Statistics for Experimentalists

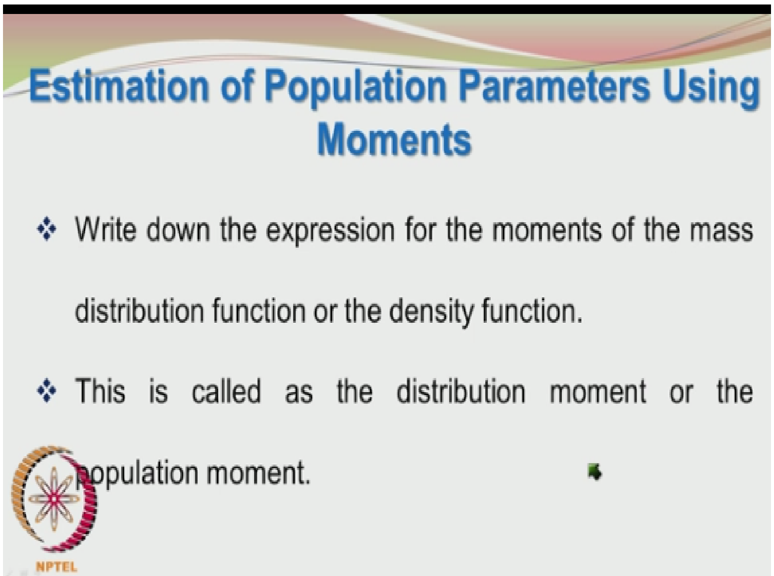
Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology – Madras

Lecture – 15

Estimation of Population Parameters Using Moments

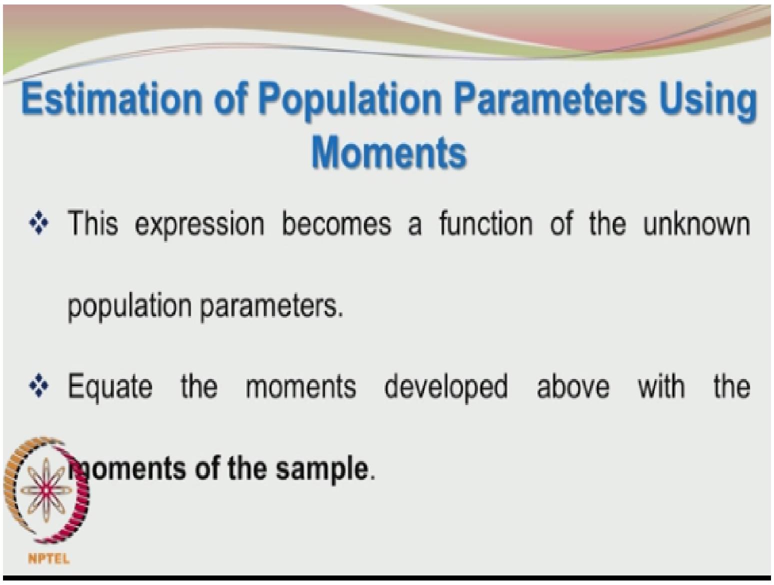
Let us estimate the population parameters using 2 different methods. The first one involves the use of moments. First, I will give a description on the method, it may sound or it may seem a bit abstract. I will demonstrate the techniques using some standard examples. So, what I request you to do is to first listen to the procedure and then, see how the parameters are being estimated using the moments method.

(Refer Slide Time: 00:15)



And then you do it yourself and see whether you get the same answer and if there are some difficulties in the middle, you can rewind and listen to the steps again. So, the first procedure is to write down the expression for the moments of the mass distribution or the density function. The mass distribution function applies to discrete random variables and the density function applies to continuous random variables.

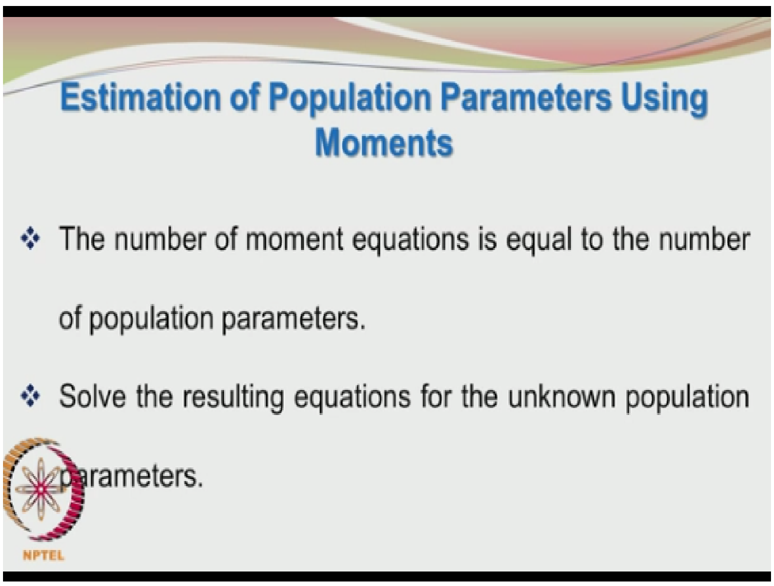
(Refer Slide Time: 02:17)



So, you have to write down the expression for the moments of the mass distribution function or the density function. So, we term it as the distribution moment or the population moment. What we are trying to do here is, first write down the expression for the population moment and then equate it with the sample moment. We have already come across the moments of the population and will be defining what is meant by the moment of the sample.

When you write down the moment corresponding to the population, it is obviously going to be a function of the unknown population parameters. So, we have unknowns on one hand. We have to relate it to the known and equate them in a suitable fashion and then estimate the parameters. What do we know? we have the sample with us. So, we equate the moments of the sample with the moments of the population and get the population parameters.

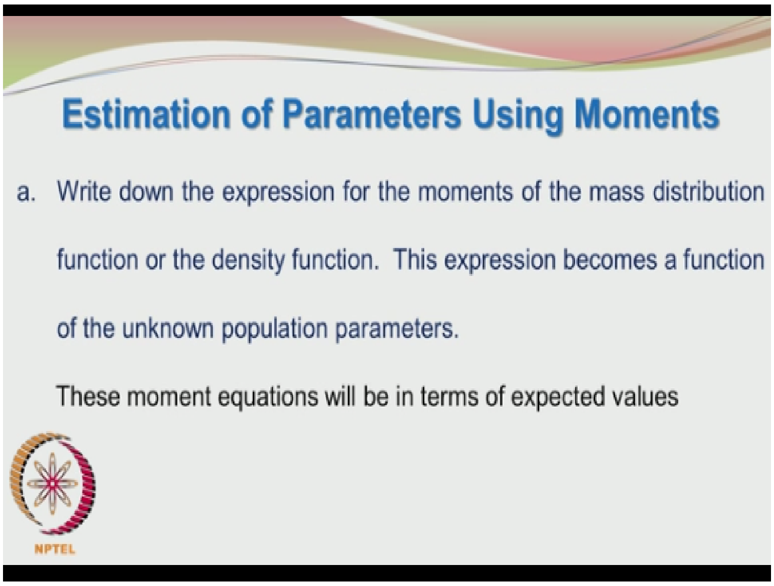
(Refer Slide Time: 03:33)



So, the concept is pretty simple. So when you write down the moments of the population, the expressions are functions of unknown population parameters. The next step is to equate the moments developed above with the moments of the sample. As of now, we are not exactly what is meant by the moment of the sample. We may not even be remembering the moments of the population.

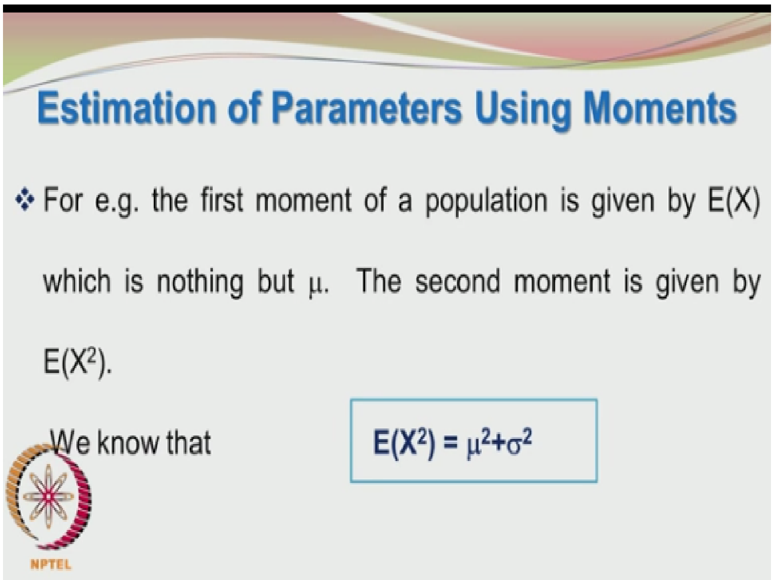
But please wait, we will be coming to them shortly. So, we have let us say 2 parameters which are to be estimated from the population, then we need to write down 2 moment equations, so that we have 2 equations and 2 unknowns which may be solved. The first step is to write down the expression for the moments of the mass distribution function or the density function.

(Refer Slide Time: 03:54)



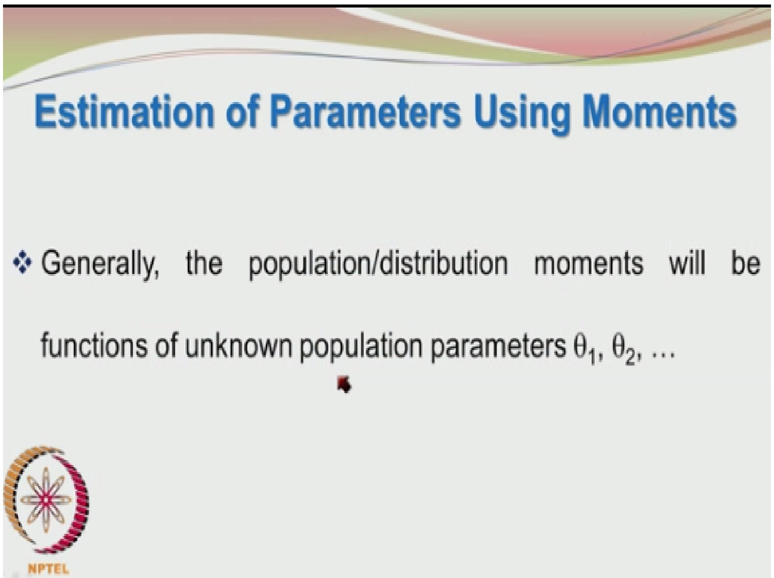
So, we know that the ordinary moments and the central moments are defined in terms of expectations, okay. For example, the mean of the population was written down as expected value of x. The standard deviation or variance sigma square was written down as expected value of x – mu whole square. So here we are talking about moments and they are defining the population mean and the population variance.

(Refer Slide Time: 04:48)



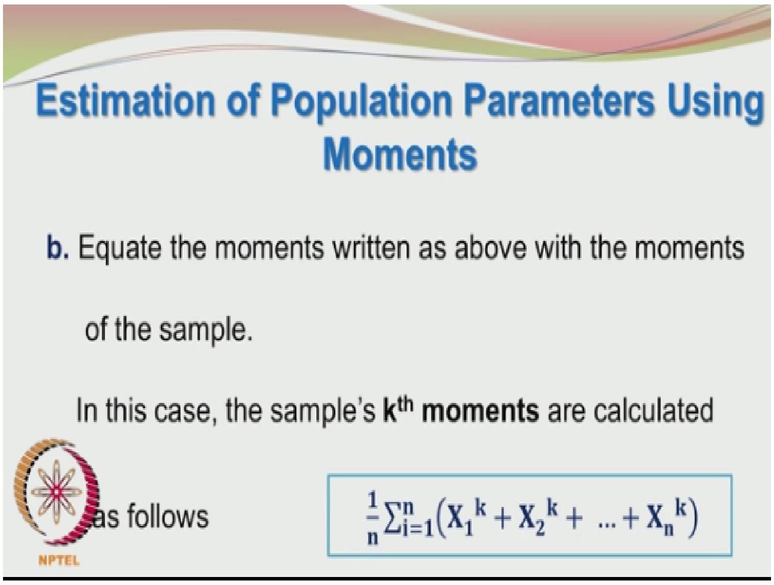
The first moment of a population is given by E of X which is nothing but mu. This is an ordinary moment which is taken about 0. The second moment about 0 would be of X square. We know that E of X – mu whole square = sigma square, but we saw in the first example set that E of X square can be written down as mu square + sigma square. If you had forgotten, you may kindly refer to the first example set to see indeed so.

(Refer Slide Time: 05:30)



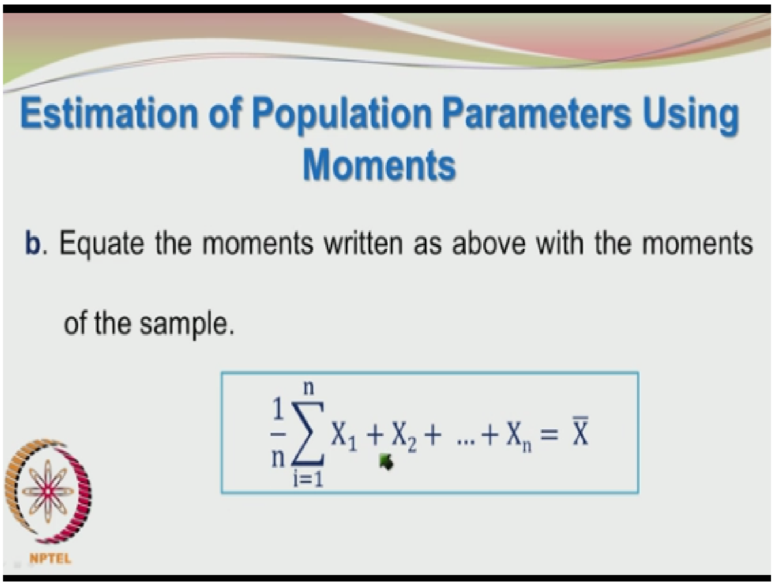
The population or the distribution moments will be functions of unknown population parameters theta1, theta2, and so on. So, we have to write down that many number of moment equations first, so that we have the same number of equations as that of the unknown population parameters. The main difficulty is sometimes the moment equations may have a combination of these parameters, okay, they may not be explicit.

(Refer Slide Time: 06:04)



As I said earlier, we have to equate the moments written as above for the population with the moments of the sample. The sample's kth moments are calculated as depending upon the k value in the moment equation. We have X1 to the power of k, X2 to the power of k, so onto Xn to the power of k. So, these are all summed over the n random variables chosen in the sample.

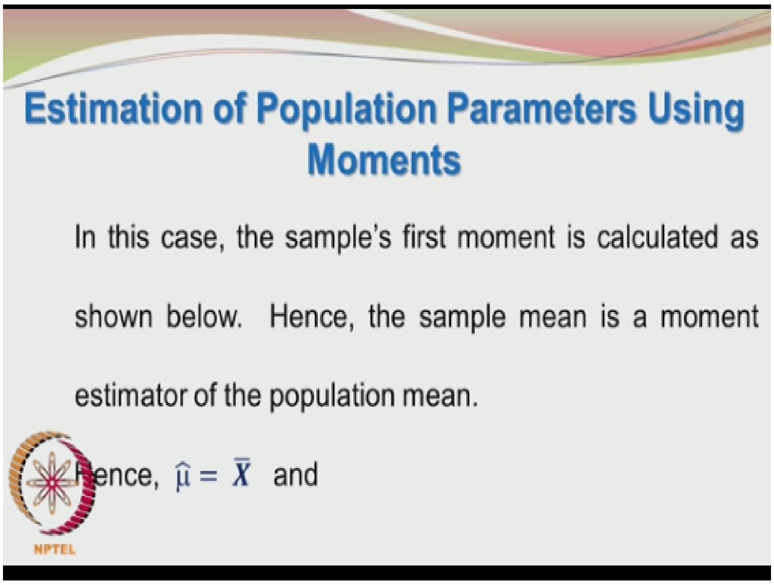
(Refer Slide Time: 06:44)



Please note that we are using the sample to find the moment. So, we can equate the moments written as above with the moments of the sample. Let us take k = 1, the first moment of the population was expected value of X which was X bar and that is related to the first moment of the sample which is given by 1/n sigma i =1 to n, X1 + X2 + so onto Xn. I will make a small correction here, the subscript has popped out, right.

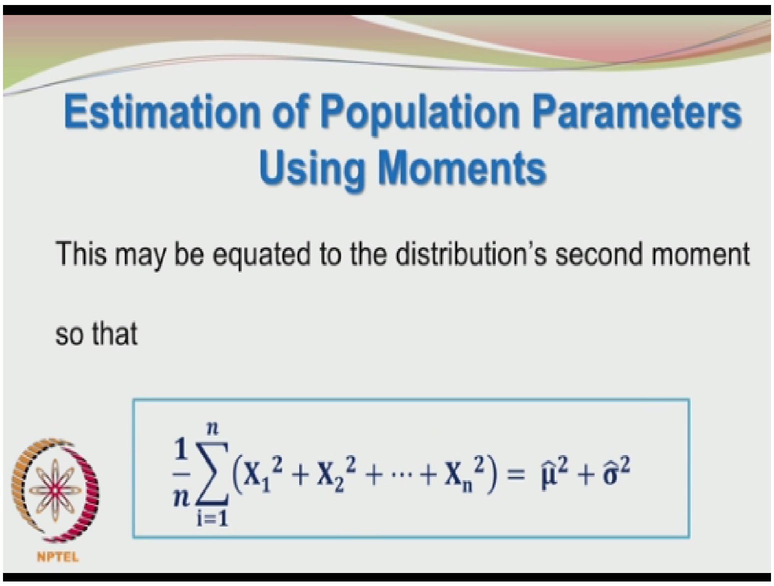
So we see that X1 to the power of 1 + X2 to the power of 1 + so onto Xn to the power 1, i = 1 to n * 1/n = X bar. So, the sample mean is a moment estimator of the population mean, very interesting. Mu hat = X bar. What we then do is, we still have not estimated the population variance sigma square, we write down the second moment. The sample's second moment is calculated as 1/n sigma i = 1 to n, X1 square + X2 square + so onto Xn square.

(Refer Slide Time: 07:39)



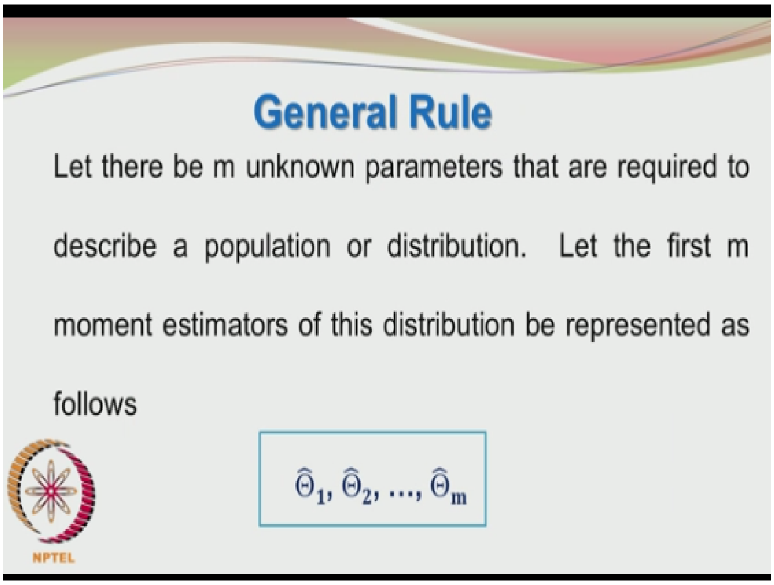
This is the sample's second moment. This may be equated to the distribution's second moment expected value of X square. The expected value of X square we saw just a moment back as mu square + sigma square. Hence, we can equate the sample's second moment with expected value of X square and express it in terms of the parameters to be estimated which is mu hat square + sigma hat square.

(Refer Slide Time: 08:26)



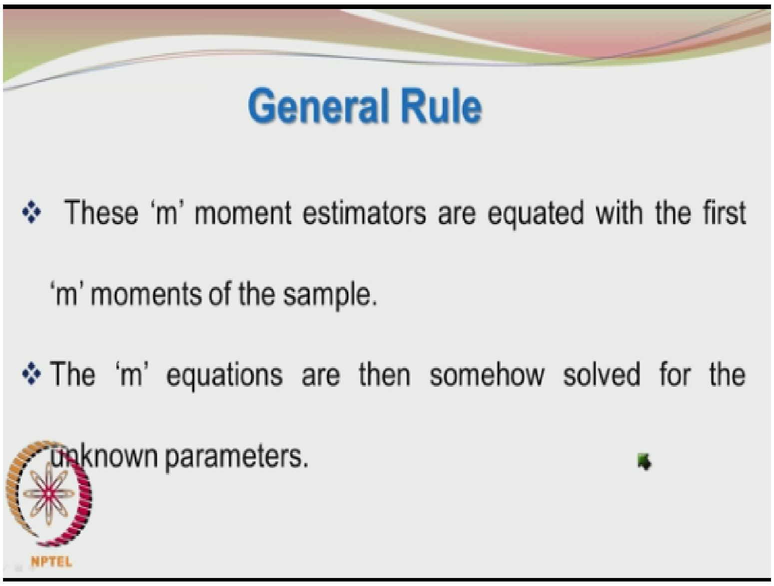
So, I have just given you the procedure. So as a general rule if there are m unknown parameters that are required to be estimated from a population, we can write down the first m moment estimators of this distribution and they are written down as theta1 hat, theta2 hat, so onto theta m hat. These 'm' moment estimators are equated with the first 'm' moments of the sample.

(Refer Slide Time: 09:03)

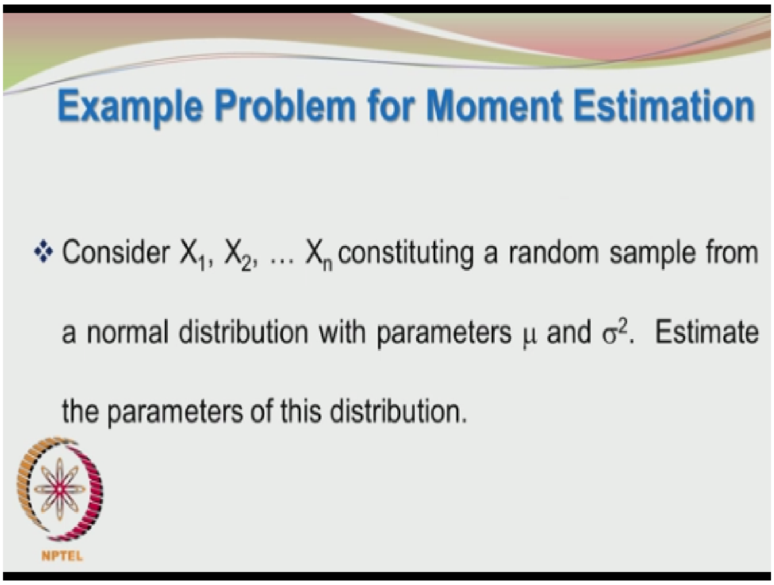


Then, we somehow solve the 'm' equations to get the unknown parameters. Let us demonstrate this with the simple example. Here, we have a random sample comprising of X1, X2, so onto Xn. The population parameters are mu and sigma square. Based on the random sample, we have to estimate the parameters mu and sigma square. We may be tempted to right immediately that the sample mean X bar = mu hat which is the estimated population mean.

(Refer Slide Time: 09:29)

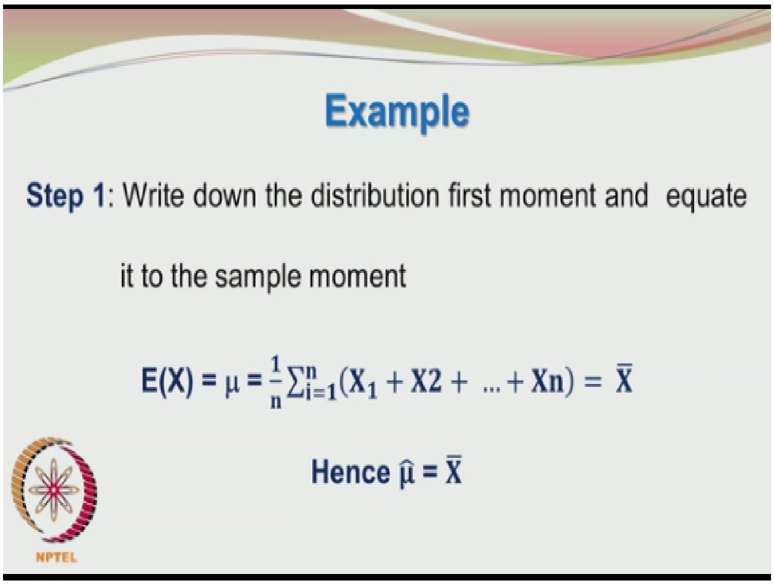


(Refer Slide Time: 09:46)



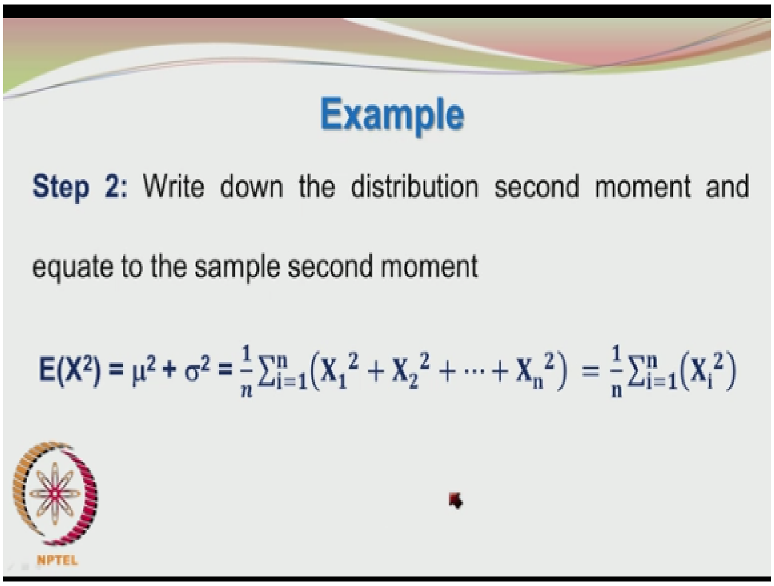
And we may also tempt to write down the sample variance S square = sigma hat square, okay. We saw that the sample mean and the sample variance are unbiased estimators of the population mean and the population variance, but we are now going to see the moment method and let us see whether the 2 assumptions we made namely mu hat = X bar and sigma hat square = S square, whether these 2 assumptions are indeed correct.

(Refer Slide Time: 11:20)



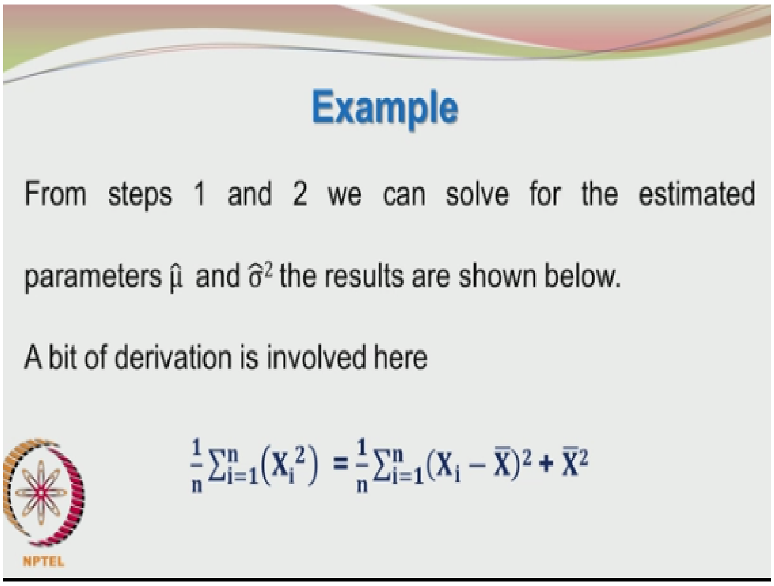
Better not to make any assumption beforehand without proper verification. So, the expected value of X, the random variable X will be given by X1 + X2 + so onto Xn divided by n. I will just put the subscripts back in its place. So, expected value of X = mu which is = sigma Xi/n and that was equated to the first population moment X bar and we indeed have mu hat = X bar.

(Refer Slide Time: 11:54)



The estimated population mean = sample mean. Now, we write down these second moment of the population E of X square and that = mu square + sigma square. Here, we have 1/n sigma i equals 1 to n, X1 square + X2 square + so onto Xn square. Here, we are just putting k = 2 because we are dealing with the second moment. That is expressed concisely as 1/n * sigma i = 1 to n Xi square. Now, we write down 1/n sigma i = 1 to n Xi square in the following way.

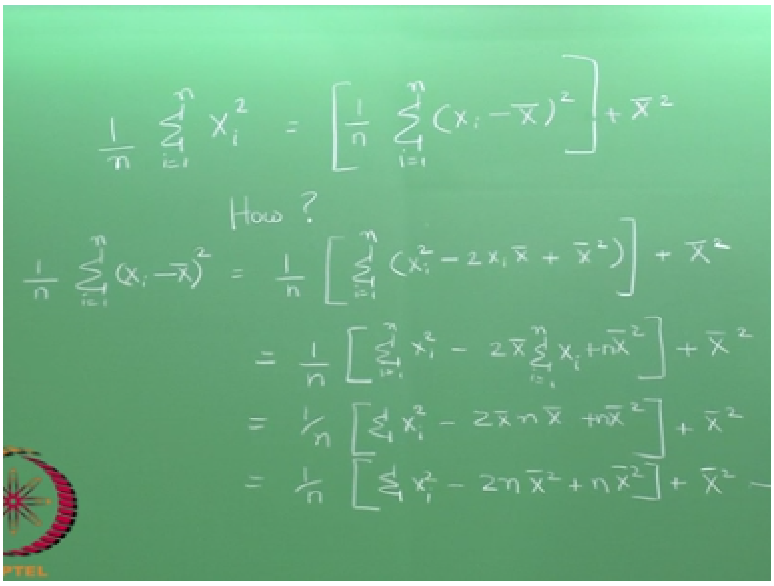
(Refer Slide Time: 12:27)



Write down 1/n sigma i = 1 to n Xi square in the following way. We are doing some mathematical jugglery to get to the final answer, okay. So, this can be written as 1/n sigma i = 1 to n, Xi - X bar square + X bar square. So just verify this, it is not difficult, the summation applies only for the term Xi – X bar square. So, essentially we have written 1/n sigma i = 1 to n Xi square in terms of this quantity + X bar square, okay.

The proof is pretty straight forward, so we will be looking at this rather simple derivation. I hope you were also interested or curious enough to work it out by yourself. We saw standard derivations commonly encountered in this field of analysis. So, what we are trying to see here is 1/n, i equals 1 to n Xi square may be written as the sum of these 2 terms how is it possible.

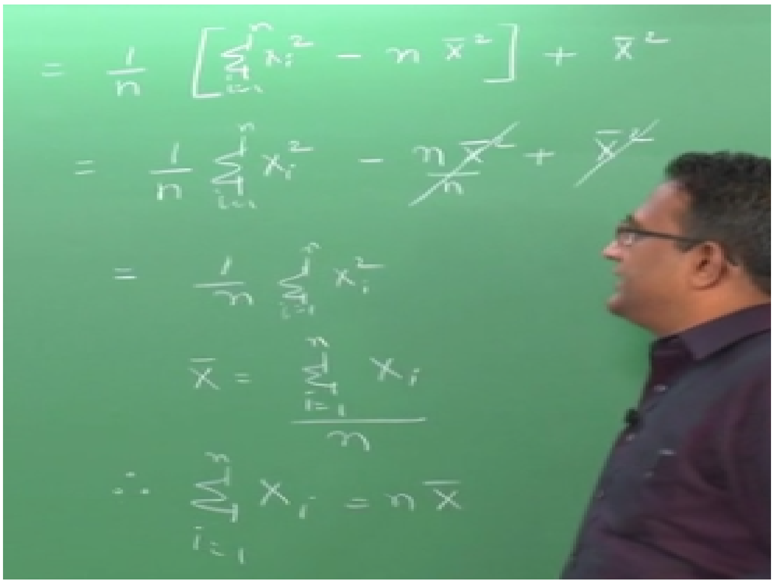
(Refer Slide Time: 13:31)



What we do is we expanded, we take 1/n outside sigma i = 1 to n, Xi square – 2Xi X bar + X bar square + X square, which is outside the bracket. Here, this can be written as again 1/n, we take the summation now inside sigma Xi square – this 2 and X bar are constants. So, they may be taken outside the summation sign, so you have 2X bar, sigma i = 1 to n, Xi + nX bar square.

So, you get 1/n * sigma Xi square – 2X bar * nX bar because we know that X bar = sigma i = 1 to n Xi/n. Therefore, sigma Xi, i = 1 to n is nX bar. So, this is written as nX bar. So, you have -2X bar * nX bar and that becomes -2nX bar square. Here, you are summing the X bar square term n times. So, this becomes nX bar square. So, this becomes -2n X bar square + nX bar square.

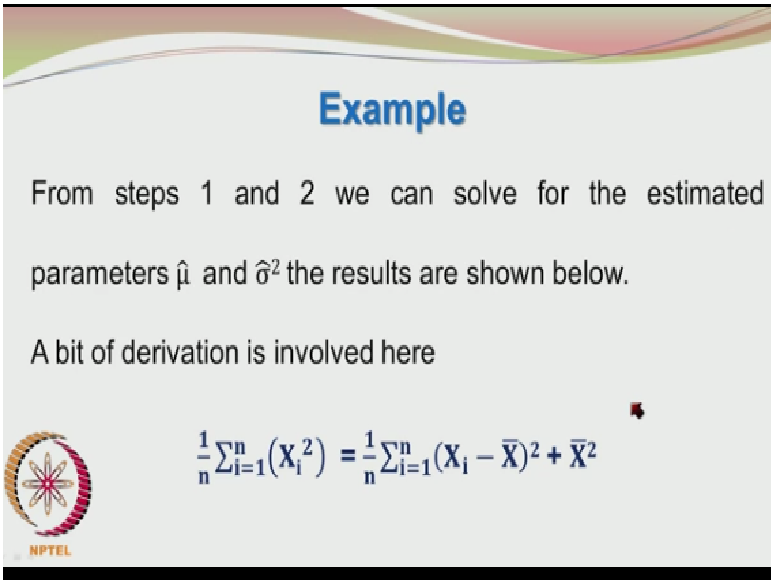
(Refer Slide Time: 15:01)



Again, you have X bar square here. I have drop the in this is of the summation, i = 1 to n, i = 1 to n. that may be added, okay. So, crossing the t's and dotting the i's, we have added the, in this is we have 1/n sigma i = 1 to n Xi square – nX bar square + X bar square and so this becomes X bar square. There is also +X bar square, -X bar square and +X bar square will cancel out and you are left with 1/n, i = 1 to n, Xi square.

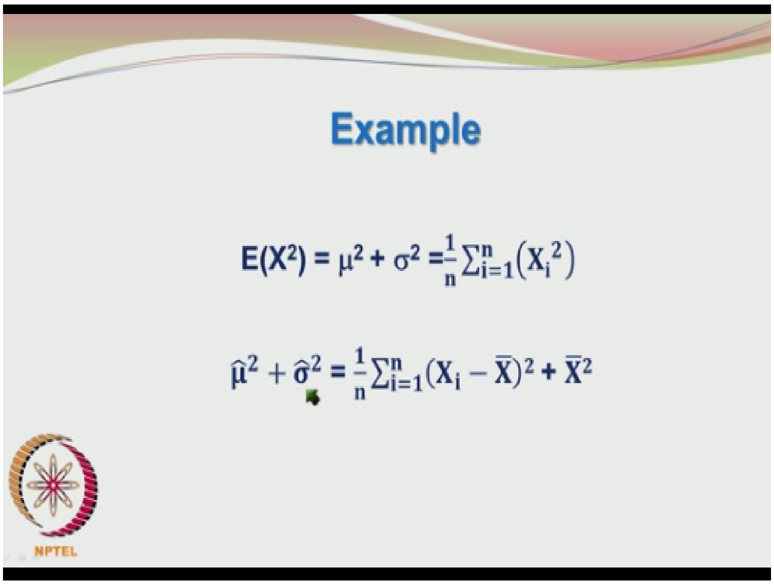
So, this is where we started and this is where we have ended. The point is I am saying that this is equivalent to this particular expression. So, this background we know also that E of X square = mu square + sigma square and i = 1 to n which is the second sample moment that = 1/n sigma i = 1 to n, Xi – X bar whole square + X bar square. But, we recently found out that mu hat square = X bar square.

(Refer Slide Time: 16:50)



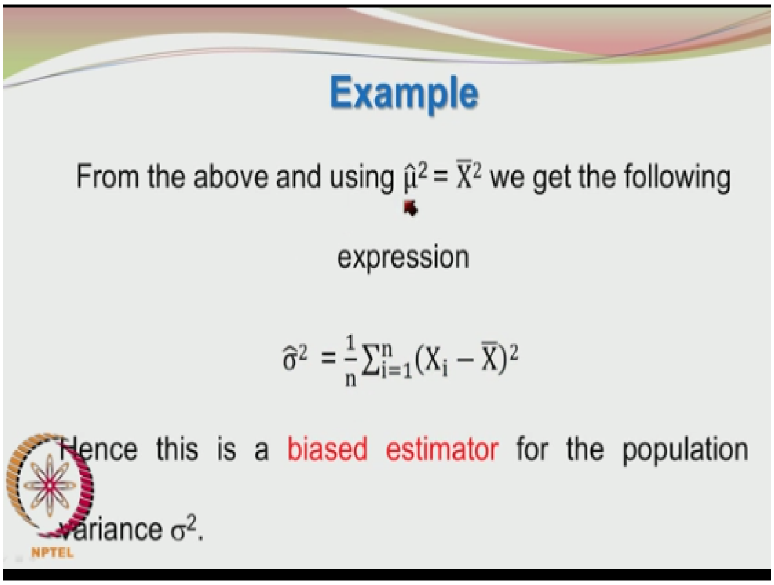
So, this X bar square will cancel out with this mu hat square leaving sigma hat square to be = to this. So mu hat square + sigma square = 1/n sigma i = 1 to n, Xi – X bar whole square + X bar square and we know that mu hat square is nothing but X bar square that is from our first sample moment, equated to the population first moment result. So once this cancels out, we are left with sigma hat square = 1/n sigma i = 1 to n, Xi – X bar whole square.

(Refer Slide Time: 17:29)



A very interesting result some of the square of the deviations from the sample mean divided by n, but we know by now that it should not be n, but rather than n – 1, if we were to take the sample variance, but we are using n here from the method of moments. So from the above, I am using mu hat square = X bar square. We get sigma hat square = 1/n sigma i = 1 to n, Xi – X bar whole square.

(Refer Slide Time: 18:24)

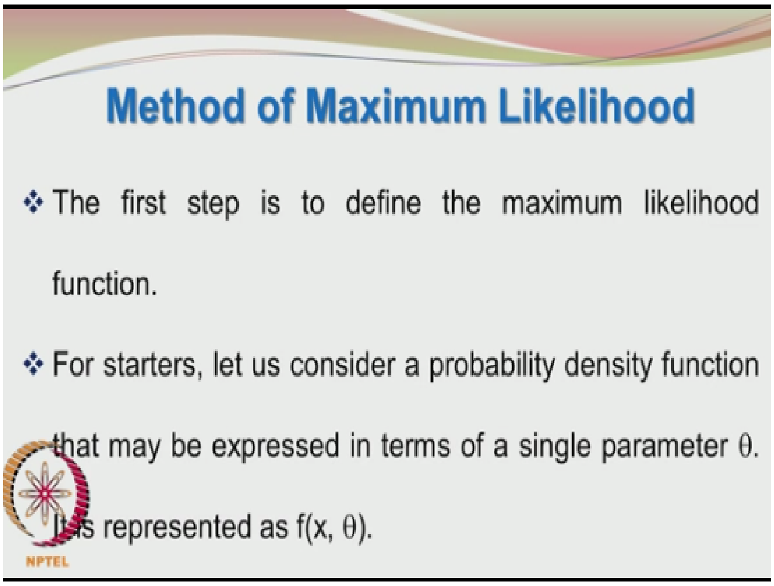


This becomes a biased estimator for the population variance sigma square. If we had used 1/n – 1 sigma i equals 1 to n Xi – X bar whole square, then we would have got the unbiased estimator, 1/n – 1 sigma i = 1 to n Xi – X bar whole square = S square which is the sample variance and the sample variance based on n-1 in the denominator represents the unbiased estimator of the population variance sigma square.

On the other hand, the method of moments led us to the expression 1/n sigma i equals 1 to n Xi – X bar whole square as the estimate of the population variance, as the estimator of the population variance. Obviously, this becomes then a biased estimator, but even though it is a biased estimator, if you take a sufficiently large sample size, the difference between n and n-1 will becomes small and so, we do not really have to worry about the biased in the estimator.

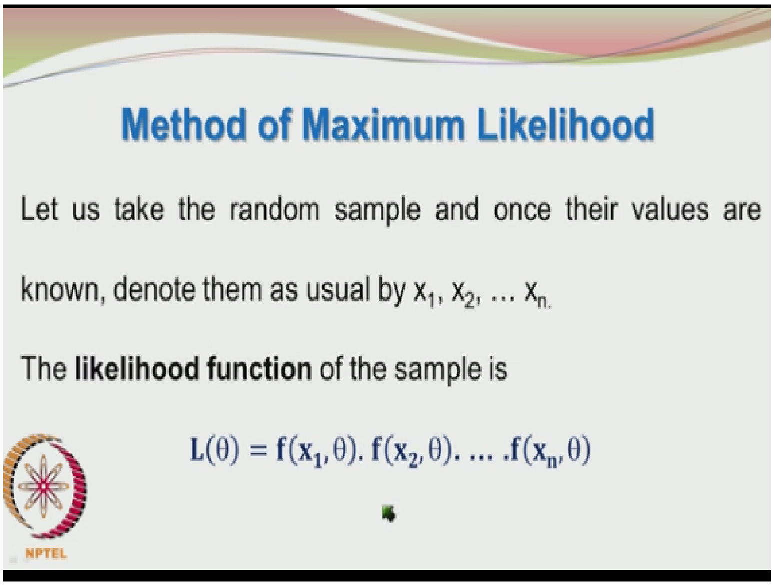
So, again we see the merits of having a large sample size. Now, we will go to the next technique, the method of maximum likelihood. You may feel that I have left the method of moments a bit to abruptly, but will be shortly doing an example set where both the method of moments and the method of maximum likelihood will be demonstrated using suitable. The first step is to define the maximum likelihood function.

(Refer Slide Time: 19:52)



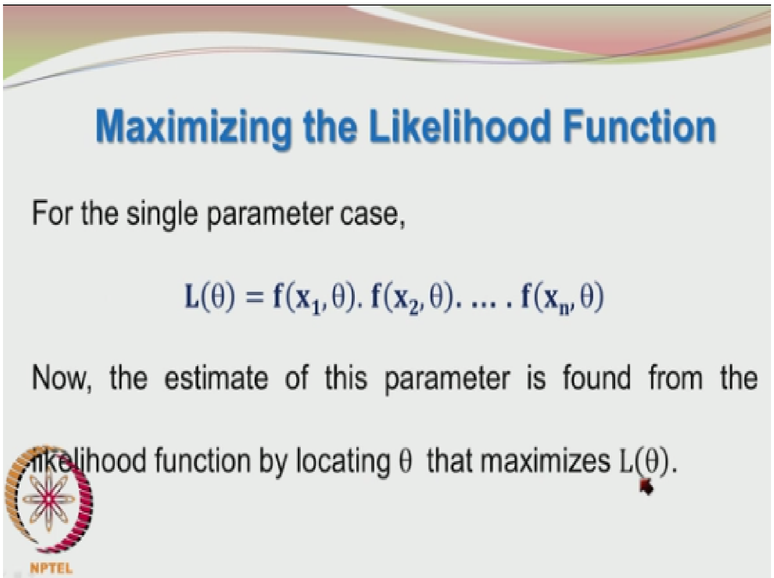
Let us do this with single parameter, okay. Assume that even if there are 2 parameters in the population, the first parameter is unknown and the second parameter is known. It is a fictitious case, but this is mainly meant for demonstration purposes. Then, we will take the more general case involving 2 unknown parameters. So, let us represent the probability density function in terms of the variables theta and X.

(Refer Slide Time: 21:08)



F of X, theta as the probability density function. So, we will take a random sample and once their values are known, the moment you have taken a random sample, you are going to do the measurements, okay that is the purpose of taking the random samples. A sample is available to you and you are going to take the height, weight or their marks in a particular subject or if it is a specimen from a industrial production unit, you may be subjecting the specimen you have drawn a sample to certain test compressive strength and strain limit and things like that.

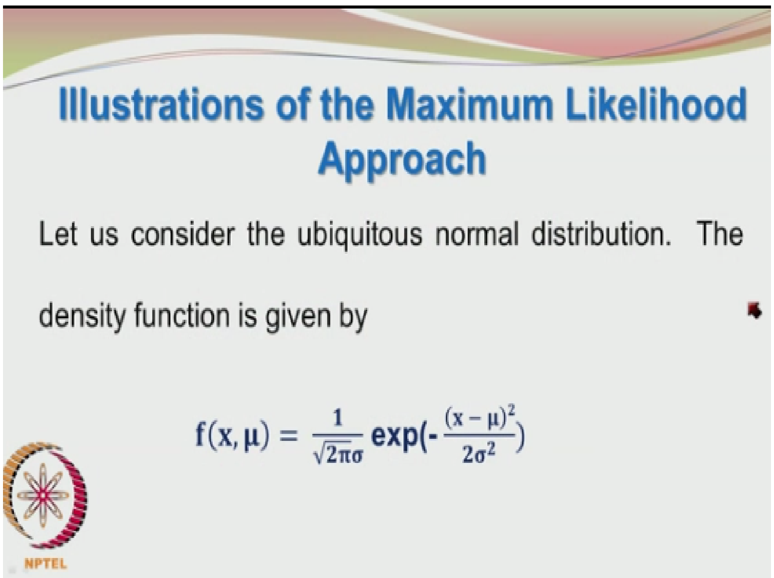
(Refer Slide Time: 23:28)



So, you are going to denote them as small x1, x2, to xn. The small x values denote the values taken by random variable x1, random variable x2, so onto random variable xn. The random variable is denoted by capital X and the value taken by the random variable is denoted by small x.

Now, we define the likelihood function of the sample to be L a function of theta, where theta is the single unknown parameter as the product of f of x1, theta * f of x2, theta, so onto f of xn, theta. So, what we are doing here is in the probability density function, we are plugging in the random sample value. Let it be not a number, but we will put it in a more general case, okay.

(Refer Slide Time: 25:14)



Even though x1, x2, and so on have up to xn have taken values, we are putting it in a general case and finally we can substitute the values there, okay. So, as of now let it be generally x1, x2, so onto xn. So, likelihood function L of theta = f of x1, theta, f of x2, theta and so onto f of xn, theta, okay. Do not be in a hurry to plug in the actual values of the random samples here.

Now, this is a function and as the name maximum likelihood implies what we are trying to do here is to maximize this particular function, okay. So, we have f of x1, theta * f of x2, theta so onto f of xn, theta that function will be differentiated with respect to the unknown parameter theta, okay. We want to find the parameter which will maximize this likelihood function.

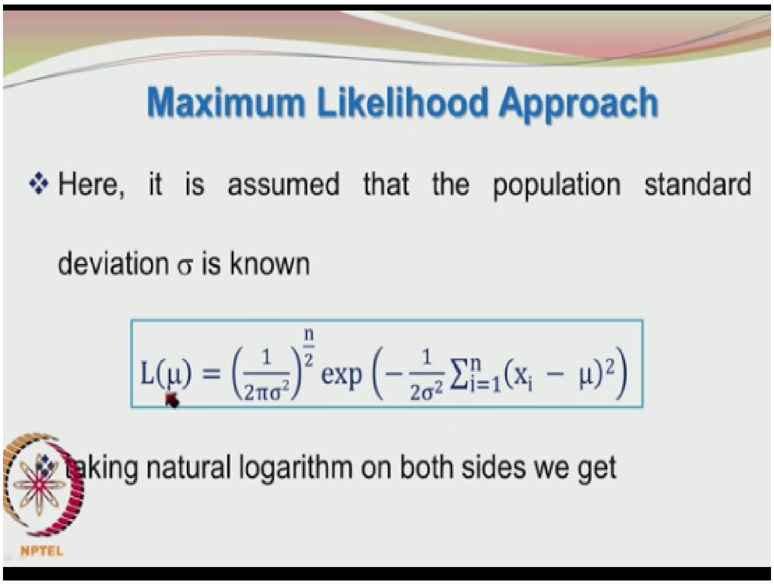
So, we are not going to differentiate with respect to x, which x you will use x1, x2, or so onto xn. No, you are going to maximize the function with respect to parameter theta and so, you have to differentiate with respect to the parameter theta. Many of us are very used to maximizing a function or minimizing a function by differentiating that function with respect to x.

So, we may be tempted to do the same thing here, but we have to actually differentiate with respect to the unknown parameter or parameters theta, okay. There can be more than one parameter and when you are having more than one parameter, you have to partially differentiated the likelihood function with respect to each of the unknown parameters, but let us not be in a hurry, we will come to that a bit later.

First let us take the simple case involving a single parameter and we will differentiate the likelihood function with respect to this parameter. So, the density function f of x, mu equals 1/root 2 pi sigma exponential – x – mu whole square/2 sigma square represents the very commonly, very frequently encountered normal distribution, okay. Now, we are putting mu because we assume sigma square to be known and we take only mu to be the unknown parameter which is to be estimated.

So, this is for demonstration purposes. In several classes from now on, we will also be assuming that sigma square is somehow known to us. This is a kind of an artificial construct because mu and sigma square are both unknown. Sigma square represents the spread, spread about what? Spread about mu. So, if you do not know mu, how will you find out the spread? Anyway for the time being, we will assume that the sigma square is known to us.

(Refer Slide Time: 26:31)

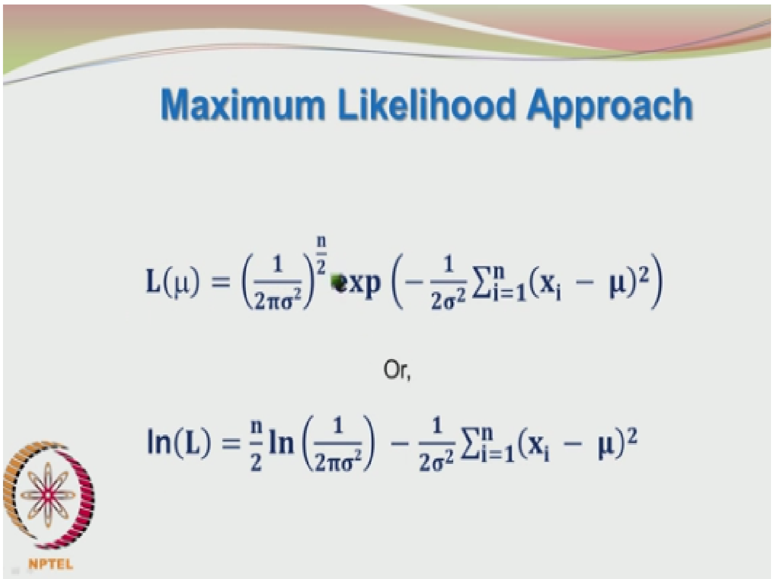


And mu is unknown for the purpose of demonstration. Right now what we can do is, we cn define the maximum likelihood function in terms of mu. What we do is, we have taken a random sample of size n and we plug in x1 here for the first random variable value, then we plug in x2, we plug in x3, so onto xn. So, we will be having n such functions. When you multiply all these functions f x1, mu * f x2, mu * so onto f of xn, mu.

So, this can be represented compactly by 1/2 pi sigma square to the power of n/2, this 2 came because of the square root, we are writing square root of 2 pi sigma as square root of 2 pi sigma square and when it is multiplied n times, we get 1/2 pi sigma square to the power of n/2, the exponential term is again very interesting, 1/2 sigma square is common. When we are multiplying exponential terms, the argument gets added up.

And so here, we are going to have i = 1 to n Xi – mu whole square. So, we have to differentiate this function directly with respect to mu because mu is the unknown parameter or to make life easier for us, we can take the natural logarithm of this particular equation. We take the natural log on both sides. So, we get L of mu = 1/2 pi sigma square to the power of n/2 exponential -1/2 sigma square sigma i = 1 to n Xi – mu whole square.

(Refer Slide Time: 28:14)

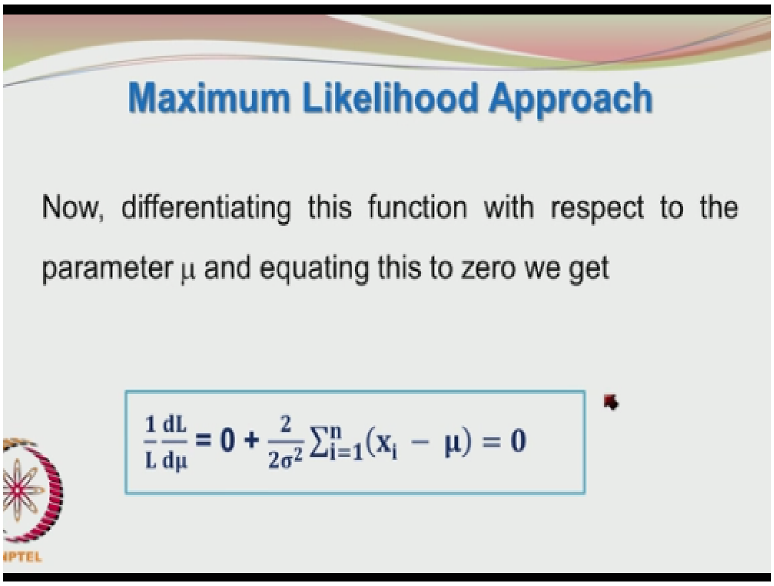


This is the likelihood function. After you have taken the natural log, you will get n/2 * ln of 1/2 pi sigma square and then you will have ln of E power – this term, ln of E power any term is = that term itself, ln E power p = p ln E, which = p. So, that would be equivalent to or rather = -1/2 sigma square sigma i = 1 to n Xi – mu whole square, okay. So, this is where we are getting the logarithm of the maximum likelihood function.

Now, we have to differentiate with respect to the unknown parameter mu and equate it to 0. To indeed, see whether the solution we find leads to the maximum value of L, we have to take the second derivative of the maximum likelihood function. We have to indeed verify. We have to verify that the root we obtained by solving this equation leads to a maximum value of L.

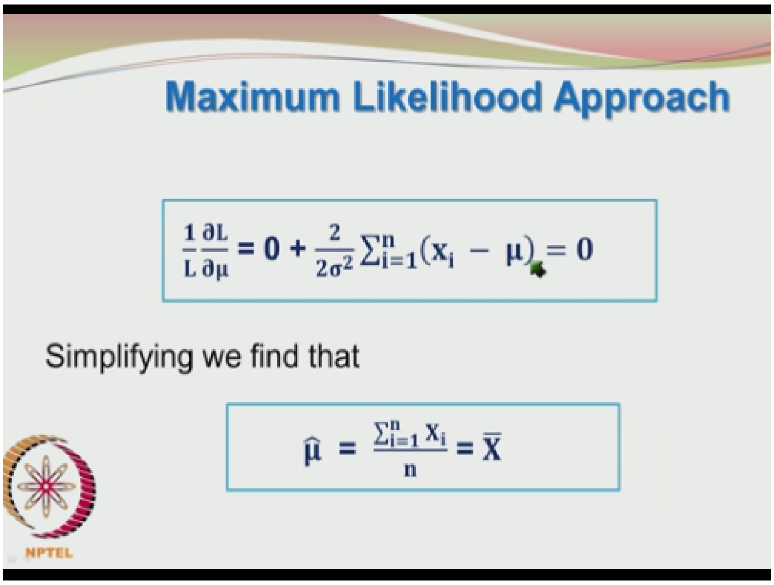
For that we know from calculus that the second derivative should be negative, okay. But, we will not be doing that. We will leave that as an exercise and we will take only the first derivative. Since, we are having a function which is depending on a single parameter instead of writing dou L/dou mu, you should actually write it as dL/d mu. So, I will just make that correction here.

(Refer Slide Time: 29:11)



So instead of dou L/dou mu, we are writing it as dL/d mu because there is only one parameter and the differentiation of the constant will become 0 and then, when you differentiate this with respect to mu, you will get, you are having a negative coefficient to mu, so you will get 2xi – mu that – will cancel out with this -, the 2 will cancel out with this 2 and then the differentiation of mu will lead to 1.

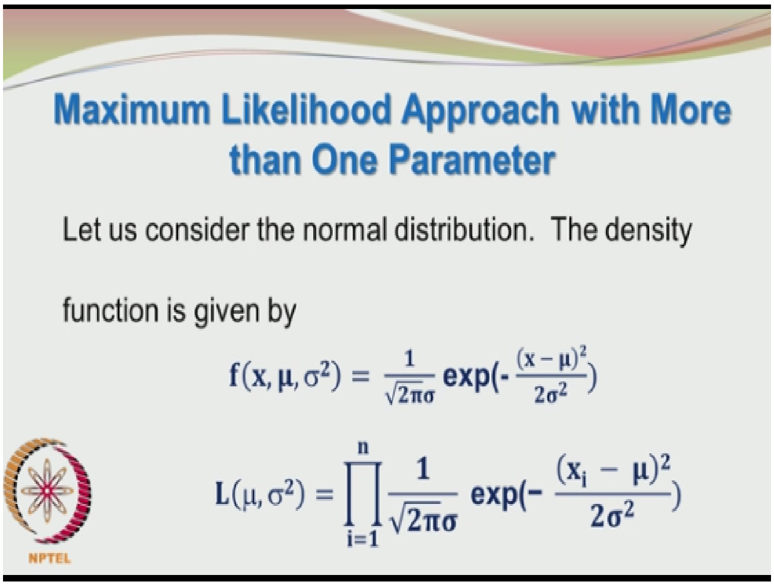
(Refer Slide Time: 31:25)



So, finally you will get the 2/2, the – has become + because you are having – of – and then sigma i = 1 to n Xi – mu = 0. Sigma square is a constant, the 2 will cancel out. You can take sigma square out because it is a constant and you are left with sigma i = 1 to n Xi – mu = 0. So, you have to essentially solve for mu. When you indeed do that mu, when it is summed n times will become n mu and this will become sigma Xi.

So, mu will be nothing but sigma Xi divided by n and that becomes the sample mean which we generalize as mu hat = X bar. So, the sample mean is an estimator of the population parameter mu hat. Now, let us look at the population described by the normal curve and we are now being general by saying that neither mu nor sigma square are known to us. Here, we stop with the first sample moment because we had only one parameter to estimate.

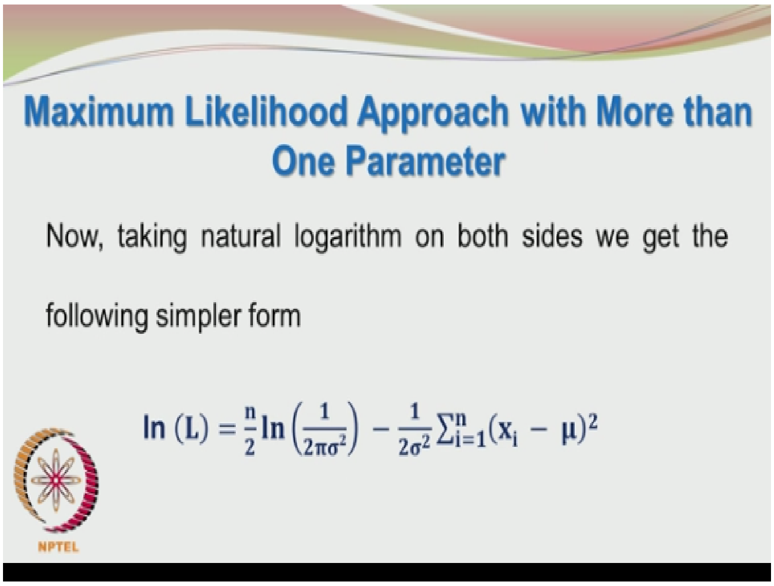
(Refer Slide Time: 32:11)



We assumed that sigma square was already known to us. Now, let us go to the second example where we have 2 parameters to be estimated and that is given by f of x mu sigma square 1/root 2 pi sigma exponential – x – mu whole square/2 sigma square. What we then do is, we take the random sample, form the values x1, x2, so onto xn, then we put L of mu sigma square in terms of the product of the distribution functions expression.

So, you have 1/root 2 pi sigma * E power –Xi – mu whole square/2 sigma square multiplied from i = 1 the first sample to i = n the last sample, okay. So, you are just multiplying it and this is the product sign. Just as you had the summation sign sigma, we are having the product sign here. So when you take the product again, this will become 1/2 pi sigma square whole to the power of n/2 and then you to multiply the exponential terms n times.

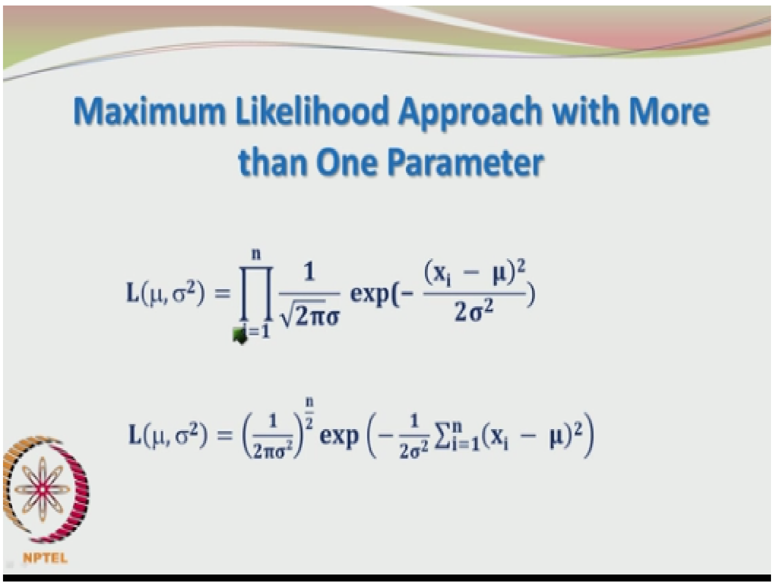
(Refer Slide Time: 33:47)



And then the arguments will be added up. Again, you will have ln of L = n/2 ln of 1/2 pi sigma square – 1/2 pi sigma square sigma i = 1 to n Xi – mu whole square. Remember, that we have 2 parameter sigma square and mu and these 2 are unknown, so we are having an expression with 2 unknown variables sigma square and mu. So, this ln of L should be partially differentiated with respect to both mu to give the first expression.

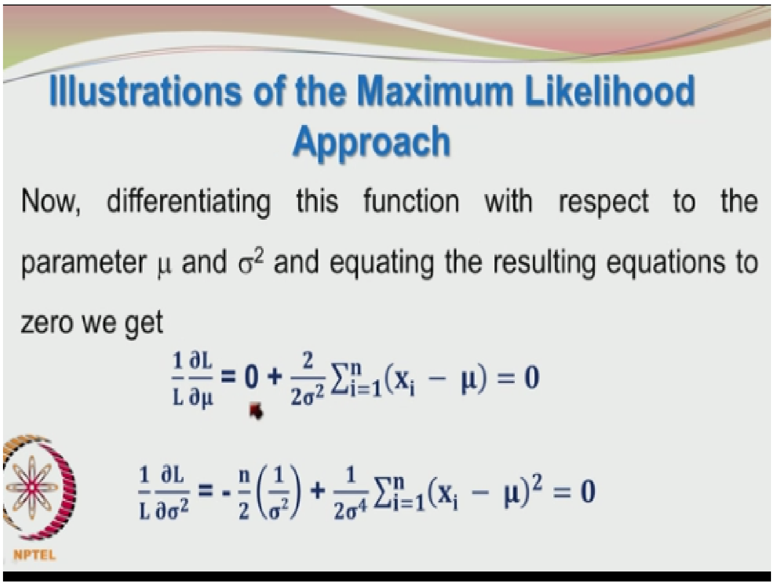
And then sigma square to give the second expression, so this is the expression we have to partially differentiate. We get 1/L dou L/dou mu = 0. Now, we apply the partial differentiation sign because there are 2 parameters to differentiate with and we are representing first we are differentiating with respect to mu, 1/L dou L/dou mu = this term and the next expression is slightly a bit more cumbersome.

(Refer Slide Time: 35:02)



You get 1/L dou L/dou sigma square, remember, you are differentiating with respect to sigma square directly. So, you have to take sigma square as a particular variable, okay. Do not think in terms of sigma square. If that is confusing to you put sigma square = p, so you will get this particular term you want to expand it, n is any constant, you can write n/2 * ln of 1 – ln of 2 pi sigma square.

(Refer Slide Time: 35:11)

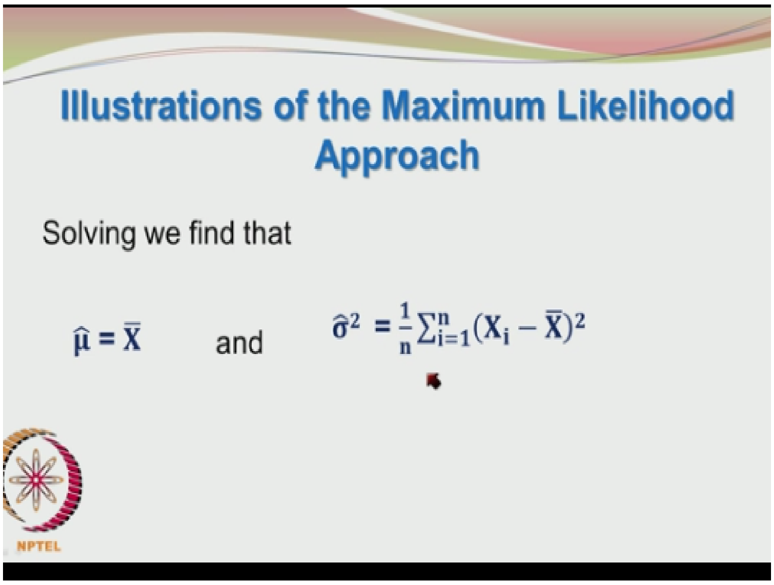


The only important term as far as the differentiation is concerned is ln sigma square. So, when you have ln of sigma square and you differentiated with respect to sigma square you will get 1/sigma square and this n/2 is outside. So, you will get after differentiation –n/2 * 1/sigma square. Similarly, you will have once (()) (36:48) settles down here, you are differentiating 1/sigma square with respect to sigma square.

It is like differentiating 1/X with respect to X. You know that the differentiation of 1/X with respect to X will lead to -1/X square, so differentiation 1/sigma square with respect to sigma square will lead to -1/sigma to the power of 4. So that – and – will get combine to become + and when you differentiate with respect to sigma square, you will get +1/2 sigma to the power 3 * sigma i = 1 to n Xi – mu whole square.

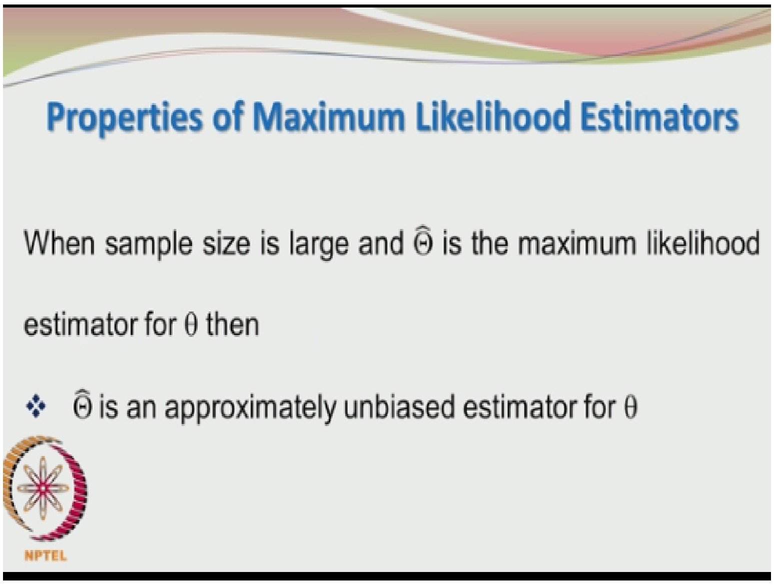
From these 2 equations, we find that mu hat = X bar and sigma hat square = 1/n sigma i = 1 to n Xi – X bar whole square. So, again we are having a biased estimator. We are using the sample and we are taking the sum of the square of the deviations of the sample random variables from the mean and divided it by n, not n-1, we are doing n. So, this expression is rather n-1 S square/n, okay.

(Refer Slide Time: 37:41)



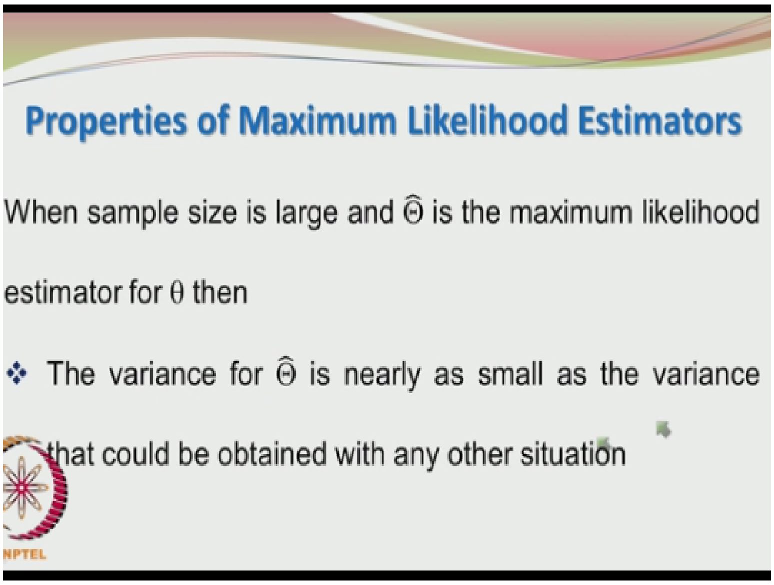
This expression is n-1 S square/n and hence, this particular expression leads to a bias in the estimation of sigma hat square in the estimation of population variance. So, what are the properties of the maximum likelihood estimators? When the sample size is large and theta hat is the maximum likelihood estimator for theta then, theta hat is an approximately unbiased estimator for theta, okay.

(Refer Slide Time: 38:36)



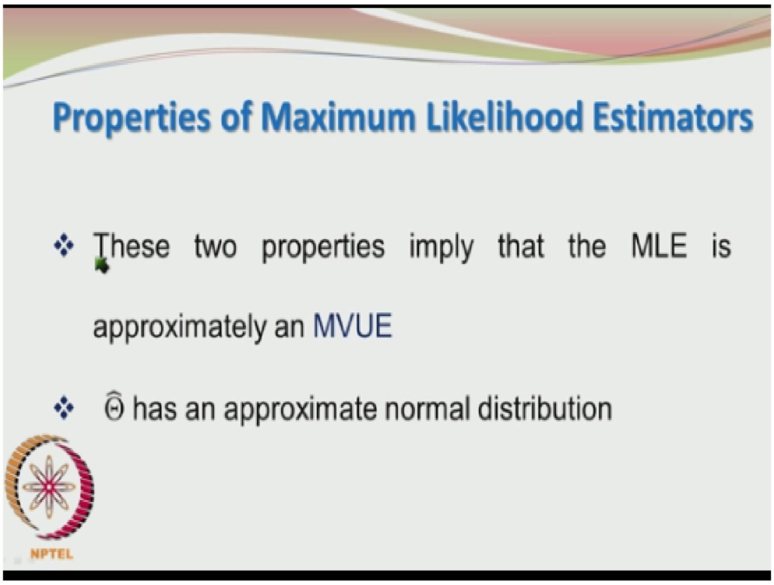
And the variance for theta hat is nearly as small as the variance that could be obtained with any other situation. These 2 properties imply that the maximum likelihood estimator is approximately an minimum variance unbiased estimator and it also tells us, we are not looking into the proof that the another important property is theta hat has an approximate normal distribution.

(Refer Slide Time: 38:59)



So, this concludes our presentation on maximum likelihood estimation. We have learnt it, it is a very interesting and useful thing to know. We, however, will not be really using this further in our discussion, but it is a very interesting thing to be aware of. It is also important to know that the properties of the sample we are taking should lead to an unbiased estimation of the population parameter.

(Refer Slide Time: 39:08)



So in our influential statistics, we are going to use the sample mean, sample variance very frequently and it is essential that we understand what their properties are. So, this discussion really helped us. We also saw that if instead of using 1/n-1 in the denominator for sample variance if we have used n that would have let to a biased estimation of the population variance which is not really very good especially in the case of small samples.

From now on, we are going more into the analysis of sample. We are going to work with samples. We are going to not work with X that much. We are going to consider X bar instead of X more and more, treat X bar as a single entity rather than a collection of n random variables even though that is indisputable, we are now going to deal with the X bar as a single entity rather than X.

And we are also going to discuss another important and interesting topic, we have discussed about the point estimation so far. Now, we are talking about the interval estimation for the population parameters. All these things will lead to useful techniques or procedures that are essential in the design of experiments and analysis of experimental data. So, we will wind up now and we will be looking at interval estimation shortly.

We will also be doing a few problems to drive home the concepts we have studied so far. Thank you.