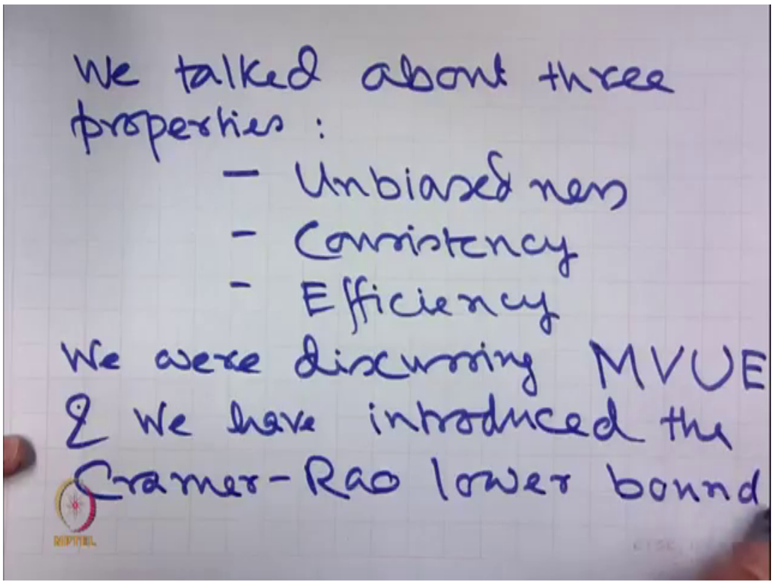
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 15
Statistical Inference

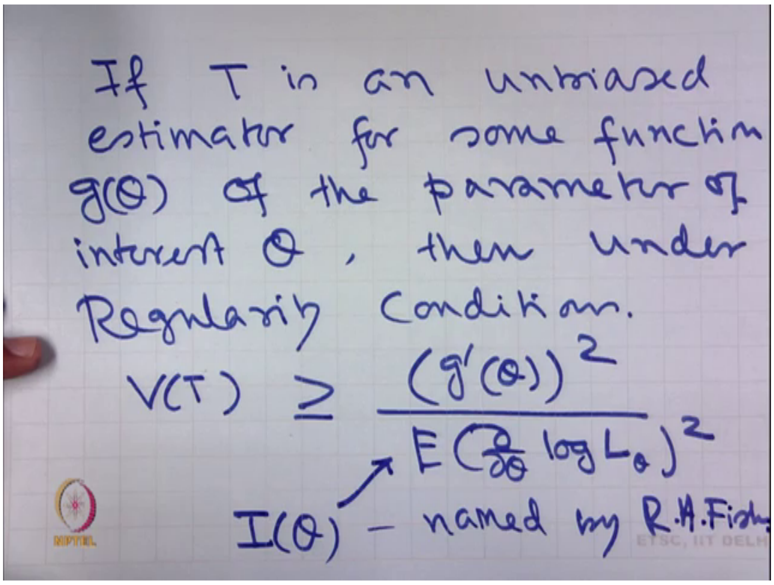
Welcome students to the MOOCS lecture on Statistical Inference. This is lecture number 15.

(Refer Slide Time: 00:37)



If you recollect in the last lecture we talked about 3 properties namely; unbiasedness, consistency, and efficiency. We are talking about minimum variance unbiased estimate and we have introduced the Cramer-Rao lower bound. What is that?

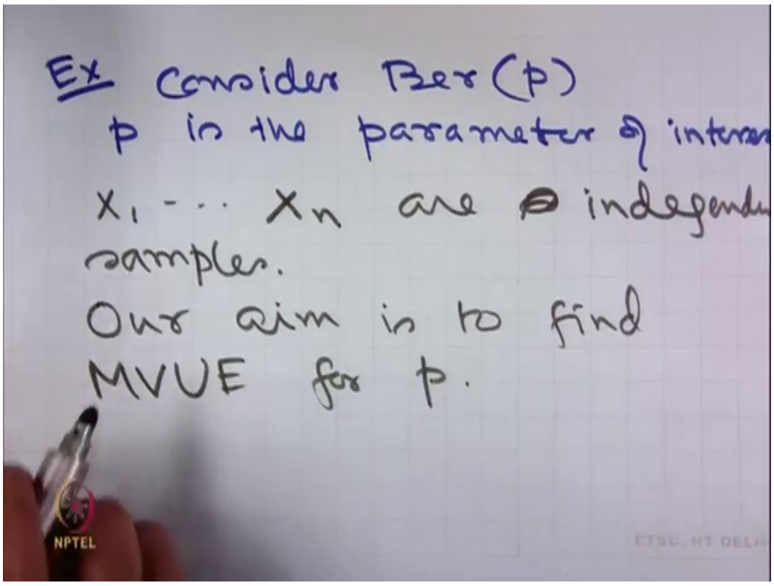
(Refer Slide Time: 01:55)



If T is an unbiased estimator for some function g theta of the parameter of interest theta, then under regularity conditions variance of T is greater than equal to g prime theta whole square divided by expected value of del del theta log of L which depends on theta whole square and this quantity is called I theta named by R.A. Fisher.

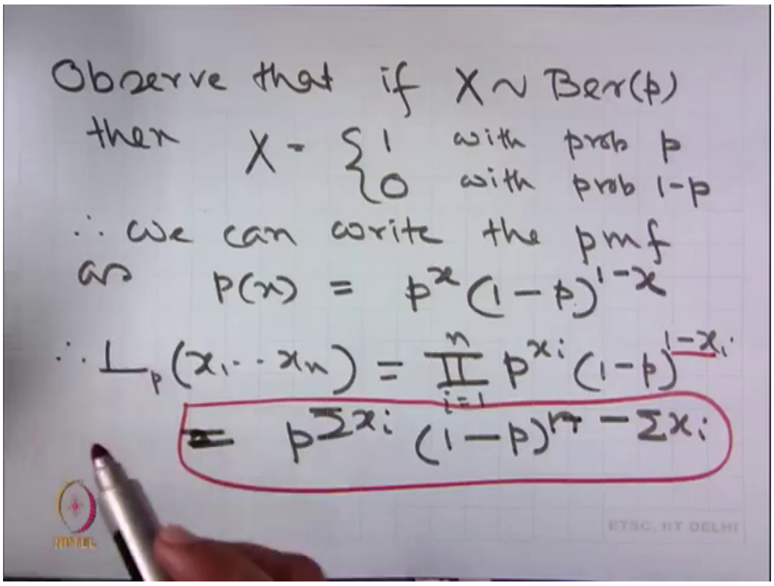
Today first I will solve a few problems on efficiency or on minimum variance unbiased estimator, then I will introduce another important property of an estimator namely sufficiency. So, some example First.

(Refer Slide Time: 04:01)



Consider Bernoulli p. So, p is the parameter of interest. Suppose I have taken sample x 1 x 2 x n are samples, independent samples, then our aim is to find the minimum variance unbiased estimator for p.

(Refer Slide Time: 05:19)

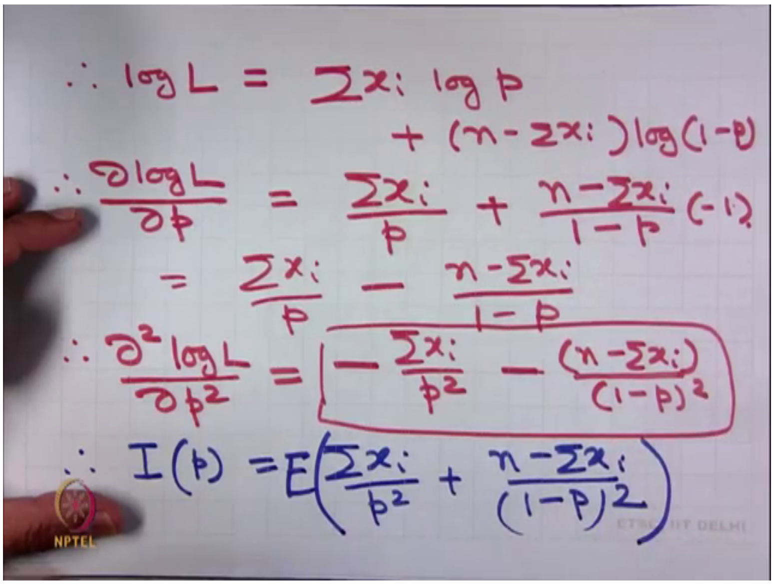


So, what we do is observe that if x follows Bernoulli p, then x takes two values; one with probability p and 0 with probability 1 minus p. Therefore, we can write the pdf or the pmf as p of x is equal to p to the power x 1 minus p whole to the power 1 minus x. So, check that if x is equal to 1, then this is p to the power 1 into 1 minus p to the power 0. Therefore, this is p and if x is equal to 0, then this becomes 1 and this gives you 1 minus p to the power 1 that is 1 minus p.

Therefore, L p of x 1 x 2 x n that is the likelihood function of the sample x 1 x 2 x n. When the parameter is p is equal to product of i is equal to 1 to n p x i 1 minus p to the power 1 minus x i is equal to p to the power sigma x i into 1 minus p to the power n minus sigma x i.

I hope this is clear to you because p to the power x i i is equal to 1 to n. Therefore, the power gets added and here it is 1 minus p to the power 1 minus x i. So, when I am adding this over n of the observations, I get n minus sigma x i.

(Refer Slide Time: 07:59)

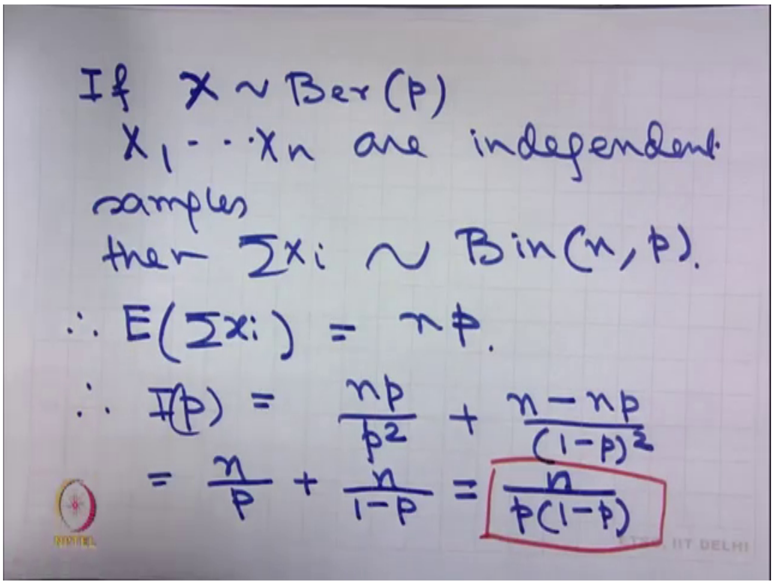


Therefore, log of L is equal to sigma x i log p plus n minus sigma x i into log of 1 minus p. Therefore, del log L del p. Remember that here theta is equal to p. The parameter that we are interested in is equal to sigma x i d log p dp is equal to 1 upon p plus n minus sigma x i upon 1 minus p into d 1 minus p dp which gives you minus 1 is equal to sigma x i upon p minus n minus sigma x i 1 minus p.

Therefore, del 2 log l del p square is equal to minus sigma x i upon p square minus n minus sigma x i upon 1 minus p whole square because this is 1 minus p to the power minus 1. Therefore, for that we get 1 minus 1 and it is 1 minus p. So, here like here like in this place here will be another minus 1. So, altogether there are 3 minuses which will keep it as minus and this is the del log L del p square.

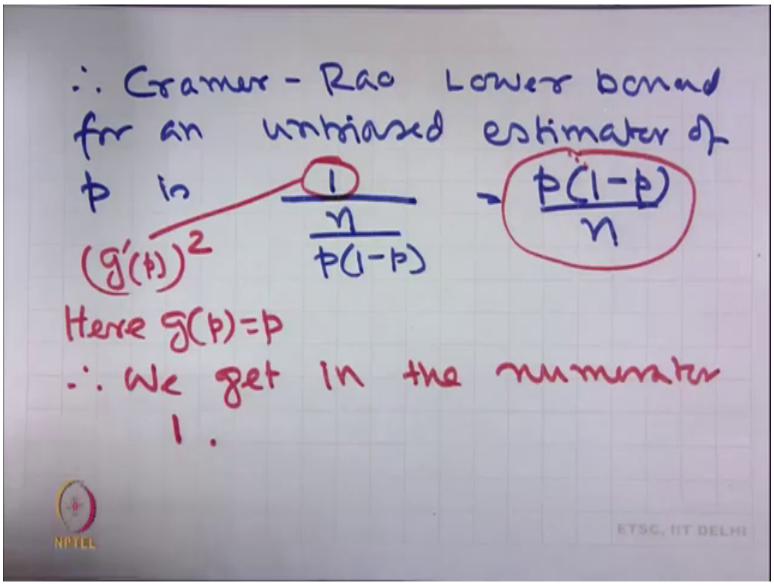
Therefore, L p is equal to minus of that is equal to sigma x i upon p square plus n minus sigma x i upon 1 minus p whole square and we take expected value of that one. So, this is equal to we know that.

(Refer Slide Time: 10:56)



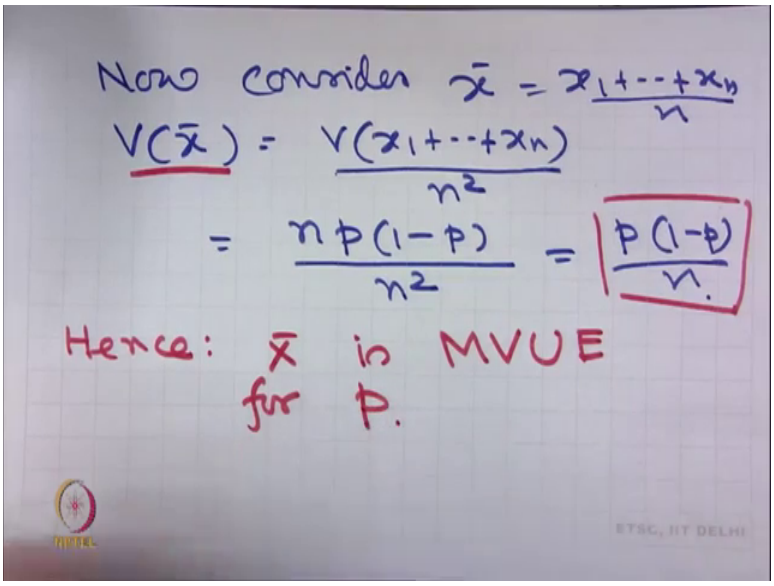
If X follows Bernoulli and x 1 x 2 x n are samples, then sigma x i there is a binomial with np. Therefore, expected value of sigma x i is equal to n p Therefore, using the linearity we can write I p is equal to np upon p square plus n minus np upon 1 minus p whole square is equal to n upon p plus n upon 1 minus p is equal to n upon p into 1 minus p. Therefore, this is my information about p that I get from the sample.

(Refer Slide Time: 12:49)



Therefore by Cramers-Rao Inequality or Cramer-Rao lower bound for an unbiased estimator of p is 1 upon n into p into 1 minus p is equal to p into 1 minus p upon n. Note that this one comes because in the numerator we are writing g prime p square. Here g p is equal to p. Therefore we get in the numerator 1. So, this is the lower bound for an unbiased estimator.

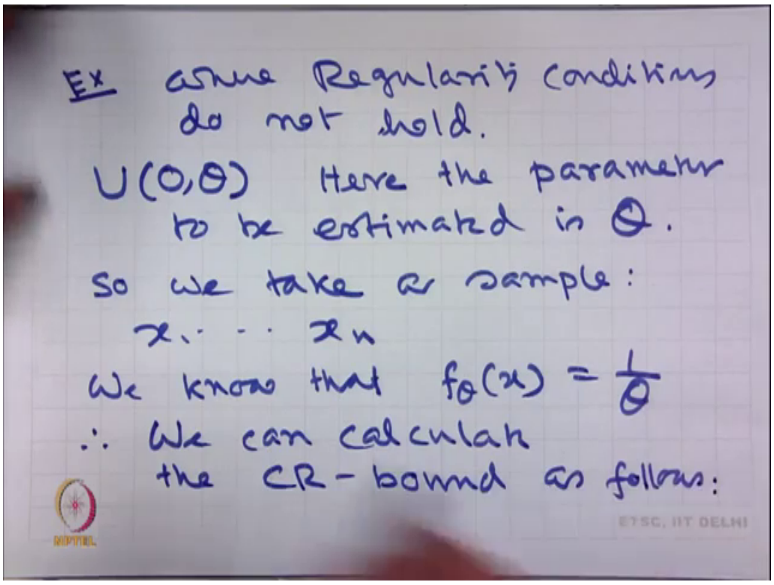
(Refer Slide Time: 14:15)



Now, let us consider X bar is equal to x 1 plus x 2 plus x n by n. Now, variance of X bar is equal to variance of x 1 plus x 2 plus x n divided by n square is equal to n into p into 1 minus p. This we know because x 1 plus x 2 up to x n is a binomial distribution. Therefore, it has variance n p into 1 minus p divided by a square is equal to p into 1 minus p upon n.

Therefore, what we find that variance of X bar is equal to p into 1 minus p upon n which is the Cramer-Rao bound as we have already seen here. Hence, X bar is minimum variance unbiased estimator for p.

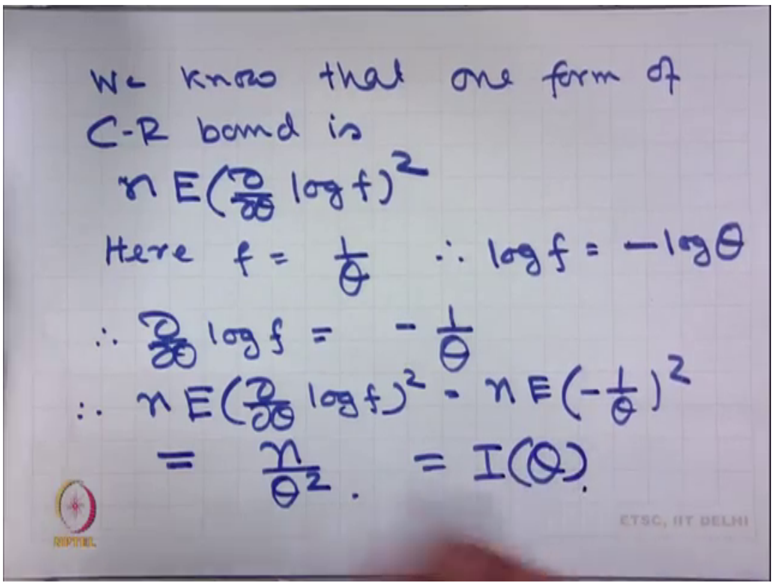
(Refer Slide Time: 15:46)



Let us now consider an example where regularity conditions do not hold. So, for example, consider uniform 0 theta. Here the parameter to be estimated is theta.

So, we take a sample x 1 x 2 x n and we know that f theta of x is equal to 1 upon theta as it is uniform in the interval 0 theta. Therefore, we can calculate the Cramer-Rao bound as follows.

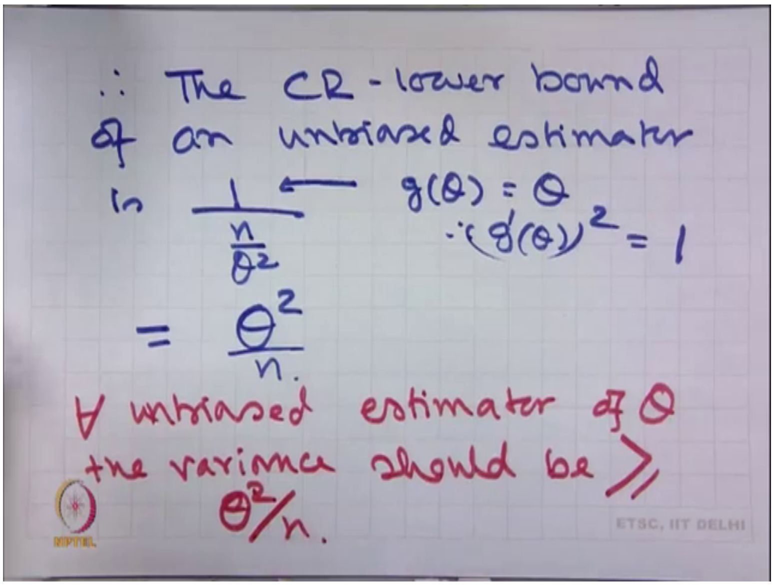
(Refer Slide Time: 17:23)



We know that one form of Cramer-Rao bound is n times expected value of del del theta log f square. Here f is equal to 1 by theta. Therefore, log f is equal to minus log theta. Therefore, del del theta of log f is equal to minus 1 upon theta.

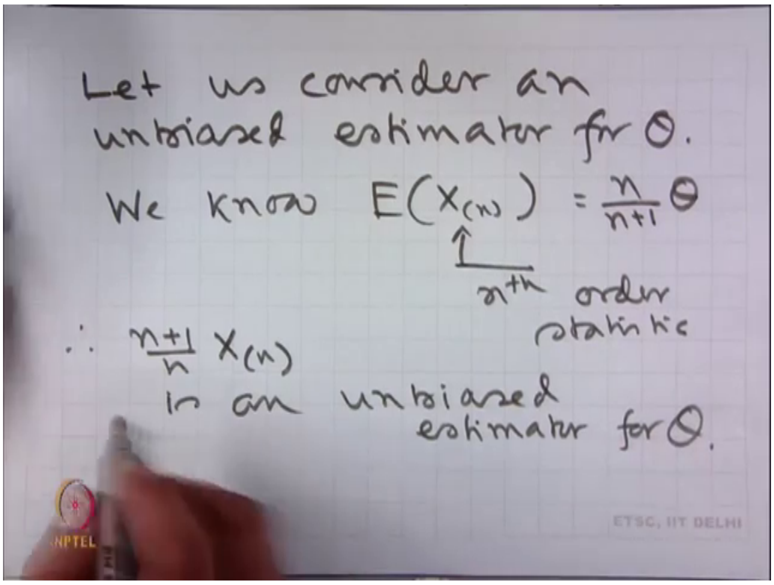
Therefore, n expected value of del del theta log f square is equal to n into expected value of minus 1 upon theta whole square is equal to n upon theta square. So, this is equal to i theta.

(Refer Slide Time: 18:53)



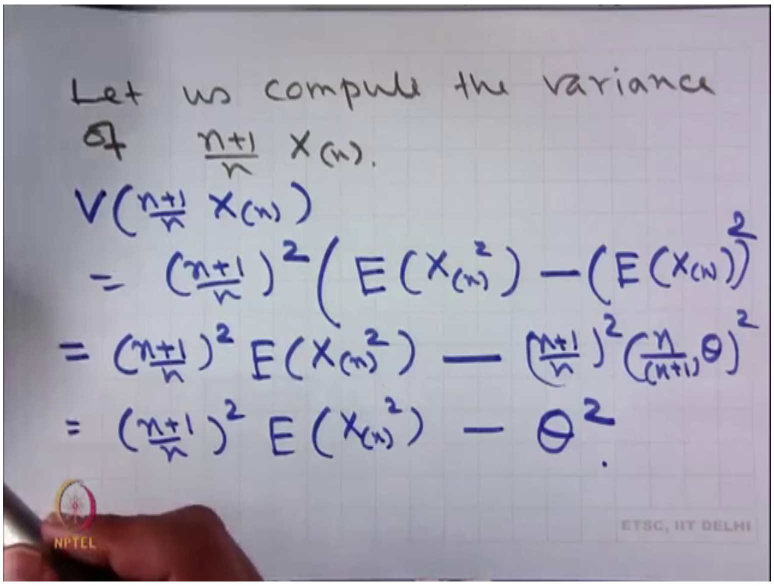
The Cramer-Rao lower bound of an unbiased estimator is 1 upon n by theta square. This one is coming because g theta is equal to theta. Therefore, g prime theta whole square is equal to 1 is equal to theta square by n. Therefore, for all unbiased estimator of theta, the variance should be greater than or equal to theta square by n.

(Refer Slide Time: 20:07)



Let us consider an unbiased estimator for theta. We know expected value of x n the nth order statistic is equal to n upon n plus 1 theta. Therefore, n plus 1 upon n x n is an unbiased estimator for theta.

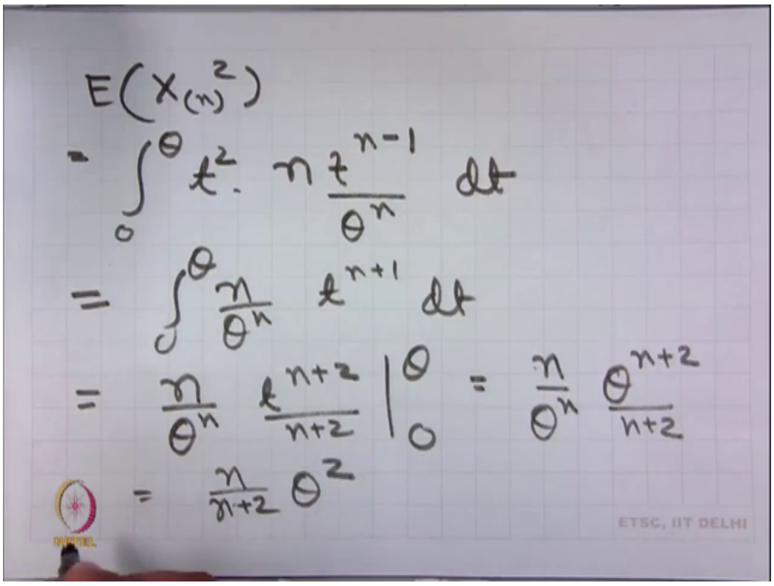
(Refer Slide Time: 21:15)



Let us compute the variance of n plus 1 upon n x n variance of n plus 1 upon n x n is equal to n plus 1 upon n whole square into expected value of x n square minus expected value of x n whole square is equal to n plus 1 upon n whole square into expected value of x n square minus n plus 1 upon n whole square into n upon n plus 1 theta whole square is equal to n plus 1 upon n whole square expected value of x n square minus theta square.

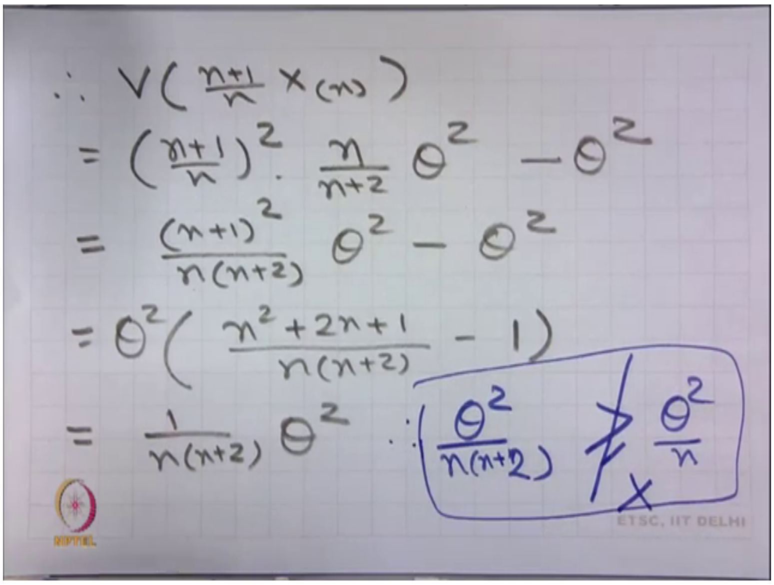
So, let us compute expected value of x n square.

(Refer Slide Time: 22:48)



This is equal to integration 0 to theta. Let the value be t. We know that the pdf of x n is equal to t square and we know that the pdf of x n is equal to n t to the power n minus 1 upon theta to the power n. And therefore, we are integrating it with respected to t. This is equal to 0 to theta n upon theta to the power n t to the power n plus 1, dt is equal to n theta power n t to the power n plus 2 upon n plus 2 theta, 0 is equal to n theta power n theta to the power n plus 2 upon n plus 2 is equal to n upon n plus 2 theta square.

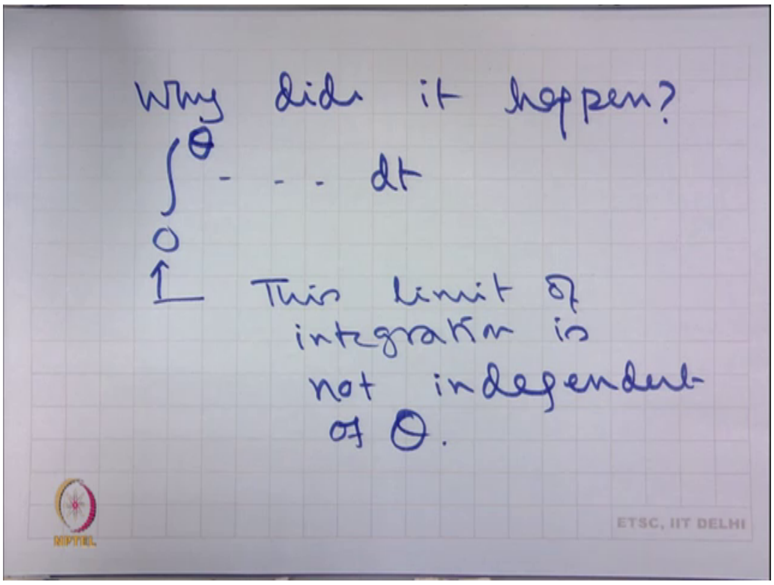
(Refer Slide Time: 24:35)



Therefore, variance of n plus 1 upon n x n is equal to n plus 1 upon n whole square into n upon n plus 2 theta square minus theta square is equal to n plus 1 whole square upon n into n plus 2 theta square minus theta square is equal to theta square into a square plus 2 n plus 1 n into n plus 2 minus 1 is equal to 1 upon n into n plus 2 theta square. And we have already found that the Cramer-Rao bound we have already found that the Cramer-Rao bound is theta square by n.

Therefore, theta square upon n into n plus 2 has to be greater than equal to theta square by n which is not correct for any positive n, because n is the number of samples and if n is equal to an integer 1 2 3. Therefore this is never valid or in other words, we have found an unbiased estimator whose variance does not obey that Cramer-Rao lower bound.

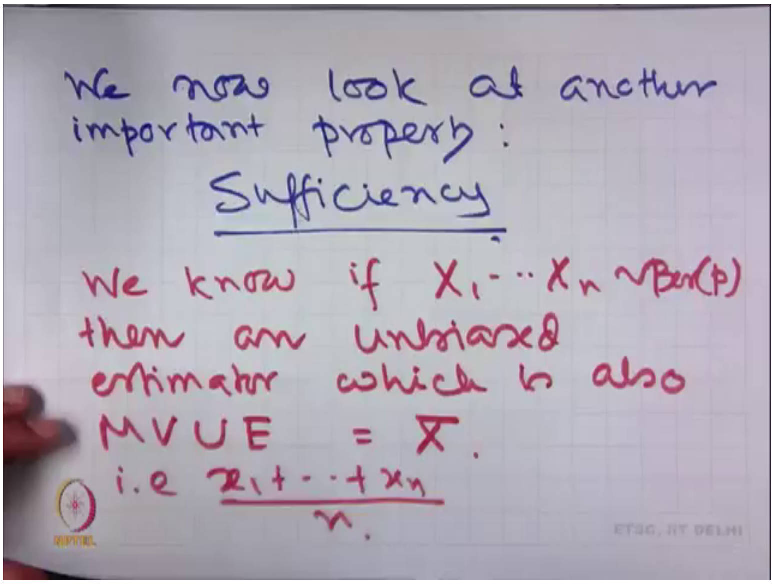
(Refer Slide Time: 26:49)



That question is why did it happen? It happens because the range of integration while computing the variance is 0 to theta dt that what we have done and this limit of integration is not independent of theta. Hence, the regularity condition does not hold and therefore, the Cramer-Rao lower bound is no more acting as a lower bound for the variance of an unbiased estimator in this case.

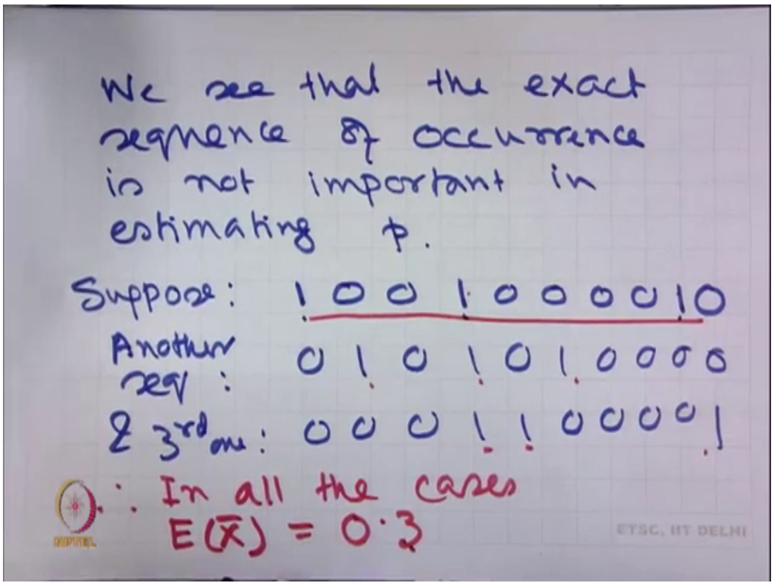
So, that is the importance of the regularity conditions because without them the minimum variance unbiased estimator bound will not be valid with that I close the chapter on efficiency.

(Refer Slide Time: 28:07)



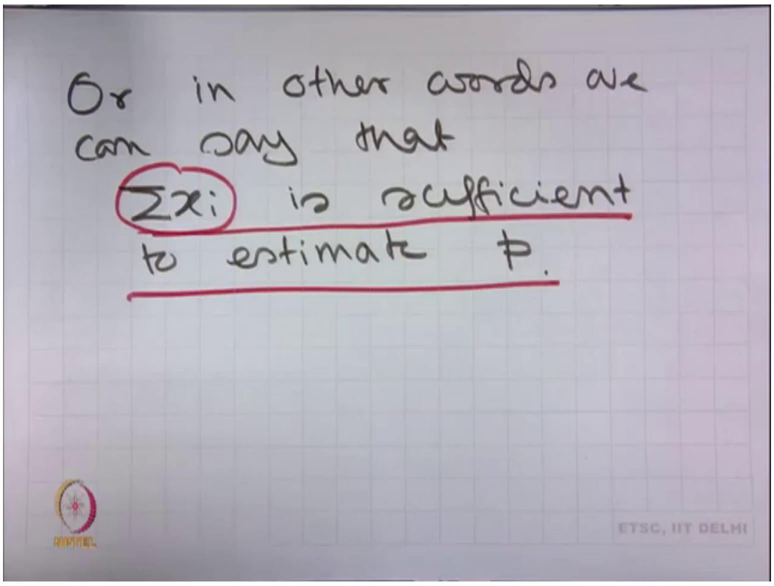
We now look at another important property, namely sufficiency. The concept is as follows. We know that if X 1 X 2 X n are from Bernoulli p, then an unbiased estimator which is also MVUE is equal to X bar that is x 1 plus x 2 plus x n by n.

(Refer Slide Time: 29:30)



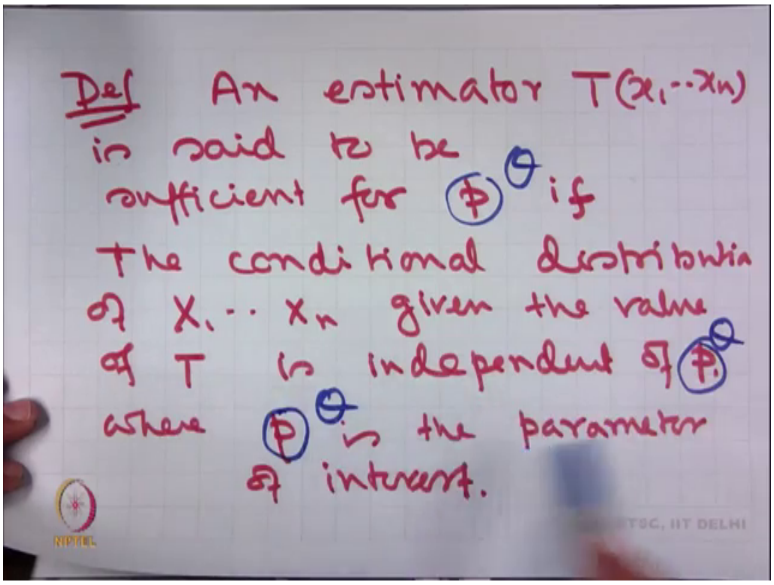
Therefore, if we observe we see that the exact sequence of occurrence is not important in estimating p. Suppose I toss the coin 10 times and my observations are 1 0 0 1 0 0 0 0 1 0, that means out of 10 tosses I have got 3 heads and rest are tails. Suppose another sequence is 0 1 0 1 0 1 0 0 0 0 and the third one is 0 0 0 1 1 0 0 0 0 1. In all the 3 cases, we find that out of 10 tosses there are 3 heads and therefore, in all the cases expected value or the estimate is X bar, then expected value of X bar is equal to 0.3 or in other words, we notice that this sample may contain a lot more information rather than just the summation of that one; namely what is the first result, what is the last result, how many times 0 and 1 occurred together etcetera. All this different informations one can find out from each sample, but while estimating the value of p, we do not really need any one of them. What I need is only the sum of the observations that is the number of heads that I found in the sample.

(Refer Slide Time: 32:35)



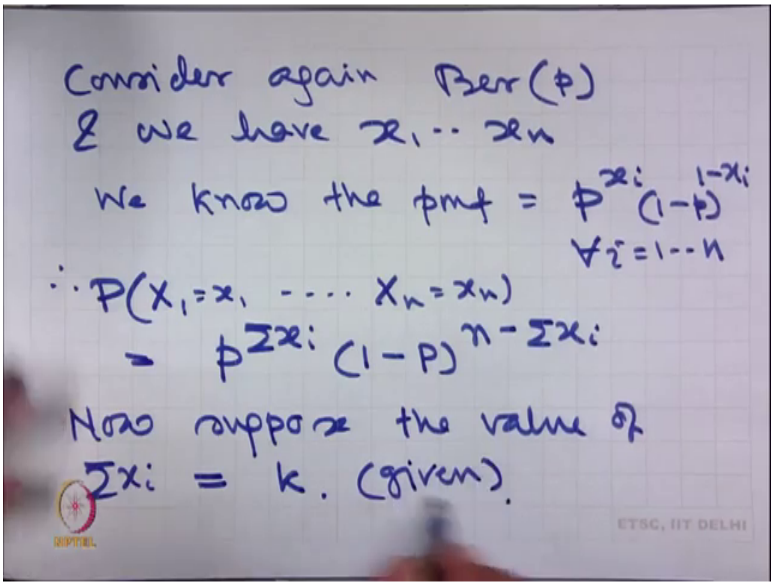
So, in other words we can say that sigma x i is sufficient to estimate p. We do not need any other information. If I get only sigma x i, then that is good enough for us to estimate p. This is the basic idea of sufficiency.

(Refer Slide Time: 33:27)



Mathematically the definition is that an estimator T x 1 x 2 x n is said to the sufficient for p. If the conditional distribution of X 1 X 2 X n given the value of T is independent of p, where p is the parameter of interest or in other words, if instead of p we call it theta, we can say that if T x 1 x 2 x n is said to be sufficient for estimating theta if the conditional distribution of X 1 X 2 X n given the value of Y is independent of theta.

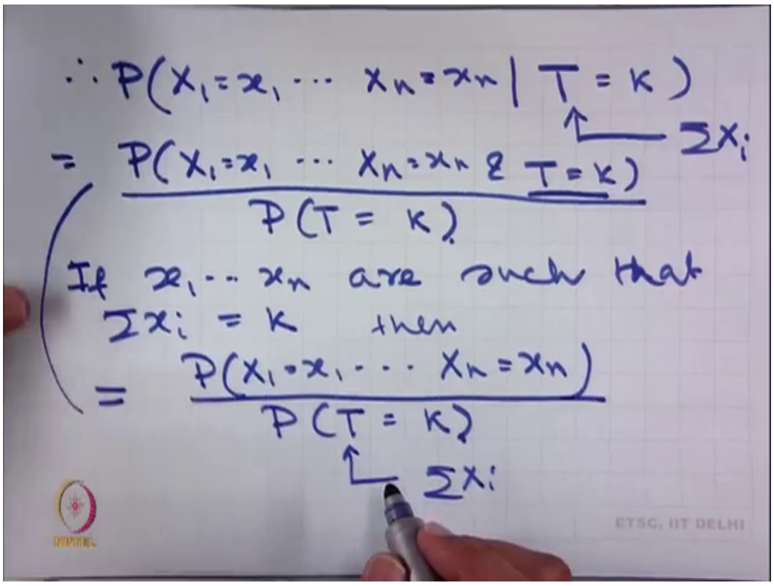
(Refer Slide Time: 35:36)



So, let us look at for Bernoulli and we have x 1 x 2 x n. We know the pdf or the pmf is equal to p to the power x i 1 minus p to the power 1 minus x i for all i is equal to 1 to n.

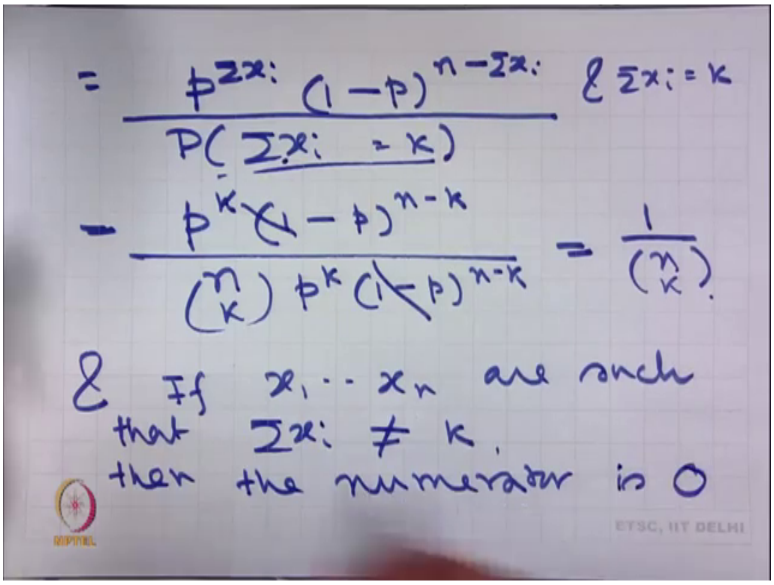
Therefore, P of X 1 is equal to x 1 X n is equal to x n is equal to p to the power sigma x i into 1 minus p to the power n minus sigma x i.

(Refer Slide Time: 37:05)



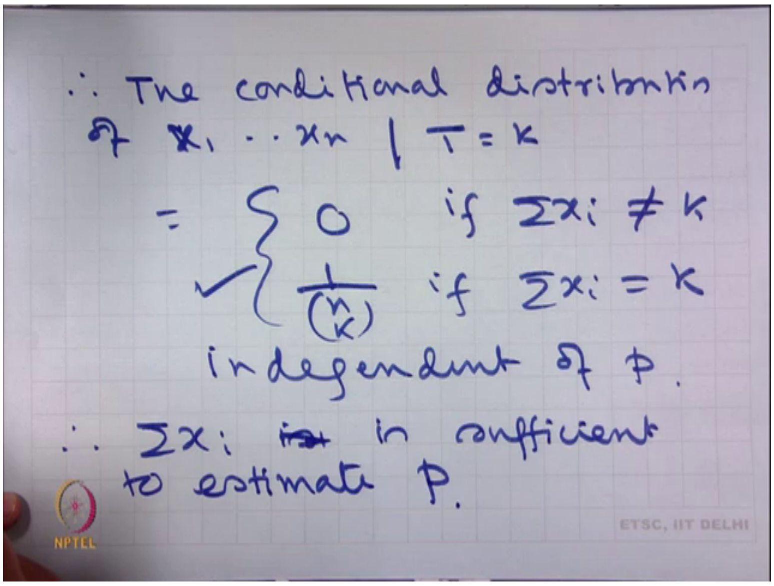
Now, suppose the value of sigma x i is equal to k given, therefore probability X 1 is equal to x 1 up to X n is equal to x n given t is equal to k. Where, t is sigma x i is equal to probability X 1 is equal to x 1 X n is equal to x n and t is equal to k divided by probability t is equal to k. What is this quantity if x 1 x 2 x n are such that sigma x i is equal to k. Then, the above is probability X 1 is equal to x 1 X n is equal to x n being I do not need to consider this because that is going to be satisfied automatically divided by probability t is equal to k. Where, t is equal to sigma x i is equal to p to the power sigma x i 1 minus p to the power n minus sigma x i and sigma x i is equal to k divided by probability t is equal to k that is probability sigma x i is equal to k.

(Refer Slide Time: 38:37)



So, this is nothing but p to the power k 1 minus p whole to the power n minus k divided by since sigma x i is now a binomial random variable with parameter n p, it will take the value k with probability n k p to the power k 1 minus p whole to the power n minus k is equal to you see that these cancels with this. Therefore, we are left with 1 upon n c k and if x 1 x 2 x n are such that sigma x i is not equal to k, then the numerator is 0.

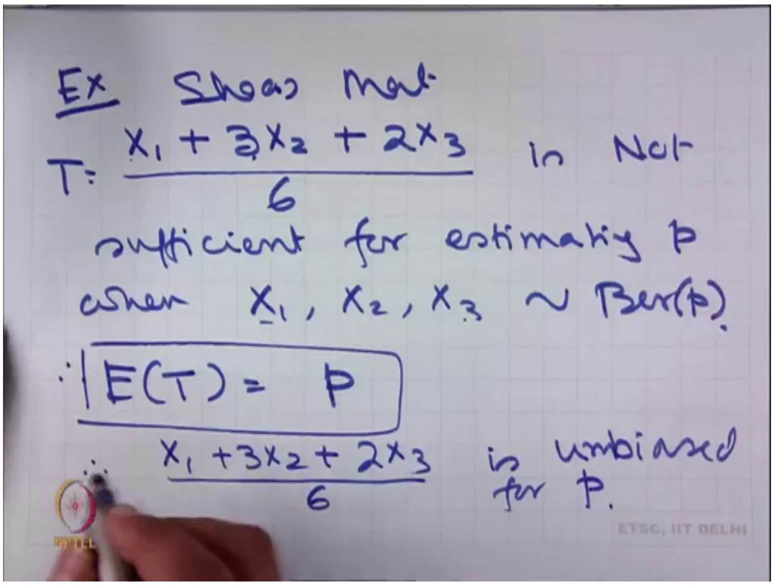
(Refer Slide Time: 40:13)



Therefore, the conditional distribution of x 1 x 2 x n given t is equal to k is 0 if sigma x i not equal to k or this is equal to 1 upon n ck if sigma x i is equal to k.

Therefore, we see that this is independent of p. Therefore, sigma x i is sufficient to estimate p and these we have are good.

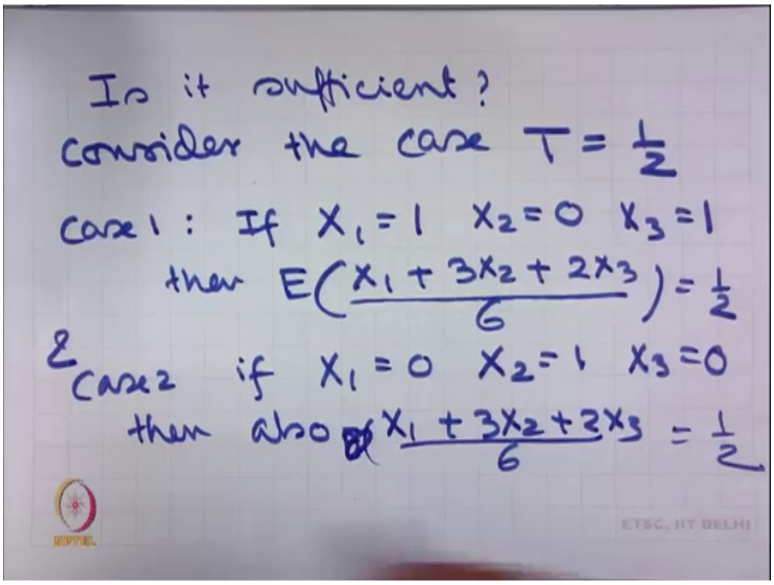
(Refer Slide Time: 41:27)



I have already shown that X 1 plus 3, X 2 plus 2, x 3 plus 6 is not sufficient for estimating p when X 1 X 2 X 3 are from Bernoulli p.

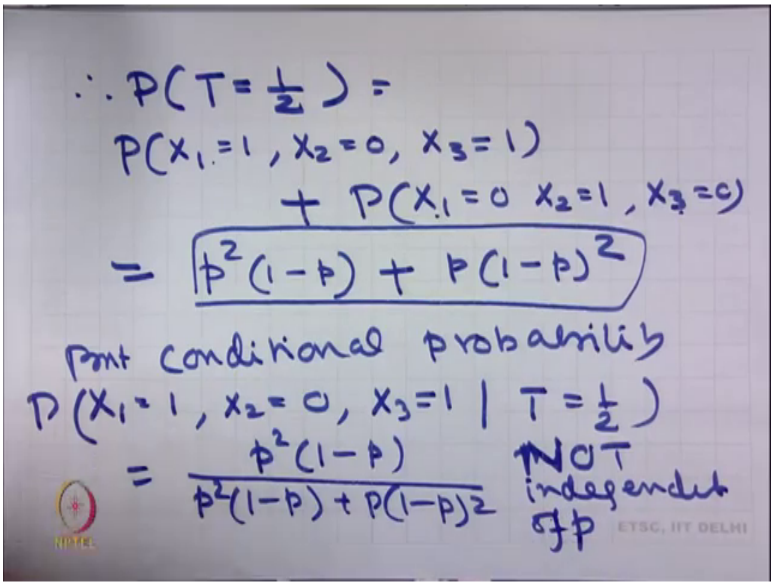
So, we have a Bernoulli distribution. We have taken 3 observations from here and my statistic is X 1 plus 3, X 2 plus 2, X 3 by 6. Let us call it T. Therefore, expected value of T is equal to p because each X i is an unbiased estimator for p. Therefore, p plus 3, p plus 2, p upon 6 is equal to p. Therefore, X 1 plus 3, X 2 plus 2, X 3 upon 6 is unbiased for p.

(Refer Slide Time: 43:13)



Question is, is it sufficient? We say that consider the case when T is equal to half. So, case 1 if X 1 is equal to 1 X 2 is equal to 0 and X 3 is equal to 1, then expected value of X 1 plus 3, X 2 plus 2, X 3 upon 6 is equal to half and case 2, if X 1 is equal to 0, X 2 is equal to 1 and X 3 is equal to 0, then also X 1 plus 3, X 2 plus 2, X 3 upon 6 is it is expected value or this is equal to half.

(Refer Slide Time: 44:50)



Therefore, probability T is equal to half is equal to probability X 1 is equal to 1, X 2 is equal to 0 and X 3 is equal to 1 plus probability X 1 is equal to 0, X 2 is equal to 1, X 3 is equal to 0 is equal to p square into 1 minus p plus p into 1 minus p whole square.

So, that is the probability that the statistic T is going to take the value half, but a conditional probability X 1 is equal to 1, X 2 is equal to 0, X 3 is equal to 1 given T is equal to half is equal to p square into 1 minus p upon p square into 1 minus p plus p into 1 minus p whole square and this is not independent of p. Therefore, we see that even if this is an unbiased estimator X 1 plus 3, X 2 plus 2, X 3 by 6, this is not sufficient to estimate the value of p, ok.

Friends with that I stop here. In the next class, I shall do some more studies of sufficiency.

Thank you.