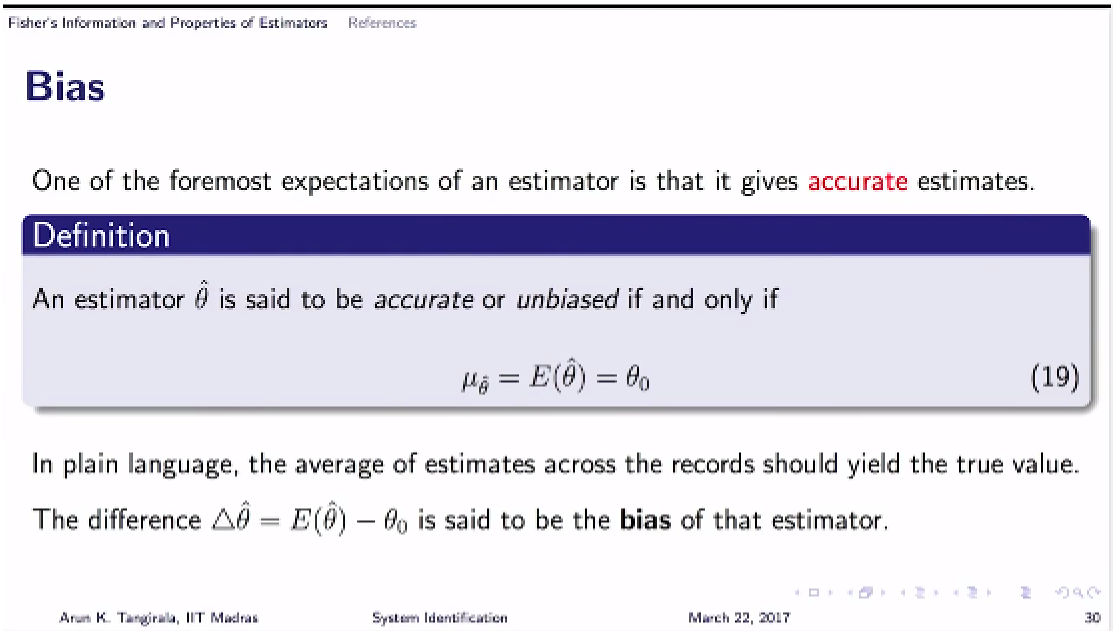
CH5230: System Identification

Fisher's information and properties of estimators

Part 04

So our focus now turns to estimator. And this is where we talked about, remember, different properties. So the first one, we talk about is Bias.

(Refer Slide Time:0:19)



I've already explained briefly in the last class, what is Bias? And then you have variance. Within this variance, we talked about efficiency.Right? And in this context we also talk of-- so let me actually write here, so this you can say a. efficiency and we talk about Cramér–Rao inequality and three-- then we move to talk about mean square error, MSE. And associated with this MSE, property is a consistency. These as we call are known as the statistical properties. These are said to be statistical properties are finite sampled properties. And then you have the asymptotic properties, [1:42 inaudible] give finite number of observations, large and large observations. How well does the estimator to-- in fact consistency actually goes and sits here. So, here is when we talk of consistency. And then we talk about distributions. Of course, you can talk of distributions even here in the finite sample properties. And from the distributions, we construct confidence intervals.

And this is where you run into hypothesis testing. These are the things that constitute a typical set of properties of estimators and the videos that I was referring to the talk about this in detail. Okay?What does bias measure? Bias measures so called accuracy of the estimator. So I going to, as I said earlier, I would like you to sit through these videos again, those videos in and listen to them in detail. I'm only going to do a quick review. The bias of these estimators, of any estimator is a measure of the accuracy, how good the estimator is, right?

And it always helps to imagine this kind of an example, where this is the target, call it theta naught, this is the truth, and here are your estimates, let us say. I am talking about the single parameter here. So if your estimates are spread around this theta naught, in such a way that their average works out to be theta naught, then we say it's an unbiased estimator, right? If you have another estimator, same theta naught here, but now the estimates are actually here, we are falling here either to the right or through left doesn't matter and the centre is here, let us say this is expectation of theta hat, which is not theta naught, then we say we have a biased estimator.

So on the top, the average of the estimates coincides with the truth. And once again, you should get used to this thought process. What is expectation of theta hat? How do you explain? How to explain expectation of theta hat? As a thought process what do you imagine? What does it mean by taking expectation of theta hat? What kind of a thought experiment do you imagine?

STUDENT 1: [4:50 inaudible]

Anyone expected. How do you explain expectation of theta hat? It's extremely important to understand that, to understand the properties of estimators. Anybody here?

STUDENT 2: Remember set of observations whose samples [5:10 inaudible] take, a estimator parameter and take the mean of those.

Mean of what?

STUDENT 2: Mean of the parameters that we have estimated from sampling of the observations.

From?

STUDENT 2: Sampling the observations.

Yeah, but so one, I have a set of samples, right? I get only one theta hat.How do we computethe mean? Let's say, I have 100 observations, I'm going to compute that the sample been, so that's an estimate of mu.

STUDENT 3: If multiple experiments [5:40 inaudible]

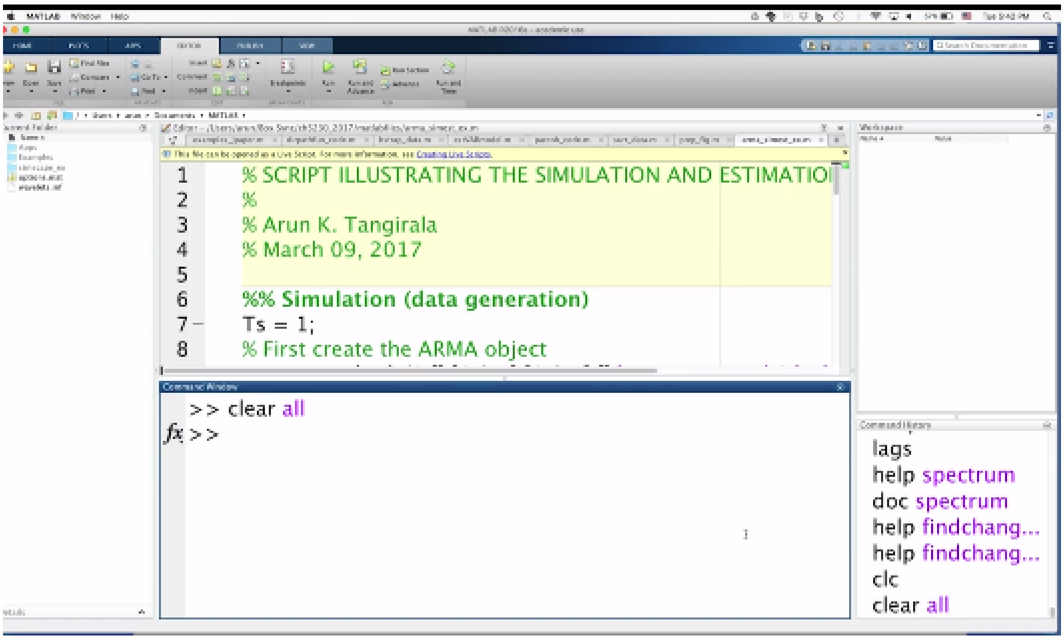
So now as I repeat the experiment, everything held constant, I repeat the experiments, I have multiple data records. This I have told earlier as well, you should get used to this thought experiment. Your repeated observations will give you repeated different data records. Obviously numerically they're going to be different. That's the nature of the random process. So from data record one, what we say realization, one realization I have one theta hat, another realization I have another theta hat.

We are assuming that each realization comes from the same process. Now you cannot apply this to processes that are non-stationary, then you're going to have a problem, right? In fact, what you're supposed to do is imagine that there are thousand sensors, billion sensors, each sensor actually giving you one realization of the process, and you're recording them in parallel. Or if you assume the process to be stationary, then you can use the same sensor, repeat your experiment, right, you will get another realization. Repeat your experiment, you will get another realization. For every realization call i'th realization I have a theta hat i. So I have theta hat I,as from the i'threalization. Estimate from i'th realization.

If you understand this concept, then all the other things are easily understood. And ultimately when I plot the theta hat, I mean all the theta hats that I have for different realizations, they may look, Gaussian, depending on the distribution. If I brought a histogram, or maybe I have a chi-square complete depends, so, let me show you this, right? Let's do this in MATLAB. As I said, let's use this time to understand. So, suppose I want to now, estimate mean of a Gaussian white noise process.

Right? And what did we prove last time? Do you remember-- if I use sample mean to estimate the mean, remember there are many ways of estimating the mean. One way is to use a sample mean. Correct? What is the other way? Sample median or mode. They are many other ways, I can use weighted mean.

(Refer Slide Time: 8:43)



Suppose I use sample mean to estimate the mean of a Gaussian white noise process. We did prove something in the last class with regards to the bias of sample mean. What did we prove?

STUDENT 4: Mean square [8:59 inaudible]

No, no. Something else we prove? What did we say about the bias in the sample mean?

Student 5: It is unbiased.

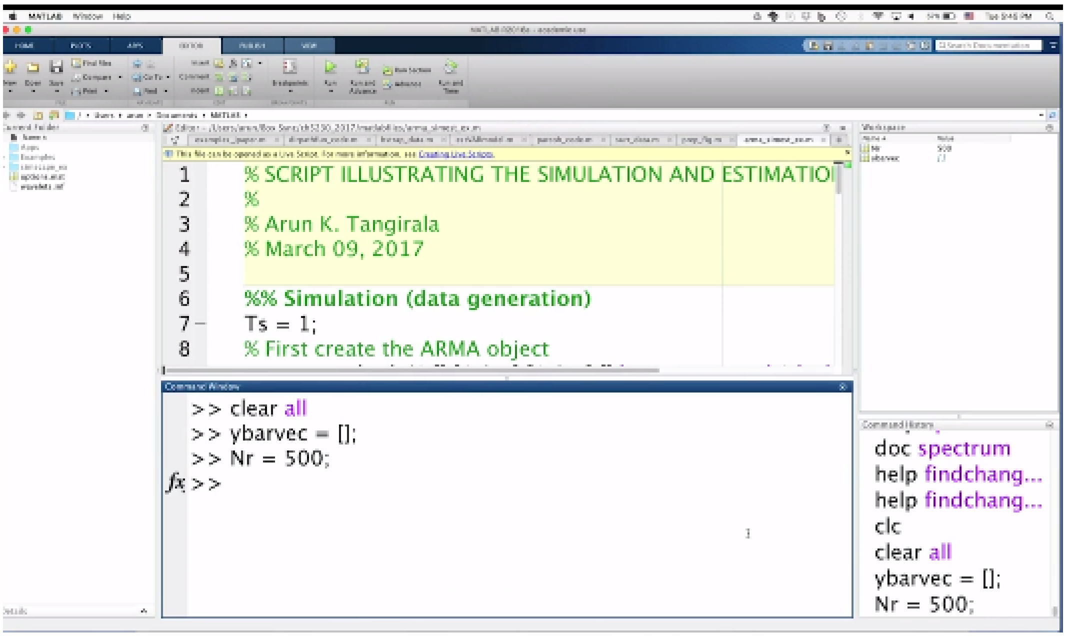
It's unbiased. So, what do we understand my unbiased? That means expectation of-- if I have y then expectation of y bar, how is white bar computed? One over n sigma y[k].What is expectation of y bar? Mu of y. Assuming y is coming out of a random process with mean mu. We proved this theoretically, right? Even if you forget it, you can straight away take the expectation and see that it is unbiased. But you have to understand what this means in practice. And unless you understand that in practice, this result will not stay with you. That is guaranteed. It will not stay with you. As soon as you walk out of this room, you'll forget it.

So let's see if I can use simulation software like MATLAB to verify this result. We may not be able to verifyit to the last decimal, but fairly well enough. So, for this what we should do is, we should get ourselves into a repetitive loop. And let us get ourselves into a for-loop. What am I going to do in one run of this loop? Generate data and calculate sample mean, okay? So, first let us actually assign here, we say ybarvec is empty. What is this y barvecgoing to be?

STUDENT6: Continue sample.

It contains the y bar i. So, let's fix the number of realizations. We say Nr, let's say I'm going to work out, let us say about 500 realizations. Usually you have to be careful with this, here because the sample mean, I can afford to run this for 500 realizations. Because estimator is very simple, but if the estimation time is going to take long, then you may have to cut down on the number of realizations.

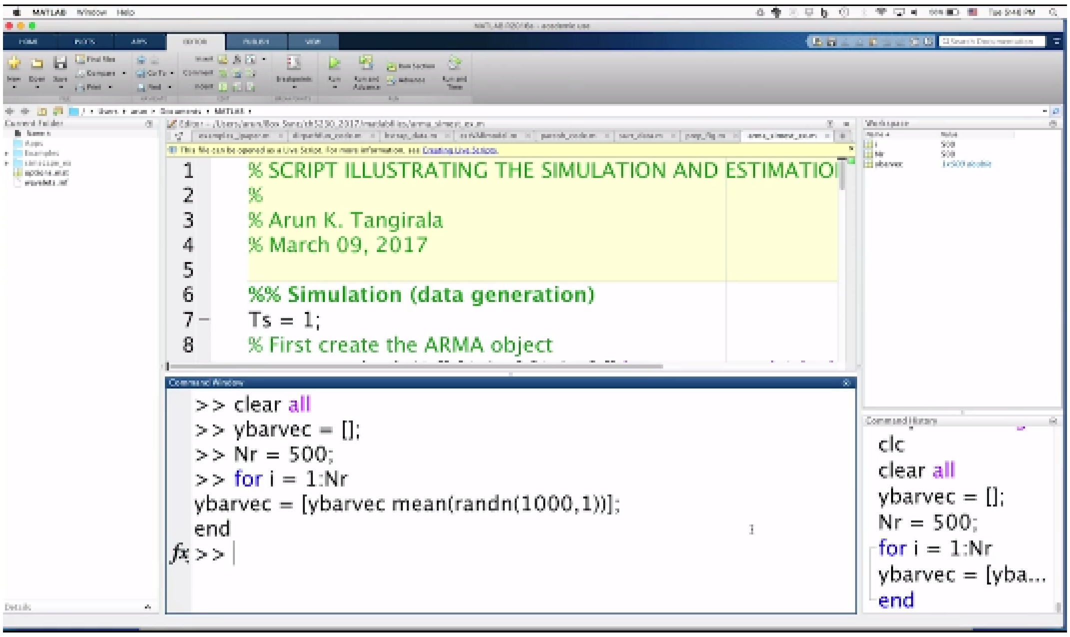
(Refer Slide Time:11:15)



Sampling mean is very simple, it's a simple mean. Okay, now what we shall do is that, we can run from 1 to Nr, right? What is this i keeping track of? Realization.Very good.

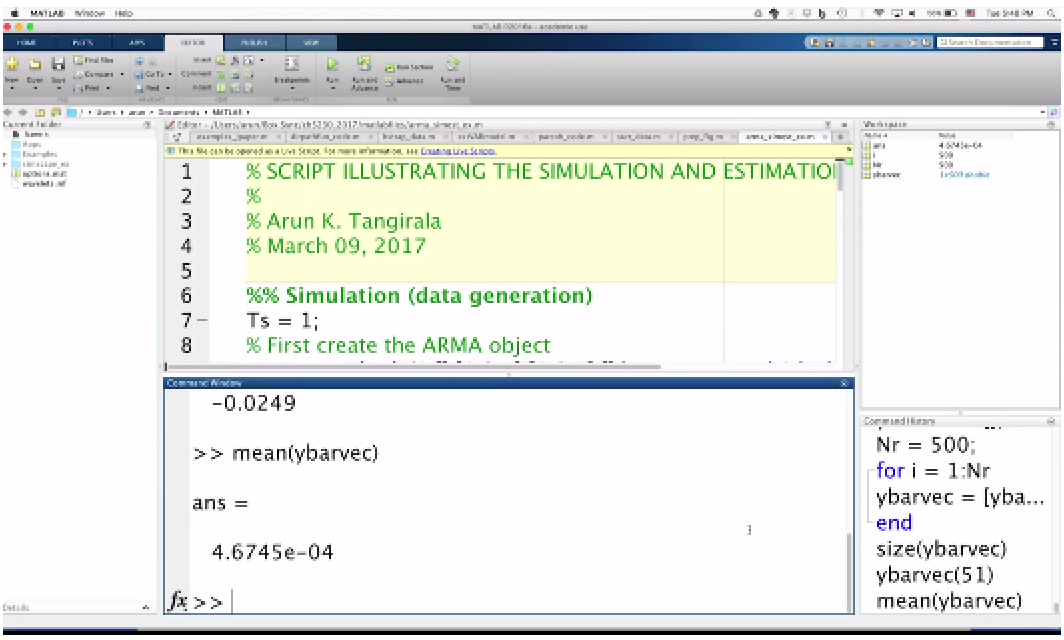
So, now I have y barvec, I'm going to do it this way. I can also do y barvec of i, either way is okay. I'm going to do this but when you follow this procedure, make sure that your y barvecis initialized always to empty matrix. Otherwise it'll just keep adding up. What am I supposed to do here? Mean of-- I'm not interested in storing the realization, let us say. I'm only interested in storing the estimate. So rand n of, let us say 1000 observations, I should have fixed n, up front but okay, let us do that. Fine.Done.Fine. I'm done. So, what should be the size of y barvec? Sorry. Find it.

(Refer Slide Time: 12:33)



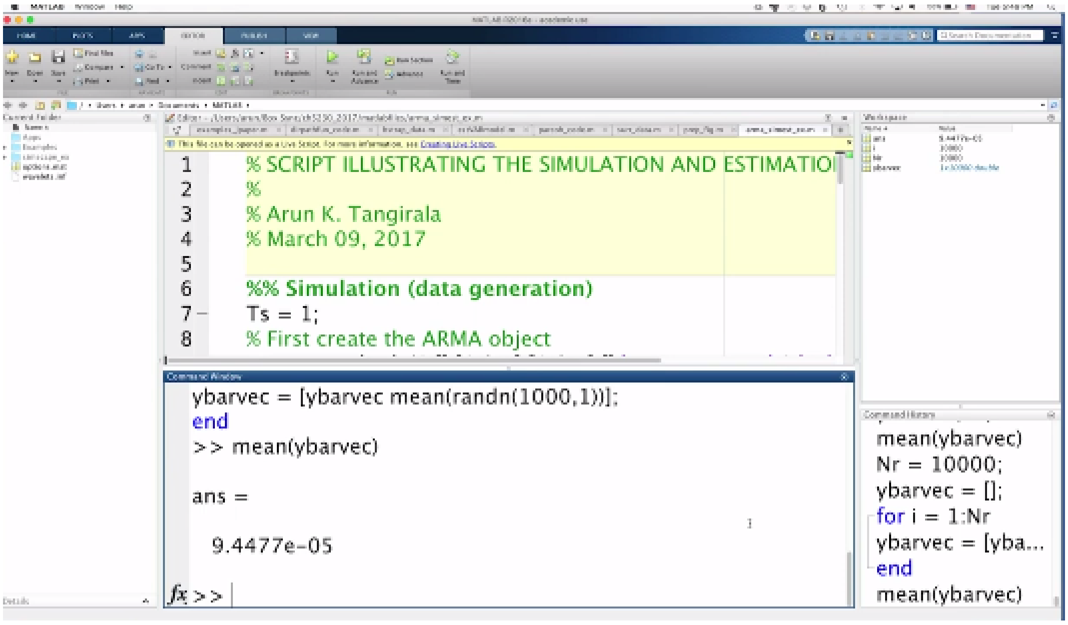
Okay. We get one by one. That's okay. We get as a because, we [12:39 inaudible] it as a row vector. Now, if I take randomly any element of this, so let us say, I picked the 51st element; this is the value, right? I can in fact; I can do many things with this now. One of the things that we would love to do is, what all is the thing that you want to do? Take mean, right? That means now we're going to take a statistical, we are not taking expectation, but we are taking a simple mean of this. That we can do, although this is not the perfect substitute for expectation. Expectation is supposed to actually take a statistical average. I'm taking a simple average. But let us say we do that. Quite small, right? Which is encouraging, so, which means now if I take the average of the averages-- again I'm cautioning this simple mean that we have taken is not theoretically the correct substitute for average expectation because what is expectation supposed to do? Take a weighted average with the probabilities, correct?

(Refer Slide Time: 14:09)



So we have replaced that to the simple average, but that's okay. Even simple average should actually get mean, what I require. So what do I observe here? This is much smaller, it's closer to zero. If I increase the number of realizations, let us say to 10,000, or maybe a million, then we can get that. So let us say, I increase it to 10,000 now, and repeat the same process. But make sure that y bar isinitialized here. Now, I have y bar, again, the same story. We will keep the number of observations fixed in a single realization. Okay, this is also is done quickly. Now you're getting there, slowly. This this is smaller than what we have seen earlier, right? It was 10 power minus 4, it's10 power minus 5.

(Refer Slide Time:15:03)

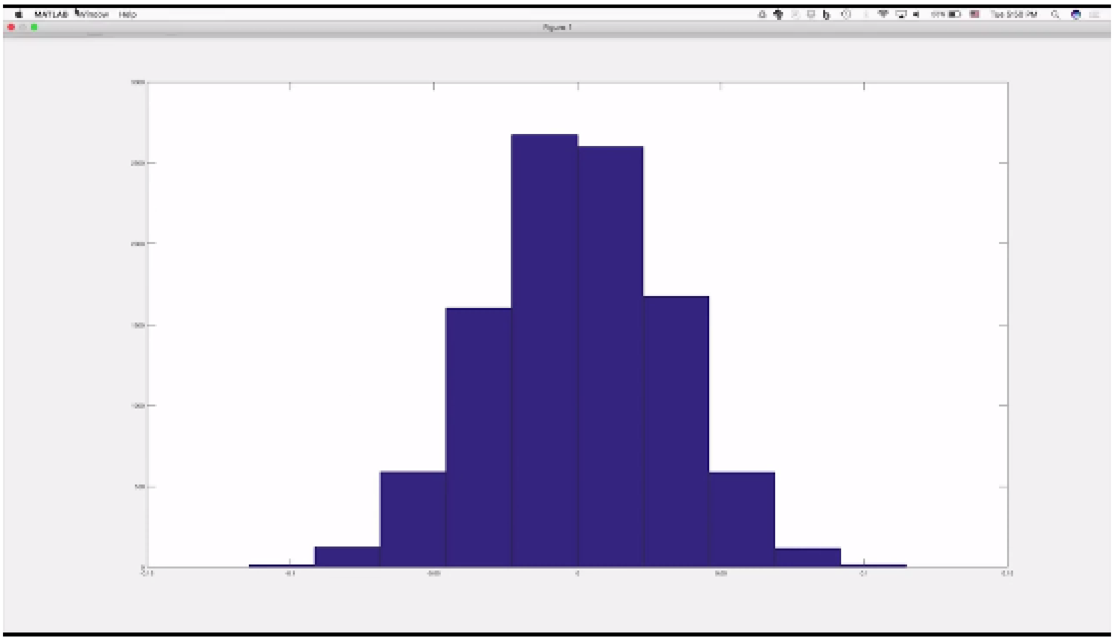


What else do we want to do with this? Something I've done on the board.

STUDENT 7: Plot.

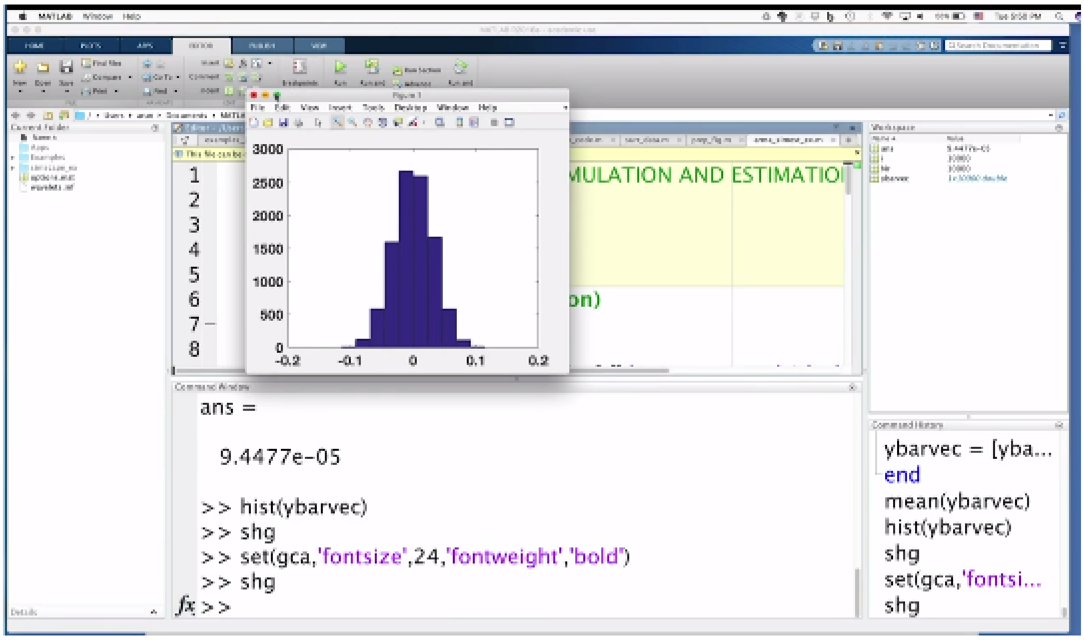
Plot what? The histogram of this, right?Because I would like to know how the estimates are distributed. Straight away, we are addressing this aspect here. The distribution is extremely useful in setting up confidence regions, we know that. So straight away plot histogram of-- there you go. What do you see here? Nice Gaussian distributed estimate, right? Sample mean is very nice and many, many respects.

(Refer Slide Time:15:54)



So it's it a Gaussian distributed and I don't know how well you can see the axis thing here, but let me increase the axis font size. To maybe 24, so that you can see nicely. Here.

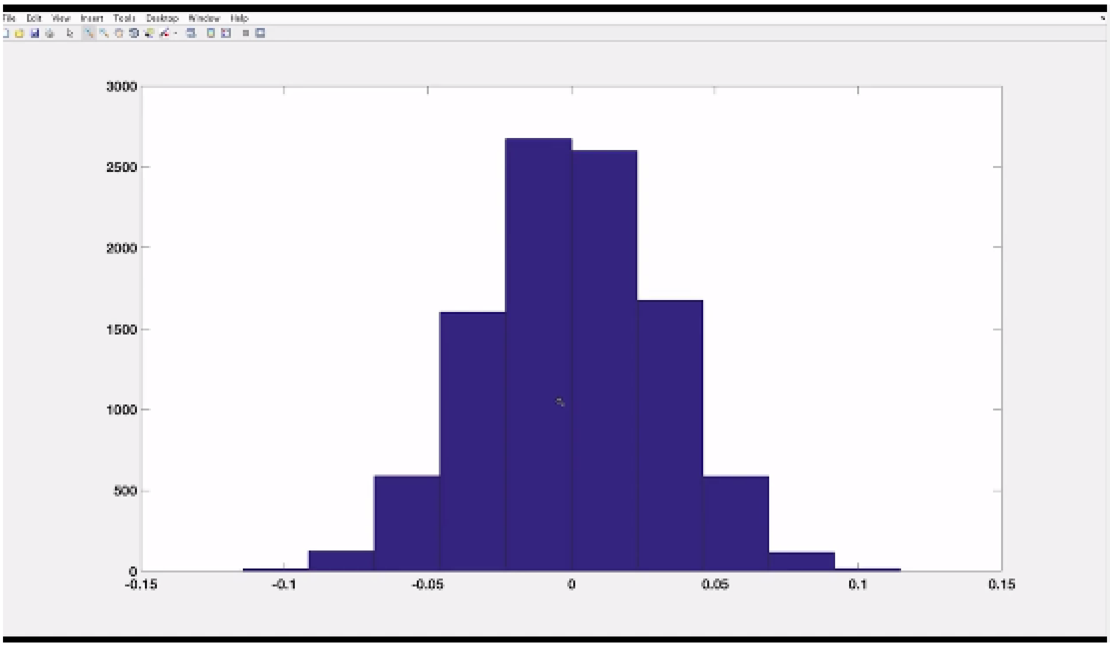
(Refer Slide Time: 16:06)



So what do you see, it's being centred around, zero. So which is what we want. That is a mark of an unbiased estimator.

On the other hand if I had probably not used one on N, but maybe just for this for the heck of it I may be used 4 over N.

(Refer Slide Time: 16:25)



Then you should expect to see bias right? What do you think? Yes or no? Theoretically itself, we can prove.

STUDENT 8: Sir, if the mean is zero, then [16:46 inaudible]

For zero mean, yes, it won't make a difference. But shall we see that? I leave it to you. I've shown you how to do this now. So, this is how given any parameter estimate, I have shown you how to do this for sample mean, tomorrow or sometime later we will be estimating parameters of a FIR model or some Arcs model, OE model doesn't matter. How do you check if the estimator is unbiased? Suppose you're estimating parameters of an OE model, how will you do it? Story is the same. You generate data, this is true simulation. Of course, performing 10,000 experiments maybe extremely difficult, unless there are 10,000 people working on the same process then it may be easier. But for a single individual to perform 10,000 experiments can take a lot, lot of time. This is the power of simulation to understand concepts.

So, if I want to later on at some point in time, verify that there is an estimator that estimates parameters of an OE model. Okay? And I want to actually check if it is unbiased or unbiased, what do I do? Generate the data; rerun the simulation and you have to ask what happens, what is different from one simulation to another simulation? The deterministic part of the process will remain the same. Like in a liquid level case study, there was a ODE governing the tank and then there is a noise that was being added. That realization will bring in the randomness. So, you generate data again and again and for every-- from every realization, you estimate the parameters of the OE model. Same model you have to fix the model. Model structure is fixed.

You have to use the same algorithm and then for individual parameters, you take the average or you can even constructed distribution and from this distribution you can either do a fitting, here we visually observed it to be Gaussian. From this you can also construct confidence regions, once you know the distribution of the theta hat, and I'll briefly talk about that tomorrow again, go through the videos. But I'll talk about that, but this is how you verify the properties of an estimator by simulation. What I discuss in the video is through theory. So sample mean, we show it to be unbiased. Then variability, I can also check variability here. What is variance? Again once you understand theta hat to be random variable, everything becomes easy.

So you have expectation of theta hat, then you have variance of theta hat, efficiency isabout obtaining minimum varianceand Cramér-Rao Inequality tells you, how you achieve? What is that minimum variance that you can achieve for unbiased estimators and that is nothing but inverse of Fisher's information. What is estimator that gets you, that minimum variance, whether such an estimator exists or not, here for example, with the simulation that we performed, we can check if variance-- what is the value of variance of theta hat. In this in the notes, we have shown, I have shown that this variance of y bar, when y[k] falls out of a Gaussian white noise process is sigma square over N. This is what we approved under what conditions?

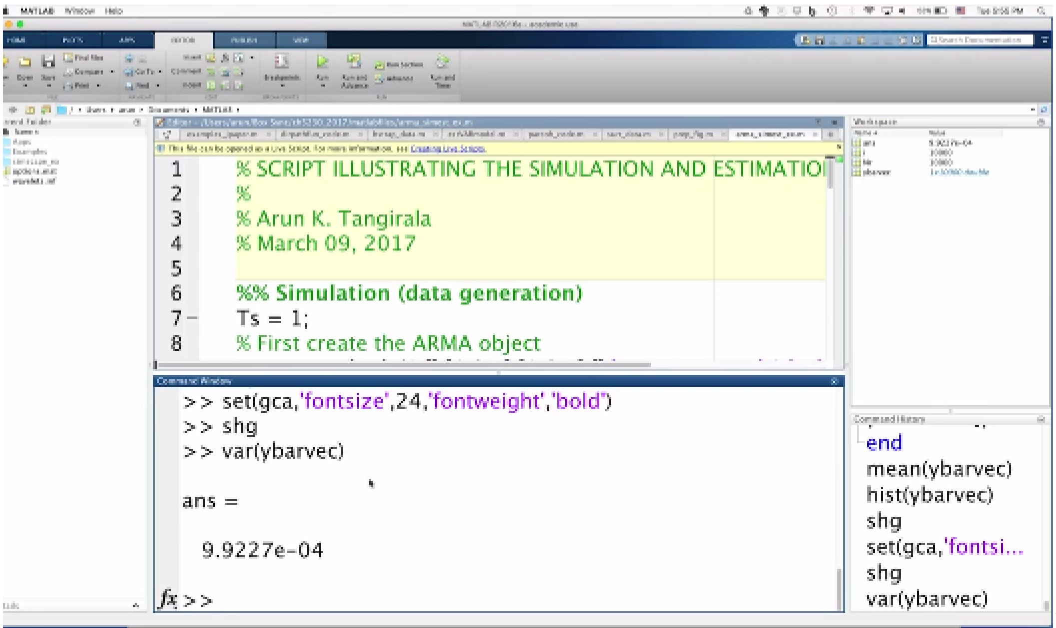
Student 9:[20:39 inaudible]

When y[k] is uncorrelated, correct? It need not be Gaussian white, uncorrelated is good enough. We can check now, if that is true, we can ask for the variability. So, variance of y barvec, again here strictly speaking, this is not the true variances is an estimate of variance of y bar. This is what we get .Roughly 10 to the power of minus 3 or minus 4?

Student 10: [21:18 inaudible]

It is 10 times. So what is it that I should expect? What is a true variance that I've used in my simulation? One. And how many observation did I use in the each realization. 1000, so variance as per theoretical calculation is 10 power minus 3. That is what I get nearly.

(Refer Slide Time:21:24)



This is the Monte-Carloway of computing the variance. The only thing is you require large number of realizations to get more and more accurate measures. But you see the difference between, doing it by simulations versus doing it by pen and paper. How do you do it by pen and paper is what is the notes. So what we'll do tomorrow is, I'll go through a few examples that explain what consistency will go through the same examples and we'll see, what consistency means and what is meant by confidence region? And then we'll get started off on methods of estimation.

Again methods of estimation as I've said early on in the lecture today, please sit through those videos which explain the methods namely Method of Moments, Least Squares, MLE and Bayesian. The

TSwill send you the links. Sit through the videos, so this week we will devote to the estimation methods. Those are generic methods. Once we have gone through that specifically in system identification-- I wouldn't say specifically, there is there are a set of methods known as the prediction error method, PEM, P-E-M which unifies all the least squares, MLE and so on. And we will focus more on prediction error methods, towards the end of this week All right? And ask what is this PEM? How does it unify the least squares, MLE, so on, so tomorrow and day after wewill spend time on some examples showing how these methods work? You will go back and watch those theory or those estimation methods in your, you know, with the access to the videos that will provide. Okay? So we'll meet tomorrow, thanks.