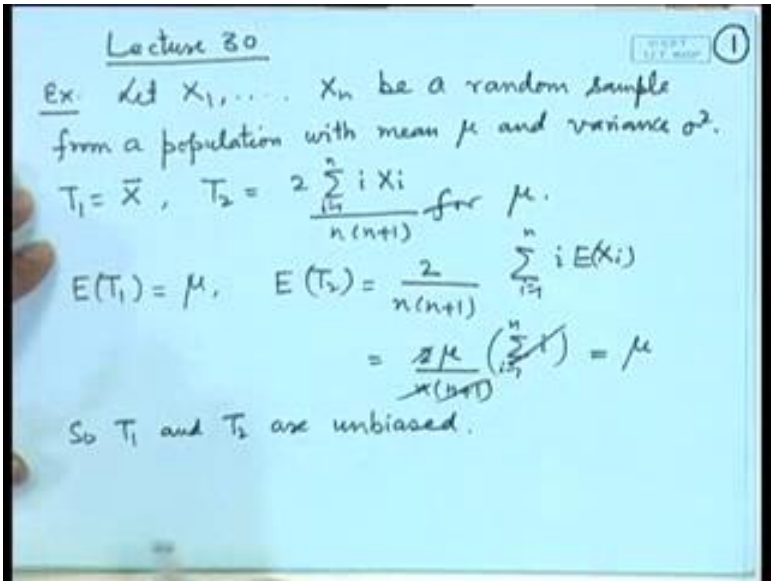
Probability and Statistics

Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 59
UMVUE, Sufficiency,
Completeness

Yesterday we have introduce the criteria that among giving estimators which estimator should be preferred for example, if T 1 and T 2 have two unbiased estimators for the same parameter g theta, then we will prefer T 1 over T 2 if variance of T 1 is less than or equal to variance of T 2. In general if I am considering any 2 estimator; that means they need not be unbiased in that case we will compare the mean squared errors and the estimator with smaller mean squared error will be preferred.

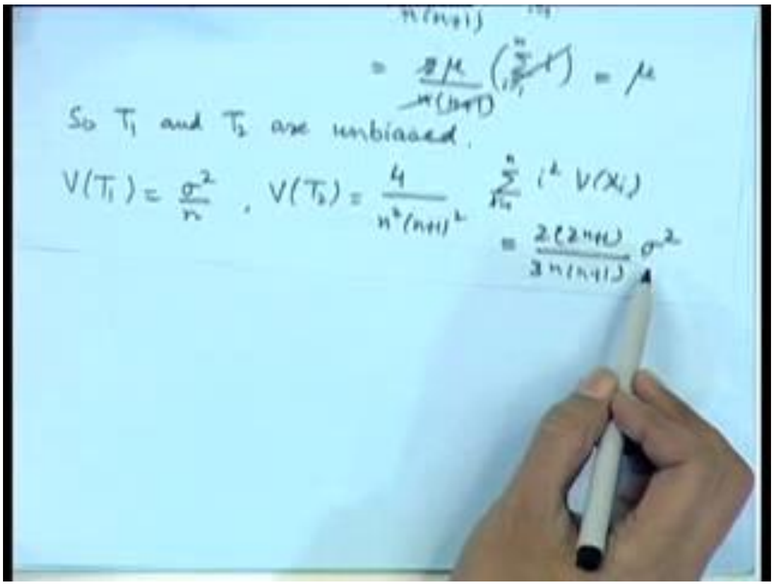
(Refer Slide Time: 00:54)



Let me give one example here suppose we have a random sample from a population with mean mu and variance sigma square. Now let us consider estimators T 1 and T 2 for mu let us see. So, what is expectation of T 1 naturally it is equal to mu; what is expectation of T 2, you can apply the linearity property of the expectation so this becomes twice divided by n into n plus 1, sigma i expectation of x i. Now expectation of x i is mu, so this reduces to 2 mu by n into n plus 1 sigma of i, i is equal to 1 to n. Now this is nothing, but n into n plus 1 by 2. So, this cancels out with this and you get that both T 1 and T 2 are unbiased estimator.

So, T 1 and T 2 both of them are unbiased.

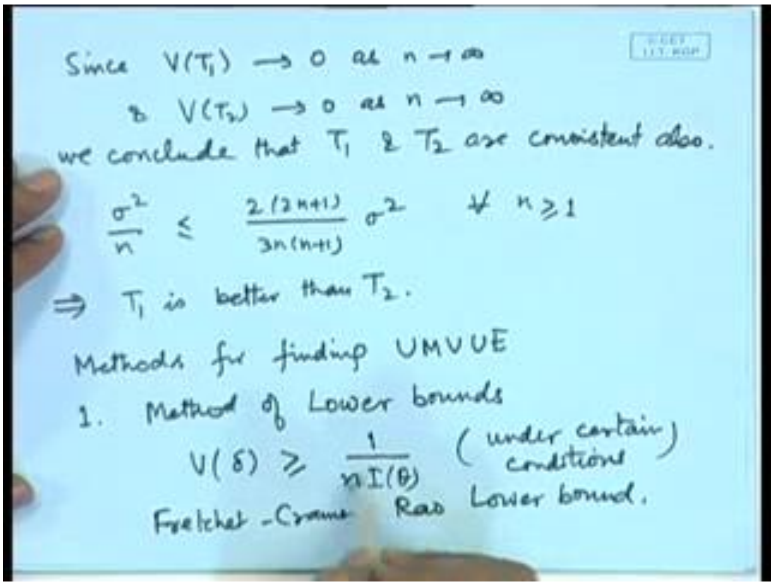
(Refer Slide Time: 01:58)



Let us look at variances. So, what is variance of T 1? Variance of T 1 is sigma square by n we have already shown that the variance of the sample mean is equal to sigma square by n, let us consider the variance of T 2. Now variance of T 2 because of the independence it becomes 4 by n square into n plus 1 square, sigma I square variance of x i. Now variance of x i is sigma i square and this is sigma of sigma i square from 1 to n that is the sum of the first squares of first an integers, that is n into n plus 1 into 2 n plus 1 by 6.

So, after simplification this quantity turns out to be twice into 2 n plus 1 divided by 3 n into n plus 1 sigma square. So, now, the question is that we can also check the consistency here for example, both of them are unbiased and variance of T 1 goes to 0 as n tends to infinity, variance of T 2 also goes to 0 as n tends to infinity.

(Refer Slide Time: 03:10)



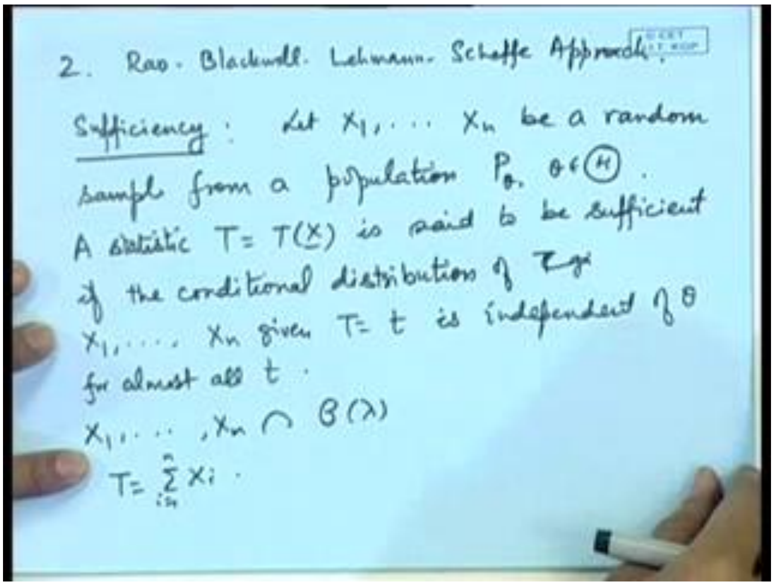
So, since variance of T 1 goes to 0 as n tends to infinity and variance of T 2 goes to 0 as n tends to infinity, we conclude that T 1 and T 2 are consistent.

So, we have 2 estimators both of them are unbiased both of them are consistent also. So, again which one you will prefer? So, we compare the variances you can see easily that sigma square by n is less than or equal to 2 into 2 n plus 1 by 3 n into n plus 1 sigma square for all n greater than or equal to 1. Actually for n is equal to 1 the 2 sides will be equal. So, this implies that T 1 is better than T 2; now the question comes that among a set of given estimators we can find by comparing the variances or the mean squared errors, but in the first place how to find the best among them.

So, because the total set of estimators is infinite. So, we need certain other methodology there are 2 methods for finding out the unbiased methods for finding UMVUE, one method is the method of lower bounds; under certain given conditions variance of an unbiased estimator is greater than or equal to a prescribed number it is 1 by n times I theta this is under certain conditions this is called Frechet Cramer Rao lower bound.

So, if there is an estimator which will have this variance equal to this that will be naturally minimum variance and biased estimator, then later on the generalizations of this Frechet Cramer Rao bound have been done and we have the bounds when we have multi parameters situation when we can use higher order derivatives etcetera so, but for application of these lower bounds certain conditions need to be satisfied and the bounds may not always be obtain.

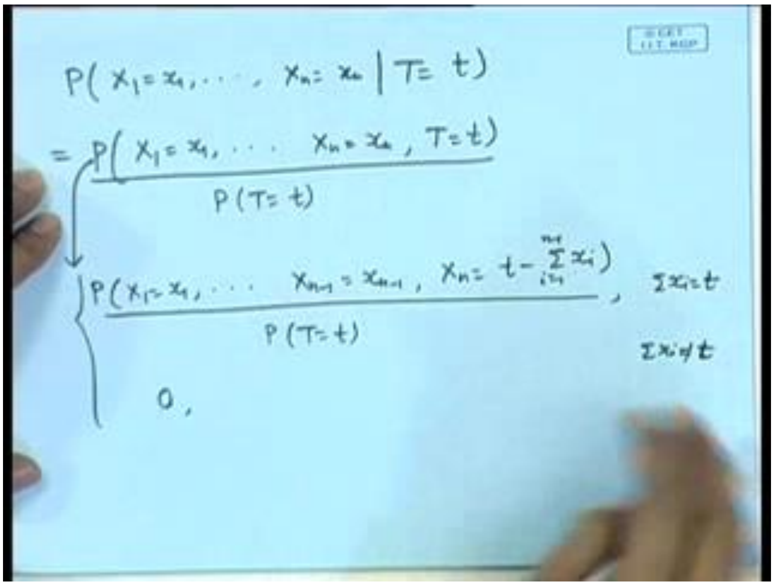
(Refer Slide Time: 06:04)



There is another approach that is called Rao Blackwell Lehmann Scheffe approach we introduce 2 concepts that is of sufficiency and completeness so firstly, we define what is sufficiency.

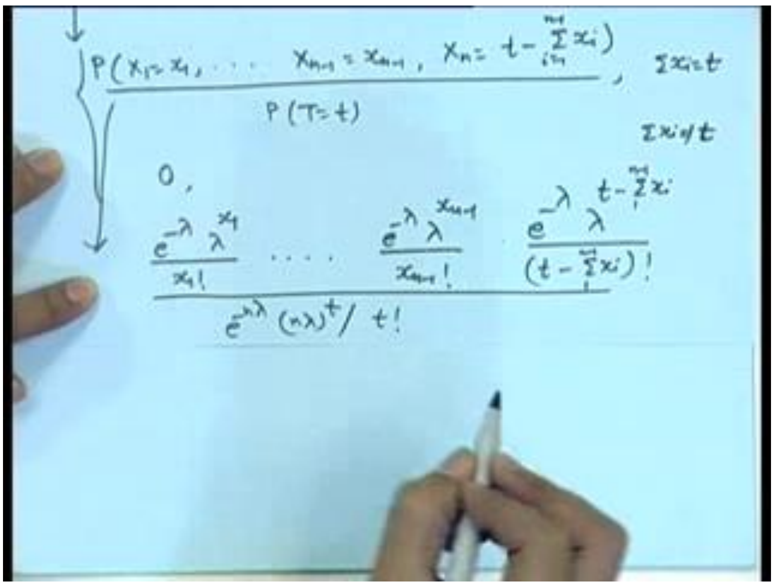
So, we have the regular model that X 1, X 2, X n random sample from a population say P theta, theta belonging to step theta; then a statistic T is said to be sufficient if the conditional distribution of X 1, X 2, X n given T is equal to T is independent of theta for almost all T. Let us take an example here suppose I consider X 1, X 2, X n follow say poisson lambda distribution, let us define T to be sigma of x i, i is equal to 1 to n.

(Refer Slide Time: 08:08)



Let us consider the conditional distribution of X 1, X 2, X n given T is equal to t. So, this is equal to probability of X 1, is equal to X 1 and so on X n is equal to X n T is equal to T divided by probability of T is equal to T. Now here T is sigma x i and we know that it follows poisson n lambda. So, the denominator quantity can be written; how to find out the numerator quantity we simplify this, we can write it as probability of X 1, is equal to X 1 and so on, x n minus 1 is equal to x n minus 1 and x n is equal to T minus sigma x i, i is equal to 1 to n minus 1, this is valid if sigma x i is equal to T otherwise it is defined to be 0, because it is conditional on t so for every value of T we have to determine this.

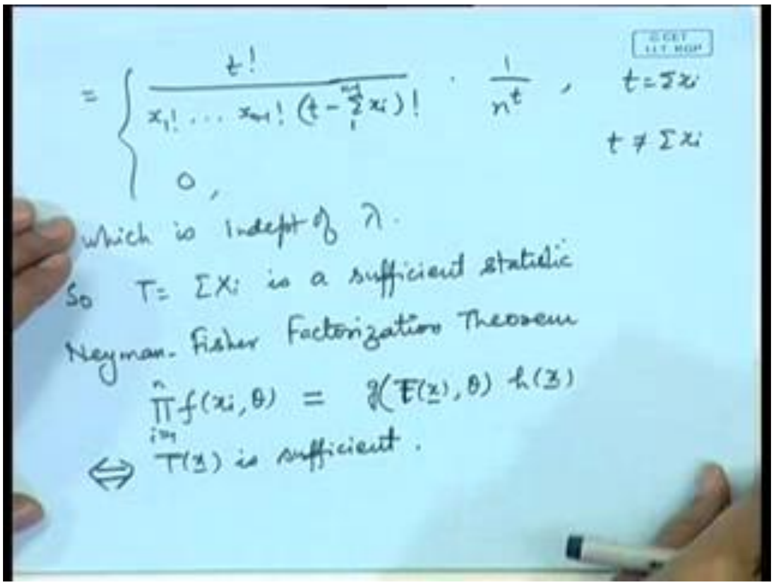
(Refer Slide Time: 09:28)



Now, the numerator quantity can be determined because X 1, X 2, X n are independent poisson lambda variables. So, we can substitute these values here; e to the power minus lambda, lambda to the power X 1, by X 1, factorial and so on and the last term will be e to the power minus lambda, lambda to the power T minus sigma x i, 1 to n minus 1 divided by T minus sigma x i, 1 to n minus 1 factorial and the divided by e to the power minus n lambda, n lambda to the power T divided by T factorial.

You can easily see that e to the power minus lambda term cancels out because we have n terms here and in the denominator we have e to the power minus n lambda, the powers of lambda that is lambda to the power t here and in the denominator we have lambda to the power t. So, that also cancels out.

(Refer Slide Time: 10:26)



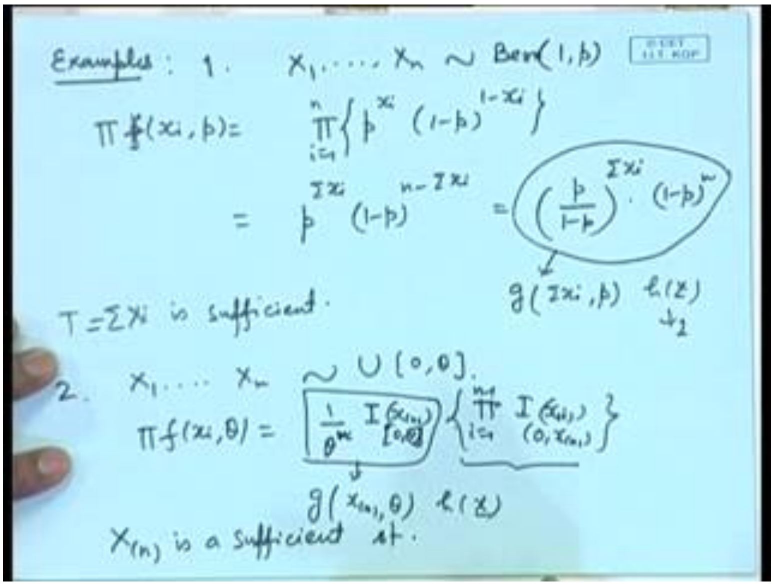
So, we are left with t factorial divided by x 1, factorial and so on x n minus 1 factorial t minus sigma x i, 1 to n minus 1 factorial and 1 by n to the power t, when t is equal to sigma x i and it is equal to 0, if t is not equal to sigma x i.

Now, you can see here this term does not depend upon lambda; again independent of lambda. So, we conclude that T that is sigma X i is a sufficient statistic, the role of sufficiency is quite important in a statistical inference. In fact, it means that we can generate an alternative sample say X 1, prime X 2 prime X n prime given t is equal to t; that means, whatever information about the parameter can be drawn from the sample X 1, X 2 X n all of that is contained in sigma x i; that means, there is no additional information in X 1, X 2 X n which is not there in sigma x i, this allows us to make the data compact because we need not keep record of all the individual observations rather we keep record of only the sufficient statistics.

Now, this method of proving that sigma X i sufficient involves finding out the conditional distribution, and which may be quite cumbersome for various problems; and another things is that here we have to guess also that what would be a sufficient statistic. So, there is another result which is known as Neyman-Fisher Factorization Theorem, which allows us to figure out what will be a sufficient statistic in a given problems. I will not a state the theorem in a full form rather we look at the practical accept theorem.

We write down the joint density that is product of f x i theta, I is equal to 1 to n; this is a joint probability density function of X 1, X 2, X n, if this can be factorized as g t x theta into h x, where the first term depends upon x i is only through t and the second term is free from theta then we say that this implies and implied by the T x is sufficient. The proof of this involves slightly major theoretic consideration, so we skip the proof, but this is a very practical way of obtaining sufficient statistics.

(Refer Slide Time: 13:43)



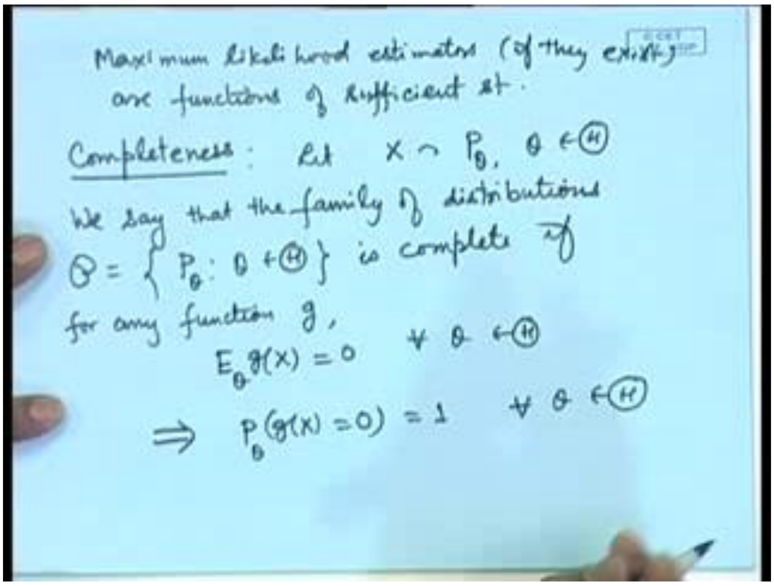
So, let us look at the applications of this let us consider say X 1, X 2 X n follows Bernoulli distribution. So, the joint distribution here product i is equal to 1 to n, p to the power x i, 1 minus p to the power 1 minus x i, that is p to the power sigma x i, 1 minus p to the power n minus sigma x i, this we can write as p by 1 minus p to the power sigma x i multiplied by 1 minus p to the power n. Now you can see here this term is a function of sigma x i and p alone and h x we can take to be 1.

So, this proves that sigma x i is sufficient; let us look at the practical aspect of it, if we have conducted n Bernoulli trials, we may will be interested and we have want to draw certain inference on the proportion of the success, then you can see that sigma x i is actually the number of successes here. So, that gives the full information about p, we do not have to keep track of individual x i, Suppose we consider say uniform distribution, then the joint density is equal to 1 by theta to the power n. Now one may say that if we write like this then where is the variable coming in which will be sufficient, but this is not a complete description because for complete description we need to write down the range of the variables which is each of the x i is from 0 to theta.

So, we can write it in the terms of indicator function, that x n is from 0 to theta and remaining x i they are between 0 to x n product i is equal to 1 to n minus 1. So, this part we can consider as g of x n and theta and this part we can consider as h x. So, by factorization theorem we conclude that x n is a sufficient statistic; let us also correlate with the discussion that we had in the previous lecture about the maximum likelihood estimators. The derivation of the maximum likelihood estimator involved the full probability density function or probability mass function of X 1, X 2, X 1, which we termed as the likelihood function, and that function we maximized with respect to the parameter.

Now, if you look at the factorization then this term does not play a role because if I take l n for example, then this term will be separated at out and the maximization problem reduces only to the maximization of this function. So, naturally theta will become a function of T x alone.

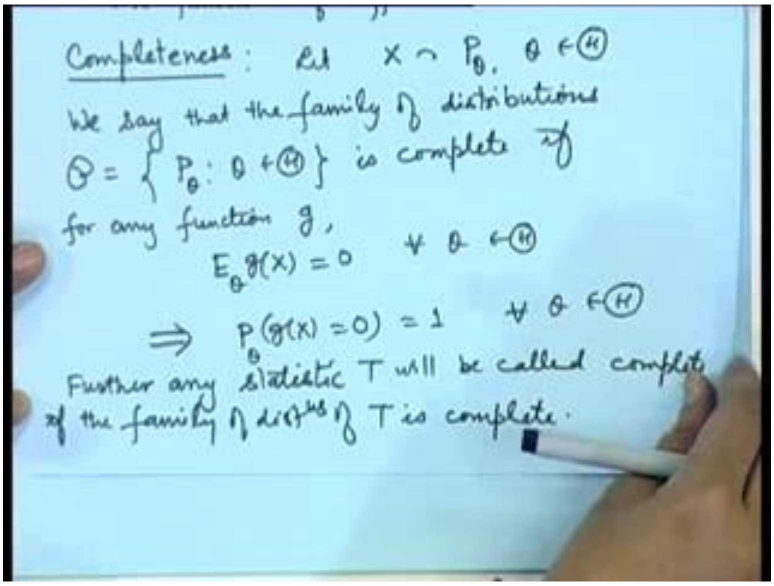
(Refer Slide Time: 17:18)



So, we conclude here that maximum likelihood estimators if they exist are functions of sufficient statistics. So, that brings as the importance of the sufficiency; that means, whatever inference we draw finally, we can restrict attention to the sufficient a statistics.

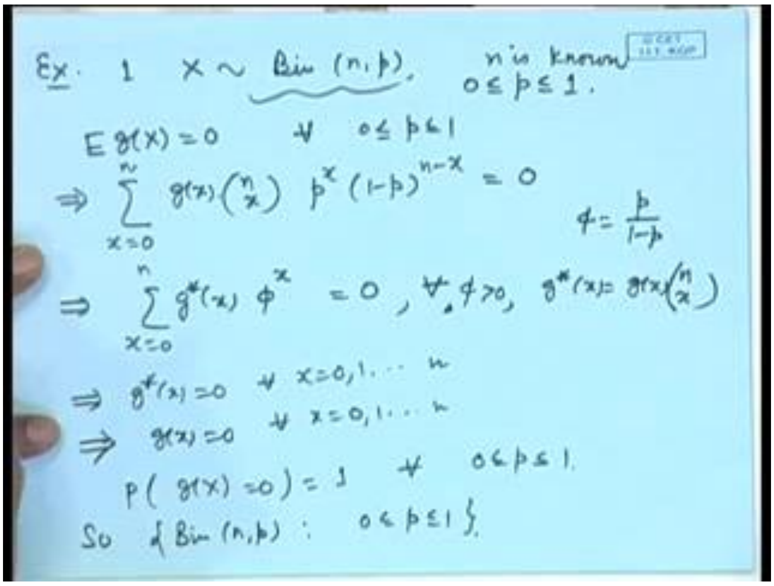
We will look at further examples of this letter; let me introduce another concept called completeness. So, let X follow a distribution say P theta, theta belonging to theta. So, we say that the family of distributions P is equal to P theta, is complete if for any function g, expectation of g x is equal to 0 for all theta implies probability of g x is equal to 0 is equal to 1.

(Refer Slide Time: 19:08)



Further any statistic T will be called complete if the family of distributions of T is complete. So, let me give the example here and explain that what is the meaning of this, what we are saying is that whenever expectation of g X is 0 that function itself is 0; that means, the only unbiased estimators of 0 are 0 itself; now this is a very important statement and let us see that why this is true for various distributions.

(Refer Slide Time: 20:03)



So, let us take say X follows binomial n, p where n is known and p is the parameter. Let us look at expectation of g X is equal to 0, now this condition is equivalent to g x, n c x, P to the power x, 1 minus p to the power n minus x is equal to 0 for x is equal to 0 to n. We may write this as we may introduce a term called say phi is equal to p by 1 minus p then this term I can write as sigma g star x, phi to the power x is equal to 0, for x equal to 0 to n, where g star I have written as g x into n c x.

Now, the left hand side is a polynomial of degree n in phi and we are saying that it is vanishing for all phi. So, the polynomial will vanish identically on an interval provided all its coefficients vanish; that means, g star x is 0 for all x is equal to 0, 1 to n which implies that g x itself is 0 for all x is equal to 0, 1 to1. So, probability that g x is equal to 0 will be 1 for all p. So, this family of binomial distributions is a complete family of distributions.