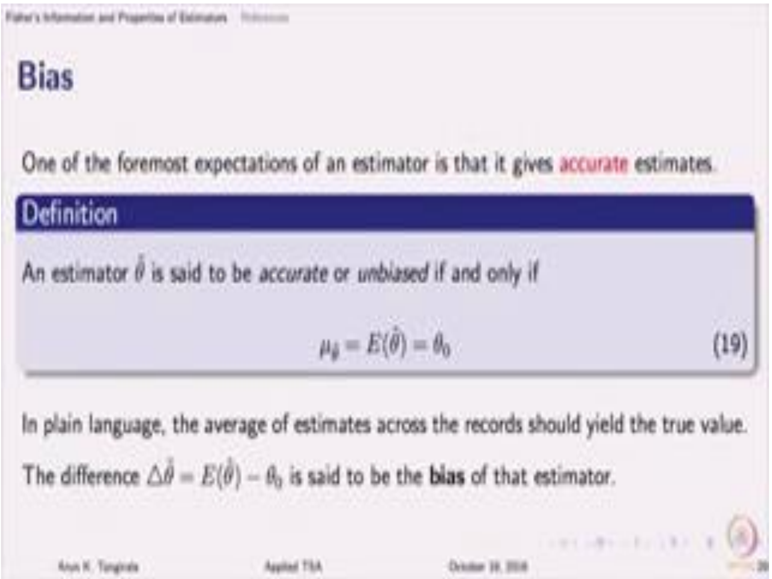
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 85
Lecture 37C - Goodness of Estimators 1 -4

Now, we have to turn our attention to the estimator; now we are coming to the main part the device that actually produces the estimate. At this point we will assume that the data is informative, we cannot handle everything right when we were talking of Fisher's information, he said I do not care how I am estimating I am only going to focus on data. But now we are going to ask if the we are going to assume that the data is informative and qualify the estimator and as I have remarked two lectures ago, one of the first qualifiers of an estimator is it is bias and bias looks at the averaging property of the estimator, remember your theta hat is a random variable and bias is simply the difference between the average of theta hat across the data space, across the outcome space and the difference between that and the true value.

(Refer Slide Time: 01:04)



So, if the estimator is unbiased then expectation of theta hat is going to be the same as the true value. Is this of interest to me in practice? Yes, although I cannot verify in general there are some estimators for which theoretically you can estimate for example, you take the sample mean, we have already shown that the sample mean is an unbiased estimator, how do you show that? You start with the expression for estimator; what is the estimator? 1 over n sigma y k and suppose I want to ask whether this estimator is biased or not? All I have to do is evaluate the expectation of y bar, plug in the expression for y bar and make an assumption what assumption do I make here?

Student: (Refer Time: 02:02).

That y k is stationary; after having assumed that I can prove that the estimator is unbiased. So, the only assumption that I have to make to claim that the sample mean is an unbiased estimator, is that y k is stationary no other assumption is required, sample mean is always going to be an unbiased estimator regardless of whether y k is correlated white or whether it comes out of a Gaussian distribution it does not matter, as long as it is stationary we can claim that the sample mean is an unbiased estimator, it is as simple as that.

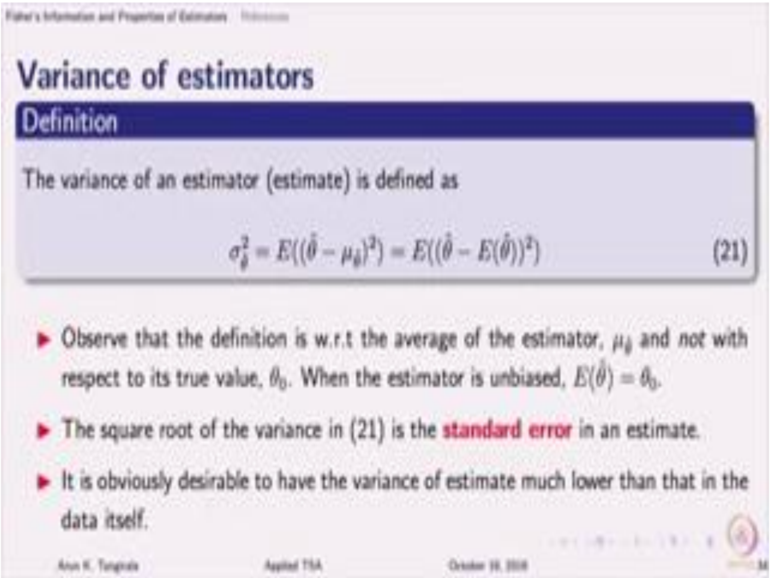
But it is not going to be this easy for any other estimator. Here it is very easy because it is a simple linear estimator, suppose I were to ask you is a sample median and unbiased estimator would it be more difficult than this or as easy as this, what is involved in sample median? There is a sorting operation. So, I have to evaluate the expectation of a sorting operation, which is not as straightforward as simple addition because expectation operator is a liner operator I could easily take the expectation operator past the summation, but I may not be able to take past the expectation operator through sorting operation because sorting is a non-linear operation correct.

So, it may not be easy to prove in general the unbiasness of any estimator by hand; what is the natural request? Today you take request to bootstrapping methods, which of course, we would not discuss in this course, but it that is something that you want to look up in the literature, the modern ways of qualifying the properties of an estimator is through bootstrapping. The idea in bootstrapping is to generate artificial realizations, after all how do you verify? If suppose I ask you to do by simulation, how would you do is suppose you did not know how to evaluate the bias or lag of lag thereof by hand, how would you verify if an estimator is unbiased or not? Generate the different data records, various realizations, compute the estimate for each such realization and then take the average of all of that, since you are the creator in simulation you know the truth, compare the average with the truth and make sure that you generate many many realizations so that your answer is as accurate as possible.

But in practice I will not have the luxury to generate many realizations, I with great difficulty I perform an experiment in my entire PhD right maybe a few experiments, but I cannot perform infinite experiments. Bootstrapping comes to your rescue and shows you how you can generate artificial realizations from the data, today it is all about artificial, artificial flowers everywhere. So, it is artificial realization world, but there is a whole theory associated with bootstrapping, you cannot just do a bootstrapping just like that and then expect things to miraculously work.

So, bootstrapping allows you to generate artificial realizations and then you verify the properties of the estimator.

(Refer Slide Time: 05:20)

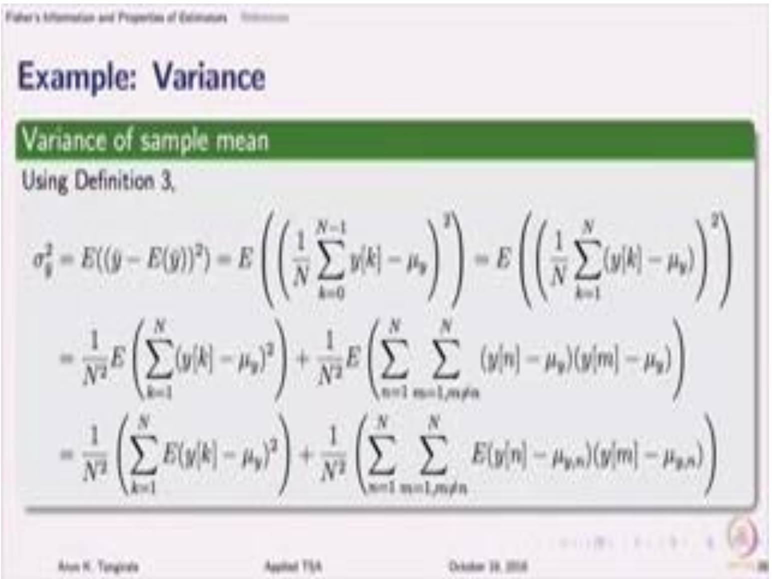


And this is not only true for bias even for variance right I am going to skip this, the variance of the estimator again as the same story behind it, we have already defined what is variance. Variance is nothing but, the spread of the estimate, around its own average, across data records; the moment you see expectation of something, you should imagine different data records.

So, again the thought process that should go on in your mind is as if you have performed many many experiments and for each experiment you have generated from each experimental record you have generated you have calculated a theta hat and now you are looking at the spread of theta hat across experiments. But the only point that you have to keep in mind is you are calculating the spread with respect to its own average, not with respect to the truth. So, the variance can be calculated in general also

Now, the square root of this variance is called the standard error in any estimate, that is a very that is a standard what is the difference between error and standard error? The error in any estimate is simply the difference between the theta hat and theta naught that is the error. This is standard error, why this is standard error? This is what you expect to see across data records, now you are averaging this error in some sense across data records; the error in a single estimate is simply theta hat minus theta naught that will never be zero, the here this is a standard error and this standard error is the error at averaged in some sense across data records and this can be driven to zero. So, let me just conclude with a simple example, I will talk about the uses of variance later on let me just conclude the class with an example.

(Refer Slide Time: 07:07)

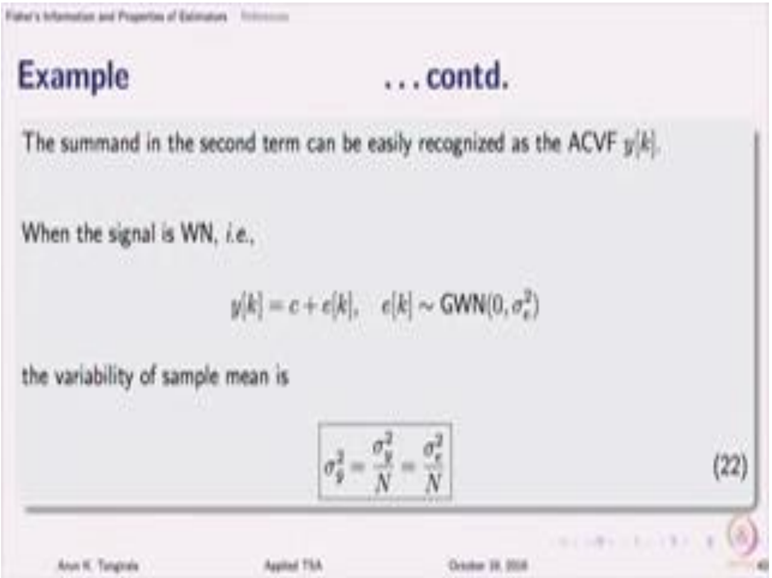


So, this is the expression now we want to now calculate, what is the variability that I see in sample mean? The sample mean is one that lends itself very nicely not only to calculations, but also to theoretical analysis. So, we return to the sample mean and now ask what is the variance in the sample mean? Yes the expressions look a bit intimidating, but if you work out they are fairly straight forward, it is just expansion and patiently working through them.

Look at the last expression there are two terms that you see, what is the first term, what is the first term work out to? In the summation you have expectation of y k minus mu to the whole square, what is that? No, no expectation of y k minus mu to the whole square what is it? It is a variance of y right and you have n such variances. So, the entire summation divided by N square works out to be sigma square by n.

Now, come to the second term, the second term you see you recognize the innermost term to be auto covariance right here we are not assuming y k to be white noise yet, if y k is white noise if white noise pro falls out of a white noise process, the second term vanishes otherwise the variance depends on the auto covariance structure.

(Refer Slide Time: 08:38)



So, for a white noise process what is the variance? Sigma square by n that is what is the standard expression that you see in all statistics textbooks, that the variability in the sample mean for a random sample; random sample would mean uncorrelated observations is sigma square by n.

So, the lesson that we learn from this expression is the variability in the sample mean even though it is derived for a white noise process, is dependent on two things what are those? The variability in the data itself, how the data varies across experiments and how many observations you have collected; which is the one that I as an experimentalist I have in control the sample size, I do not have a control on the variability in the data that is all fixed the moment I use a fixed sensor and so on. The process and measurement mechanism will fix it what the good news is as n goes to infinity, the variance can be driven to 0, which means the standard error can be driven to 0; when I can do that for an unbiased estimator note carefully when I can drive the variance to 0 for an unbiased estimator, we say that the estimator is consistent meaning the sample mean or that estimate will converge to the truth, but it will converge to the truth only when n goes to infinity.

So, that is one form of define defining consistency, but anyway we will talk about consistency later on in the next class which is next week, but this is good news for experimentalists, which says that as you increase the number of samples, the error in your estimates standard error in your estimate will fall down and this is what you are looking for in one of the key properties that we are looking for in an estimator. As I increase the sample size the estimate should get better and better.