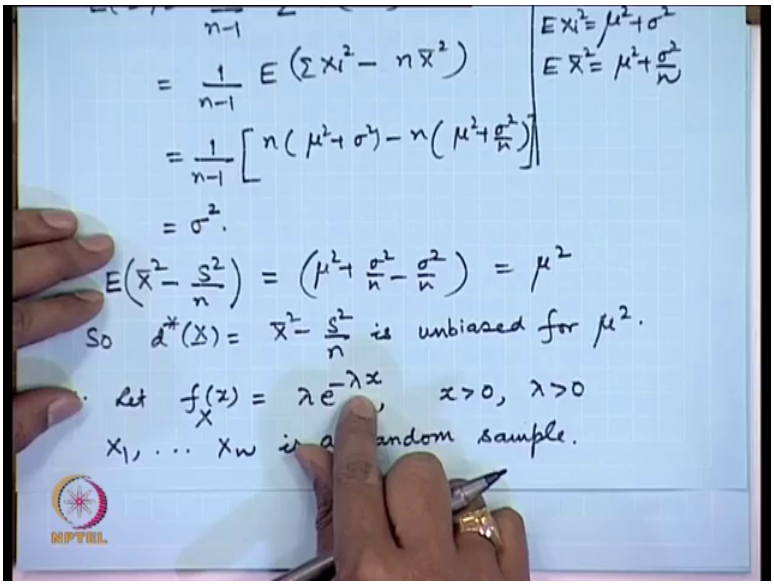
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 04
Basic Concepts of Point Estimations-II

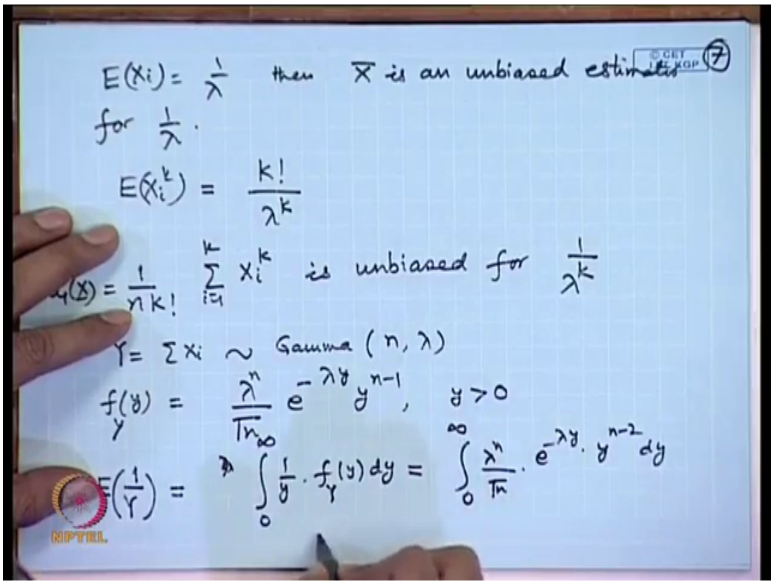
Sometimes we may be interested in lambda itself which is the rate of this.

(Refer Slide Time: 00:25)



Now, in that case we may have to do little bit of introspection here because, if we are considering direct moments I am getting the powers of 1 by lambda. If we want to estimate lambda itself then it suggests that we may have to consider reciprocal of X or X bar. Now, that is a curious thing here, if I consider expectation of 1 by X in the exponential distribution here that is not existing; that means, I cannot obtain averaging in the way that I have done in these two cases.

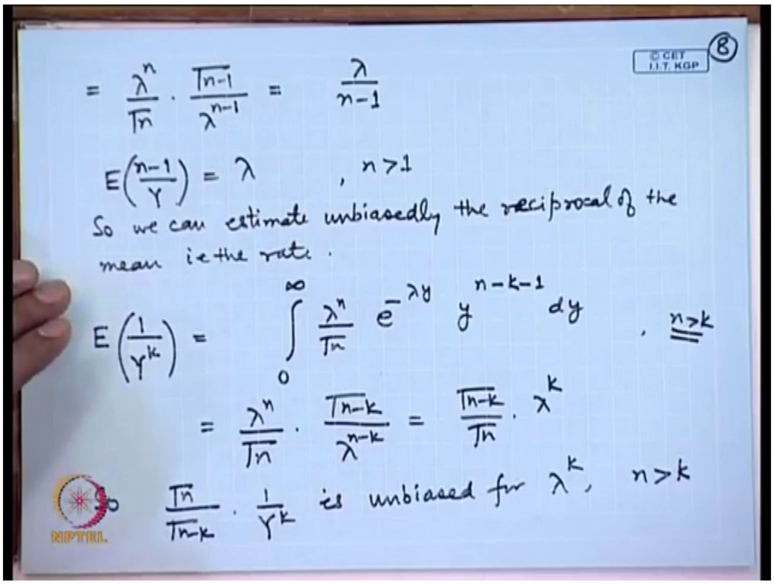
(Refer Slide Time: 01:03)



On the other hand if we consider say the distribution of say sigma X i that is having a gamma distribution with parameters n and lambda. That means, if I want to write down the density of Y that is equal to lambda to the power n by gamma n e to the power minus lambda y, y to the power n minus 1, where y is greater than 0.

Now, let us consider expectation of say 1 by Y, now that is equal to lambda to the power well 1 by y into this density from 0 to infinity. So, here let me substitute here 0 to infinity this form for this density. So, I will get lambda to the power n by gamma n e to the power minus lambda y, y to the power n minus 2 d y.

(Refer Slide Time: 02:17)



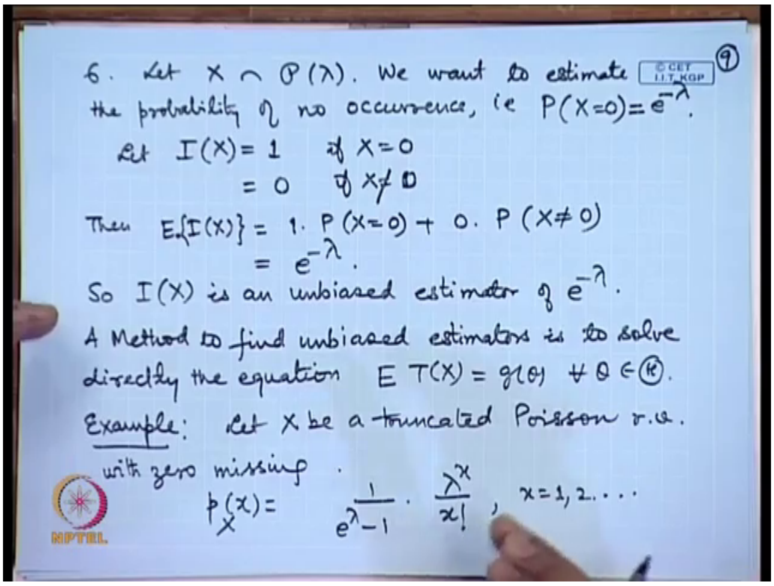
Now, that is equal to lambda to the power n by gamma n, now this can be evaluated using the gamma function formula. So, it becomes gamma n minus 1 divided by lambda to the power n minus 1 that gives us clearly lambda divided by n minus 1. So, what do we conclude? Expectation of n minus 1 divided by Y is equal to lambda of course, here n has to be greater than 1 otherwise this value will not exist. So, if I have more than one observation from an exponential distribution, I can estimate even the reciprocal of the rate, which was not possible if I am considering only one observation.

Because, expectation of 1 by X does not exist, but here I am considering n observations, where n is greater than 1, then expectation of n minus 1 by Y that is sigma X i that is equal to lambda. So, we can estimate unbiasedly the rate the reciprocal of the mean. The rate that is the reciprocal of the mean, not only that if we want to now consider some powers of lambda that also can be considering the corresponding powers of Y in the denominator. For example, if I consider say expectation of 1 by Y to the power k, now that becomes lambda to the power n by gamma n e to the power minus lambda y, y to the power n minus k minus 1 d y.

So, if n is greater than k then this is a gamma function and we can straightforwardly evaluate it as gamma n minus k divided by lambda to the power n minus k that is equal to gamma n minus k by gamma n lambda to the power k. So, after adjustment of this coefficient we get gamma n by gamma n minus k 1 by y to the power k is unbiased for lambda to the power k, if n is greater than k. So, in most of the typical estimation problems the a structure of the moments gives the unbiased estimators for the usual parametric functions.

So, the usual parametric functions I mean either the moments or some linear functions of the moments or some other type of functions. For example, in this case we have considered even the reciprocal of the moments and we are able to evaluate but of course, we use the a structure of the gamma distribution here. Sometimes the parameter to be estimated may not be in the form of a moment or its or any really it may not have any relation with the moments. In that case you may have to identify the kind of parameter that we are having let us take one example here.

(Refer Slide Time: 05:57)



Say X having Poisson distribution and we want to estimate in many of the Poisson problems because, Poisson distribution generally denotes the arrival rate in a service queue etcetera. Therefore, it is of interest to the organizers or the service providers to know when there will be no arrivals. Accordingly, they can provide the or you can say distribute the service personnel in such a way that when there is a slack period, that is they can estimate that these much period there will be no arrivals or no customers or no persons in the queue. Then the person which who is supposed to do the duty there he can be a posted elsewhere or that slot may be kept free also.

So, we need to estimate say the occurrence of 0 or the probability of no occurrence that is probability say X is equal to 0. So, we want to estimate this. Now, in the Poisson distribution this is equal to e to the power minus lambda. If I am considering Poisson lambda distribution the probability of X equal to 0 is e to the power minus lambda. Now, you can easily notice that this is not a moment. In fact, if I expand it I will get it as a series 1 minus lambda plus lambda is square by 2 factorial and so on. And, if I substitute the corresponding estimates one by one then we are not sure of the convergence of the series.

So, it is not a good idea to substitute directly in the expansion of e to the power minus lambda. However, we can notice something like this. Let us consider say I X the indicator function as 1, if X is equal to 0 and it is equal to 0 if X is equal to 1, if X is not equal to 0. Then what is expectation of I X that is equal to 1 into probability of X is equal to 0 plus 0 into probability X not equal to 0. So, this cancels out and we get only probability X equal to 0 which is our required parameter. So, here the indicator function of the set where X equal to 0, that itself becomes an unbiased estimator for the parameter or you can say the probability of e to the power minus lambda.

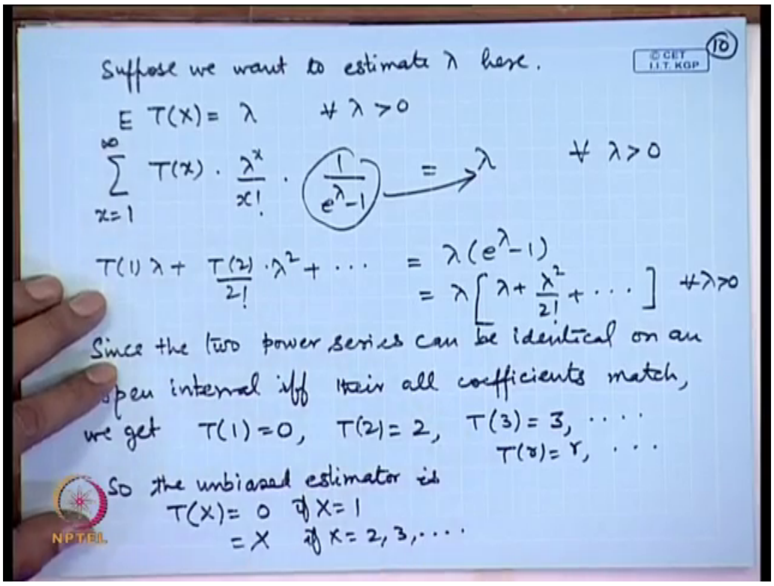
Of course, at this point you may raise the question that this is not a proper estimator or it may not be very informative, in the sense that I just conduct the trial once and use the estimator as 1; if X is equal to 0 otherwise I use it as a 0. So, this is not a very good estimator for the function which is actually lying between 0 and 1. So of course, that question remains and we may do further ramifications of this. So, right now we are able to obtain an unbiased estimator. So, we may look at the methods of finding out the unbiased estimators here. Right now, I have told two cases: one is that we may use moments or reciprocal of the moments or some functions of the moments.

Another thing could be to use the form of the parameter which is coming in the form of the probability; so, we took the indicator function of that set. So, let us consider some more example here. We may directly write down the equation expectation of T X is equal to the g theta. Now, by using some methods of analysis we may be able to solve this equation; that means, we need to get the solution for the function T X here which function T X will satisfy this equation. So, that is another one so, let me describe this thing. A method to find unbiased estimators is to solve directly the equation expectation of T X is equal to g theta for all theta.

Let us take one example here, let X be a truncated Poisson random variable with zero missing; that means, the probability mass function is of this form say 1 by e to the power lambda minus 1 lambda to the power x by x factorial for x equal to 1 2 and so on. A standard Poisson distribution is x equal to 0 1 2 and so on. So, suppose 0 is missing this type of situation may arise, where we know from the given setup that the assumptions of the Poisson process are satisfied for the arrival distribution.

However, in the physical setup it may turn out that when we are actually doing the sampling we cannot record the occurrence of 0. So, in that case the distribution which will be actually recorded will be of this form. Now, suppose we are considering again say estimation of lambda.

(Refer Slide Time: 12:43)

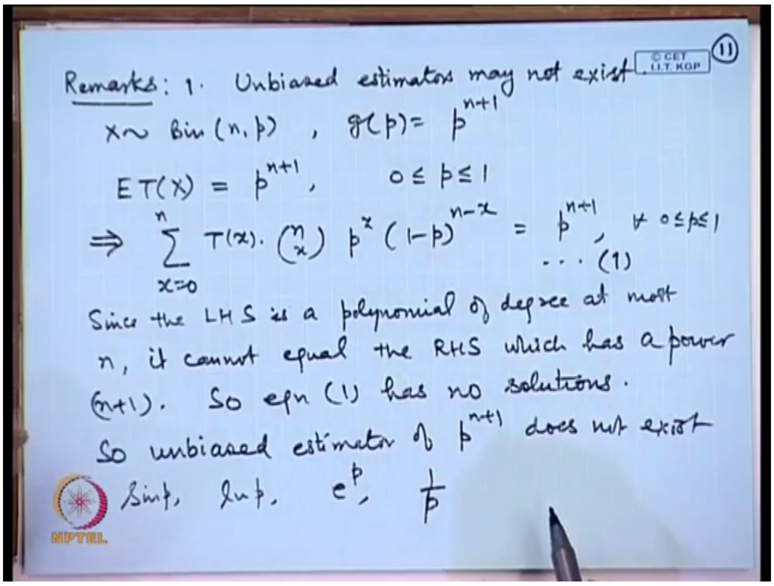


So, suppose we want to estimate here; now notice here that here the expectation of X is not equal to lambda. So, we cannot directly use the moment here. So, we write down an equation expectation T X is equal to lambda for all lambda greater than 0. So now, substitute here T x where x is equal to 1 to infinity lambda to the power x by x factorial 1 by e to the power lambda minus 1. This is equal to lambda for all lambda greater than 0, let us elaborately write down this equation. So, I take this term to this side, this becomes lambda into e to the power lambda minus 1 and then we expand this. So, I get T 1 lambda plus T 2 by 2 factorial lambda square and so on. That is equal to lambda into e to the power lambda minus 1 which I can expand; this becomes lambda plus lambda square by 2 factorial and so on.

Now, this is statement I am writing for all lambda greater than 0; the left hand side is a power series in lambda, the right hand side is a power series in lambda. So, the two power series are equal on a open subset of a real line if and only if all the coefficients are equal. So, if we do that then we can compare the coefficient. Since, the two power series can be identical on an open interval if and only if their all coefficients match we get; so, you compare the coefficient of lambda on the right hand side there is no coefficient of lambda. So, we get T 1 is equal to 0. Now, T 2 the coefficient of lambda square is 1 here, the coefficient of lambda square is T 2 by 2 factorial.

So, T 2 becomes 2 factorial that is 2 then coefficient of lambda cube on the left hand side will be T 3 by 3 factorial and here it is 1 by 2 factorial. So, T 3 by 3 factorial is equal to 2 factorial; that means, T 3 is equal to 3 factorial by 2 factorial that is equal to 3 and so on; that means, in general I will get T r is equal to r that is r factorial divided by r minus 1 factorial. So, the unbiased estimator is T X is equal to 0 if X is equal to 1 it is equal to X if X is equal to 2 3 and so on. So, here we have seen that by solving an equation directly one can obtain unbiased estimators. However, this type of technique may not be very useful in continuous distributions etcetera.

(Refer Slide Time: 16:15)

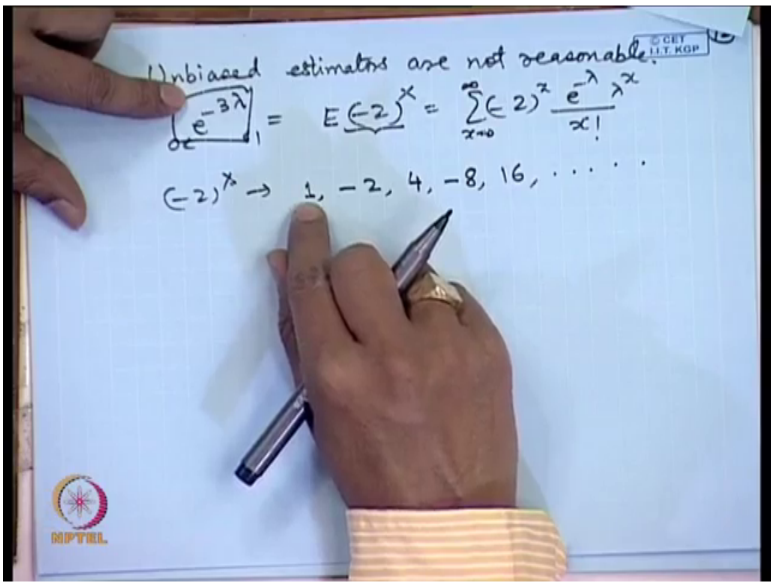


Let me give a brief remark on the existence of unbiased estimators. It may turn out that in some situations there may not be any unbiased estimators, unbiased estimators may not exist. So for example, consider say X following binomial n p. I want to estimate say p to the power n plus 1, let my g p be equal to p to the power n plus 1. So, if I consider say T to be an unbiased estimator; so, I will write expectation of T X is equal to p to the power n plus 1, for p in the interval 0 to 1. This means n c x p to the power x into 1 minus p to the power n minus x.

Now, notice this equation here on the left hand side you have a polynomial in p which is having degree at most n. Because, the maximum power that p can have is p to the power n or here 1 minus p to the power n whereas, on the right hand side you have p to the power n plus 1. So now, the 2 polynomials can agree on an interval if and only if all their coefficients agree, but here that does not seem to be possible. Therefore, this situation or you can say this equation has no solutions.

Since, the left hand side is a polynomial of degree at most n, it cannot equal the right hand side which has a power n plus 1. So, this equation let me call it 1, equation 1 has no solutions. So, unbiased estimator of p to the power n plus 1 does not exist. See similarly we may consider a function say sin p, we may consider say l n p, suppose you consider e to the power p, suppose we consider 1 by p. In all these cases unbiased estimators will not exist.

(Refer Slide Time: 19:21)



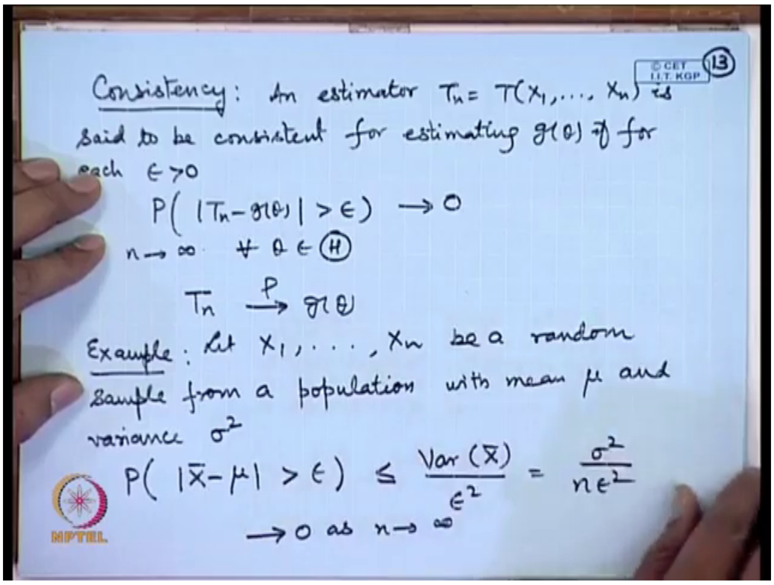
There may be yet another type of situation, that unbiased estimators are not reasonable. By reasonable I mean that if I say my parameter lies between 0 to 1 then my estimator should also take values between 0 and 1. If I say my parameter is positive then my estimator should also take positive values. If I have my parameter to lie in a given range say from minus m to m then my estimator should also be between minus m to m.

So, these are some physical constraints that the estimator must satisfy. So, there may be some situations where unbiased estimator actually does not satisfy this. I gave you one example for estimation of the probability of 0 occurrence. So, here you can see e to the power minus lambda, this is lying between 0 to 1 whereas, the estimator is taking value either 0 or 1. So, this is not a very proper estimator here. We may consider another example say I want to estimate say e to the power minus 3 lambda.

Now, if I consider expectation of minus 2 to the power X in Poisson distribution then it is equal to minus 2 to the power x e to the power minus lambda, lambda to the power x by x factorial. So, if you sum this you get e to the power minus 3 lambda. So, now you see the values of this minus 2 to the power X, it will take values corresponding to X is equal to 0 it is taking value 1, corresponding to X equal to 1 it is taking value minus 2, corresponding to X equal to 2 it is taking value 4, corresponding to X equal to 3 it is taking value minus 8 16 and so on.

So, you can see here this is never taking values, which is prescribed for this, this parameter lies between 0 to 1 e to the power minus 3 lambda because lambda is a positive parameter here. But, the estimator is taking absurdly different values a starting from 1, then minus 2 4, minus 8 and so on. So, this is not a reasonable estimator. So, in unbiased estimation we may have to be careful that the estimator should be reasonable; more about unbiased estimation we will take up in the next class.

(Refer Slide Time: 21:49)

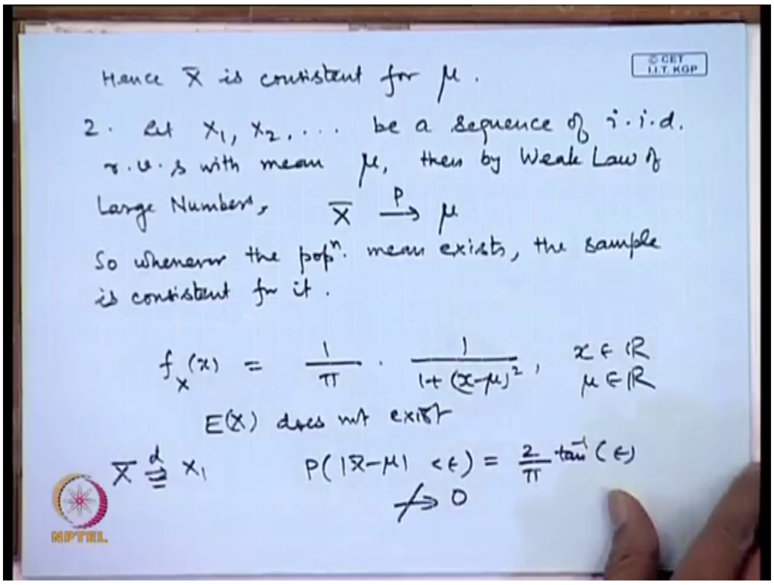


Now, we take up another desirable criteria that is called consistency. So, this is a large sample property, by a large sample property we mean that if n is large what is the behavior of the estimator. So, an estimator T n that is equal to T of X 1 X 2 X n is said to be; so, here we are showing dependence upon n here that n observations are used. So, this is said to be consistent for estimating g theta if for each epsilon greater than 0; probability of modulus T n minus g theta greater than epsilon this goes to 0 as n tends to infinity for all theta. Actually in probability theory, when we discussed the concept of convergence of random variables this is equivalent to saying that T n converges to g theta in probability.

So, this means here essentially that as n increases; that means, if I have a sufficiently large sample then the probability that my estimator is quite close to the true value of the parameter. Because, I am saying that the probability of this being greater than epsilon is actually almost negligible. So, there is a very high probability that in a large sample my estimator will be almost equal to the or it will be very close to the true value of the parameter. So, this is actually a essentially a large sample property, you can say it is a slightly relaxed property compared to the unbiasedness.

So, let us take an example here let X 1 X 2 X n be a random sample from a population with mean mu and say variance sigma square. Let us consider probability of modulus X bar minus mu greater than epsilon, then by Chebyshev’s inequality it is less than or equal to variance of X bar by epsilon X square that is equal to sigma square by n epsilon square. Now, you can see that this quantity goes to 0 as n tends to infinity; that means, if I am assuming the mean and variance then the sample mean is a consistent estimator for the population mean; X bar is consistent for the population mean mu.

(Refer Slide Time: 25:11)

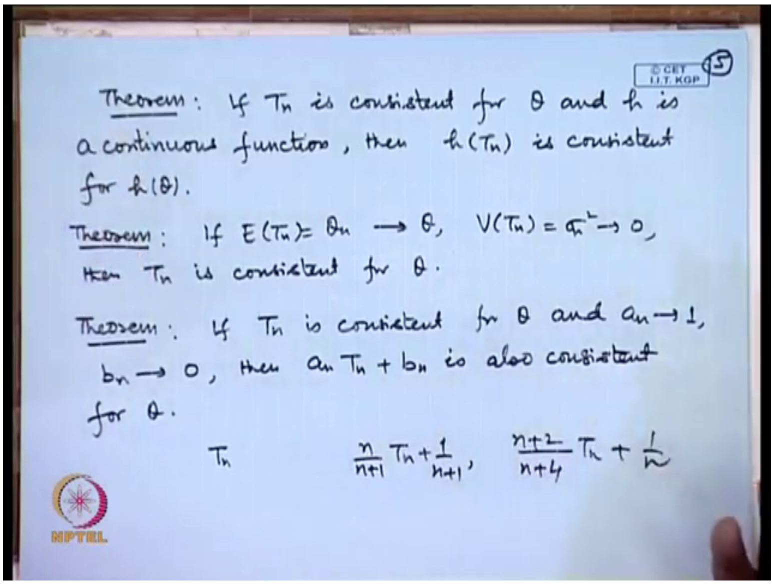


Another way of looking at it could be through the weak law of large numbers. Let us consider let X 1 X 2 and so on be a sequence of independent and identically distributed random variables say with mean mu, then by weak law of large numbers X bar converges to mu in probability. So in fact, the existence of the second moment is not required. Here I used Chebyshev’s inequality that is why I assumed sigma square here, but for actual weak law of large numbers that is not required. So, in general whenever the population mean exists the sample mean is an consistent estimator for the population mean. So, like unbiasedness criteria the consistency for the sample mean is also a very nice property that it is holding.

So, we can say that whenever the population mean exists, the sample mean is consistent for it. You may say that this property may be a trivial, but actually it is not so. There are distribution such as Cauchy distribution, see if I consider Cauchy distribution then we know that here the mean does not exist expectation X does not exist. In fact, if I consider the distribution of X bar that is same as X 1 that means, in distribution these two are same. So, if I look at the probability of modulus X bar minus mu less than epsilon that is equal to basically 2 by pi tan inverse epsilon.

So, this does not go to 0 because there is no role of n here. So, here the criteria of consistency fails to hold.

(Refer Slide Time: 27:53)



There may be also cases where for certain range of parameter the first moment will exist, in some other range it will not exist. Therefore, consistency of the sample mean or the unbiasedness of the sample mean will be holding only in that region. Another important property which is satisfied by the consistency is the invariance. So, we have the following result, that if T n is consistent for say theta and h is a continuous function, then h T n is consistent for h theta. So for example, here I am mentioning say T is say consistent for theta and I am looking at say estimation of theta square, then T square will be consistent for theta square.

Now, you can notice here the difference from the unbiasedness here. In unbiasedness this type of invariance was not there except for the linear invariance. If I am considering say 1 by T, then if T is unbiased for theta 1 by T is not necessarily unbiased for 1 by theta. In fact, in most of the cases it will not be whereas, in consistency this will be true. Another important result in consistency is that, if expectation of T n is equal to theta n which actually converges to theta and variance of T n converges to 0, then T n is consistent for; that means, it need not be unbiased, but in limit it is unbiased. And, if the variance of the estimator is a negligible or it becomes negligible as the sample size increases then T n becomes consistent for theta.

So, as I was mentioning this is slightly relax property and it is quite helpful in the large samples that many estimators which may not look very reasonable from the point of view of unbiasedness etcetera, but they become alright for the consistency property. Similarly, if T n is consistent and a n is a sequence of numbers which goes to 1, b n goes to 0 then a n T n plus b n is also consistent for theta. That means, if I say T n then I can consider say n plus n by n plus 1 T n, I may consider say n plus 2 by n plus 4. So, these are all consistent because these are all going to suppose I put here plus 1 by n plus 1 plus 1 by n etcetera then they are also consistent estimators for the parameter theta.

So, today we have discussed two important criteria: one is unbiasedness and another is consistency. However, it is important to know what are the methods for determining the estimators or how to find out the desirable estimators. So, in the next lectures we will be covering important methods for finding out the estimators and then we will come back further to the topic of criteria; that means, we will look at the efficiency of the estimators which estimator should be used. Suppose there are more than one estimator available for the same problem, satisfying the same criteria then what extra criteria should be introduced to choose one over the other. So, these and other topics we will be covering in the forthcoming lectures.