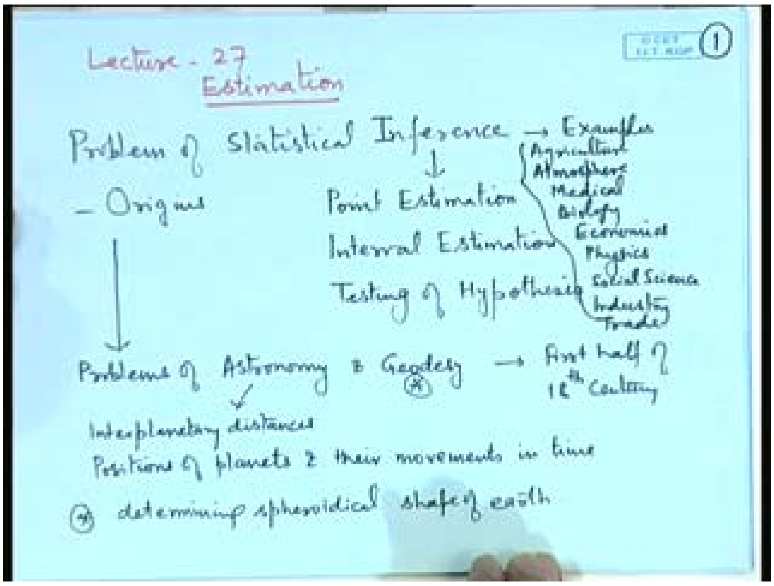
Probability and Statistics
Prof. Dr. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Module No. #01
Lecture No. #27
Estimation-I

Today, I will introduce the problem of statistical inference. So, far we have concentrated on discussing the concepts of probability, the concepts of statistical distributions, various kind of discrete and continuous distributions, multivariate random variables and their distributions, and we also looked at the concept of sampling distributions.

(Refer Slide Time: 00:45)



Further, I described something called descriptive statistics; that means, when a data is given then, how do we plan to analyze it; that means, how to present that data graphically or we draw certain basic, say basic characteristics such as measures of central tendency, measures of dispersion or variability from that data; however, all of this is actually to be utilized for drawing inference on populations. So, what is the population problem of inference? So, for example, a government is interested that how much will be the average heat production in the coming year? How much will be, say the production of sugar in the country? How much will be the production of a particular commodity? How much will be the production of say cotton? How many farmers or what percentage of land is utilized for farming of a fruits?

In atmospheric sciences scientists are worried about, what is the average temperature likely to be in the month of January or in the year 2010, is it going to be more than the 2009?. In medical sciences, we are interested about the occurrences of diseases. So, what is a estimated number of people, who will be affected by a certain kind of disease and what will be the effect on the longevity of the people by that disease? In biology, in economics, in physics, in social sciences, in industry, trade and commerce, in almost every area of human activity, we come across such situations or such problems.

Now, one may question that why do we have to use statistical methods here. For example, if I am looking at say occurrences of disease or say agricultural production, then where is the statistical thing coming into picture or suppose I am measuring the say diameter of a star in universe, in a far away galaxy, then where is statistics coming into picture. The statistics comes into picture that, although we may feel that the diameter of a star is deterministic and one should be able to get an exact figure of it, but we do not have methods of getting that value exactly. So, certain formula will be used and in that formula certain ingredients will be there, which will be measured by certain instruments repeatedly. Now, that measurements, the process of taking measurements introduces certain errors, which we assume are random or statistical in nature. And therefore, when we draw any inference based on those measurements, the inference becomes a statistical inference. And therefore, this entire topic or you can say the entire subject, the needs that we use correct methodology of a statistical inference. So that the conclusions drawn from that data are correct. So, that brings on to the focus the problem of a statistical inference.

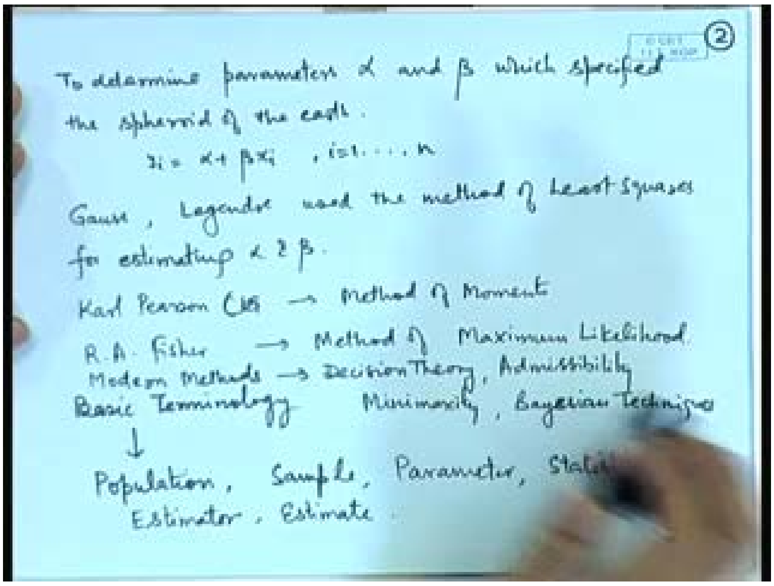
Primarily speaking, the problem of inference can be divided into two portions: one is called the problem of estimation and another is the problem of testing of hypothesis. For example, if we want to actually get a value that what is the average longevity of the people of India or people of a particular country, then we actually do not know the value of what we want? And therefore, we actually get a value based on a sample. So, this is called the problem of estimation; that means, to get the value. Now, that the estimation itself can be split into two parts: one is to get an actual value, suppose I say the value of average longevity is or average age is 65 years, then we are assigning a single value for the characteristic to be estimated. This is called the problem of point estimation.

On the other hand, we may not give the exact value but we may give an interval of the values and say that, with a certain confidence or certain probability, the given value lies in that interval. For example, we may say that the average age of a person in India is from 62 to 68 years with 95 percent of confidence. This is called the problem of interval estimation or confidence intervals.

On the other hand, sometimes we would like to test a fact. For example, a new drug has been introduced in the market for creating a certain disease. Now, the manufacturing company which has introduced the drug will certainly like to know, that whether the new medicine is more effective than the previous one. So, if I say p 1 is the proportion of people which were treated earlier and p 2 is the proportion of the persons which are treated now successfully, then whether p 2 is bigger than p 1. This type of judgment; that means, to tell on the basis of the sample whether p 2 is bigger than p 1 or p 2 is less than p 1, etcetera, this is called the problem of testing of hypothesis. So, broadly we divide the problem of statistical inference into three parts: one is the problem of point estimation, another is the problem of interval estimation and another is the problem of testing of hypothesis. There are various other facets of statistical inference like prediction, sequential inference and other things but they can be considered to be following up from here. So, these are the basic, you can say facts of the or basic points of statistical inference.

Let me give some historical facts about how the problem of statistical inference was initially studied. So, it seems out to have been have origins in the problems of astronomy and geodesy in the first of half of eighteen century, when many scientists where finding out like distances between the stars; that is interplanetary distances, the positions of the stars, their shapes, how do they move with the time; that means, for example, mercury take this much time to rotate around the sun or it takes this much time to rotate on its axis and all those kind of statements; that means, the problems in astronomy are in geodesy.

(Refer Slide Time: 08:08)



So, for example, some of the earliest measurements were made on to check whether the spherical shape of earth, to determine that thing. And so, it turned out that the data is of the form that we have observations x 1, x 2, x n and y 1, y 2, y n and they are related with the equation y i is equal to alpha plus beta x i, which today we know, as an equation of a simple linear aggression model. So, these are earliest occurrences of this model. So, the famous mathematicians Gauss and Legendary, they used the method of least squares for finding out the values of alpha and beta. So, you can say that the method of least squares is probably one of the oldest methods for finding out the estimates of parameters. Towards the end of nineteenth century, Karl Pearson introduced the method of moments and minimum chi square, for estimating parameters. In the beginning of the twentieth century, R A fisher, he introduced the method of maximum likelihood. In fact, as I have already mentioned, he is credited to be the, you can say initiator of most of the methods of modern statistical inference which we use today. So, he was the one who introduced the concept of maximum likelihood estimation. In the mid twentieth century Abraham Wald introduced the some decision theoretic methods and methodology such as admissibility, minimaxity and Bayesian techniques in a statistical inference.

Now, let me introduce the basic terminology to be used in a statistical inference. The first term is a population. So, a statistical population is a collection of measurements in which we are interested. So, for example, we are interested in estimating the average per capita income of persons in a state, then there may be a household survey or there may be a survey of people in different organizations and the incomes of individuals are noted. So, in this particular case the statistical population is the measurements corresponding to the individual incomes. If we are interested in the average longevity of persons, then suppose we are considering a particular state or a particular country, then the total lifespan of each person of that country or that state will constitute the statistical population.

If we are interested to study the yield of wheat in the state of Punjab, then corresponding to each plot of land where the wheat is grown, if we look at the total output or yield of the wheat from each of the plot, then those values will be considered the statistical population for this purpose. So, a statistical population is a collection of measurements with respect to certain characteristic, which we are interested to study.

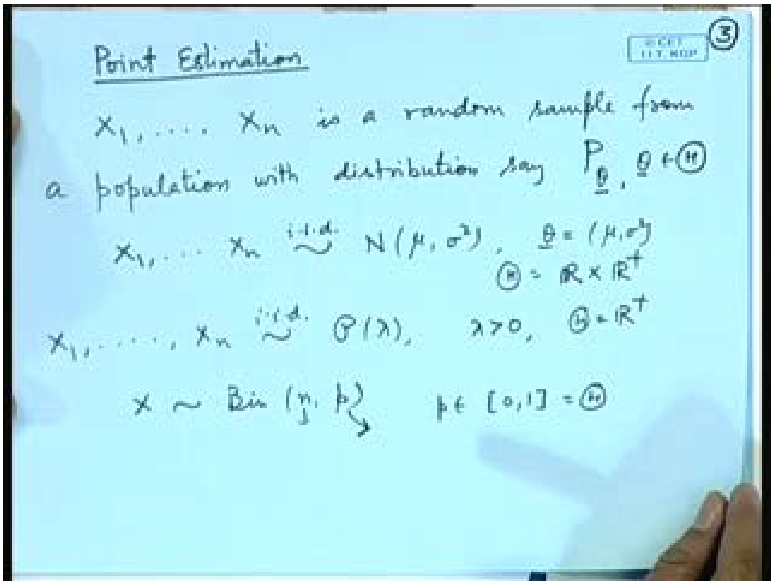
Here, one thing I would like to mention, that it is not necessary, that all the time we will have to look at only numerical value. Sometimes, it may be in the form of yes, no or some answers which we can call attribute data. For example, if we are looking at preferences of people for a certain opinion, whether they have a positive opinion about certain issue. So, they may answer is yes or no. So, corresponding to each person you will be noting down the data, yes or no and you may put it as values say 0 or 1.

You may record the persons who are say possessing a certain characteristics say an IQ greater than hundred or below hundred persons, whose average incomes are above say a particular level or below a particular level or we may classify them in according to four different levels: very poor, lower middle class, upper middle class, and say higher income group. So, we may assign for each person or each household according to the level of income that the person is having the values say 1, 2, 3, 4 or 0, 1, 2, 3, etcetera. So, this is qualitative data or attributes data and in statistical population one also studies such data.

Next is sample. So, what is a sample? The exact definition of sample is, that sample is a subset of population. So, in general when we want to study any characteristic about the population, it is requiring the entire measurement; that means, complete enumeration of the population. So, which is not feasible. So, for example, if you are studying say the household, per household expenditure on say medical expenses in a particular town, then it will require going to each household and get the monthly expenditure on the medical; however, this may not be feasible. So, the best solution for various such enumeration problems is to take a representative sample from the population and draw the inferences based on that. So, the concept of sampling techniques or sampling methodology is widely developed in statistics. So, here we assume that the sample has already been selected and we will draw the inferences based on that. So, sample, critically speaking is a subset of the population and we will assume that it has been randomly selected.

A parameter of a population is the characteristic in which we may be interested in. So, for example, when we talk about the population of say incomes, then we may be interested to know the range, for example, what is the difference between the maximum salaried imply and lowest salaried imply. If we are interested in the say yields of different states, for say wheat, then per hector wheat production may be in a particular state is much higher corresponding as compared with the other one. So, we may be looking at the averages, the maximum value, the minimum value, the variability, the median value. So, these characteristics of the population, they are termed as parameters.

(Refer Slide Time: 17:32)



So, since we are interested to know about the characteristics of the parameter populations; that means, parameters, the statistical inference problem relates to either finding out an estimate or you can say point estimator or an interval estimator for the parameter or to test about those parameter values.

Now, at our disposal, we have a random sample say x 1, x 2, x n. Now, whatever we want to draw our inference from x 1, x 2, x n, we will be using certain function of that. So, for example, if I say, I wanted to use find out the average height and from the sample I take the average and I use it as an estimate. So, that means, I have used a function of the sample observations. So, these sample observations if we make a function out of that, that is called statistic. So, the statistic will have different uses. For example, like I can use them to make a point estimator, I can use them to make a confidence interval, we can use them to create a test. So, we may use it as a test statistic. So, when I use it as, to estimate certain parametric function, then it is called an estimator and the realized value of that is known as an estimate.

So, now let me introduce the basic features of estimation. So, let me concentrate on the problem of point estimation. Now, to begin with, I mentioned that there are several mathematicians or statisticians who gave some methods of the estimation. For example, I mentioned the word least squares estimates, the method of maximum likelihood, the method of moments, the minimum chi square method, etcetera. So, each of these methods is based on certain concept or you can say certain theory that why this is desirable method. Now, the question is that they may give different values of the estimators or they may give the same values of the estimators, then the question comes that how do you distinguish that which one should be used? So, for that purpose we introduce certain criteria of estimation. So, before going to give the actual methods of estimation, let me introduce certain criteria.

So, in any statistical inference problem, the model is like this. That we have x 1, x 2, x n is a random sample, from a population with distribution say P theta, theta belonging to say script theta. Let me explain this. Normally, we will be talking about sentences such as x 1, x 2, x n is a random sample from poisson lambda distribution. x 1, x 2, x n is a random sample from normal mu sigma square distribution. So, what is the meaning of this? The meaning of this, is that in the inference problem, we assume that the determination of the statistical model has already been done; that means, the problem is already specified. For example, if I am saying it is estimation of say average longevity, the estimation of average temperatures, etcetera, the problem has already been identified by the person who is going to use it. It may be a government agency, it may be a commercial organization, etcetera and then the statistician has already determined the parametric model for that; that means, if we are talking about average heights, then the statistician has determined that this population follows a normal distribution; that means, if we have a large data set from the, our target group and we have taken the heights and then we make a histogram and a frequency curve, and we find that it looks like a normally distributed random variable.

Therefore, the problem to, for inference is now to draw certain inference on the parameters of the population; that means, what could be the value of mu, whether mu is equal to 0 or mu is less than or equal to certain value, whether sigma square is known value or unknown value, etcetera; that means, we are going to do a testing or confidence interval or point estimation about the parameters of the population; that means, when I am saying point estimation or testing, etcetera, we are talking about parametric inference.

So, there are two types of inferences: we have parametric inference and non-parametric inference. So, where the non-parametric inference will arise? When we are unable to determine the model from which the data has come from; that means, we may not be able to say that, it is normally distributed. So, this could be in several ways. For example, the data is too haphazard or the data is too less or we are not having sufficient experience to determine, the data is coming from which population. Then, there are certain methods which we call distribution free methods or non-parametric inference.

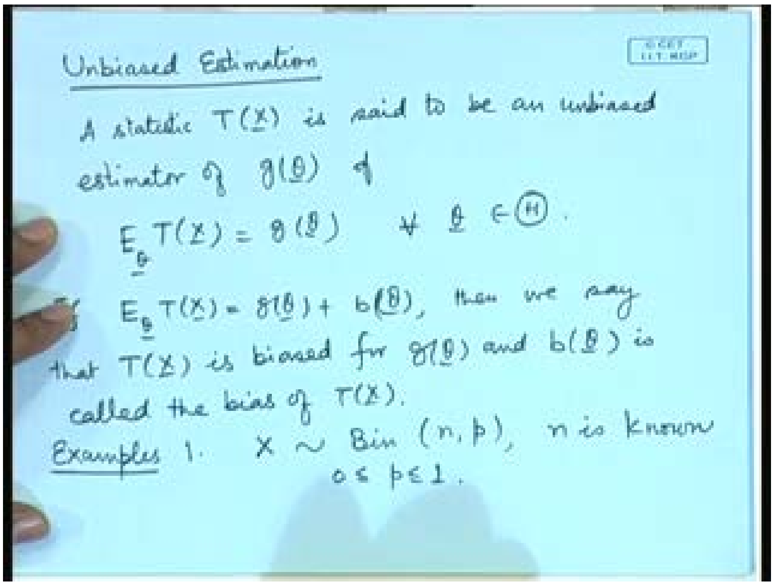
In this particular topic, we will be restricting our attention to parametric inference; that means, we assume that the model is coming from a certain population P theta. So, P we already know, that what distribution it could be. Only thing to be determined is that parameters we may not know. So, the statements such as x 1, x 2, x n is a random sample from normal mu sigma square. So, we go back to our terminology which we used in the distribution theory, that we write that x 1, x 2, x n are independent and identically distributed random variables with normal mu sigma square distribution. So, here what is theta, theta is equal to mu sigma square. Now, what is this script theta? It is the set of all possible values of the parameter.

For example, if I am saying normal mu sigma square, then mu varies from minus infinity to infinity and sigma square is positive; that means, here theta is your R, that is the real line cross R plus, that is the positive half of the real line. We may say x 1, x 2, x n follow poisson lambda distribution. So, here lambda is a positive parameter. Therefore, my parameter space is R plus. If I say, x follows say binomial n p distribution, I know what is N because, I know in how many trials, I am looking at for the number of successes. So, the parameter could be P and we may say that P belongs to the interval say 0 to 1. So, this is the parameter space in this situation. So, depending upon the different parametric model the distribution and the parameter space will be specified.

So, in any inference problem, we start with this model, that we have a random sample from a given population. So, the meaning of that is, that we have identically and independently distributed random variables from a given population. And our objective is to make certain inference about the parameters of the population in the form of point estimation, interval estimation or confidence interval. So, now, for the time being, we restrict attention to the problem of point estimation.

Now, one of the first concepts in the point estimation can be as a layman, 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 unbiasness.

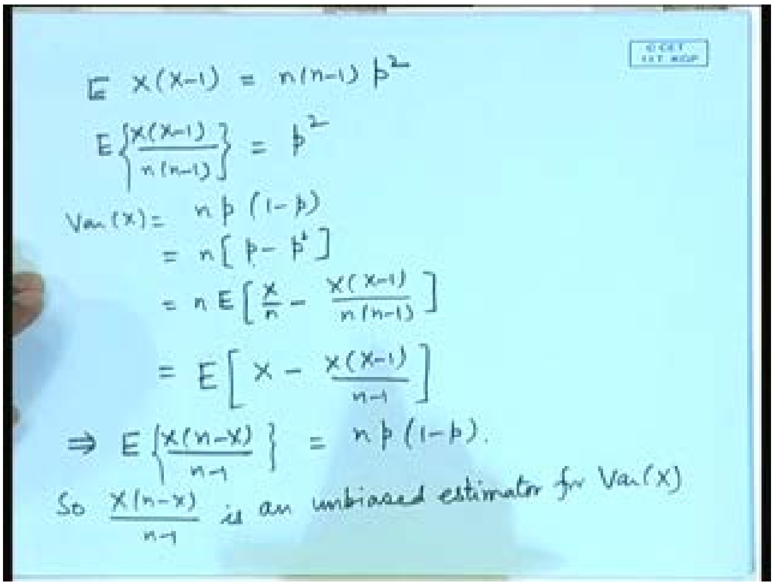
(Refer Slide Time: 25:19)



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 x1, x 2, x n or you can say T x. So, we will use this notation for 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 in. 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 say 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 the 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.

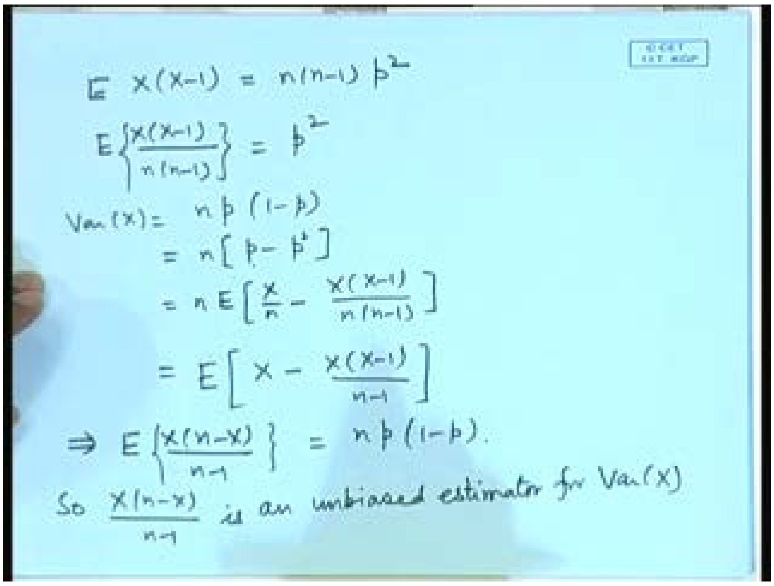
(Refer Slide Time: 29:03)



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 it is 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 write it 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 variable T, because in the population I may be interested in estimating the variable T also. So, here we are able to derive an unbiased estimator for that.

(Refer Slide Time: 31:05)

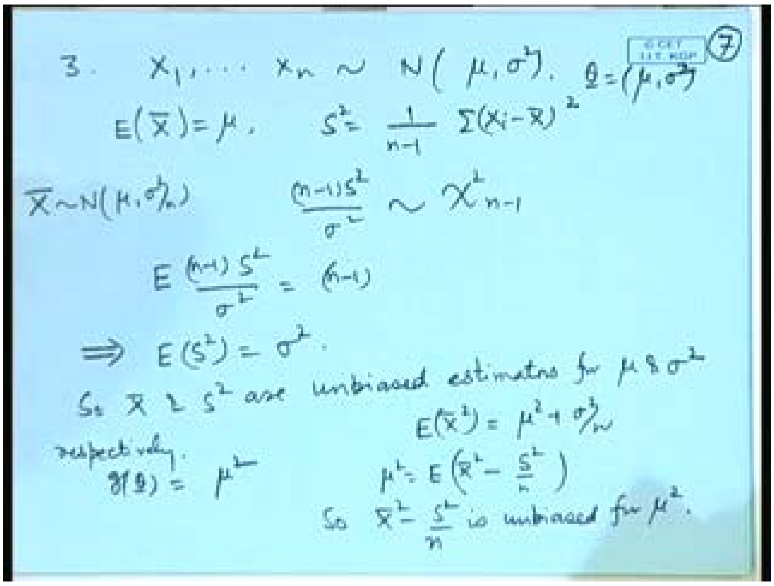


Let us take another problem. Let x 1, x 2, x n follow poisson lambda distribution. So, here lambda is the parameter. Suppose, I want to estimate lambda itself, then I may use say x 1. So, expectation of x 1 is lambda. One may suggest using x bar, that is 1 by n, sigma x i, then expectation of x bar is 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 arrivals 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 personnel may not be employed. 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 1 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.

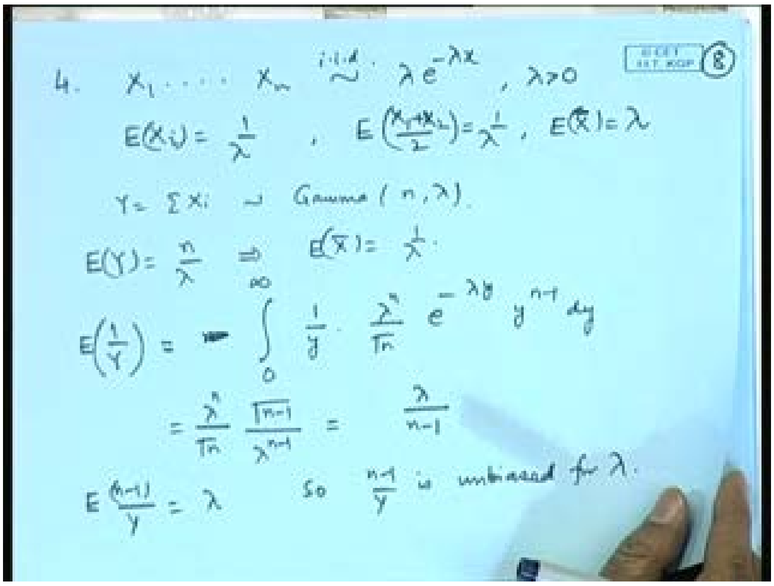
Let me take say x 1, x 2, x n a random sample from say normal mu sigma square population. If I am interested to estimate mu, I may use say x bar. So, expectation of x bar is equal to mu. Now, we know here that variance is sigma square and suppose, I am interested to estimate that, then I may make use of say S square; that is 1 by n minus 1 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 1 degrees of freedom.

(Refer Slide Time: 33:48)



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 say in mu square say. So, suppose my g theta, where theta is mu sigma square and I am interested 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 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: 36:30)



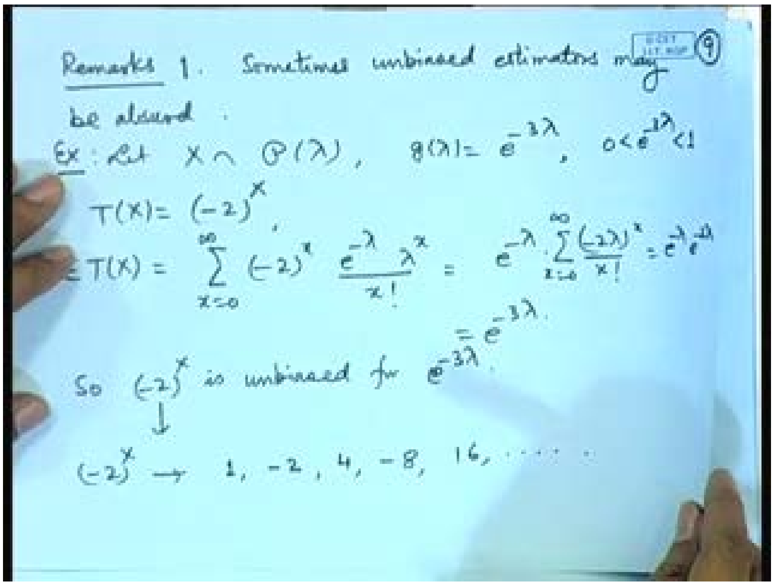
Let x 1, x 2, x n follow say exponential distribution, I may be interested to estimate the mean here, I may be interested to estimate say lambda here. So, if I am interested to estimate say mean, I may consider expectation of x i, that is equal to 1 by lambda. So, I may consider expectation of x 1 plus 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 choose among these.

If I am 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 the 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 to the power n minus 1 dy, 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 had introduced this lambda as the arrival rate in the poisson process or I had introduced a term called instantaneous failure rate or the hazard rate. So, lambda was the hazard rate. So, if we 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. Statistically 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: 40:01)

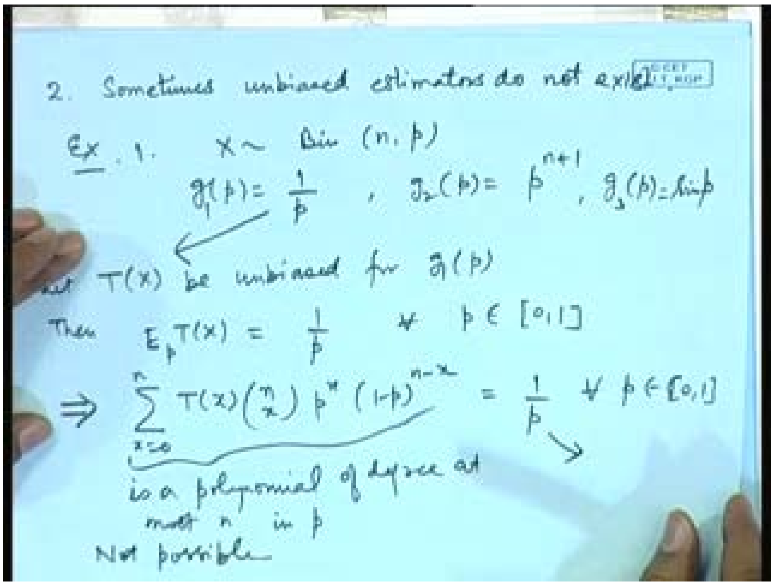


However, it is not necessary that all the time the concept of unbiased estimation may be useful. Sometimes, unbiased estimators may be absurd. Let me give an example. So, let x follow poisson lambda. I am interested in the parametric function say e to the power minus 3 lambda. Since, lambda is positive, you 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 as y 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 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 can take values 0 1 2 and so on, because x is a poisson random variable. So, it will take non-negative integral values, if I take x equal to 0, this is 1. If I take equal to 1, I get minus 2. If I take x equal to 2, it is 4. If I take x 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 estiman is between 0 to1. 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 become negative whereas, mu square is always positive. So, this may again give an absurd estimator.

(Refer Slide Time: 43:18)

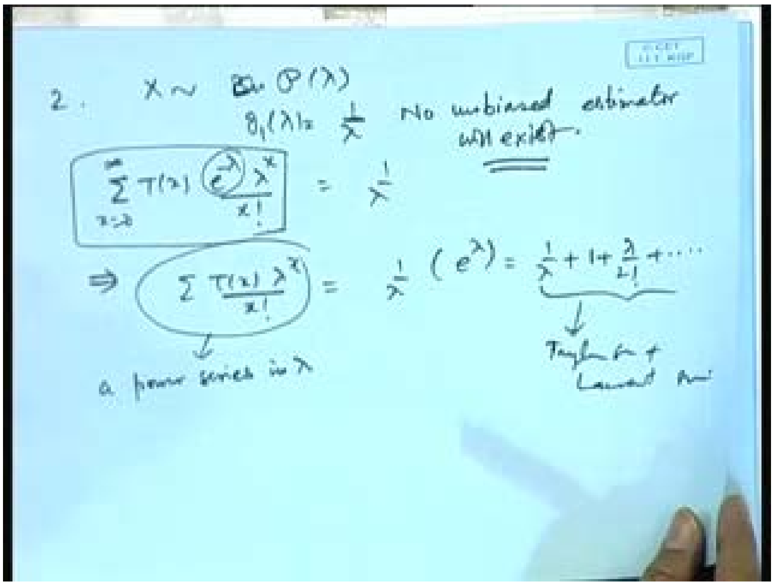


Sometimes, unbiased estimates do not exist. Let us take this binomial situation and I may be interested to estimate say 1 by p; that is the reciprocal of the probability of success. I may be interested to estimate say p to the power n plus 1 or I may be interested to estimate say sin of p. Let us see. 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 is 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 utmost 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.

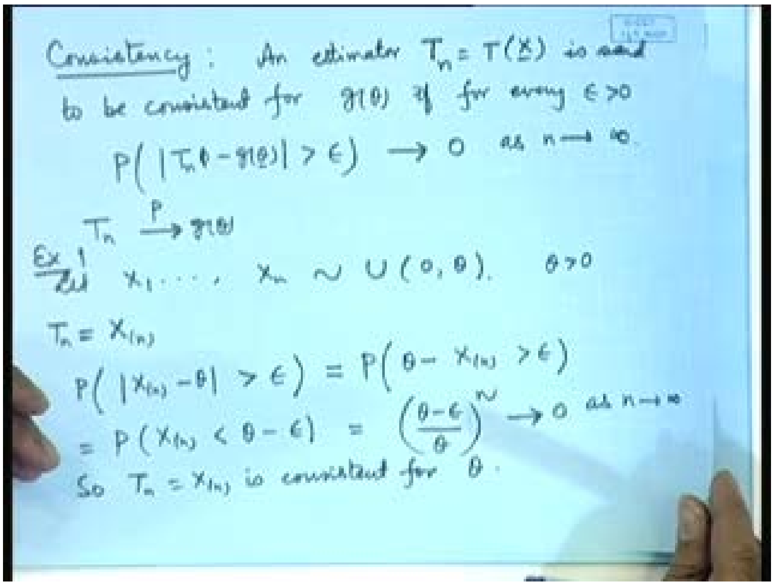
Similarly, if I put say p to the power n plus 1 on the right hand side, again it is not possible. Because, left hand side is a polynomial of degree utmost and on the right hand side, you have a term of degree n plus 1. Similarly, sin p has an infinite expansion. So, that can never be equal to this finite 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: 46:00)



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 estimator exists.

(Refer Slide Time: 47:46)



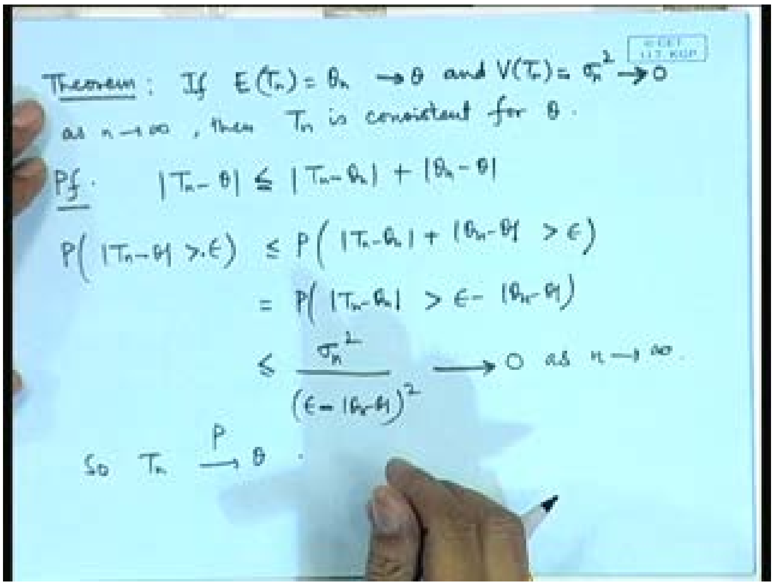
Now, let me introduce another concept that is called consistency. So, an estimator, I will use a notation now T n, T x. Now, x is x 1, x 2, x n. I am putting here n to denote the dependence that there are n observations used here. So, an estimator T n is said to be consistent for say g theta, if for 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 prescribed quantity, this probability must go to 0 as n tends to 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 the estiman becomes close. So, in the unbiasness, we said 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 the estiman becomes close.

So, let us see some example. Let me take say x 1, x 2, x n follow uniform 0 theta estimation. Now, I may be interested to estimate the parameter theta which is 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 x i's lies between 0 to theta. So, this x n also lies between 0 to theta. So, this x n minus theta, modulus value 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 largest 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.

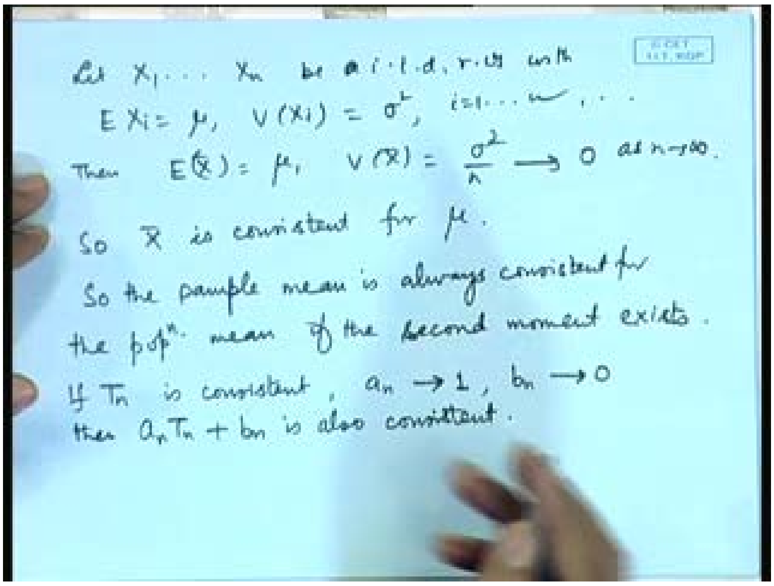
Now, in general, who win the consistency may be slightly more difficult than the unbiasness. In the sense that, in proving consistency we need to look at the actual probability distribution and look at the probability of a certain event whereas, in the expectation you look at the full range. So, for certain distribution, this may not be very convenient. And therefore, some sufficient conditions are helpful. We have the following result.

(Refer Slide Time: 52:10)



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 probability 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 Chebyshev’s 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. This result is extremely useful.

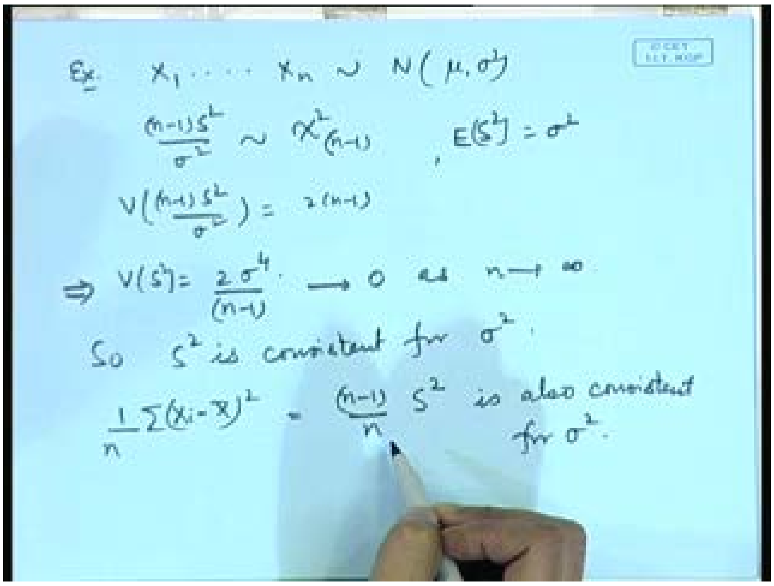
(Refer Slide Time: 54:10)



In the sense that if I am considering say let x 1, x 2, x n be a i i d random variables with 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 square by n which actually goes 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, 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 or even if the expectation does not exist. For example, in a distribution like a Cauchy 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 unbiasness whereas any change in the value of the estimator will actually describe the unbiasness 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 or the constant will actually converge to 1. So, in the long run both the things become almost the same.

(Refer Slide Time: 56:43)



Let me give an example here. 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 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 unbiased and its variance goes to 0 as n tends to 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.