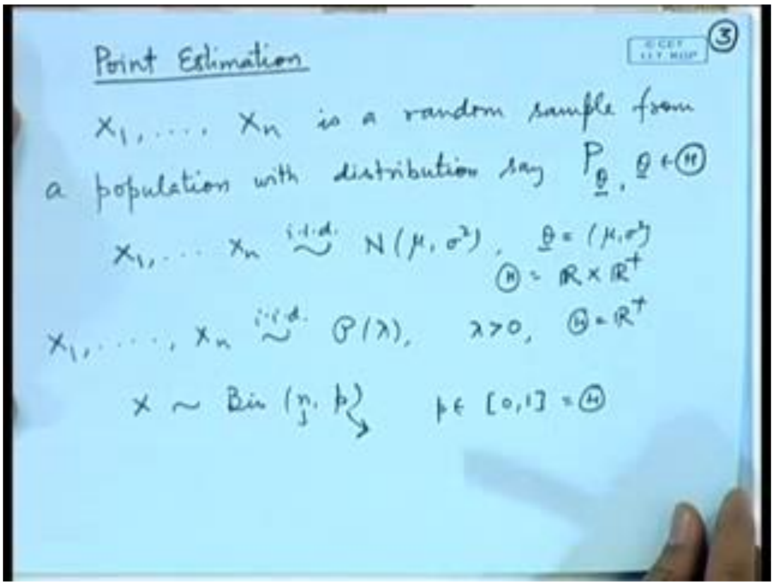
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute Technology, Kharagpur

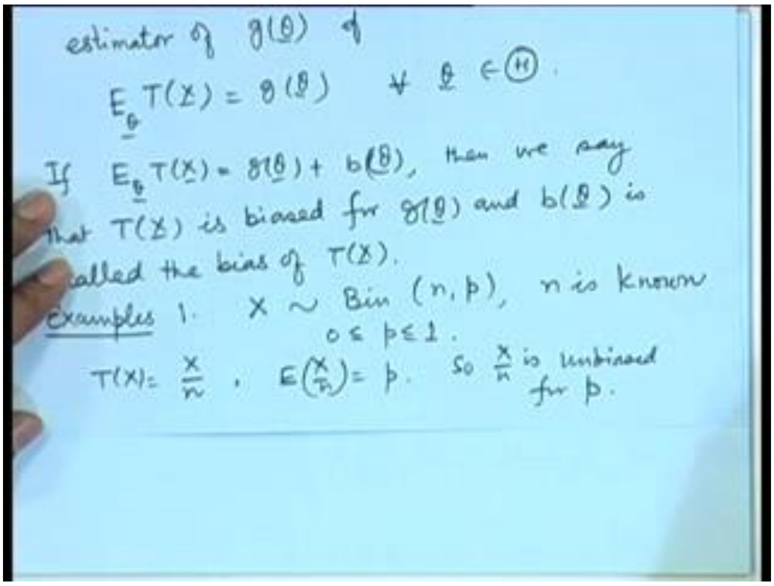
Lecture - 54
Unbiased and Consistent Estimators

(Refer Slide Time: 00:21)



Now, one of the first concepts in the point estimation can be as (Refer Time: 00:28) that when I specify that for using for estimating average heights of say persons of a community, I take a sample and I make use of the sample mean. Then the question arises; is it alright to do that; that means we are actually giving a value based on the sample. So, it may be less than the true value or it may be more than the true value, then is on the average this value equal to the true value so; that means, on the average the kind of errors that we will be making plus and minus they cancel out each other; this is the criteria of un-biasness.

(Refer Slide Time: 01:15)

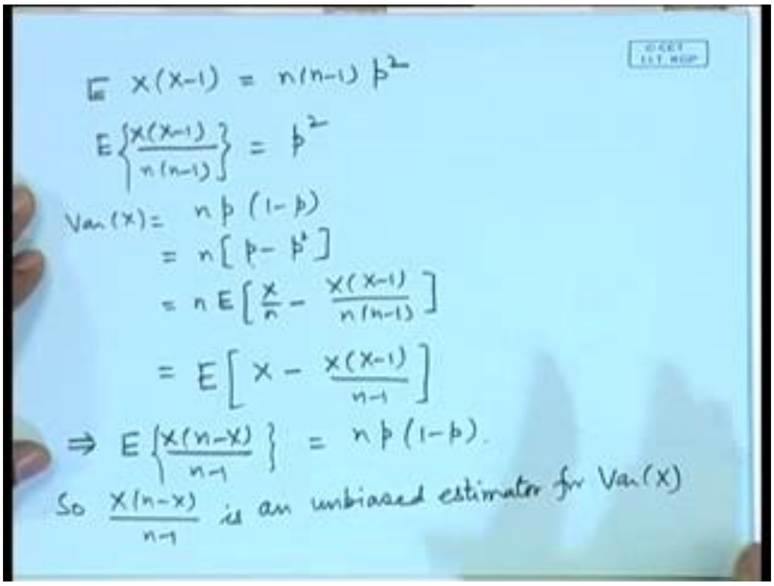


So, we have unbiased estimation, so now we have already mentioned that we will be making use of the functions of X 1, X 2, X n. So, T of X 1, X 2, X n or you can say T X. So, we will use this notation for a statistic and therefore, we will use it as an estimator. So, a statistic T X is said to be an unbiased estimator of g theta, now I am writing a parametric function because if I have certain parameter then some function of that we will be interested for example, I may be interested in mu; I may be interested in sigma square, I may be interested in sigma or I may be interested in a linear function of mu and sigma.

Here I may be interested in lambda; here I may be interested in n p etcetera. So, in general I am interested in any parametric function; if the average value of T X is equal to g theta; for all theta. So, if it is not equal then it may be equal to some value say g theta plus some b theta, then we say that T X is biased for g theta and b theta is called the bias of T x. So, let us consider certain examples, so let me take X follows binomial says n p, here n is known and p is a parameter.

So, I may be interested to estimate p because what is p; p is the probability of success or p is a proportion. So, if I consider say T X is equal to X by n; we know in binomial distribution expectation of X is equal to n p, so expectation of X by n is equal to p. So, X by n is unbiased for the population proportion; of course, it may not be that we are interested only in p; I may be interested in the variance term for example, variance in binomial is n p q that is n p into 1 minus p, I may be interested in p square.

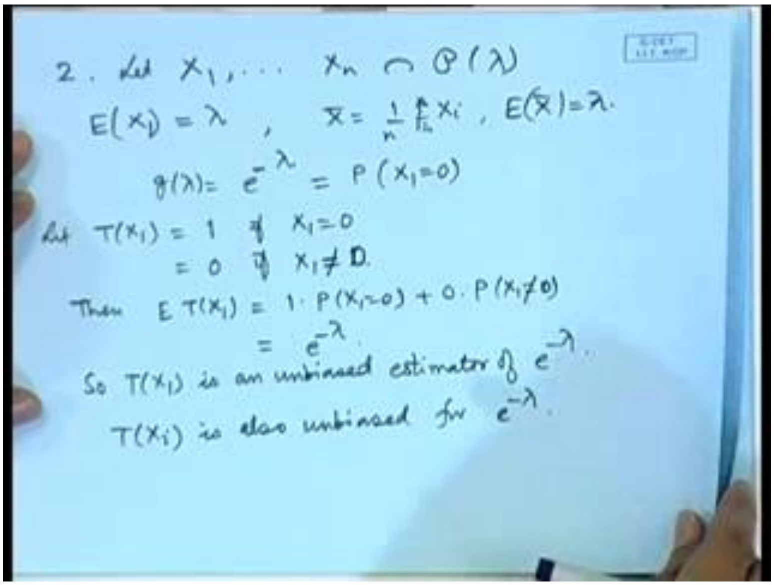
(Refer Slide Time: 05:00)



So, let us see that whether we can do that; if I consider say expectation of say X into X minus 1; then in binomial distribution we know equal to n into n minus 1; p square; that means, I have an estimate of p square here. So, expectation of X into X minus 1 divided by n into n minus 1 is equal to p square, so I have an unbiased estimate of p square. Now suppose I want to estimate say variance that is n p into 1 minus p, I can it write as n p minus p square. Now for p; I can write X by n and for p square; I write X into X minus 1 by n into n minus 1 and let me multiply by n here.

So, this becomes expectation X minus X into X minus 1 by n minus 1. So, this implies expectation of X into n minus X by n minus 1; this is equal to n p into 1 minus p, so X into n minus X by n minus 1 is an unbiased estimator for variability because in the population I may be interested in estimating the variability also. So, here we are able to derive an unbiased estimator further; let us take another problem.

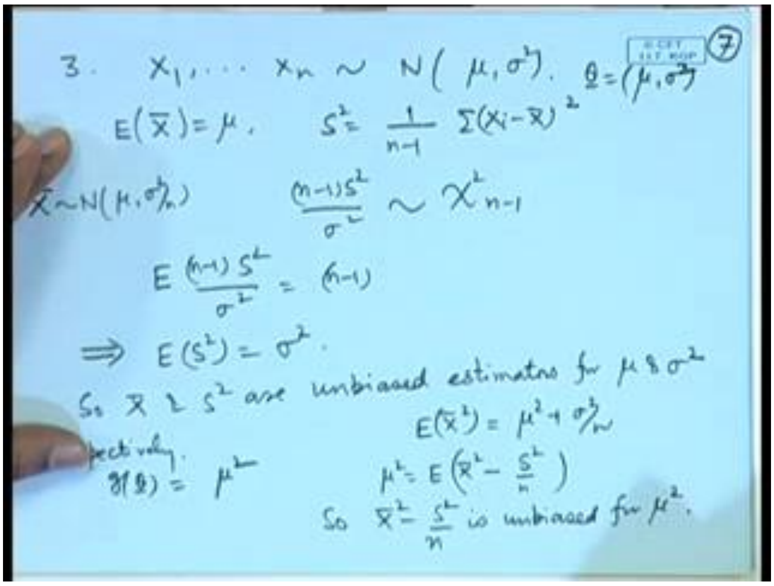
(Refer Slide Time: 07:00)



Let X 1, X 2, X n follows Poisson lambda distribution, so here lambda is the parameter. Suppose I want to estimate lambda itself then I may use say X 1, so expectation of X 1 is lambda one may suggest using X bar that is 1 by n, sigma X i then expectation of X bar is also lambda, so we can have several unbiased estimators for the same parameter. We may be interested to estimate say g lambda that is equal to e to the power minus lambda; what is this term it is actually the probability of observation being equal to 0. In Poisson case; this is important for example, if we are looking at say arrival at certain service point of customers, then it is important to know the time or proportion of the time for which there will be no customer. So, the service company or the service provider can actually plan in such a way that for the time when there are no customers, the service personal may not be implied so that they can make some savings.

So, the 0 probability is of interest; so we may create an estimator like this T X 1 is equal to 1; if X 1 is equal to 0; it is equal to 0 if X 1 is equal to 1. Then if I look at expectation of T X 1 then it will be equal to one into probability of X 1 is equal to 0 plus 0 into probability of X 1 is equal to 1 or we may put X 1 not equal to 0 rather than 1, so X 1 not equal to 0, so that is equal to e to the power minus lambda. So, we are able to create an unbiased estimator of course, one may say that T X 2 or T X I in general unbiased, so which one should be used; so we will come to this question a little later.

(Refer Slide Time: 09:48)

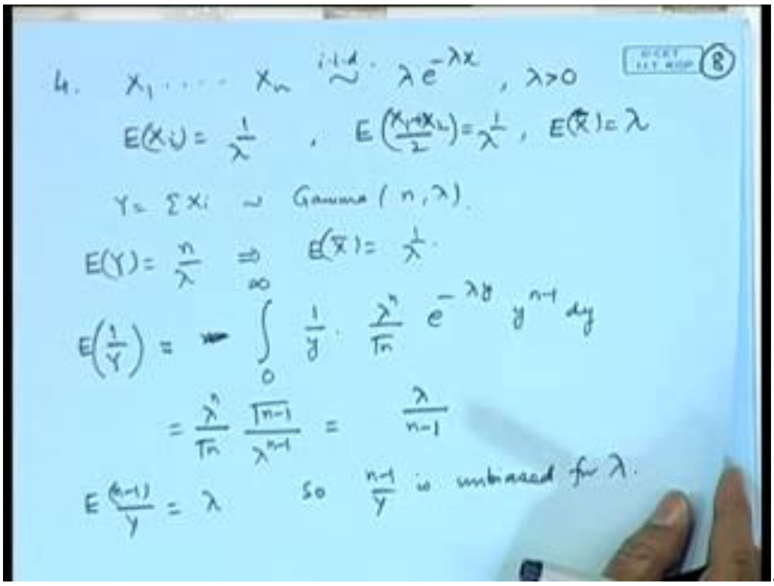


Let me take say X 1, X 2, X n a random sample from say a normal mu sigma square population. If I am interested to estimate mu; I may use a X bar, so expectation of X bar is equal to mu. Now we know here that variance is sigma is square and suppose I am interested to estimate that then I may make use of say S square that is 1 by n minus one sigma X I minus X bar whole square. I have already proved that n minus 1 S square by sigma square follows; chi square distribution on n minus one degrees of freedom.

So, if I look at expectation of n minus 1 S square by sigma square that is equal to n minus 1; this means expectation of S square is equal to sigma square. So, X bar and S square are unbiased estimators for mu and sigma square respectively. One may even be interested in certain different parametric function, in this particular case we may be interested saying mu square say.

So, suppose my g theta; here theta is mu sigma square and I am interest to estimate say mu square, then I may consider something like this; you make use of the distributional properties X bar follows normal mu sigma square by n. So, expectation of X bar square; that is equal to mu square plus sigma square by n. So, I can subtract the estimate of sigma square by n from here, so mu square becomes expectation of X bar square minus S square by n. So, X bar square minus S square by n is unbiased for mu square.

(Refer Slide Time: 12:28)

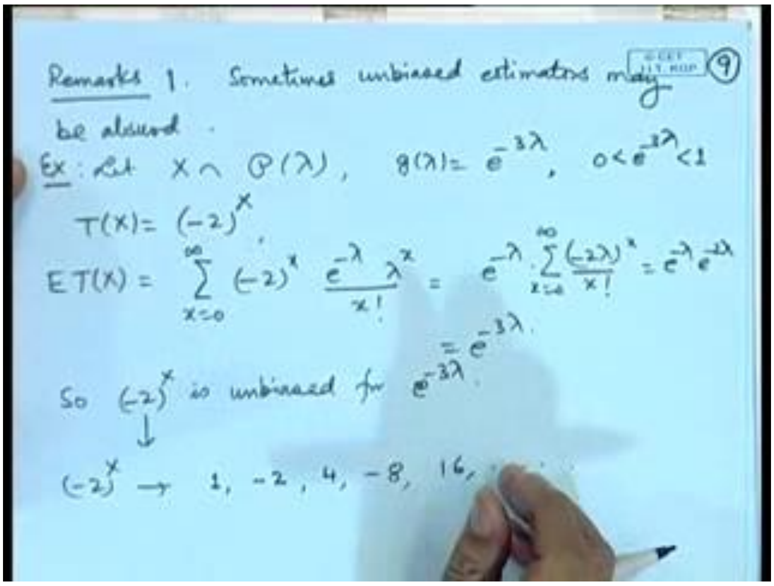


Let X 1, X 2, X n follow say exponential distribution I may be interested to estimate the mean here, I may be interested to estimate say lambda here. So, if I am interest to estimate say mean; I may consider expectation of X I that is equal to 1 by lambda. So, I may consider expectation of X 1 plus X 2 by 2 that is also 1 by lambda; expectation of X bar is also lambda. So, we will come to the question that which one we should chose among these, if I interested to estimate say lambda itself then I may consider for example, here I may define say y is equal to sigma X i and that will follow gamma n lambda. So, then we know expectation of Y is equal to n by lambda; this implies expectation of X bar is equal to 1 by lambda. I may consider the reverse, what is expectation of say 1 by Y then one can show that actually it is equal to n minus it is equal to; so, one may look at the distribution 1 by y.

Now this is gamma and lambda, so we can write it lambda to the power n by gamma n; e to the power minus lambda y, y to the power n minus 1 d y 0 to infinity, which is equal to gamma n minus 1 lambda to the power n by gamma n divided by lambda to the power n minus 1; that is equal to lambda by n minus 1. So, we get that expectation of n minus 1 by Y is equal to lambda. So, n minus 1 by Y is unbiased. Exponential distribution you may remember that I have introduce this lambda as the arrival rate in the Poisson process or I had introduced a term called instantaneous failure relate or the hazard rate.

So, lambda was the hazard rate, so if you want to estimate the hazard rate; we have an estimator for that here. So, this unbiased estimation can be done and one can actually look for the desirable estimates which are unbiased. So, they satisfy the property that their average value is equal to the true value of the parameter. Statically speaking which is a very nice concept because if we are repeating the process several times, then the errors which we make in the actual estimation are even doubt in the long run.

(Refer Slide Time: 15:58)

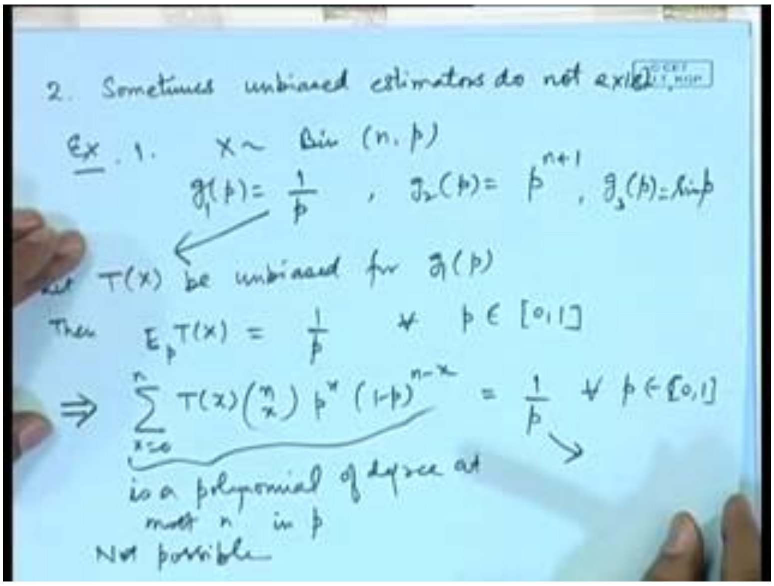


However, it is not necessary that all the time the concept of unbiased estimation may be useful. Sometimes unbiased estimators may be absurd; let me give an example, so let X follow Poisson lambda; I am interested in the parametric function say e to the power minus 3 lambda, since lambda is positive; we can see that 0 less than e to the power minus 3 lambda is less than 1. Let me define T X is equal to say minus 2 to the power x, so what is expectation of T X, it is equal to minus 2 to the power X, e to the power minus lambda; lambda to the power X by X factorial, X is equal to 0 to infinity. So, that is e to the power minus lambda; minus 2 lambda to the power X by X factorial that is equal to e to the power minus 2 lambda that is equal to e to the power minus 3 lambda.

So, minus 2 to the power X is unbiased for e to the power minus 3 lambda, but let us see e to the power minus 3 lambda, as we have seen it lies between 0 to 1, but what are the values of minus 2 to the power X; X can take values 0, 1; 2 1 and so on because is a Poisson random variable. So, it will take non negative integral values, if I take X is equal to 0; this is 1, if I take X is equal to 1, I get minus 2, if I take X is equal to 2. it is 4 if I take X is equal to 3, it is minus 8, 16 and so on.

Now, you notice here the values of the estimator are never in the interval 0 to 1. In fact, you can see for as X becomes large, the values are actually progressively increasing on the positive and the negative side; whereas my estimate is between 0 to 1. So this is an absurd type of situation. You look at another situation for mu square I gave an estimate X bar square minus S square by n, but there may be a situation where X bar is may be close to say 0 and S square may be a little larger value; in that case this may be communicative, where as mu square is always positive, so this may again give a absurd estimator.

(Refer Slide Time: 19:17)



Sometimes unbiased estimates do not exists; let us take this binomial situation and I may be interested to estimate say 1 by p; that is the reciprocal of the probability of success. I may be interested to estimate say p to the power n plus 1 or I may be interested to estimate say sin of p; let us see let I say T X be unbiased for g 1 p; then expectation of T X must be equal to 1 by p for all p in the interval 0 to 1.

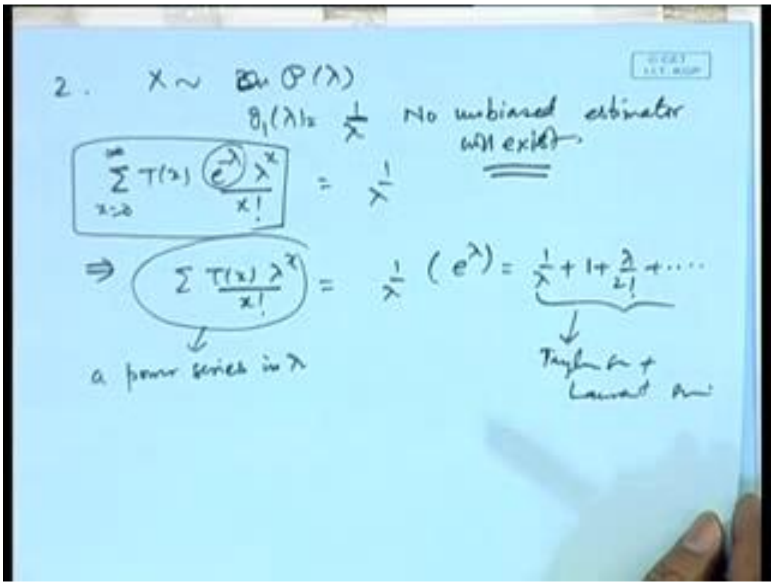
Now you see this; this left hand side term may be equivalent to T X n c X p to the power X 1 minus p to the power n minus X is equal to 1 by p for all p in the interval 0 to 1. Now left hand side this is a polynomial of degree at most n in p and this is not a polynomial term at all actually it comes in the Laurent series; this is the reciprocal term.

So, this can never be equal to this because this has to agree for all the points on an open interval, so this is not possible.

Similarly, if I put say p to the power n plus 1 on the right hand side; again it is not possible because left hand side is a polynomial of degree at most and on the right hand side you have a term of degree n plus 1. Similarly sin p has an infinite expansion, so that can never be equal to this finite polynomial expansion.

So, in a given problem it is not necessary that we will always be able to find an unbiased estimator.

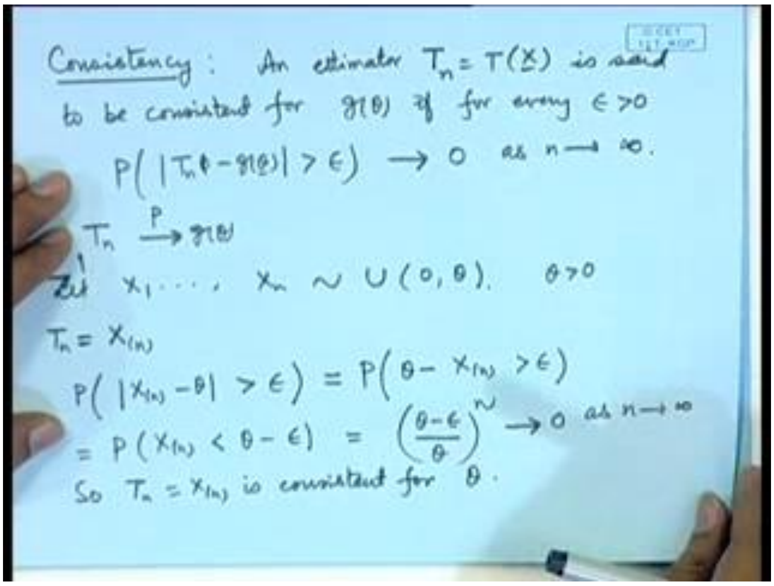
(Refer Slide Time: 21:57)



We may take another example say X follows binomial Poisson lambda and again I want to estimate say g 1 lambda is equal to say 1 by lambda, then sigma T X e to the power minus lambda; lambda to the power X by X factorial. If you look at this term, the left hand side term even if I take this to the other side this will imply sigma T X lambda to the power X by X factorial is equal to 1 by lambda into e to the power lambda which I can write as 1 by lambda plus, I can expand this 1 plus lambda plus lambda square by 2 factorial and so on.

Now the left hand side, this is a power series in lambda and the right hand side is a Taylor series plus Laurent series, so they can never be equal. So, no unbiased estimate will exist; now let me introduce another concept that is called consistency.

(Refer Slide Time: 23:52)



So, an estimator I will use the notation now T n; T X, now X is X 1, X 2, X n I am putting here n to denote the dependence that there are n observations used here. So, an estimator T n is said to be consistent for say g theta if for every epsilon greater than 0 probability that modulus T n minus g theta is greater than epsilon goes to 0 as n tends to infinity.

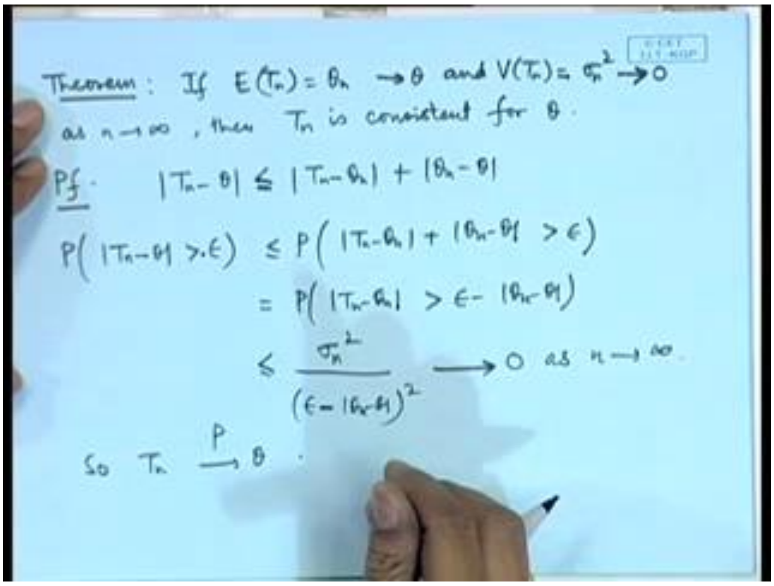
So, this means that the distance between T n and g theta becomes close as n becomes large; that means, the probability that the distance is larger than a (Refer Time: 25:00) quantity, this probability must go to 0 as n tends infinity. In convergence concept this is called T n converges to g theta in probability, so this is the so called large sample property of the estimators because what we are trying to say here is that; in the long run the estimator and estimate becomes close. So, in the un-biasness we say that the errors the positive errors and the negative errors cancel out each other.

Here we say that in the long run, the estimator and estimate become close. So, let us see some example; let me take say X 1, X 2, X n follow uniform 0 theta distribution. Now I may be interested to estimate the parameter theta which is the upper bound for the uniform distribution. So, let me take say X n; T n is equal to X n. We know the distribution of X n, so if I have to calculate probability of modulus X n minus theta greater than epsilon then what is this probability equal to; if I am saving uniform 0 theta distribution then each of the xi's lies between 0 to theta, so this X n also lies between 0 to theta.

So, this X n minus theta epsilon modulus values is actually theta minus X n, so this is equal to probability that X n is less than theta minus epsilon. We have already worked out the distribution of this larger order statistic, it is theta minus epsilon by theta whole to the power n; if epsilon is a positive number then theta minus epsilon by theta will be less than 1. So, this power n will go to 0 as n tends to infinity, so T n that is equal to X n is consistent for theta.

Now, in general proving the consistence may be slightly more difficult than the un-biasness in the sense that in proving consistency we need to look at the actual probability distribution and look at the probability of a certain event whereas then the expectation you look at the full range. So, for certain distribution this may not be very convenient and therefore, some sufficient conditions are helpful.

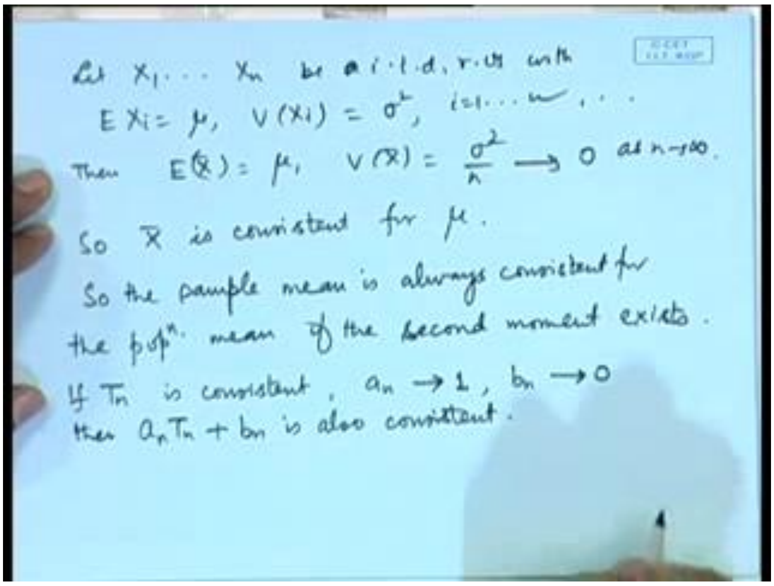
(Refer Slide Time: 28:01)



We have the following result; if expectation of T n that is equal to theta n converges to theta and variance of T n is equal to say sigma n square that goes to 0 as n tends to infinity then T n is consistent for theta. Let us look at the proof of this, so we can write this T n minus theta as equal to T n minus theta n plus theta n minus theta. So, it will be less than or equal to, so if I look at probablity of modulus T n minus theta greater than epsilon then this is less than or equal to probability of modulus T n minus theta n, which is equal to probability of modulus T n minus theta n greater than epsilon minus; if I use semi shapes inequality, it is less than or equal to sigma n square by epsilon minus theta n minus theta whole square.

Now, as n tends to infinity modulus of theta n minus theta becomes very small, so you have a non negative quantity in the denominator in fact, a positive quantity and sigma n square goes to 0, so this goes to 0, so T n converges to theta n probability.

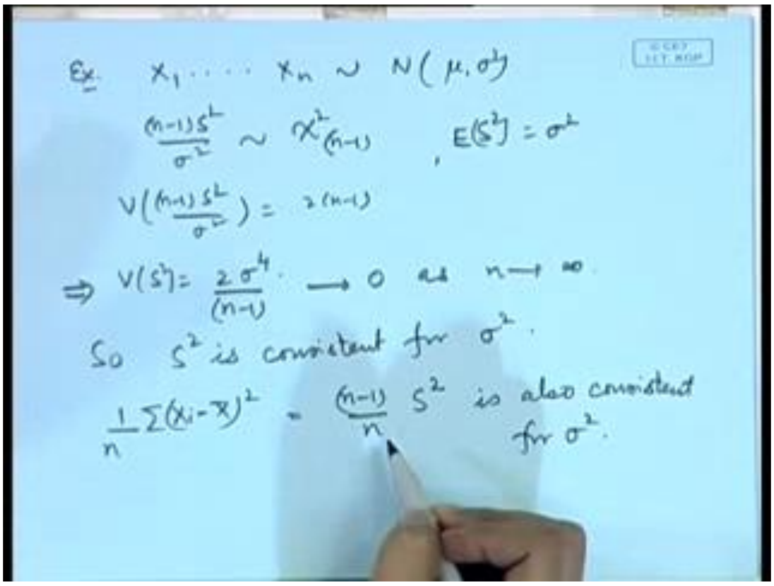
(Refer Slide Time: 30:05)



This result is extremely useful in the sense that if I am considering say let X 1, X 2, X n be a i i d random variables; which say expectation of X i is equal to mu and variance X i is equal to sigma square; then expectation of X bar is mu, what is variance of X bar; it is sigma X square by n which actually go to 0 as n tends to infinity. So, X bar is consistent for mu; that means; that if the mean and variance that is the first two moments are existing, then the sample mean is always consistent for the population mean, if the second moment exists.

Notice that this result will not be applicable if say variance does not exist is even if the expectations do not exist. For example, in a distribution like a quasi distribution, this result will not be valid. On the other hand, I can multiply by say if T n is consistent and a n is a sequence of numbers which converges to 1; b n is a sequence of numbers which converges to 0, then a n; T n plus b n is also consistent. So, unlike un-biasness where any change in the value of the estimator will actually destroy the un-biasness property, the consistency is a more you can say relaxed kind of property; that in the long run if I modify my estimator little bit, it does not make any difference because it will be simply a that coefficient are the consent will actually converge to 1, so in the long run both the things become all most the same. Let me give an example here.

(Refer Slide Time: 32:36)



In the sampling from normal population, if I have considered say n minus 1; S square by sigma square; the distribution is chi square n minus 1. So, we know variance of n minus 1, S square by sigma S square is twice n minus 1. So, variance of S square is actually equal to twice sigma to the power 4 by n minus 1 because I can take out these terms here n minus 1 square by sigma to the power 4 and I can adjust on the other side. We have already seen that expectation of S square is sigma square, so this is a unbiased and it is variance goes to 0 as n tends infinity.

So, S square is consistent for sigma square; now in place of S square I consider 1 by n sigma X i minus X bar whole square, then this is nothing, but n minus 1 by n S square then this is also consistent for sigma square because in the long run n minus 1 and n are the same; that means, n minus 1 by n goes to 1. So, we will look at various other properties and the methods of deriving the estimators in the next lecture.