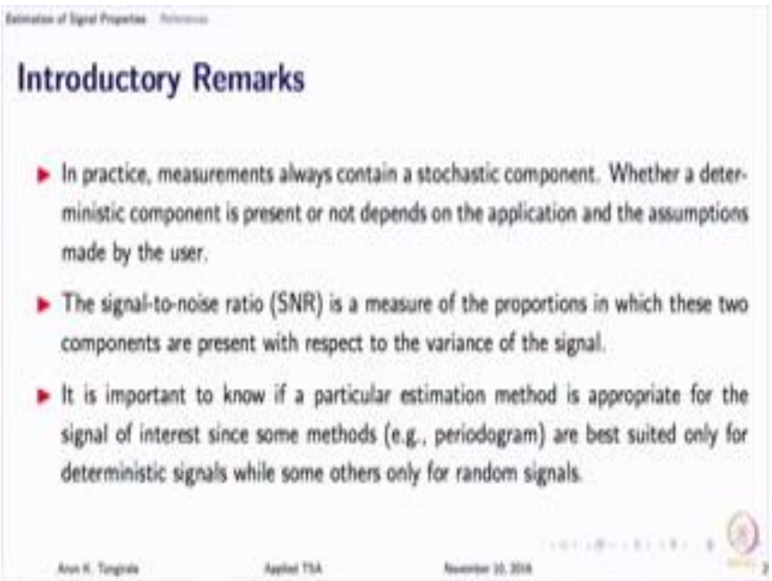
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 110
Lecture 48A - Estimation of Time Domain Statistics 1

Good morning, let us begin. As I said today, we look at the estimation of signal properties namely the cross correlation function, auto correlation function and the spectral density. So, those will be the centerpieces, of course, in the estimation of auto covariance function or even cross covariance function, we know that one has to estimate the mean apriori, if you look at the theoretical definition and before we jump into any of this, there is one very important remark that I want to make with regards to the estimation of signal properties, whenever you are estimating some property there have to be 2 points that should be kept in mind.

One, what assumptions you are making about the signal, which is whether you are assuming it to be deterministic or stochastic or a mix of both that is very important because as you will see in the estimation of power spectral density, it makes a huge difference whether you are treating the signal as a deterministic one or is a stochastic one that is primarily because a theoretical definitions are different.

(Refer Slide Time: 01:28)

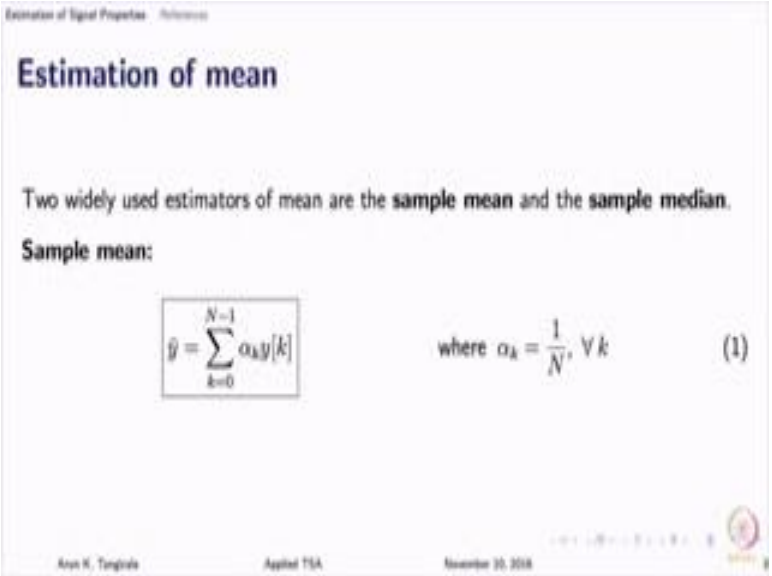


That has to be kept in mind and then there is a second point concerning estimation in practice you can use any of this formulae that we are going to see or any other expressions that you come across in the literature for estimating signal properties.

As far as a computation of that estimate or that use of that formula is concerned, there is no problem you can use for example, if you may have a non stationary signal and you can use the regular estimate of ACVF that we are been using, there is nothing wrong, the actual issue arises when you want to infer something about the true signal using the estimate that you have obtained that is when you want to connect the 2 worlds; the estimation world and the truth that is where the right assumptions have to come in and these. So, basically remember these 2 things, one, what assumptions you are making about the signal and 2 whether you are going to simply use the estimate as is or you are going to draw some statistical inferences. Typically it is the case that the purpose of estimation is for statistical inferencing, but there are some applications where people will say I do not really worry about what the truth is, I am going to use some formula to come up with the metric and I am going to use that formula for all signals. Then it is a different story.

But then unless you make some assumptions about the data generating process, you will not be able to connect the worlds of estimates and statistical inferencing and in the case of composite signals where you have deterministic plus stochastic signals, the signal to noise ratio place a significant role in your ability to uncover the truth so that we have already seen many a times. So, let us gets started I will skip the initial portions of this lecture which has got to do with estimation of mean, we have talked about sample mean quite a lot.

(Refer Slide Time: 03:24)

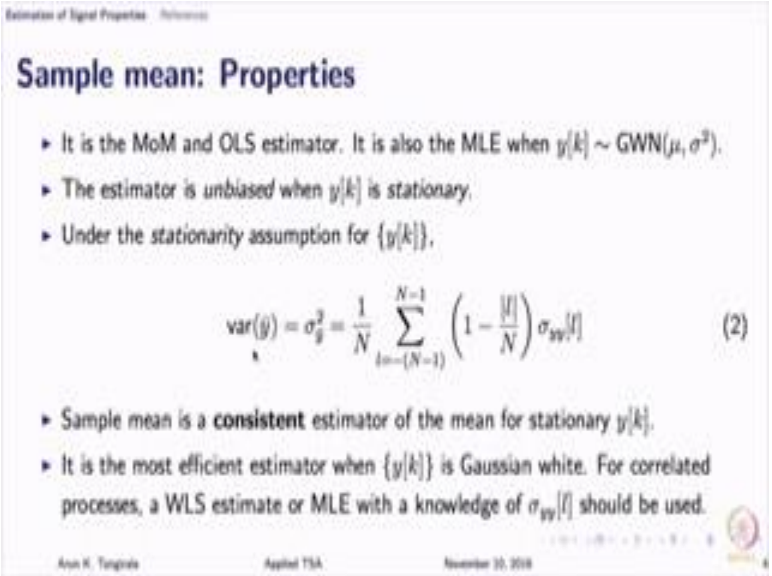


We have said sample mean is efficient, it is an unbiased estimator, it is efficient when the data comes from a Gaussian process, but it is not robust, we know that it is very sensitive to out layers.

Now, on the other hand you have sample median sample median is not as efficient as sample mean when the underline process is Gaussian since all about what is giving you the data if the data generating process is Gaussian sample median is not as efficient as the sample mean on the other hand it is quite robust. So, if you turn to robust statistics as I said there is a measure called break down point which measures how sensitive the statistic is to the presence of out layers and unfortunately sample mean is a breaks down very easily as compared to sample median.

Now, having said that if the data comes from a laplacian distribution then the sample median is fully efficient so you see the distribution; this itself tells you that the data generating process has a huge impact on the error characteristics of your estimate. So, will we will predominately work with sample mean, but may be its for completeness sake you should read through the sample median as well one point I want to make in passing is if you look at the variance of y bar.

(Refer Slide Time: 04:56)



We have derived in the class the expression for the case of white noise process uncorrelated process.

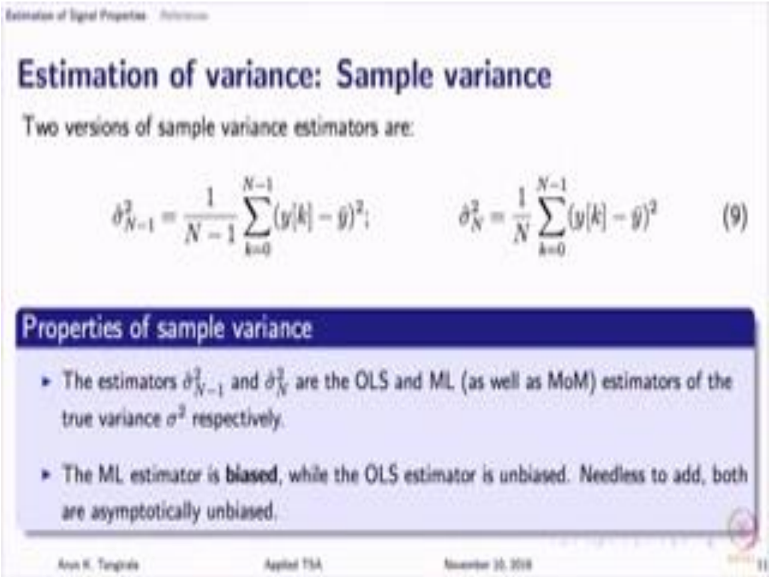
But if the data falls out of a correlated process then you can show that the variance of y bar has this expression and you can see that the correlation nature or the structure of the signal has an impact on the variance of psi bar. In fact, when the data is correlated, this sample mean is not an efficient estimate, what do you need? What kind of mean would you work with? The sample means tends from ordinary least squares right. So, if the data is correlated we know that OLS is not going to be efficient. So, what kind of estimator would you work with just common sense, the sample mean gives equal importance to all observations?

Student: (Refer time: 05:54).

You just you would have waited at least squares, in other words you constructed a weighted mean or you construct a weighted mean or you go back and use MLE, we have solved MLE for the case of uncorrelated Gaussian process that is white noise process, but if the data is correlated one has to go back to the MLE formulation, but then the challenge there would be to construct the pdf. So, unless we know how to construct the pdf or correlated cases we may not know how to estimate the mean which we will learn either today or tomorrow how to set up the likelihood for correlated signals.

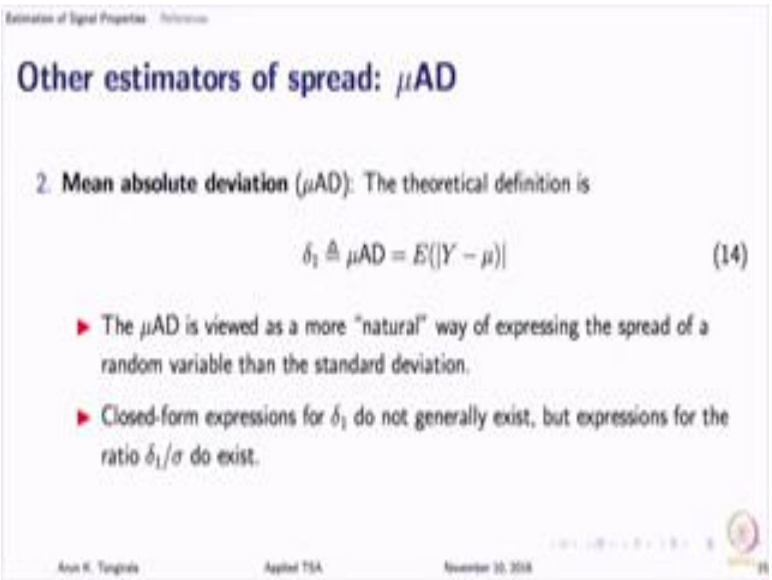
That is one point to be kept in mind; we already know that the sample mean has an asymptotically Gaussian distribution. So, I am going to go fast.

(Refer Slide Time: 06:40)



And go straight to even estimation of covariance because estimation of variance is a special case of estimation of covariance we have talked about sample variance therefore, I am skipping.

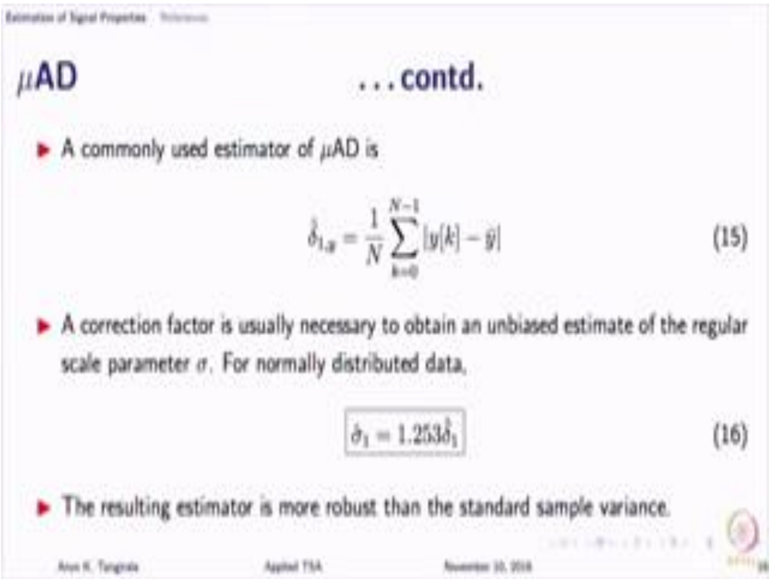
(Refer Slide Time: 06:54)



There are other estimators of variance namely interquartile range or mean absolute deviation and median absolute deviation and so on, these particularly the mean absolute deviation and the median absolute deviation they are robust estimators of the variance.

The mean absolute deviation as defined here is called mu ad it is simple the theoretical definition is it is a expectation of mod of y minus mu compare this with the standard definition of standard deviation in the case of standard deviation you take the square root of expectation of y minus mu to the whole square. So, this is one variation of in the definition of standard deviation.

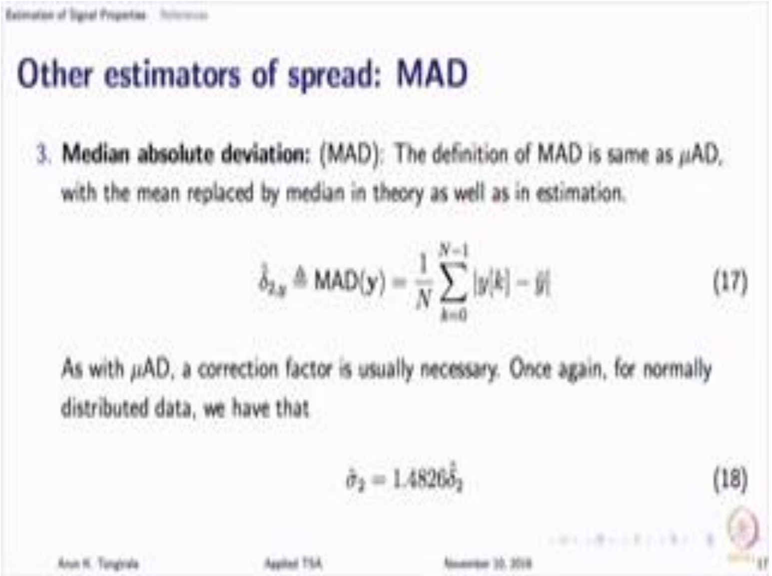
(Refer Slide Time: 07:39)



And in practice you would use this estimator, this is called mu AD estimator unfortunately, this is a biased estimator of the standard deviation of the truth standard deviation therefore, some correction factor has to be applied and I am giving you the correction factor here for normally distributed data this correction factor is determined through Monte Carlo simulations.

You simulate several realizations construct the mu AD, compare it with the truth and say well this is the difference or this is the scaling factor that has to be incorporated. So, you adjust the estimate so that you get biased unbiased estimator.

(Refer Slide Time: 08:19)



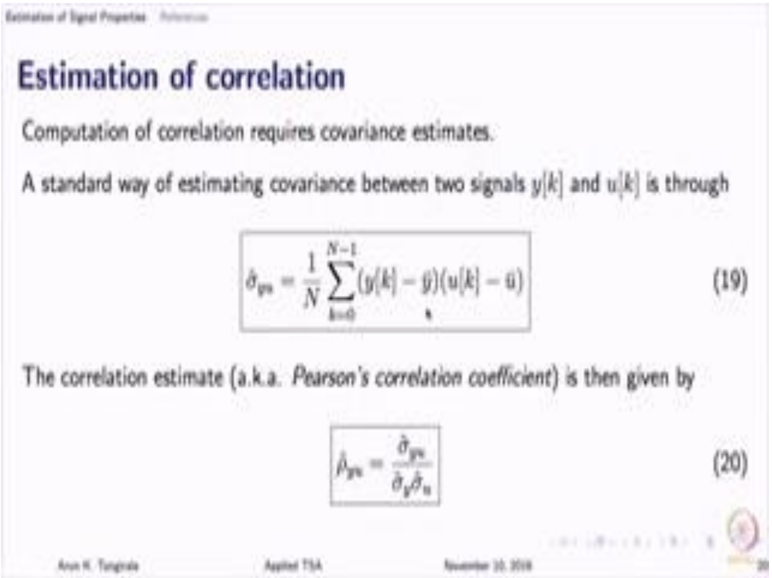
And the third one is a median absolute deviation, in the median absolute deviation case again what you do is a difference between mean and median absolute deviation is in mean absolute deviation you are calculating deviation still with respect to sample mean or with respect to mean where as a median absolute deviation you are replacing the reference point with median, but otherwise the idea remains the same.

This is further more robust, you can say in fact, this is more robust, the mu AD is not so robust because it still realize on sample mean, among the many estimators of standard deviation, the median absolute deviation known as MAD, it is a very popular estimator in robust statistics of standard deviation. We have used this in one of the applications for pipeline health monitoring where we know that we are going to encounter out layers again here this is a biased estimator therefore, a correction factor has to be applied and this correction factor is again for Gaussian distributed data.

You can see in every estimator first we refer to the true definition then we provide an expression for estimation and then we ask the standard question is it unbiased is it efficient is it consistent and so on and depending on the data generating process the estimator may be efficient consistent or biased or unbiased and so on. But what we want always is an estimator that is asymptotically unbiased essentially consistent and efficient, these 2 are extremely important the distributions are not so much of a concern, but most of the times the estimates that we worked with our asymptotically Gaussian distributor and if there is a bias you can see it is to work with bias estimators because I can always have a correction factor.

But if it is asymptotically biased then there is a serious issue and we will encounter such an estimator which is you know consist not consistent which is the periodogram. So, I will go forward and move to the most important one of the most important statistics which is correlation, correlation is first computed by estimate by estimating the covariance. So, correlation is compute estimated.

(Refer Slide Time: 10:43)



By first come estimating the covariance using the expression given here, once again you must understand that this is a biased estimator because although we have n minus n observations we do not have that many degrees of freedom never the less we preferred to use this and I will tell you why very soon.

This is in fact, a special case of the estimator that we use for cross covariance or cross correlation. So, it is fairly straight forward this is an estimator that you have been using one thing that you should keep in that you should watch out for is when you are using routines whether it is r or any other software package look up the expression that is being used for the estimator for example, var which is the routine for estimating variance in r you should go and figure out whether it uses a one over n minus one for variance calculation or one over n do you know already.

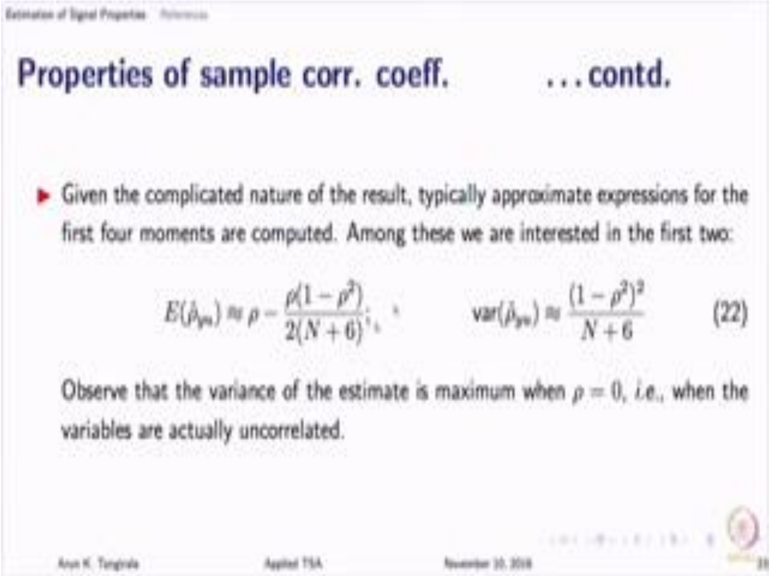
Student: n minus 1.

N minus 1, that is correct. So, you have to be careful where as c o v may use 1 over n. So, if you use c o v to compute the variance it may give you slightly different expression as compared to v a r if you want to be really keep things come and should at then you should use c o v. So, that even across cross covariance function and so on you are using c o v.

Now comes the again now that we have estimated correlation the standard questions is it unbiased is it consistent efficient and so on, we know you can show that it is asymptotically unbiased estimator of correlation. In fact, when it comes to correlation it really does not matter whether you use one over n or 1 over n minus 1 that does not matter because as you can see in the expression it cancels out. So, what is more important to us now is consistency you can prove consistency we are not going to prove that and the most in the last, but the most important thing is a distribution because we want to be able to construct confidence regions and therefore, conduct hypothesis test.

Now, the studies on the distribution of correlation coefficient are quite involved and this what is studied at least about a century nearly a century ago people started looking at estimation sorry at the distribution of rho hat and in those days you would not have any simulators and so on, at least to get a feel. So, it was a tough problem, but it turns out I am going to skip the covariance matrix it turns out that the only approximate expressions are available.

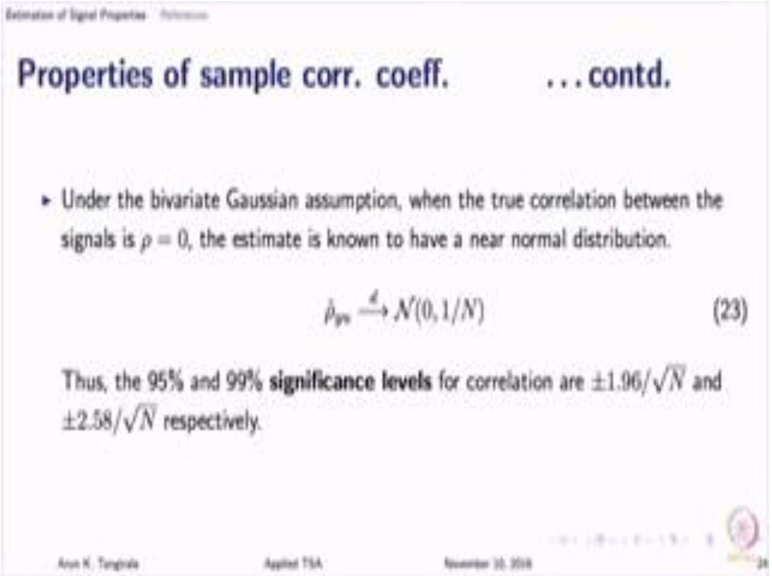
(Refer Slide Time: 13:23)



For example, what I am giving you here are the typically the first four moments of the estimate are computed your estimate is a random variable what you want is a p d f, but approximate expressions are available only for the first four moments and among the first four moments I am giving you the expectation and the variance.

As n goes to infinity what happens do you get an unbiased estimate or a biased estimate biased unbiased. So, it is an asymptotically unbiased estimator and variance goes to 0 as n goes to infinity therefore, it is a consistent estimator right because consistency requires that both the bias and the variance go to 0 as n goes to infinity. So, both of them go to 0 we are convince that at least rho hat is a consistent estimator of a rho now you should also observe that the variance of the estimate is maximum when n equals 0 when the rho true rho is 0 right which means when there is no correlation between 2 variables you going to have maximum error in the estimate, ideally we want it to be the other way round, but we do not know, so that is the issue with this correlation estimate.

(Refer Slide Time: 14:56)

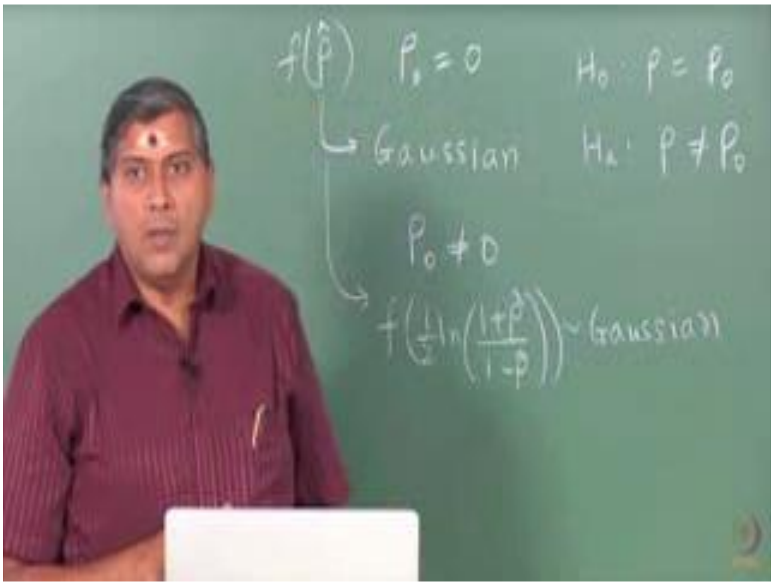


Now, it turns out when it comes to distribution of correlation coefficient as I said determining the property the distribution property of rho hat is not easy, but it turns out that when the true rho is 0 by the way if you look at the expression of rho hat, do you think it is a linear estimator or a non-linear estimator when we talk of estimators and linearity of estimators linearity with respect to what observations. So, what do you think if you look at the rho hat is it a linear estimator, it is clearly non-linear now for non-linear estimators as with non-linear functions the distribution properties will change with as a true value can change.

Whereas the linear estimators behave differently the linear estimator the distribution is essentially the same regardless of the true value. So non-linear estimators the distribution property is depend on the true value whereas, for linear estimators they do not the distribution property is more or less remain the same regardless of the truth that is one disadvantage of working at non-linear estimators. So, it turns out in this case when the true correlation is 0 asymptotically it was shown that the correlation estimate follows a Gaussian distribution again, once again we are back into the Gaussian world with the mean 0 which is good because we know then it is asymptotically unbiased.

This distribution expression that you see in equation 23 is only valid for the true rho being 0.

(Refer Slide Time: 16:48)

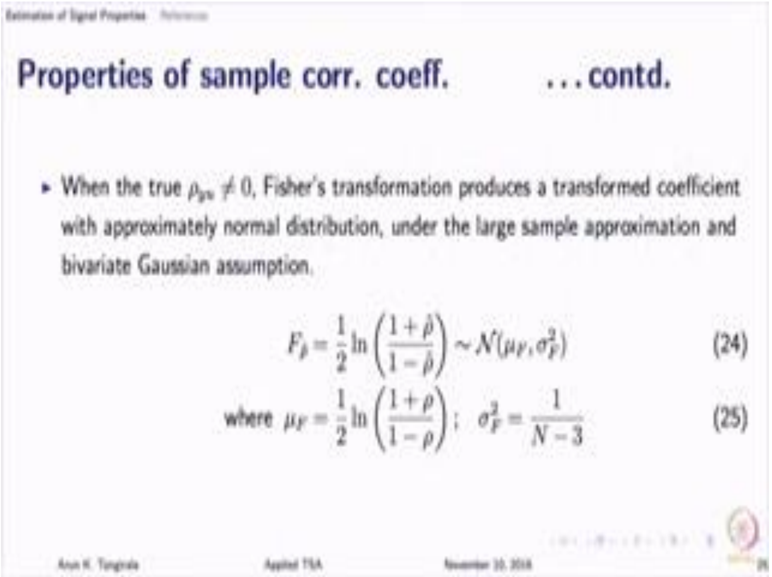


When you are testing a null hypothesis that the true rho is 0 versus the alternative you can use this distribution and then you know carry on or you can use this distribution to construct the confidence regions and because hypothesis test of these type are called significance test all you have to do is you can draw significance bands for example, here the 95 percent significance band for rho hat would be 1.96 by square root of n what; that means, is if the true correlation is 0 and I were to compute estimates of correlation then 95 percent of the times the correlation estimates would be within the band 1.96 over plus or minus 1.96 over root n that is what we mean by significance band.

On the other hand, you can use the confidence region; they lead to the same thing. So, often you see associated with significance test, there is a significance band like you see for your ACVF and PASCF and so very soon, we will talk about ACF and PCF. So, the 95 percent significance level for correlation, the moment you say significance level, you should think of this hypothesis test and use those to conduct this hypothesis. So, if your estimate falls within this band, either plus or minus 1.96 by root n or plus or minus 2.58 by root n then you can you cannot reject the null hypothesis only when it outside this band you can reject the null hypothesis and conclude that correlation is the true correlation is non 0.

But suppose you want to estimate a different you want to conduct a different type of hypothesis test some postulated correlation then unfortunately you cannot use this distribution where rho naught is not equal to 0 in that case you have to use what is known as.

(Refer Slide Time: 18:45)



This was Fisher's contribution again you have to use what is known as fishers transformation what fishers showed is when the true correlation is non zero. The estimate does not follow a Gaussian distribution asymptotically only the transformed estimate and the transformation is given here follows - a Gaussian distribution again asymptotically and that transformation is here half of logarithm 1 plus rho, rho hat over 1 minus rho hat and that was considered a break through work and this is used widely now in statistics right for several decades.

Of course there are more regress results available in the literature and so on, I am just giving you the most important once and once that are widely used. So, to summarize when correlation the true correlation is 0 rho hat follows a Gaussian distribution when the true correlation is not 0 transformed rho hat follows a Gaussian distribution this is a mark of a non-linear estimator this is trademark feature, they do not have the same distribution necessarily regardless of the true values, you have any question? Sorry!

Student: (Refer Time: 19:57).

What is not very clear?

Student: Why did we (Refer Time: 20:03) take a distribution (Refer Time: 20:04).

Is a data, is falling out of it.

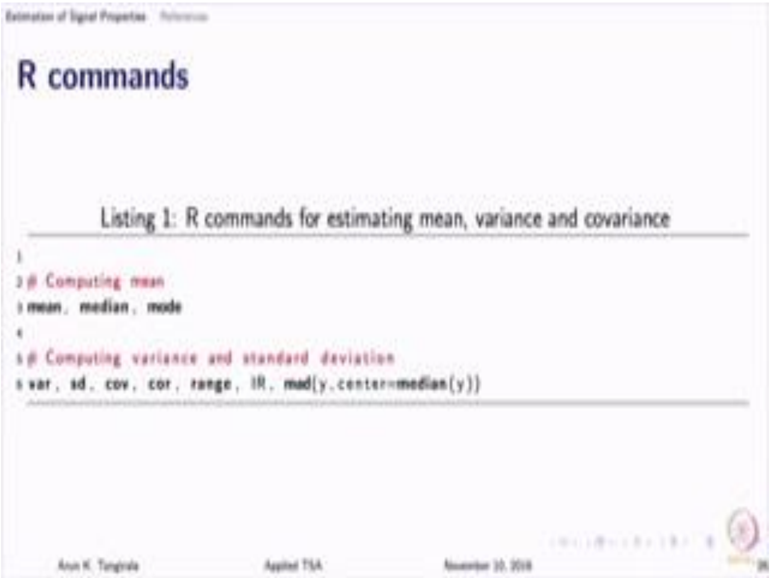
Student: Regression is non 0 when it is (Refer Time: 20:11).

No, no, no, no, no, we had worried about the distribution of rho hat. So, we are asking what is f of rho hat f of rho hat depends on the truth if you say truth is rho; rho naught if rho naught is 0 f of rho hat has a Gaussian shape its belongs to the Gaussian family if the true correlation is not 0 then f of transformed that is half of lone one plus rho hat over one minus rho hat followers a Gaussian.

The distribution of rho hat depends on the truth which you do not know, but that is where hypothesis test are going to come handy right. If you want to conduct significant stress on correlation always the significance stress are of these type for any parameter then you should use straight away this result. If on the other hand you are conducting hypothesis test of the type rho equals rho naught and rho is not equal to rho naught where rho naught is not 0 anymore then you cannot use this result you have to use the distribution result given on the screen which is that the transformed rho hat follows a Gaussian distribution. So, it is cautioning you it is saying that the distribution properties of rho hat depends on the truth and therefore, be careful alright.

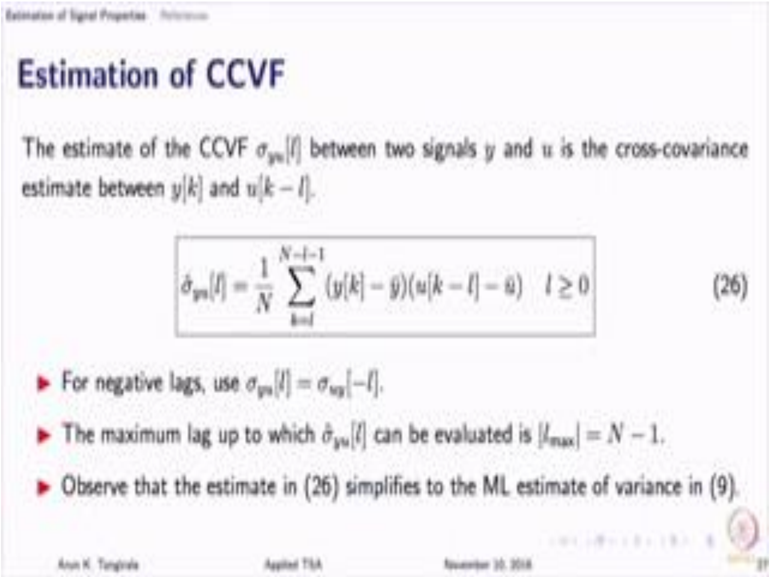
But all of this is actually embedded in your course in your test I think in r you have c o r test for example, which computes which conducts the correlation test for you

(Refer Slide Time: 21:56)



These are some of the routines in r relevant to what we have just discussed mean, median mode and var as I said sdcov, but you should be careful with every routine that you use please look up the help and find out what expression for estimation is being used.

(Refer Slide Time: 22:16)



Now, we move on to the most important quantity perhaps in time series analysis which is a covariance function of course, I am giving you now the expression for the estimating the cross covariance function which is a generalized version of generalization of ACF to the bi variate case.

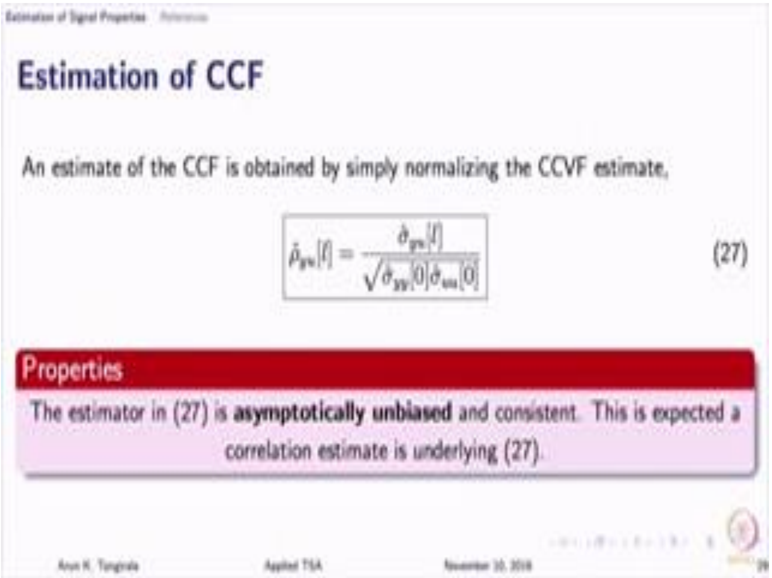
The; obviously, you can see from this expression that this cross covariance bares estimator bares very strong resemblance to the one that we have used for covariance. So, if you substitute l equals 0 you will recover the estimator for covariance as usual this cross covariance estimator is a biased estimator CCVF estimator, the biased estimator. But it is asymptotically unbiased it is consistent and one of the things that has to be kept in mind when you develop or you are working with an expression for an estimator for some quantity some parameter the estimator should have the same theoretical properties of the that the parameter has. So, we know for example, cross covariance function is asymmetric theoretically the estimator should also be asymmetric.

When it comes to auto covariance function for example, there is an important property that the auto covariance function satisfies, which is that of symmetricity for sure and.

Student: Positive semi definite.

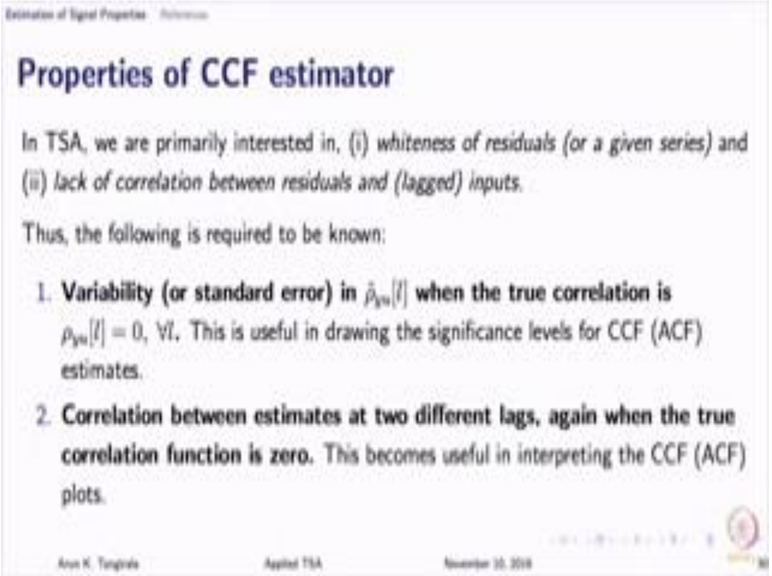
Positive semi definite or non negative definite the estimator should also have that property if we does not then you have to unfortunately throw it away you cannot use it and that is a case you going to be with the ACVF. So, again you derive the cross correlation here.

(Refer Slide Time: 24:04)



By basically working with the cross covariance estimates and the variance estimates of u and y, now, as I said these estimators is asymptotically unbiased. In fact, I do not have to say it is asymptotically unbiased when I said it is consistent it is understood it has to be.

(Refer Slide Time: 24:26)



Now, the when it comes to using cross correlation function what do we use a cross correlation function for we are we use it to determine if one series has some information to predict the other one using a linear model.

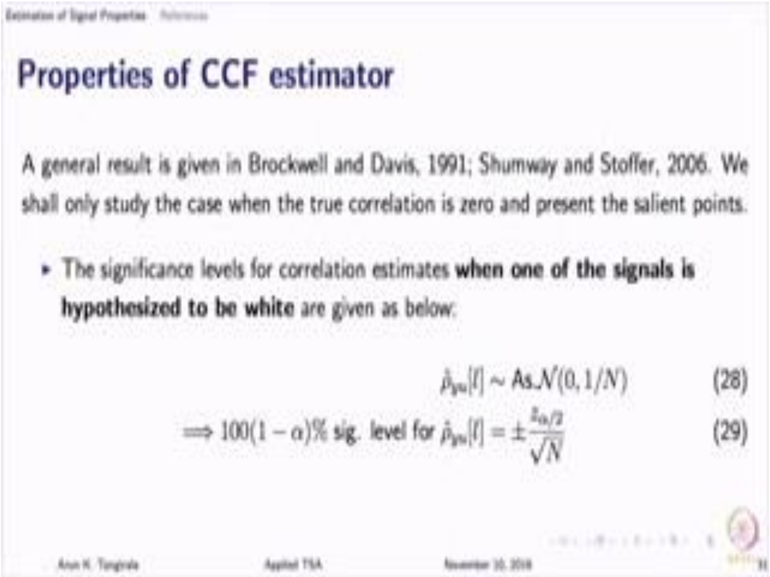
Essentially you are testing for correlation. So, if I am given 2 series let us say 2 prices of 2 commodities then I can use a price of one commodity to predict the other and so on. So one of the standard tests that we want to conduct after estimating is whether the true correlation is 0. So, we need to worry about that case and also since we are computing cross correlation at different lags we have to worry about the correlation between the estimates one is correlation within the series.

But since now I am computing correlation at different lags I have to worry about the correlation between the estimates because what we ideally want is the estimate at 1 lag should not influence estimate at another lag, but that may not be possible because there are going to be common terms correct. But we want to at least see if the true when the true correlation is 0 whether the estimates are uncorrelated at 2 different lags and what is a standard error because you know when we are doing a single time series analysis that is univariate time series analysis.

The first step in time series modeling is whiteness stressed we want to see if the series is white; that means, if the auto correlation is 0 at all lags likewise in bi variate analysis the first step in modeling is to determine if the true correlation is 0. So, the true correlation is

0 is of at most interest always not only in bi variate time series analysis, but also in system identification. So, the results that I am reporting here are given for the case of one of the series being white and being uncorrelated with the other one.

(Refer Slide Time: 26:38)



See, there are many results available when it comes to properties of this cross correlation and the expressions are quite intermediating.

But we are not interested in all of them if you are then you can refer to this textbooks, but what we are particularly interested in is the case where one of the series is un well 2 series are uncorrelated and one of the series is white I will tell you will become clear why we are requiring that one of the series be white. So, if that is the case let us assume that the case is that there is no correlation between the signals and one of the signals is white then you can show that the cross correlation function has an asymptotically Gaussian distribution with mean 0 same something that you saw earlier with the case of correlation. The properties of the correlation essentially depend on correlation function depend on the true correlation.

So, here we are saying if the true correlation is 0 at all lags which correlation cross correlation and one of the series is white why we are insisting that as I said will become clear with the example that I am going to bring up then rho hat follows a Gaussian distribution with mean 0 and variance 1 over n. Which means when I draw the cross correlation when I compute a cross correlation and sorry plot the estimates I would draw significance levels 95 percent significance level or 99 percent significance level and see if the estimates are within that band. If all the estimates are within that band at least up to a certain lag the one of the things with this correlation function is you have to determine up to what lag you want to compute typically, we restrict ourselves to 20, but then if the series has a stronger correlation then we go look beyond. Here, we use this result to draw the significance band.