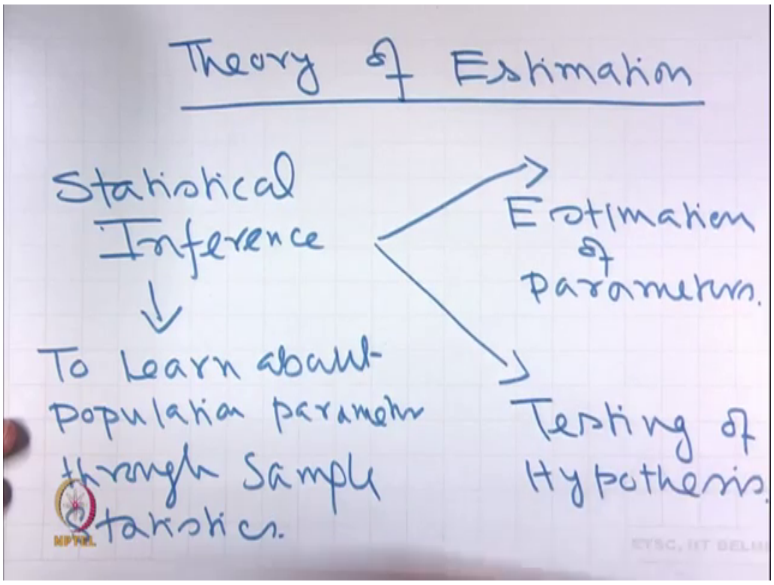
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 12
Statistical Inference

Welcome students to the MOOCs lecture on Statistical Inference. This is lecture number 12.

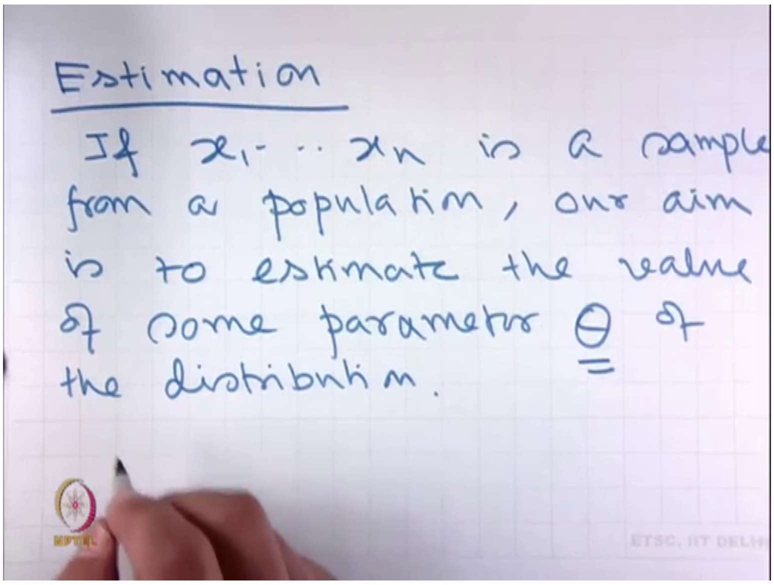
(Refer Slide Time: 00:30)



And today what I will start is called Theory of Estimation. As I told you in my earlier classes that the most important aspect of statistics is statistical inference, to learn about the population parameter with the help of sample statistics. And I have mentioned that there are two basic ways of statistical inference, one is estimation of parameters. Here, the underlying assumption is that the form of the distribution is known, and our job is to identify the parameters of the distribution.

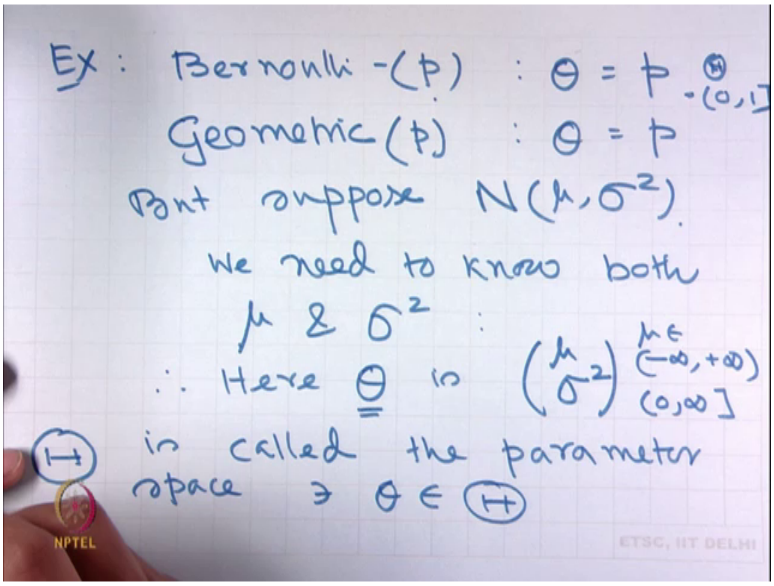
And the other one is testing of hypothesis. Here we want to check whether the sample gives us enough evidence, so that some hypothesis can be accepted. Here, by hypothesis we mean some pre assigned values of the parameters.

(Refer Slide Time: 02:52)



As you can understand now I am going to do on estimation. So, what is an estimator? If x 1 up to x n is a sample from a population, our aim is to estimate the value of some parameter theta of the distribution. So, what is theta, theta is something, so that if we know the value, we can understand the distribution completely.

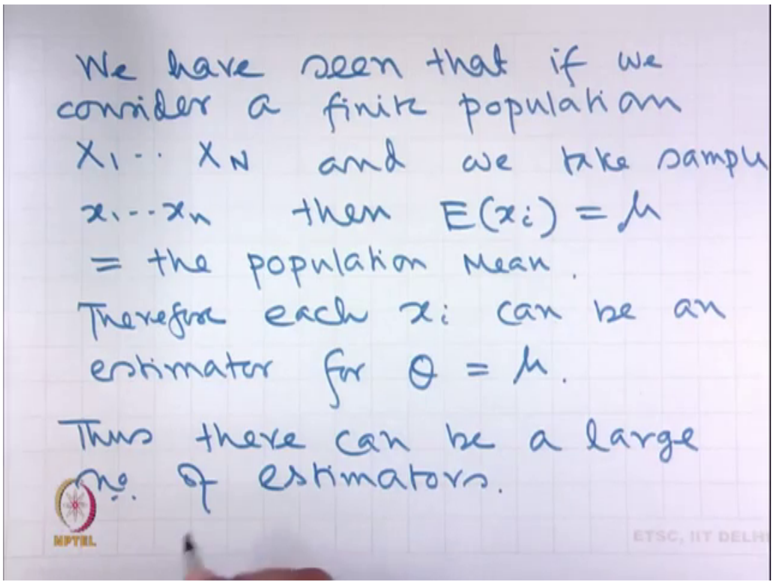
(Refer Slide Time: 04:09)



For example, Bernoulli distribution; here P is the only unknown parameter. If we know the value of P, then we know what the distribution is. So, in this case theta is equal to P. Similarly, for geometric here also the parameter is P, therefore theta is equal to P, but suppose we consider normal mu sigma square. Here, in order to know the distribution completely we need to know the value of both mu and sigma square. Therefore, here theta is mu sigma or say sigma square ok, or in other words we are looking at a bivariate parameter. You often find the term capital theta is called the parameter space such that theta can belong to theta.

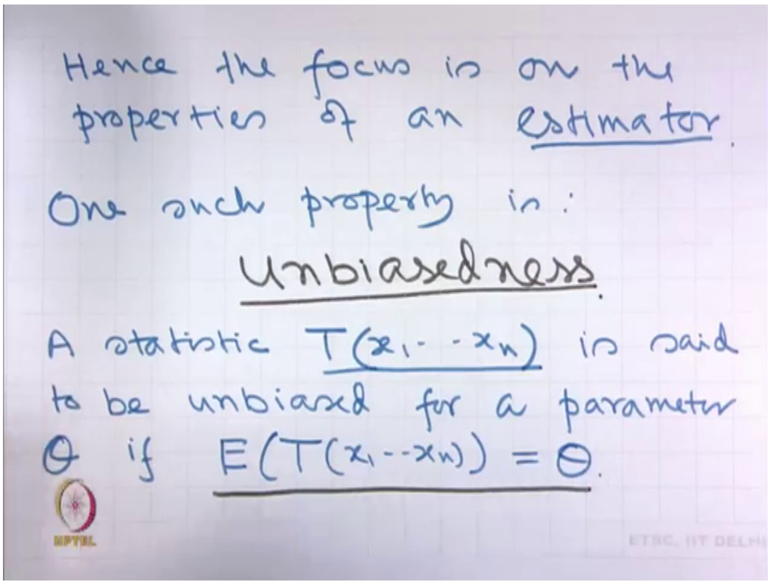
For example, when we are looking at P the capital theta is equal to 0 to 1 on the real line. When we are looking at mu, mu can belong to minus infinity to infinity or on the real line, sigma square may belong to 0 to infinity on the real line, so that is the parameter space to which that parameter of the distribution can belong to.

(Refer Slide Time: 07:00)



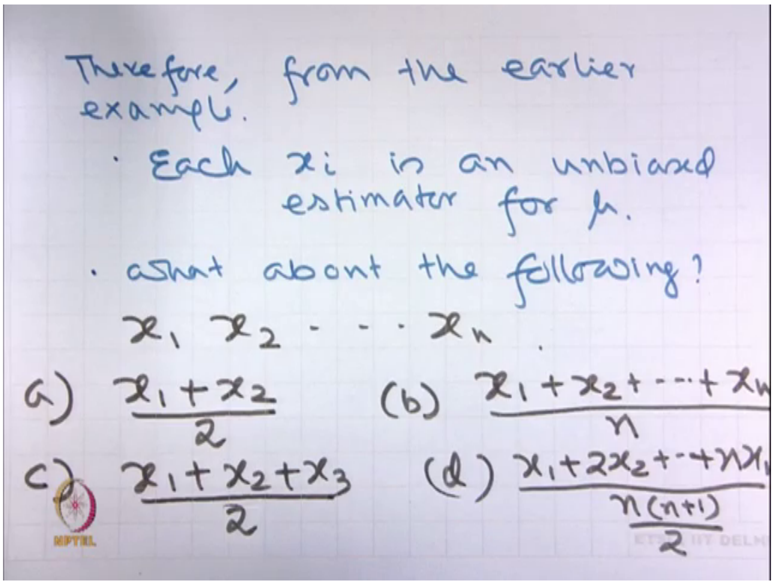
We have already seen some estimators, if you remember; we have seen that, if we consider a finite population X 1, X2 up to X N. And we take sample x 1, x 2 up to x n, then expected value of each x i is equal to mu is equal to the population mean. Therefore, each x i can be an estimator for theta is equal to mu. Thus, there can be a large number of estimators. Hence, the question comes, which of the different estimators we should choose for our purpose.

(Refer Slide Time: 08:58)



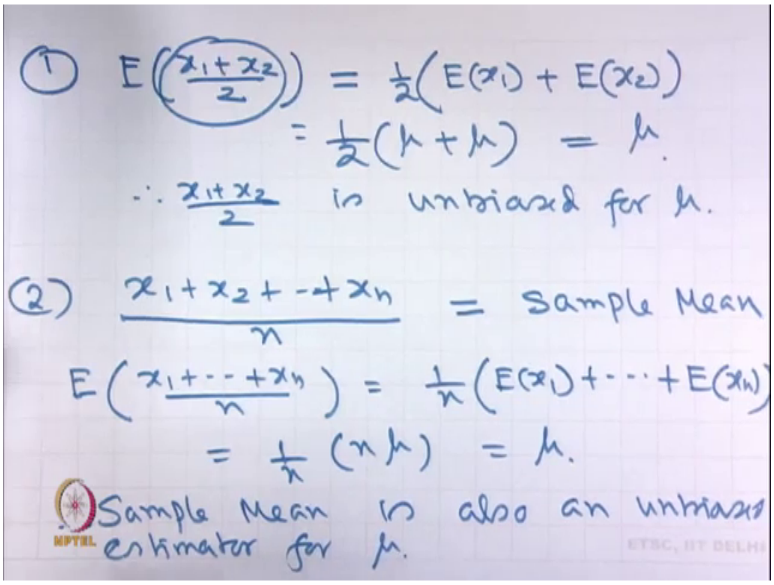
Hence, the focus is on the properties of an estimator. So, if an estimator satisfies those properties, then we consider it to be better suited for our purpose. One such property is you have already seen unbiasedness. A statistic T x 1, x 2, x n what does it mean, it means that I have taken the samples x 1, x 2, x n. And T is the function that is defined on the sample. So, once the sample is taken we can compute T. It is said to be unbiased for parameter theta if expected value of T x 1, x n is equal to theta.

(Refer Slide Time: 11:17)



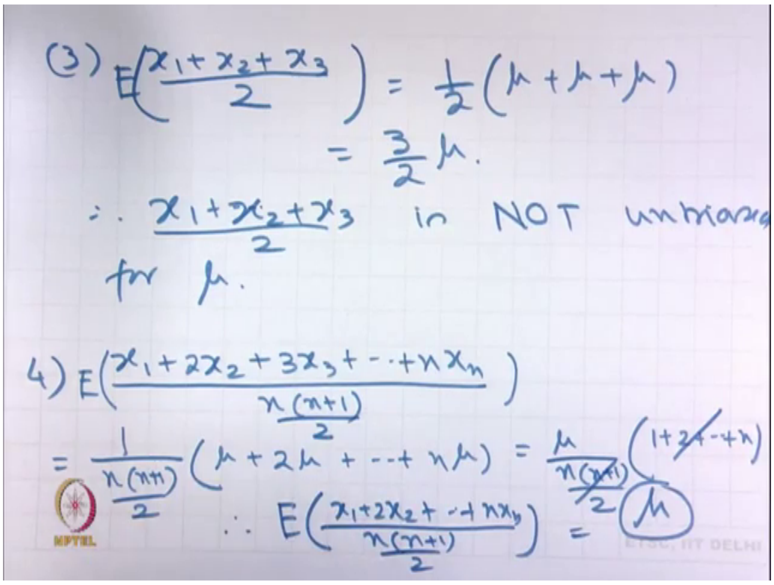
Therefore, from my earlier example each x i is an unbiased estimator for mu. What about the following I have the sample x 1, x 2, x n. So let me consider, x 1 plus x 2 by 2. b, x 1 plus x 2 plus x n by n. c, x 1 plus x 2 plus x 3 by 2. d, x 1 plus 2 x 2 plus n x n by n into n plus 1 by 2.

(Refer Slide Time: 13:18)



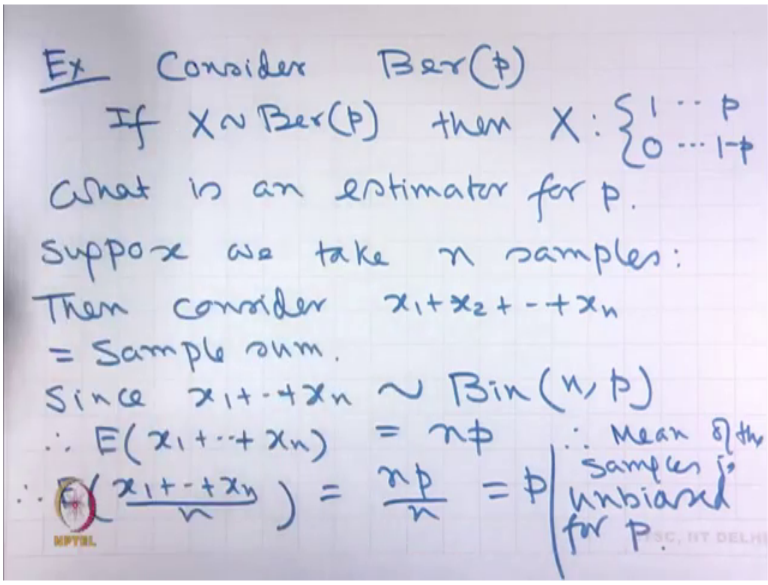
Let us see their expectations, expected value of x 1 plus x 2 by 2, because of the linearity is equal to half of expected value of x 1 plus expected value of x 2 is equal to half times mu plus mu is equal to mu. Therefore, x 1 plus x 2 by 2 is unbiased for mu. 2, x 1 plus x 2 plus x n by n or in other words we are looking at sample mean. Its expectation is equal to 1 by n times, expectation of x 1 plus up to expectation of x n is equal to 1 by n times n times mu is equal to mu. Therefore, sample mean is also an unbiased estimator for mu.

(Refer Slide Time: 15:22)



X 1 plus x 2 plus x 3 by 2 you can now easily understand that its expectation is equal to half into mu plus mu plus mu is equal to 3 by 2 mu. Therefore, x 1 plus x 2 plus x 3 by 2 is not unbiased for mu. Number 4, x 1 plus 2 x 2 plus 3 x 3 up to n x n divided by n into n plus 1 by 2. Its expectation is equal to 1 upon n in to n plus 1 by 2 multiplied by mu plus 2 mu plus n mu is equal to mu n into n plus 1 by 2 into 1 plus 2 plus n. And we know that the sum of first n natural numbers is n into n plus 1 by 2, therefore this cancels with this. Therefore, the expectation of x 1 plus 2 x 2 up to n x n upon n into n plus 1 by 2 is equal to mu. Therefore, this is also an unbiased estimator for the population mean mu.

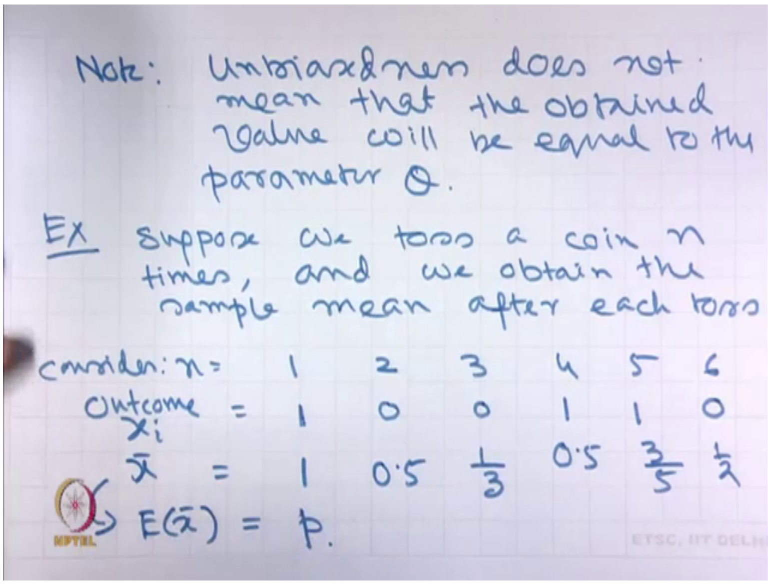
(Refer Slide Time: 17:38)



Another example consider Bernoulli P what is a Bernoulli distribution we know that. If I toss a coin, then P is the probability of head or mathematically. If x follows Bernoulli of P, then x can take two values 1 with probability P and 0 with probability 1 minus P. What is an estimator for P. Suppose, we take n samples, then consider x 1 plus x 2 plus x n is equal to sample sum.

We know that x 1 plus x 2 plus x n is distributed as binomial with n comma. Therefore, expected value of x 1 plus x n is equal to n P. Therefore, expected value of x 1 plus x 2 plus x n by n is equal to n p by n is equal to p. Therefore, mean of the samples is unbiased for P. Let me, warn you on a point.

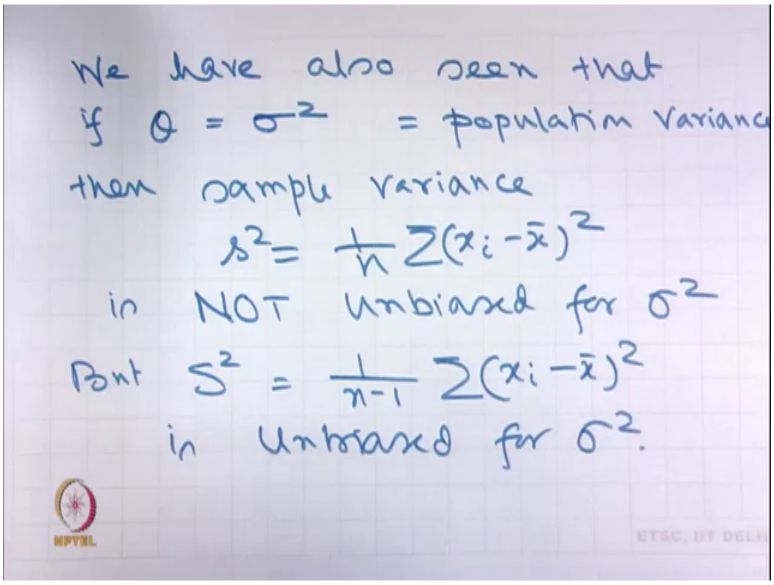
(Refer Slide Time: 20:23)



Unbiasedness does not mean that the obtained value will be equal to the parameter theta. Example suppose, we toss a coin n times, and we obtain the sample mean after each toss. So, consider n is equal to say 1, 2, 3, 4, 5, 6. Suppose, I am looking at 6 tosses, and outcomes are that is x i are say 1, 0, 0, 1, 1, 0. Therefore, x bar is equal to 1, because it is the mean of one sample it is 0.5 at this point, at this point it is 1 by 3, at this point it is again 0.5 or half, at this point it is 3 by 5, at this point it is again half.

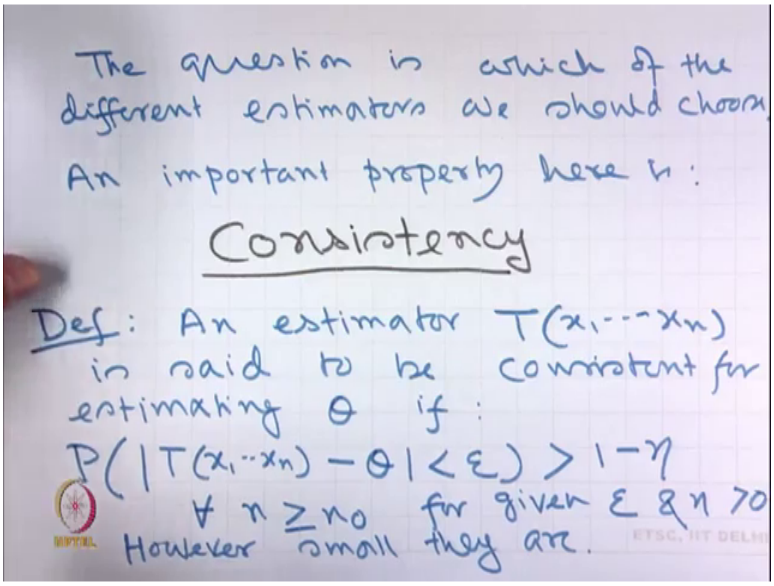
So, depending upon how many tosses you have actually done, the obtained value of the statistic can change. It is not mandatory that they will be equal to the actual value of the parameter, but what we are saying is that, if we consider its expectation, that is going to be equal to P. I hope the distinction is clear.

(Refer Slide Time: 23:31)



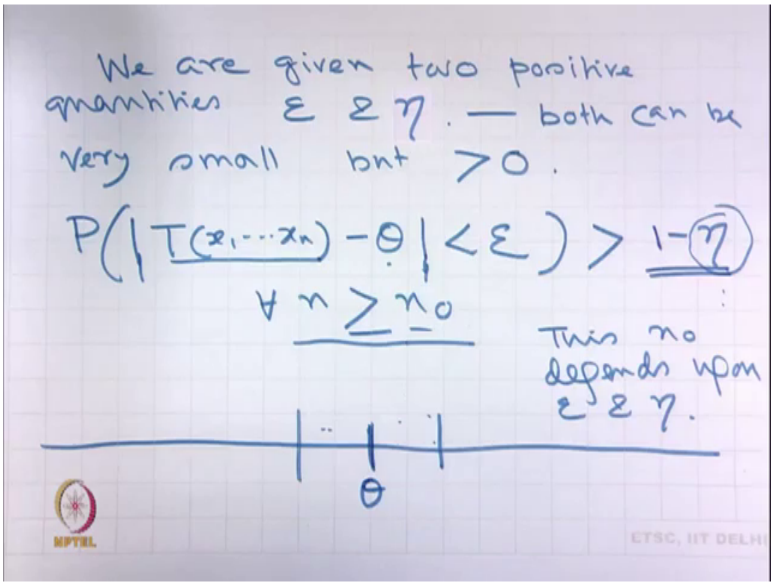
We have also seen that, if theta is equal to sigma square is equal to population variance, then sample variance a square is equal to 1 by n sigma x i minus x bar whole square is not unbiased for sigma square. But, S square is equal to 1 upon n minus 1 sigma x i minus x bar whole square is unbiased for sigma square. Therefore, given a parameter, if we want to estimate it, we can get many different statistic. Such that each of them is an unbiased estimator for the population parameter under consideration, which we generally denote as theta.

(Refer Slide Time: 25:24)



Therefore, the question is, which of the different estimators we should choose. An important property here is consistency definition, an estimator T x 1, x 2, x n is said to be consistent for estimating theta. If probability T x 1, x 2, x n minus theta less than epsilon is greater than one minus eta for all n, greater than equal to n naught for given epsilon, and eta greater than 0, however small they are. Let us, understand what it means.

(Refer Slide Time: 28:08)

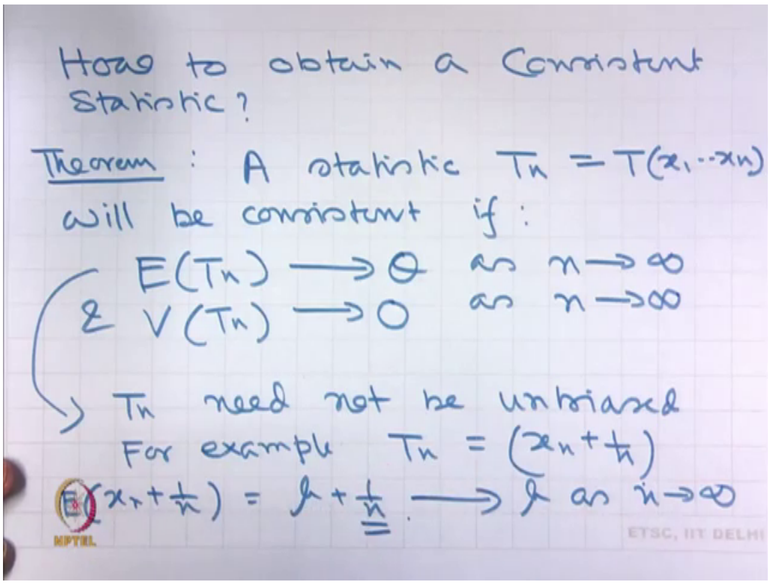


We are given two positive quantities epsilon and eta. They can be very very small, but greater than 0. So, the definition says, that probability T of x 1, x 2, x n the obtained value of the statistic from the sample of size n minus theta less than epsilon. So, this is the event. We have the n samples, we have calculated the statistic, and we are saying that, this statistic should be very close to the actual value of the parameter. How much close that, the absolute difference between them is less than epsilon that, probability can be made arbitrarily large that is that is greater than equal to 1 minus eta, for all n greater than n naught.

So, what it means suppose this is the actual parameter theta. And we have given an epsilon bound around it. And we are saying that, if we keep on taking the samples, then there will be an integer n naught. Such that if the sample size is greater than n naught, then the probability that the obtained value of the statistic will remain within this interval, that probability is going to be arbitrarily large that is it can be made as large as you want that is it is greater than 1 minus eta that, and eta can be, however small you want.

Obviously, this n naught depends upon both epsilon, and eta. It is not mandatory that the same n naught will work for all values of epsilon and eta, but given epsilon and eta we can choose an n naught, or we can find an n naught. Such that, if the number of samples is more than n naught, then the probability that this sample mean, sample statistic will be very close to the parameter that probability is going to be very very high.

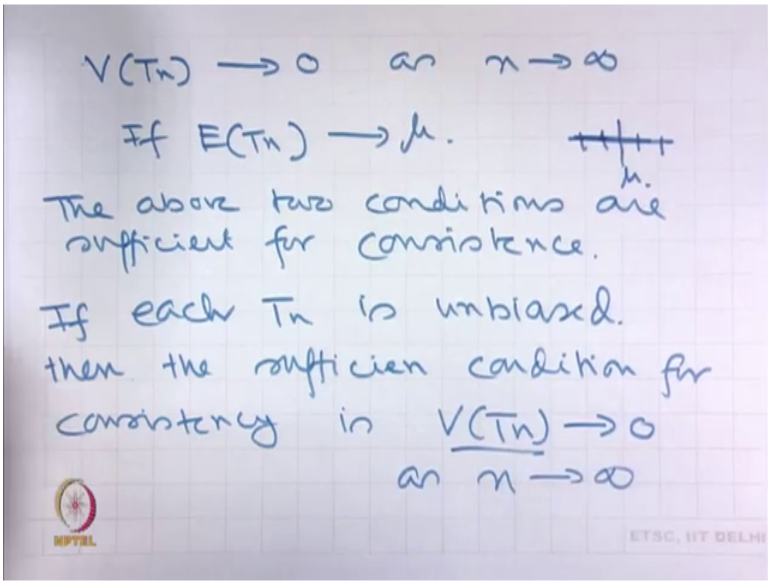
(Refer Slide Time: 31:46)



The question is, how do we obtain a consistent statistic? So, I give you a theorem. A statistic T n is equal to T of x 1 up to x n will be consistent. If expected value of T n goes to theta, as n goes to infinity, and variance of T n goes to 0, as n goes to infinity. So, first thing we note that, T n need not be unbiased.

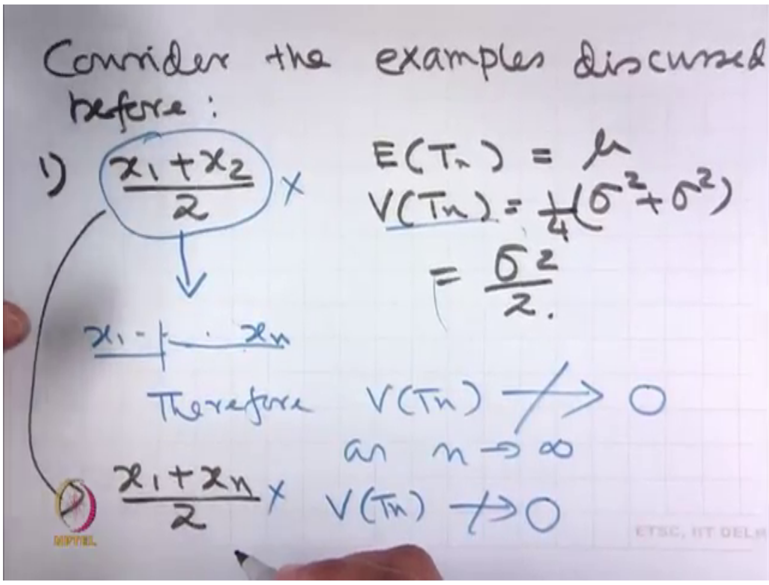
For example, T n is equal to x n plus 1 by n. What is the expected value, is this mu plus 1 by n, and these goes to mu, as n goes to infinity, because this is the bias, this is the quantity by which it is different from the parameter or the intended parameter mu. Therefore, x n plus 1 by n is not an unbiased estimator for mu for any n, but as n goes to infinity; its expected value converges to mu. Therefore, in order to be consistent unbiasedness is not necessary.

(Refer Slide Time: 34:35)



But, we have to also look at that variance of T n should go to 0, as n goes to infinity. We know that, variance gives us the measure of dispersion. So, if expected value of T n converges to the mu, then as if variance of T n goes to 0, then what we can say that, it is coming within arbitrary closeness of the parameter mu. So, the above is a sufficient condition for consistence. If each T n is unbiased, then we are even better off. So, we need to check the variants of T n is going to 0, as n is going to infinity.

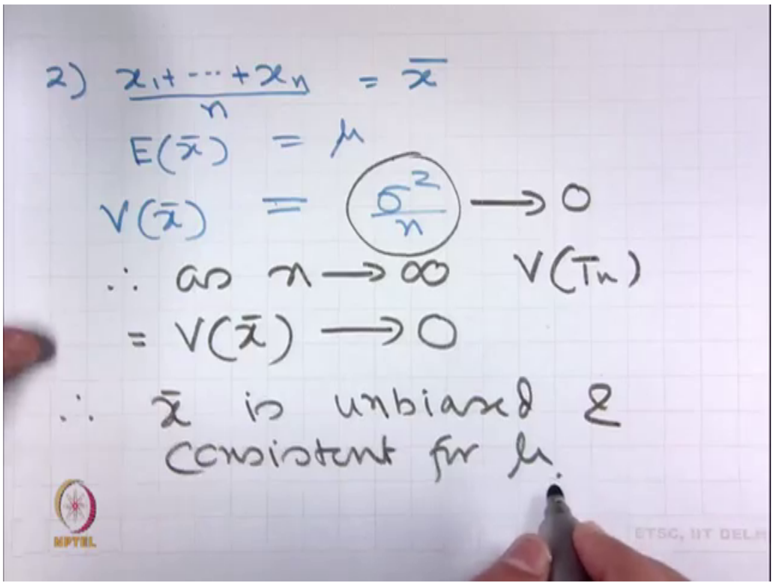
(Refer Slide Time: 36:36)



Let us now consider, we had x 1 plus x 2 by 2, it is unbiased. And variance of T n is equal to 1 by 4 sigma square plus sigma square is equal to sigma square by 2. First, you notice that, even if we have chosen samples x 1 up to x n, we are trying to estimate the parameter on the basis of only the first two samples, actually here we are not using the remaining samples. Therefore, the variance of T n will not change, even if we take many different samples, it will remain sigma square by 2. And therefore, variance of T n does not go to 0, as n goes to infinity.

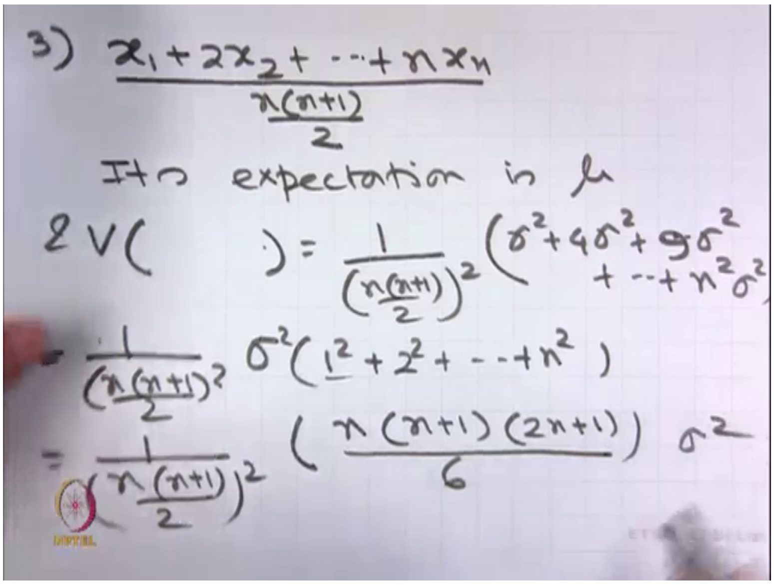
Suppose, instead of x 1 plus x 2 by 2, I would have chosen x 1 plus x n by 2. Then, this is also unbiased, also I am taking care of the large sample, that I have taken, but here also variance of T n will not go to 0. Therefore, neither this, nor this is a consistent estimator of mu in this case.

(Refer Slide Time: 39:01)



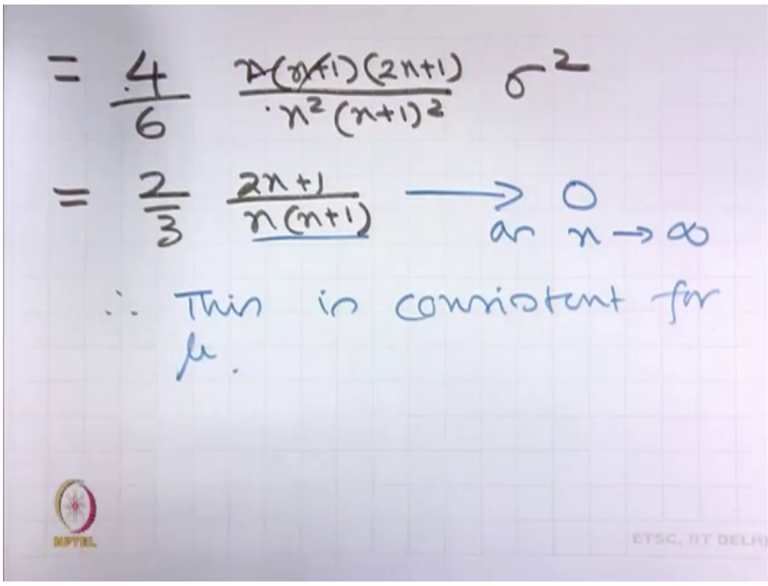
The second example was x 1 plus x n by x 1 plus x 2 plus x n by n is equal to x bar. Expected value of x bar is equal to mu. Therefore, the first condition is automatically satisfied. Variance of x bar we know is equal to sigma square by n. Therefore, as n goes to infinity, variance of T n, which is nothing but x bar goes to 0, because sigma square by n sigma square is fixed, therefore as n increases, this goes to 0. Therefore, by the above theorem, x bar is unbiased, but that is not important, and what is important is that, it is consistent for mu.

(Refer Slide Time: 40:36)



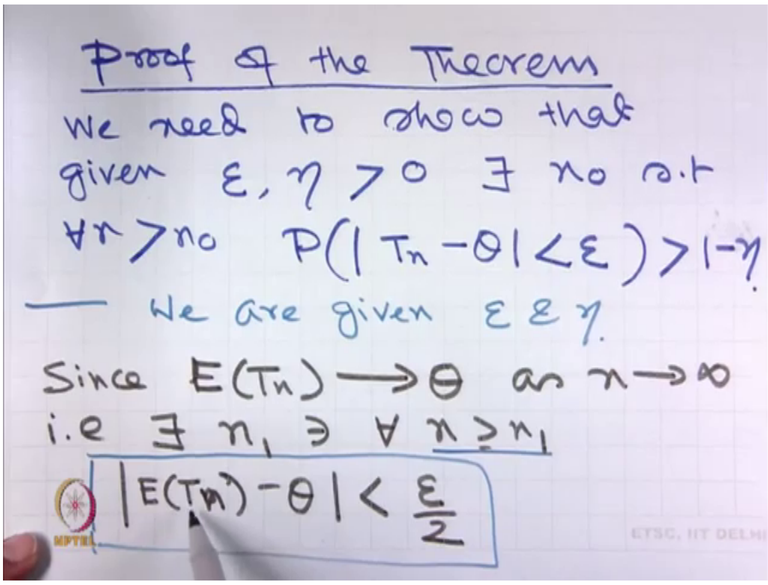
Let me consider this also x 1 plus 2 x 2 plus n x n upon n into n plus 1 by 2. What is the variance, its expectation is mu, and variance is equal to 1 upon n into n plus 1 by 2 whole square into sigma square plus 4 sigma square say plus 9 sigma square plus n square sigma square is equal to 1 upon n into n plus 1 by 2 whole square multiplied by sigma square into 1 square plus 2 square plus up to n square is equal to 1 upon n into n plus 1 by 2 whole square into sum of square of first n natural numbers is equal to n into n plus 1 into 2 n plus 1 by 6 multiplied by sigma square.

(Refer Slide Time: 42:13)



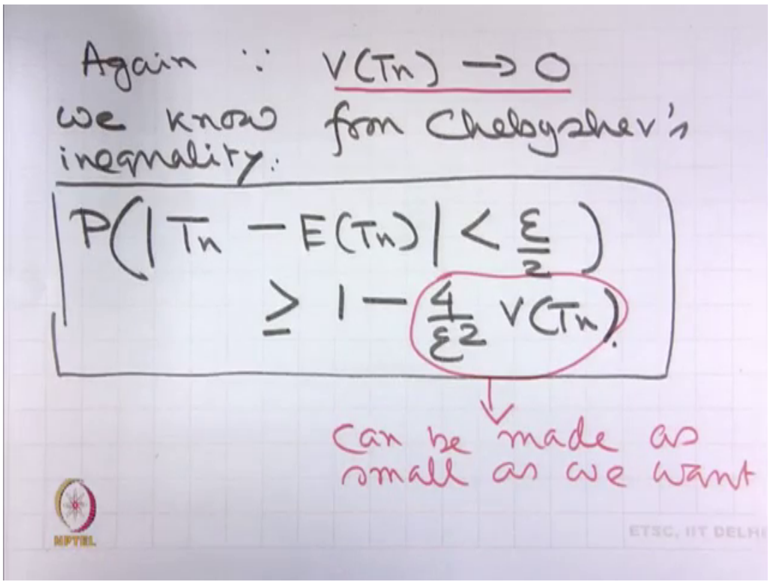
Which is equal to 4 by 6 n into n plus 1 into 2 n plus 1 divided by n square into n plus 1 square sigma square is equal to 2 by 3 into one of them cancels 2 n plus 1 upon n into n plus 1. Since, the numerator is linear in n, but the denominator is quadratic in n, we know that its limit is 0, as n goes to infinity. Therefore, this is consistent for mu.

(Refer Slide Time: 43:22)



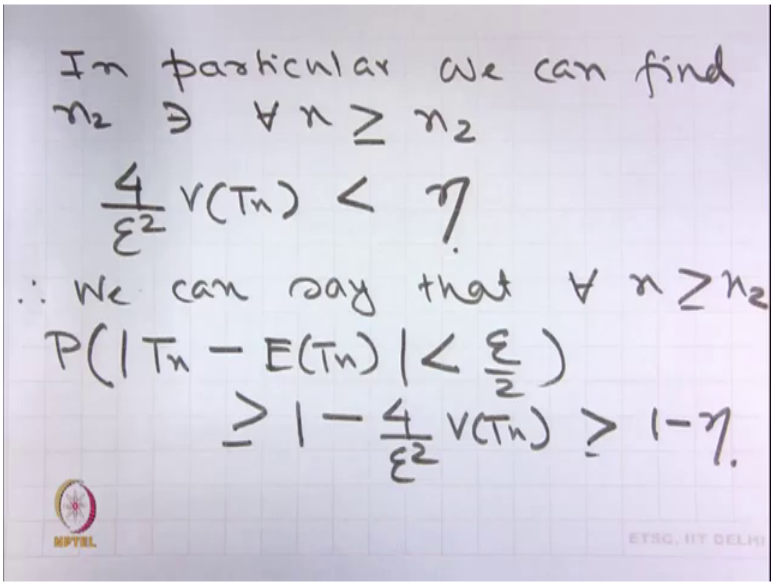
We need to show that, given epsilon and eta greater than 0, there exist n naught, such that for all n greater than n naught probability modulus of T n minus theta less than epsilon is greater than 1 minus eta. So, this is what we will have to prove, so we are given epsilon and eta. First thing that is given is expected value of T n is converges to theta, as n goes to infinity that means, there exist n 1 such that for all n greater than equal to n 1 modulus of expected value of T n minus theta less than epsilon by 2. So, we have been given an epsilon, we are trying to identify an n 1, such that for all n greater than equal to n 1. This difference expected value of T n minus theta less than epsilon by 2.

(Refer Slide Time: 45:44)



Again, since variance of T n is going to 0. We know from Chebyshev's inequality, probability modulus of T n minus expected value of T n less than epsilon by 2 is greater than equal to 1 minus 4 epsilon square into variance of T n. This is known because of Chebyshev's inequality. Therefore, what we get it, that as variance of T n is going to 0. This quantity can be made very very small right.

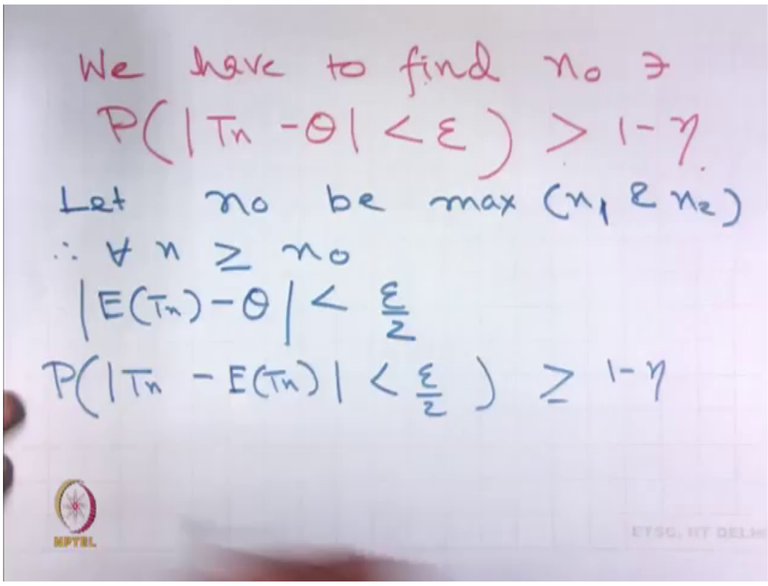
(Refer Slide Time: 47:35)



In particular, we can choose or we can find n 2, such that for all n greater than equal to n 2 4 by epsilon square into variance of T n is less than eta. This is possible, because this quantity is going to 0, and eta is fixed, that is given to us. Therefore, we can say that for all n greater than equal to n 2. Probability modulus of T n minus expected value of T n less than epsilon by 2 is greater than equal to 1 minus 4 by epsilon square into variance of T n, which is greater than equal to 1 minus eta.

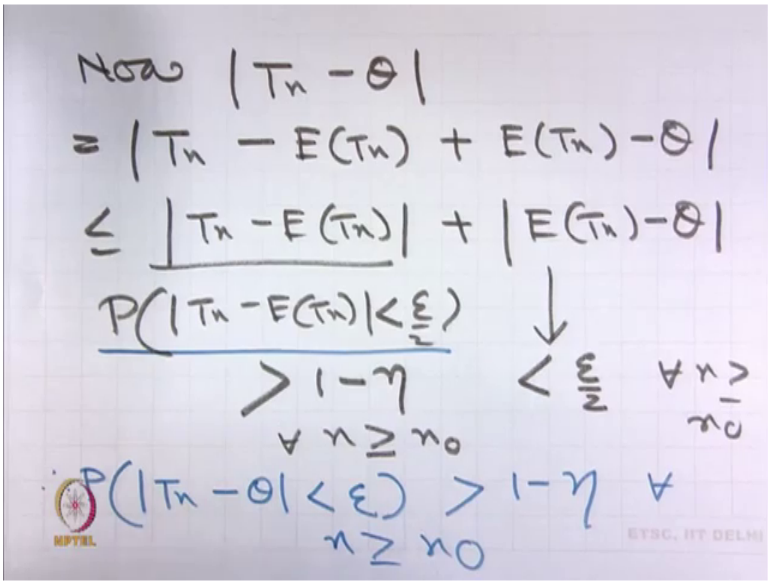
Therefore, what we got, so we have an n 1, such that for all n greater than equal to n 1, expected value of T n minus theta is less than epsilon by 2. There is an n 2, such that for all n greater than equal to n 2, probability T n minus expected value of T n less than epsilon by 2 is greater than 1 minus eta.

(Refer Slide Time: 49:33)



Or we have to find n naught, such that probability modulus of T n minus theta less than epsilon is greater than 1 minus eta. This is what we have to find. Let n naught be maximum of n 1 and n 2. Therefore, for all n greater than equal to n naught, we have expected value of T n minus theta is less than epsilon by 2, also we have probability modulus of T n minus expected value of T n less than epsilon by 2 is greater than equal to 1 minus eta.

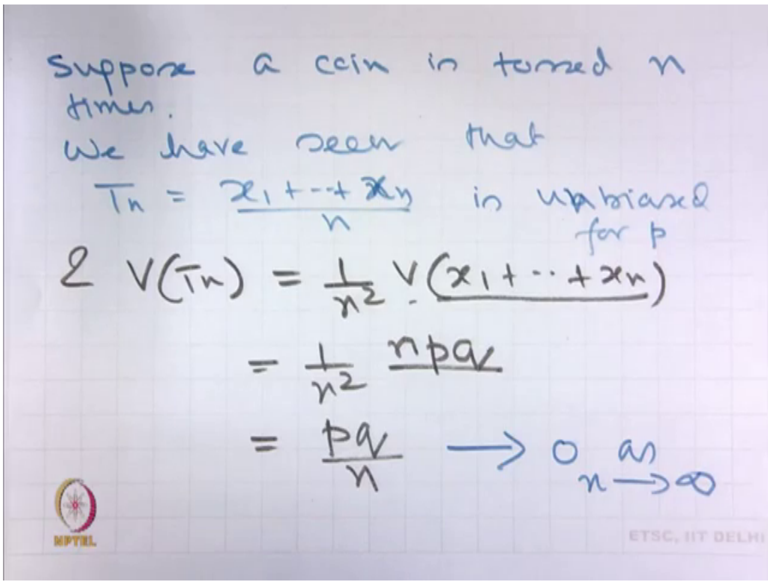
(Refer Slide Time: 50:48)



Now, modulus of T n minus theta is equal to modulus of T n minus expected value of T n plus expected value of T n minus theta, which is less than equal to modulus of T n minus expected value of T n plus modulus of expected value of T n minus theta. This is less than epsilon by 2 for all n greater than equal to n naught. This probability less than epsilon by 2 is greater than 1 minus eta for all n greater than equal to n naught.

Therefore, from these two, we can see that probability modulus of T n minus theta less than epsilon is greater than 1 minus eta for all n greater than equal to n naught. So, these proves the sufficiency of the condition that expected value of T n has to go to theta, and variance of T n has to go to 0, as n goes to infinity. Then the T n is going to be a consistent estimator for the parameter theta.

(Refer Slide Time: 52:50)



Suppose, a coin is tossed n times, we have seen T n is equal to x 1 plus x n by n is unbiased for p. And variance of T n is equal to 1 by n square into variance of x 1 plus x 2 plus x n is equal to 1 by n square into n p q. This we know, because it is a binomial random variable, its variance has to be n by n p q, therefore this is equal to p q by n. Therefore, this goes to 0, as n goes to infinity. Therefore, this sample mean is not only unbiased, it is consistent for estimating p.

Ok students, I stop here. In the next class, I shall examine some more properties of an estimator.

Thank you.