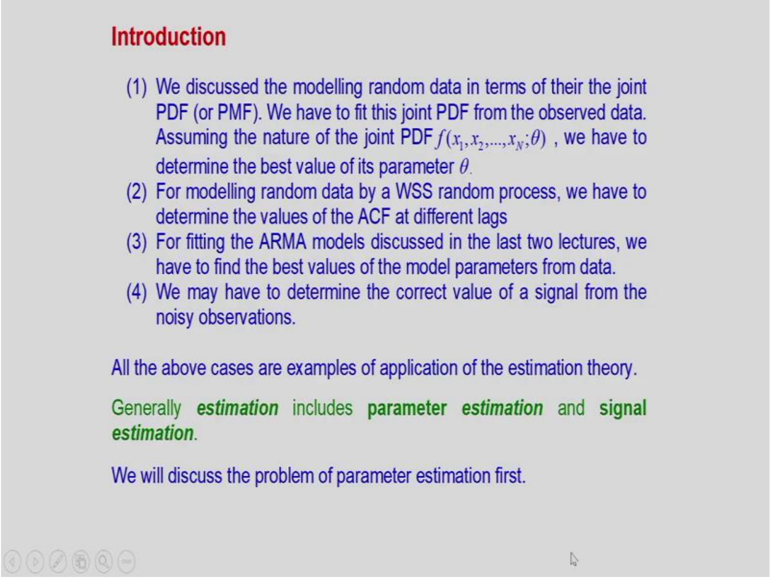
Statistical Signal Processing
Prof. Prabin Kumar Bora
Department of Electronics & Electrical Engineering
Indian Institute of Technology - Guwahati

Lecture 8
Estimation Theory 1

Hello students welcome to lecture 9 on estimation theory on 1. In this lecture, I will discuss the basics of in estimates on theory and some of the important properties of estimators. Let us start with the introduction.

(Refer Slide Time: 00:58)

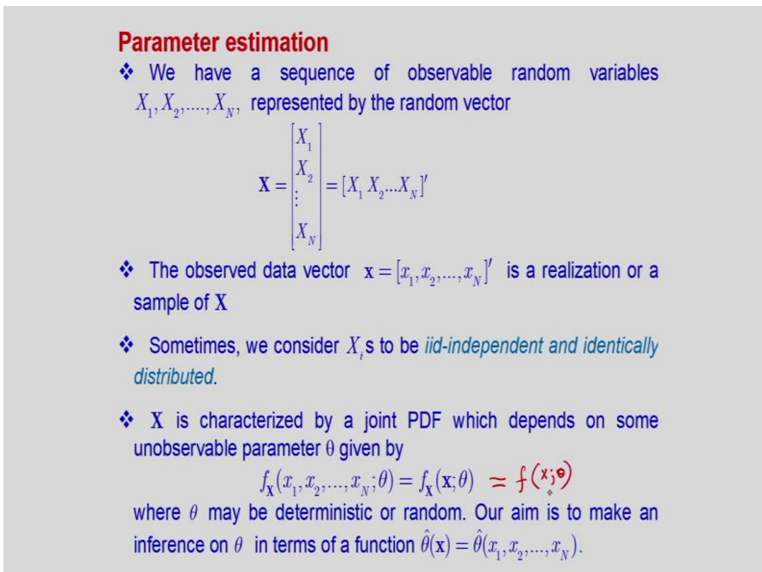


We discussed the modeling of random data in terms of their joint PDF or PMF indicates of the discrete we have to fit this joint PDF from the observed data. Assuming the nature of the joint PDF suppose we nodine f of what is the PDF and there is an unknown parameter theta, we have to determine the best value for this unknown parameter data. So, that way estimation comes for modeling random data by a WSS random process, we have to determine the value of the autocorrelation function at different lags.

Therefore, estimates are no autocorrelation function is to be done for fitting the ARMA models discussing in the last lectures, we have to find the base values of the model parameters from data. This is again estimation of the model parameters; we may have to determine the correct value of a signal from the object noisy data. So, noisy signals from the observed noisy signals. So, that this is also a case of estimation signal estimates of signal from noisy observations.

All the above cases are examples of application of the estimation theory particularly, this is an example of signal estimation and these 1, 2 and 3 are examples of parameter estimation. And we can do also signal estimation by using parameter estimation. For example, by putting the ARMA model we can estimate the signal so that way estimation includes 2 classes parameter estimation and signal estimation. We will discuss the problem of parameter estimation.

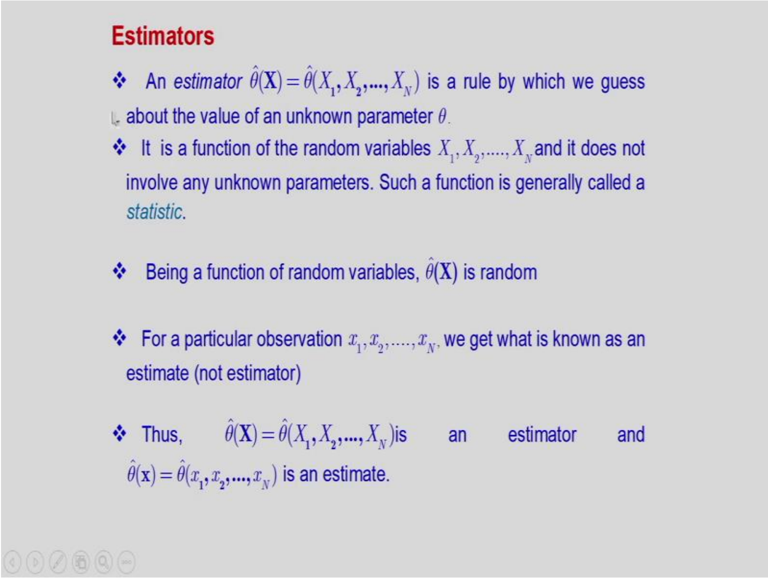
(Refer Slide Time: 03:03)



We have a sequence of observed random variables X1, X2 up to XN there are N random variables represented by a random vector this is the representation it is represented the column vector and which you can write us the row vector transpose. So, that we observed data are modulus random variable. Now, particular the observed data what we observed the observed data vector small x factor that is small x1, x2 upto small xN transpose is a realization for the samples of x because random vector.

And whatever we observed at a particular time, these constitute a realization party sample up x. So, small x is a realization big X capital X is a random vector sometimes will consider exercise to be iid that is also 1 important concept independent and identically distributed all X1, X2 and XN are independent and each of them will have the same distribution access characterized by a joint PDF whose depends on some unobservable parameter theta given by this we can write the joint PDF in terms of a parameter theta and this in the vector will write like this and will omit this x also. So, this will be simply f of x theta.

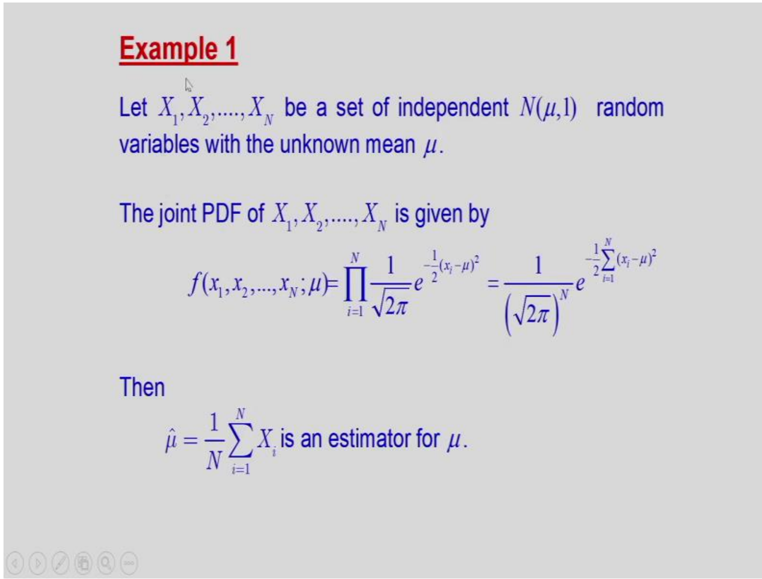
(Refer Slide Time: 04:57)



Now, what is an estimator, an estimator theta hat x = theta hat X1, X2 up to XN and is there a rule by whose we guess about the value of an unknown parameter theta. So we will guess about the unknown parameter theta using this rule, it is a function of the random variable X1, X2 up to XN, and it does not involve any unknown parameters. This functional relationship does not involve any unknown parameter. Such a function is generally called a statistic.

So, an estimator is a statistic being a function of random variable theta hat X is also random. So, that way this is a random variable, theta hat X is a random variable for a particular observation of x1, x2 upto xN, we get what is known as an estimate, not estimator of the parameter theta thus when we consider in terms of random variable theta hat x that equal to X1, X2 upto XN is an estimator. And theta hat of small x factor that is theta hat of small x1, x2 upto xN is an estimate. So, this we have to distinguished an estimator and estimate.

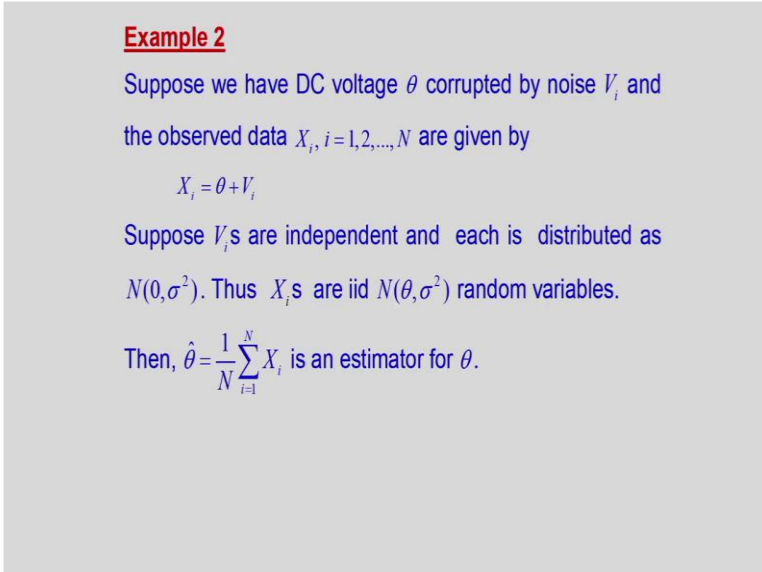
(Refer Slide Time: 06:29)



Let us give an example 1 let us X1, X2 upto XN be a setup independent normal mu 1 random variable with unknown mean mu parties normal distribution mu is unknown the joint PDF of X1, X2 upto XN is given by this is the joint pdf and that will be productive individual marginal PDFs so, that product i going from 1 to N of 1 / root of 2 pi into e to the power -1/2 of x i - mu whole squared.

So, using the property of exponential concern, we can write it in terms of a sum here. So, this is the joint pdf of data. And here unknown quantities mu and if we take mu hat = 1 / N summation X i, i is going from 1 to N is an estimator for mu. It does not involve any unknown parameter and it is a function of the random variables. So it is an estimator for mu.

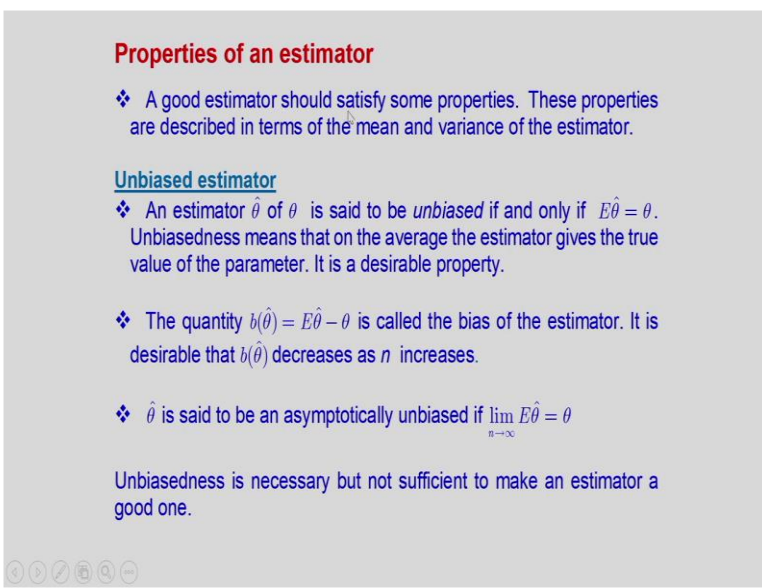
(Refer Slide Time: 07:42)



But question will arise how we can get estimator. Example 2 this is a practical example. Suppose we have a DC voltage theta corrupted by noise Vi and the observed X i, i = 1, 2 up to N are given by X i = theta + Vi. Now, we assume that Vi S are independent and each is distributed as normal 0 sigma squared, it is 0 mean sigma squared, the variance thus X i s, because X i is the sum of theta + Vi, this is the 0 mean so, X i s are iid and it is normal.

Theta sigma squared random variable so it is normal distribution with mean theta and variance sigma squared. Then again, since it is a normal distribution, this theta hat = 1 / N into summation i to 1 to N of X i is an estimator of theta.

(Refer Slide Time: 08:59)

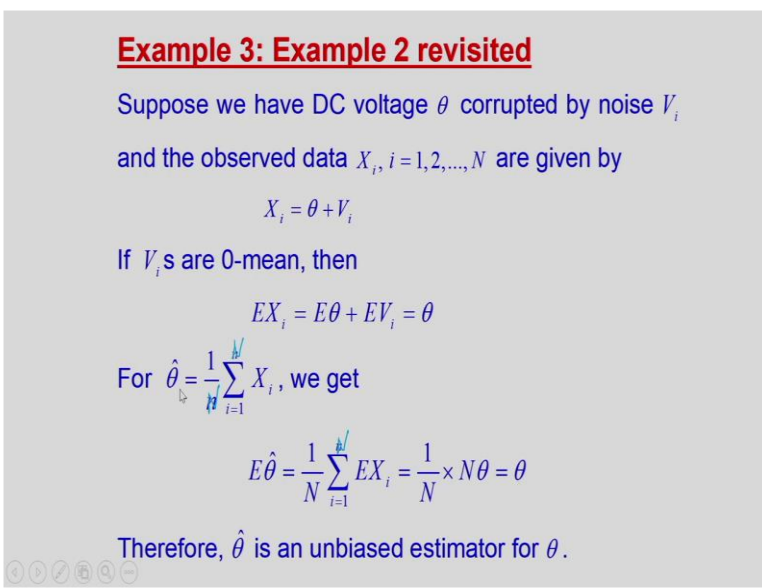


Let us discuss the desirable properties of an estimator. A good estimator should satisfy some properties, these properties are described in terms of the mean and the variance of the estimator. A part will discuss unbiased estimator and estimator theta hat of theta is said to be unbiased if and only E of theta hat = theta unbiasedness means that on the average the estimator gives the true value of the parameter.

So on the values of estimator gives the 2 value of the parameter, it is a desirable property. Now, if it is not an unbiased estimator supposed theta hat is not an unbiased estimator then b of theta hat = E of theta hat - theta is called the bias of the estimator. So, this quantity difference between expected value and the crew value is called a bias of the estimator, it is desirable that b theta hat decreases as an increases eventually go down to 0.

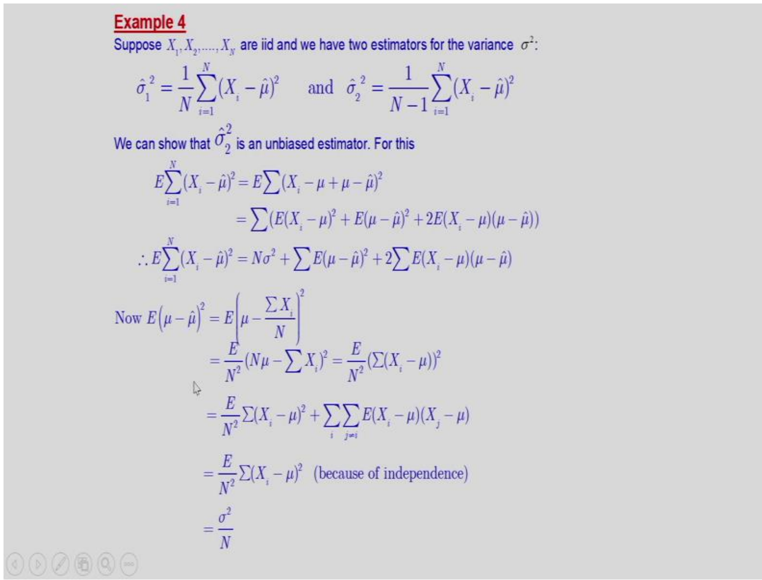
Theta hat is said to be an asymptotically unbiased estimator if limit up E theta hat into infinity is equal to theta. So, as we have large number of theta, then E theta hat will become close to theta unbiasedness is necessary, but not sufficient to make an estimator a good one we have to consider other properties.

(Refer Slide Time: 10:51)



Will give another example, example 3 example 2 revisited. Suppose we have a DC voltage theta corrupted by noise Vi and the observed data X i, i = 1, 2, upto N are given by X i = theta + Vi if Vi is 0 mean then E X i will be = E of theta + E of Vi = theta. For theta hat = 1 / N summation x i, i is going from 1 to N that is the estimator we are considering E of theta hat will be = 1 / N into summation i going from 1 to N of E of X i. Now, E of X i = theta. Therefore, this will be = 1 / N and into N theta that will = theta, therefore this theta hat is an unbiased estimator of theta.

(Refer Slide Time: 11:56)



We will consider another example this is the estimator for the variance suppose X1, X2 upto XN are iid independent and identically distributed. And we have to estimate the part sigma squared. Sigma squared is the variance, there are 2 estimators, 1 is sigma 1 hat square that = 1 / N summation i going from 1 to N of X i - mu hat whole squared. So, this is 1 estimator and 2 estimator is same as the summation, but that is = 1 / N - 1 and here 1 / N.

And so that way we have 2 estimator sigma 1 hat squared and sigma 2 hat squared. We can show that this estimator sigma 2 hat squared is an unbiased estimator, but this is not an unbiased estimator for this we will consider what is your Xi – mu hat whole squared part? Will determine that Xi – mu hat whole squared i going from 1 to N that will write in terms of mu. So, that way x i - mu + mu – mu hat whole squared.

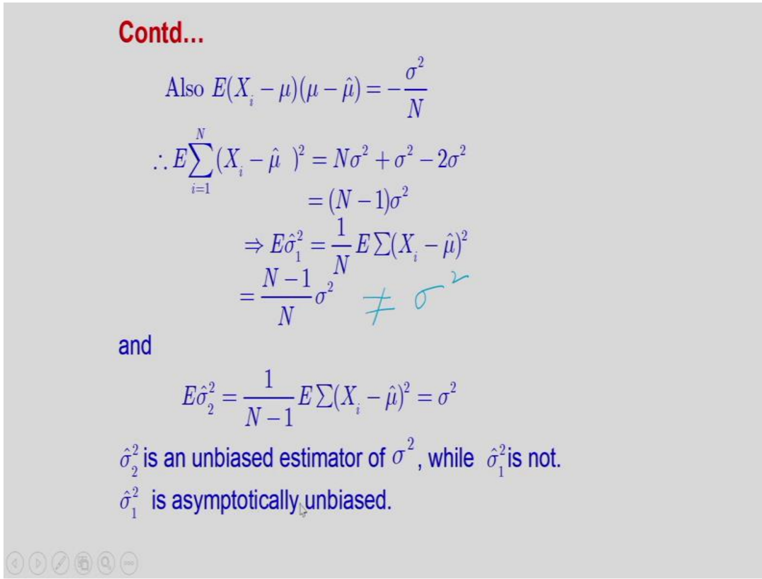
So, this quantity now we can expand first term we have X i - B whole squared, the second term will be your mu – mu hat whole square and then crossed terms will be there twice you have X i - mu into mu – mu hat. So, I know that this is iid, so, that way they are N such terms because of the summation. So, we can write this expectation the E = first term will be N sigma squared and second term will be summation up this quantity here mu – mu hat whole square and third term will be summation of this quantity of x i - mu into mu – mu hat.

So, let us see what this term will equal to so, this is your mu – mu hat whole square, this expression is important for us. So, your mu – mu hat whole squared that is billion of mu hat = your mu – summation x i / N whole square. So, if I take N, then because of this squared E of, and mu - summation Xi whole square / N square and this we can write as your summation X i - mu whole square / N square.

Now, this is again a summation that summation we can write in terms of the individual squadrons E of summation of X i - mu whole square and then all cross terms we have to consider, because these are cross terms X i - mu into X j - mu in this form. Therefore, joint expectation of this quantities will be 0, because of independent exercise and X j are independent therefore, this expression will become 0. So, therefore, this will be simply E of summation of X i - mu whole squared divided by N squared and this is equal to sigma squared by N.

And because it is identical with variance will be same for all X i and this time will be = E X i - mu whole squared will be sigma square so that way we will get sigma squared here divided by N squared it will be sigma square by N. So, this is an important observation therefore, this quantity E of mu – mu hat whole square = sigma square by N this result will be used later on also we have to remember this result.

(Refer Slide Time: 16:12)



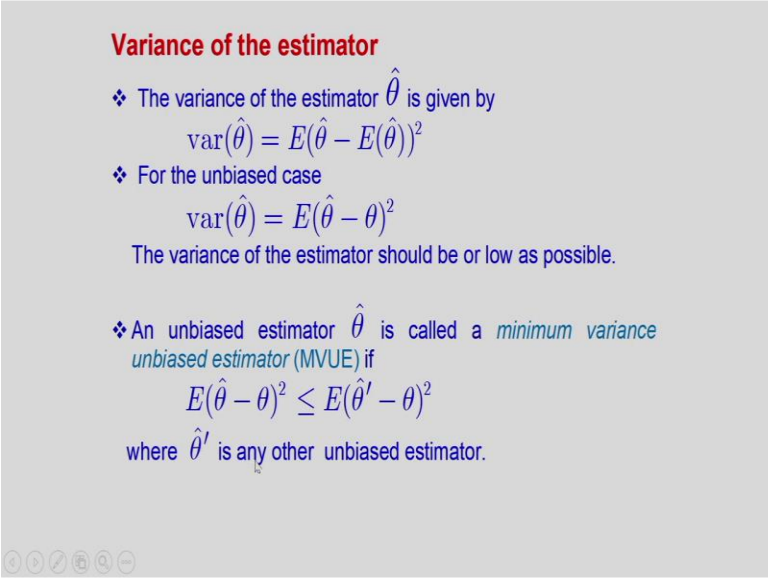
And similarly, we can do X i - mu into mu – mu hat now, mu hat to involve all the random variables, but only in the case where it is X i, then only this expectation will become non-zero otherwise this expectation will become 0. So, that way this expression will become - sigma square by N and therefore, we have to determine this expression, your E of summation X i - mu hat whole square, i going from 1 to N, this = N sigma square then this term plus this term.

So, we can write this is E of summation I going from 1 to N of X i – mu hat whole squared = N sigma square + sigma square – 2 sigma square. So, that way it will become because this is sigma squared, this is - 2 sigma squared. So, that way it will become N - 1 sigma squared. So, this will imply that now, if we have to find out that E of sigma 1 hat square that is scaling factor is 1 / N into this quantity. So, that way it will become N - 1 / N into sigma squared.

So, this is not equal to sigma squared so, if I consider this E of sigma 1 hat square that is not equal to sigma squared therefore, it is not an unbiased estimator. But if we consider sigma 2 hat squared = 1 / N - 1 into E of summation X i - mu hat whole square = sigma squared therefore sigma hat whole square that will be unbiased estimator. So sigma 2 hat squared is an unbiased estimator of sigma square while sigma 1 hat squared is not.

That is why in determining the sample variance we always divide by N - 1 to make it an unbiased estimator. But as N tends to infinity this quantity will become as N tends to infinity this quantity will become 1, therefore sigma 1 hat squared is asymptotically unbiased.

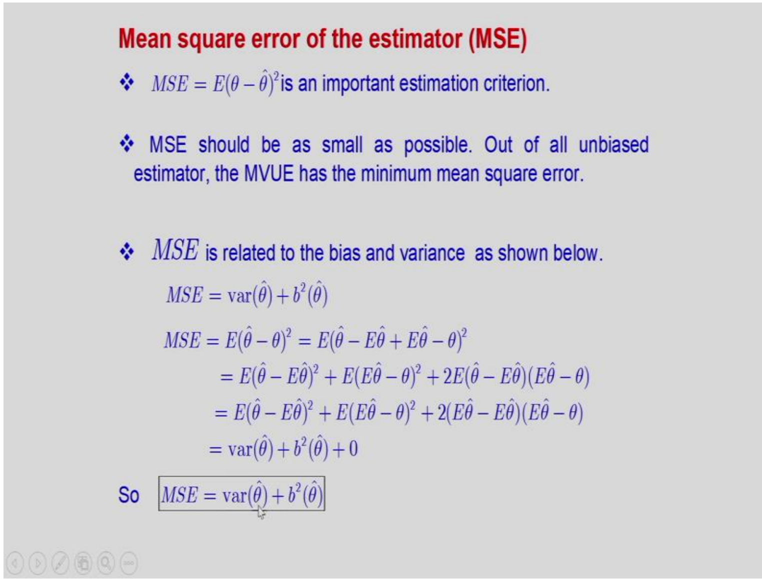
(Refer Slide Time: 18:48)



So we consider the main of the estimator will consider the variance of the estimator. The variance of the estimator theta hat is given by variance of theta hat = E of theta hat – E of theta hat whole square. That is the variance this is the random variable and it is mean. So, that way this is the variance for the unbiased case this variance of the theta hat is will be simply of theta hat – theta whole square because E of theta hat = theta.

The variance of an estimator should be as low as possible and unbiased estimator theta hat is called a minimum variance unbiased estimator MVUE if E of theta hat – theta whole squared that is variance of theta hat is less than equal to E of theta hat thus - theta whole square that is variance of theta hat thus so variance of theta is less than equal to variance of theta hat thus were theta hat thus is any other unbiased estimator.

(Refer Slide Time: 20:00)

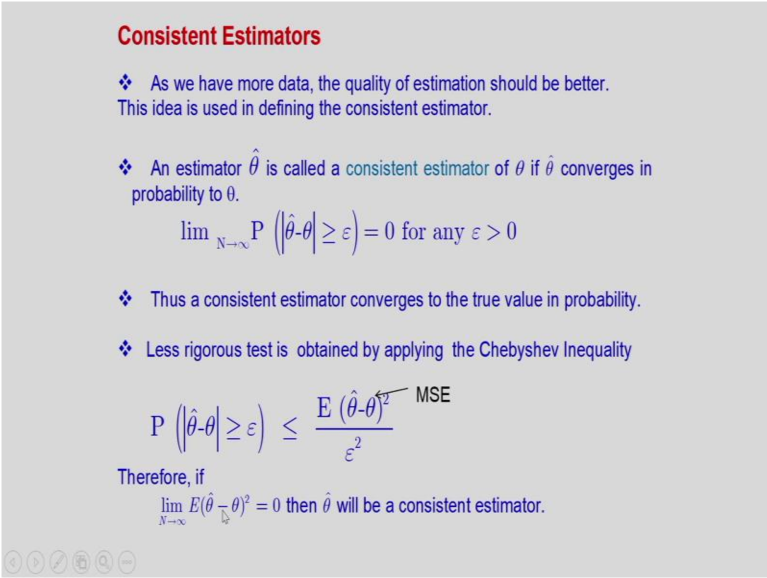


So we discussed mean variance another term mean square error MSE. MSE = theta - theta hat whole square, minimizing the MSE is an important estimation criterion. MSE should be as small as possible out of all unbiased estimators the MVUE has the minimum mean square error so in the case of unbiased estimator. MVUE will have the list MSE now, MSE is related to bias and variance as shown below MSE = variance of theta hat + bias of theta hat square.

This you can probably MSE = E of theta hat – theta whole square by definition that is equal to now we can write E of theta hat – E of theta hat + E of theta hat – theta square we have subtracting E of theta hat and adding E of theta hat. Now we expand it E of theta hat – E theta hat whole square E of E theta hat - theta whole square + twice a term 2 of theta hat - E theta hat into E of theta hat - theta. Now, this term, first term is the variance of theta hat.

And second time, because it is a constant quantity it will remain as E of theta hat – theta only. And therefore, this quantity will be biased of theta hat squared and the third term because E of theta hat - E of E of theta hat that will be same as the E of theta hat because this is a constant quantities or this term will become 0 therefore, this will become 0. So therefore what we get MSE = variance of theta hat + biased of theta hat. So this is 1 important relationship and when this quantity will become 0, bias 0, and in this case MSE = variance of theta hat.

(Refer Slide Time: 22:19)

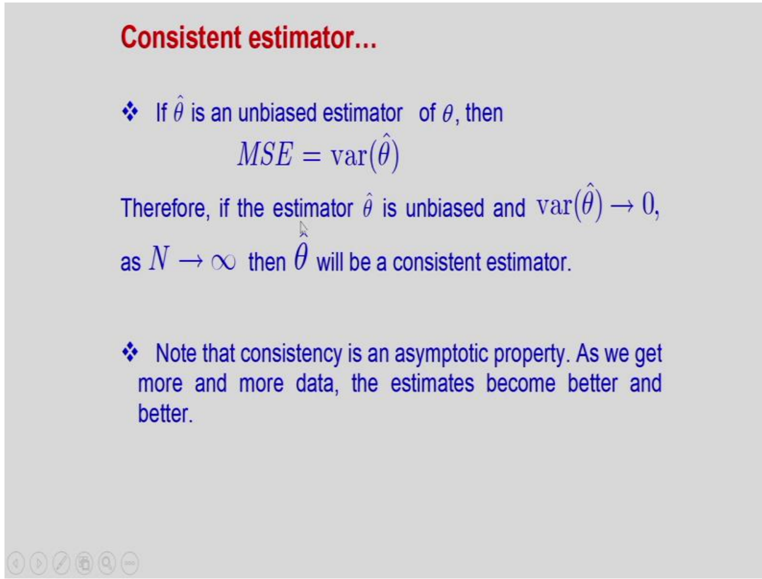


Now we will discuss consistent estimators as we have more data the quality of estimations will be better. This idea is used in defining the consistent estimator. So this estimator is a good estimator, if we have large amount of data and estimator, theta hat is called a consistent estimator if theta hat converges in probability to theta. So converges in probability that is defined in this way. So this is the limit as N tends to infinity of the probability of mod of theta mins theta greater than or equal to epsilon.

So, that means probability of the deviation as N tends to infinity that would go down to 0 for any epsilon greater than 0. So that we have defined a consistent estimator thus, a consistent estimator converges to the true value of data in probability less rigorous this because here we have to determine the probability. So they are less rigorous test is obtained by applying this Chebyshev inequality.

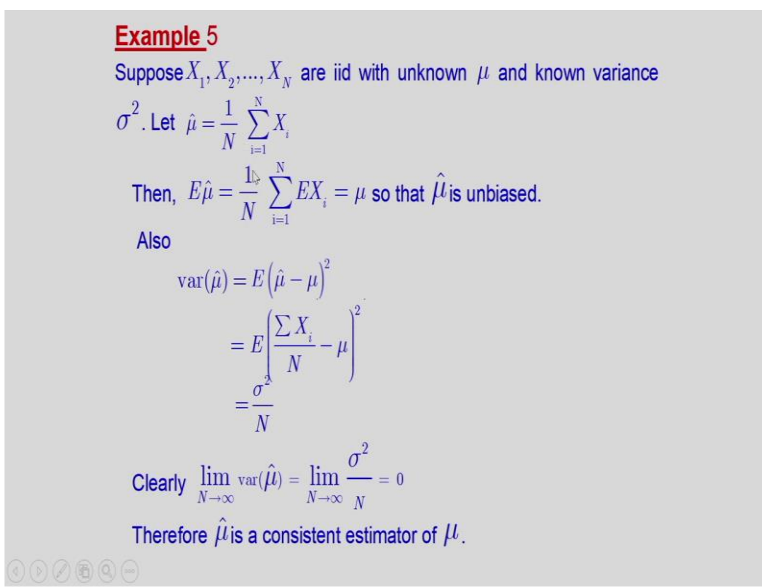
Now probability of this derivation is less than equal to E of theta hat - theta hat whole squared, that is nothing but the MSE divided by epsilon square. So therefore, if this quantity will be 0, then this will also become 0. Therefore if limit of E of theta hat - theta whole square as N tends to infinity = 0, then theta hat will be a consistent estimator. So this is the consistent estimator, this is the test for consistent estimator will be using.

(Refer Slide Time: 24:15)



Now if theta hat is unbiased in the case MSE is same as variance of theta hat does if the estimator theta hat is unbiased and variance of theta hat goes down to 0 as N tends to infinity, then theta hat will be a consistent estimator. Note that consistency is an asymptotic property. As we get more and more data the estimates become better and better that is the concept behind consistent estimator.

(Refer Slide Time: 24:48)

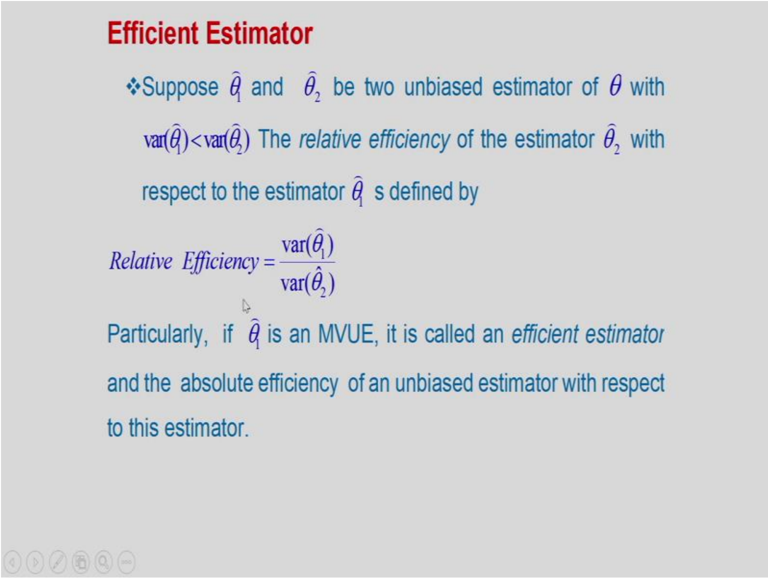


Let us consider 1 example X i upto X N are iid with unknown mu known variance sigma square and suppose mu hat = 1 / N into summation X i, i going from 1 to N. So, mu of hat will be = summation E of X i / N. So, that way it will be mu so that mu hat is unbiased and variance of mu hat is unbiased. Variance of mu hat now, this variance we have that equal to variance of mu hat =

E of mu hat - mu whole squared that we have already determined in our example 4. So, that way, this will be equal to a simply sigma square / N so, variance of mu hat = sigma squared / N because X i that iid.

So, as limit N tends to infinity, this variance will become 0. So, limit of variance of mu hat as N tends to infinity that = limit of sigma square by N, N tends to infinity that is 0 therefore, mu hat is a consistent estimator mu. So, that is sample mean is not only unbiased, but it is also consistent. That means, if we have more and more a number of data, then variants will go down to the 0 and we will get the true value of mu. So, that is the idea behind consistent estimators.

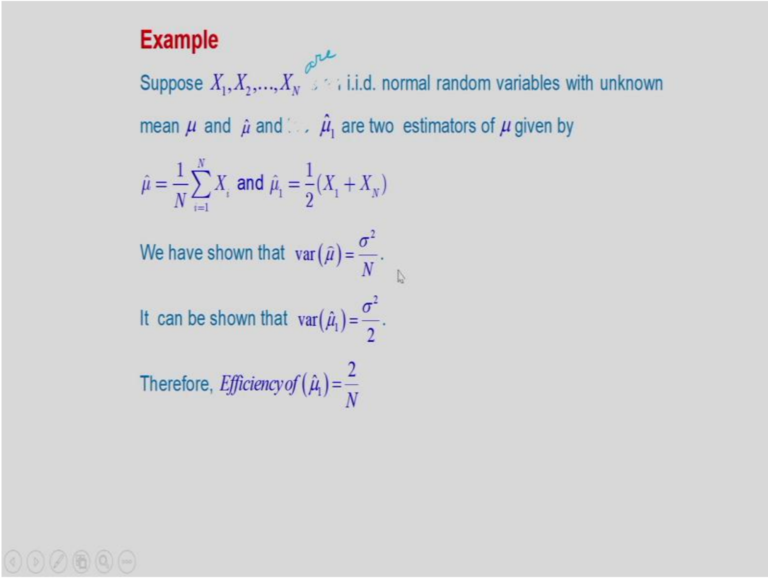
(Refer Slide Time: 26:32)



I will introduce another term efficient estimator supposed theta 1 hat and theta 2 hat are 2 unbiased estimator of theta, with variance of theta 1 hat < variance of theta 2 hat. This is unbiased estimator. The relative efficiency of estimator of theta 2 hat with respect to estimator theta hat 1 is defined by this ratio. Variance of theta 1 hat divided by variance of theta 2 hat this variance is less therefore, this number will be less than 1 particularly if theta 1 hat is an MVUE.

Then theta hat will be called an efficient estimator and the absolute efficiency of an unbiased estimator which respect to this estimator. So, this is a relative efficiency, but absolute efficiency of an unbiased estimator will be determined which respect to the MVUE. So, here and this will be the estimator with minimum variance. Then we will call this efficiency at the absolute efficiency.

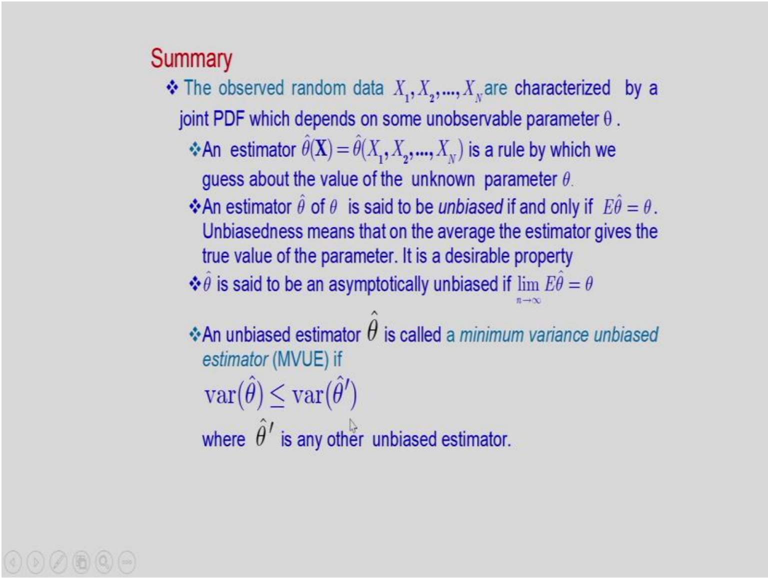
(Refer Slide Time: 27:46)



Now will give another example, suppose X1, X2 upto XN are iid random variables with unknown mu and mu hat and mu 1 hat are 2 estimators of mu is given by mu hat this is the sample mean as it was well, and mu 1 hat suppose we defined that equal to half of X1 + XN. And now both are unbiased estimator and it see that variance up mu hat, this quantity sample means is equal to sigma square / N. And we can do that by using the same formula N = 2 variance of mu 1 hat will be = sigma squared / 2.

Therefore, efficiency relative efficiency will mu 1 hat will be 2 / N if I consider of this quantity divided by this quantity, then we will get 2 / N so, this is the relative efficiency of estimator mu 1 hat which respects to mu hat.

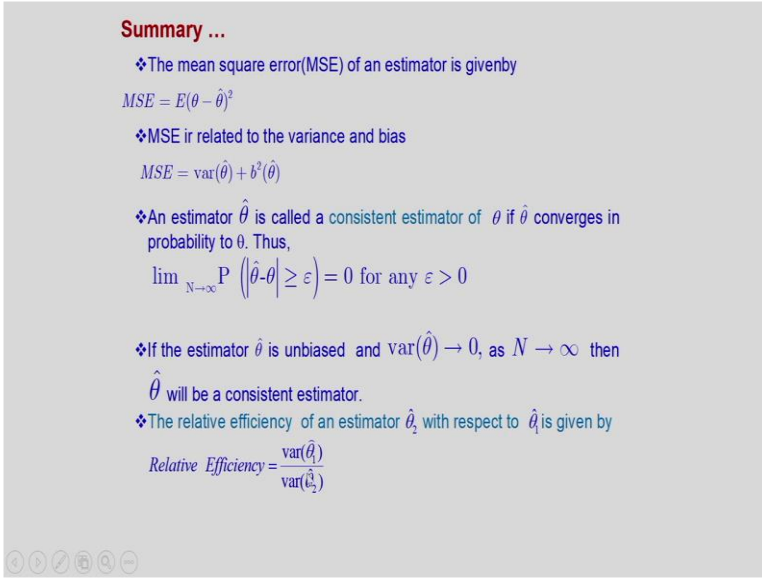
(Refer Slide Time: 28:54)



Let us summarize the observed random data X1, X2 upto XN these are the random variables are modeled as random variables and are characterized by joint PDF with depends on some unobservable parameter theta an estimator theta hat X = X1, X2 upto XN is a rule by which we guess about the value of the unknown parameter theta and estimator theta hat of theta is said to be unbiased, we introduce what is an unbiased estimator.

So unbiased if and only if E of theta hat = theta unbiasedness means that on the average the estimator gives the true value of the parameter, it is a desirable property theta hat said to be asymptotically unbiased, if limit of E theta hat as N tends to infinity = theta. Then we discussed about MVUE an unbiased estimator theta hat is called a minimum variance unbiased estimator MVUE if variance of theta hat is less than equal to variance of theta hat thus so, where theta hat thus is any other unbiased estimator. So, in the case of MVUE variances is least minimum variance unbiased estimator.

(Refer Slide Time: 30:33)



We also discuss about the mean square error is MSE of an estimator given by MSE = E of theta - theta hat whole squared. MSE is related to the variance and bias by this relationship MSE = variance of theta hat + biased of theta hat. Then we define a consistent estimator and estimator theta hat is called a consistent estimator of theta if theta hat converges in probability to theta, thus, what we have limit and N turns to infinity of probability of derivation theta hat – theta is greater than equal to epsilon = 0 for any epsilon > 0.

So for any epsilon greater than 0 if a probability of this derivation goes down to 0 as N tends to infinity, then the estimator will be consistent. If the estimator theta hat is unbiased, and variance of theta hat goes down to 0 as N tends to infinity, then theta hat will be a consistent estimator. This is a test for the consistencies of an unbiased estimator variance of theta hat go down to 0 as N tends to infinity. The relative efficiency of an estimator theta 2 hat with respect to theta 1 hat, which has lower variance is given by relative efficiency is equal the variance of theta 1 hat divided by variance of theta 2 hat. Thank You.