Advanced Engineering Mathematics

Prof. Somesh Kumar

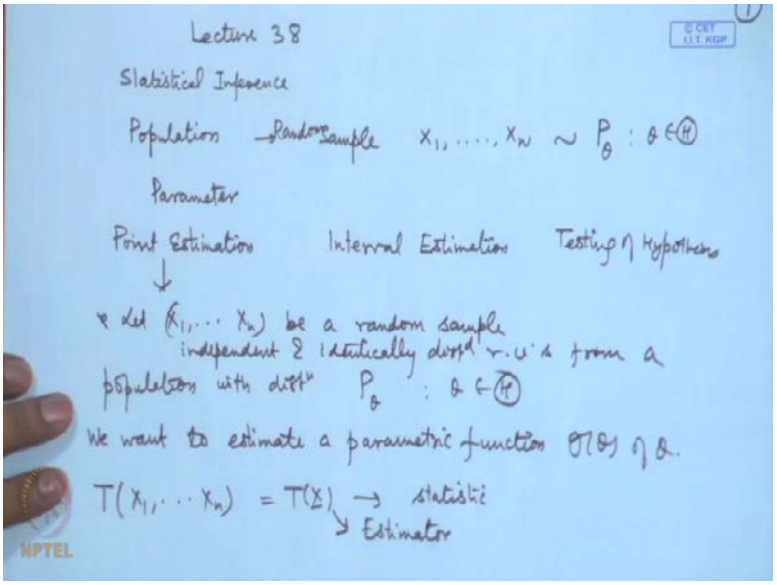
Department of Mathematics

Indian Institute of Technology, Kharagpur

Lecture No. # 39

Point Estimation

(Refer Slide Time: 00:23)



Now, I introduce the problem of a statistical inference. What is the problem of a statistical inference? In real life situation, we are ask to make certain statements, we are ask to verify the certain claims; for example, the government wants to know the expected growth rate of the economy. And for example, there will be more exports; there will be less, what will be the size of gdp and so on.

We want to know, what would be the total agricultural production in the country in the next year; we want to know what is the infant modality rate in a particular geographical region or in a particular state; we want to know, what would be the reaction of the population to a certain constitution amendment which the government wants to make; we want to know, whether the a particular diet program is helpful in reducing to the rate for certain class of people; we want to know, whether a certain new medicine will be more effective in curing a certain diseases. In all of these statements, we are concerned about

certain numerical measurements or attributes. For example, in the problem of finding out the effectiveness of certain medicine, we may like to know that if the medicine is given to a target group of patients, then whether more patients get cured, because of this compared to the previously used medicine or not that means, whether the proportion has increased.

Here, the attribute is whether the percent gets cured or not. Whereas, if we are looking at say modality rate, then we may be interested in the number, for example, if in a particular year or in a particular month, this many children are borned then per 1000 of children how many children or wife, after say 1 month or after 1 year or after 5 years. We may look at the average age of a population, for example, in India we say that, the average age or average longevity of a merely 63 years or average longevity of the general population is 62 years or we say in Japan, the average longevity 77 years.

So, all of these statements are concerned about, certain think called numerical measurements, which we call population; so, population is a collection of numerical values regarding the characteristic in which we are interested. Now, in order to say something about the population, for example, if we want to say something about average, we want to say something about variable T, we want to say something about range. Then one thing is to have the complete enumeration; but complete enumeration is not possible in most of the practical cases.

And therefore, one takes a subset of the population which we call sample; and this sample is then ordinarily called or we use the notation X 1, X 2, X n which will correspond to the values of the random variables, which are they are in the sample. Now, when we near the assumption that it is a random sample; that means, each unit of the population is having the same probability of getting in the sample. Then for the population, whatever distribution is there we have the same distribution for each of the observations here. So, we say each of this is having the probability distribution say P theta, theta belonging to theta. Now, when we say P theta, here theta denotes the parameter of the population.

So, in general the parameter of the population is unknown, for example, when we say normal mu sigma square; so, normal mu square has two parameters mu and sigma square. If, I say Poisson lambda distribution, then lambda is the parameter of the

population or the distribution. So, in general the problem of inference relates to make certain a statement about the unknown parameter of the population; now this statement could be of several forms, one is to give a value for that parameter. For example, we want to know, what is the arrival rate in a service q at a railway reservation counter?

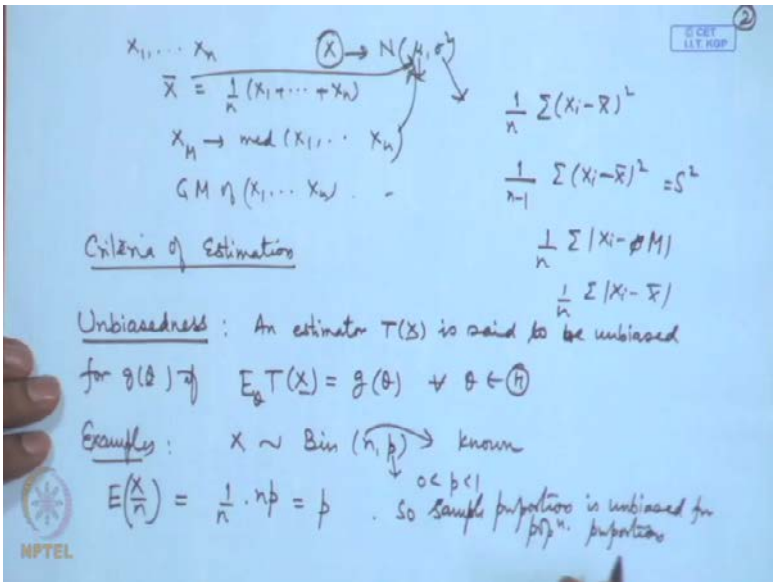
So we want to know, what the value of lambda is; so, this is called the problem of point estimation or a problem of estimation. In place of one value if we want to give an interval of the values. For example, we may say that, the number of persons arriving between 8 am to 10 am is anything between 100 to 120 then it is an interval, than that is called the problem of interval estimation, if we want to check; that means, we have taken a sample of the patients whom a new intera has been given.

And we want to know whether, the new drug is more effective; that means, more number of people get cured or not. In that case you are checking a statement, because previously we know the proportion of the people getting cured, probably the previously it was the number was say half. Now, we want to know whether more than 50 percent of the people get cured, then we want to test something; this is called the problem of testing of hypothesis. Now, I will concentrate on the problem of point estimation.

So our model is that we have, a random sample let X 1, X 2, X n be a random sample so; that means, they are independent and identically distributed random variables from a population with distribution, I use a general notation P theta, where theta belongs to theta; these theta could be scalar or vector here. Now, based on we want to estimate a parametric function, say g theta. Now, for estimating we have to make use of the sample; that means, we will assign a function of X 1, X 2, X n which we called a statistic.

Any function of the observation is called a statistic, and we are using a tool estimate, so we call it a point estimator or an estimator of g theta. Now, let us a start with very simple example, suppose you want to estimate average heights of the adults, adult males in a given population.

(Refer Slide Time: 08:11)



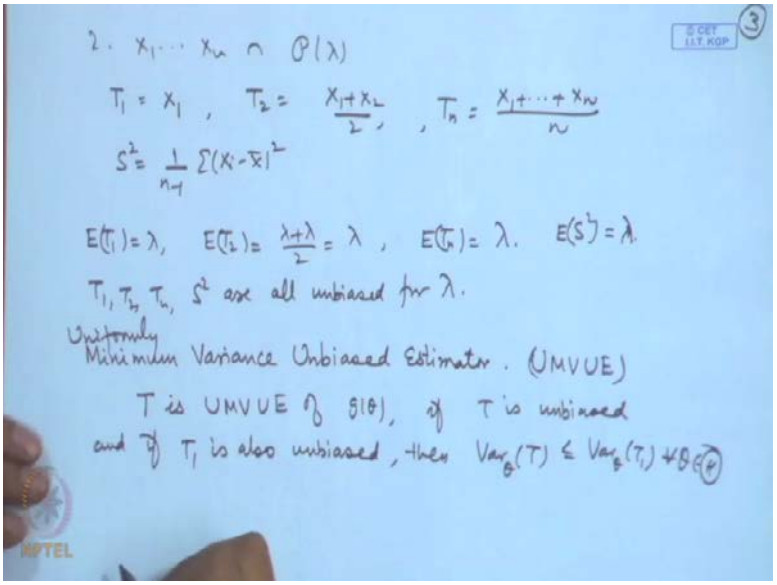
So what we say, when we are considering the population, that population could be may be say they are from normal distribution; that means, the adult heights, if are denoted by random variable X, it may follow a normal distribution with mean mu and variance sigma square, we want to estimate mu. Now, we have taken a random sample X 1, X 2, X n from this population, now one may suggest that use X bar, which is actually one of the functions of the random sample; one may suggest that X bar can be used to estimate mu but some other person may say that X bar may have certain disadvantage. For example, it is affected by the extreme values.

So, one may say use X median; that is median of X 1, X 2, X n for estimating mu. One may say use, for example, geometric mean of X 1, X 2, X n and so on. One may propose various values; similarly if we are considering an estimation of sigma is square, I may consider variance term calculated from the sample that is 1 by n sigma X i minus X bar whole square or here also, I have use the notation 1 by n minus 1 sigma X i minus X bar whole square, which we called sample variance.

So, one may suggest using this; one may use say mean deviation from the median or mean deviation from the mean, once again the question arises, which one should be used so that brings us to certain criteria of estimation. So there are various criteria of estimation; we will consider here only two of them; one is unbiasedness; so, an estimator T (x) is said to be unbiased for g theta, if expectation of T (x) is equal to g (theta), for all

theta. physically if we want to interprets this is statements; it means that on the average T (x) must be equal to g theta; that means, if I consider all possible samples and then if I take the average of the T (x) value calculated from the all the samples, then it should be equal to g theta. Let us consider examples here, suppose I have an observation from binomial distribution; that means this is the number of success conducted in n trial here; n is of course known here. The problem is of estimating p, where p lies between 0 to 1, than one may consider say expectation of X by n, and then naturally this is equal to 1 by n into the expectation of that is the n p, that is equal to p.

(Refer Slide Time: 12:05)



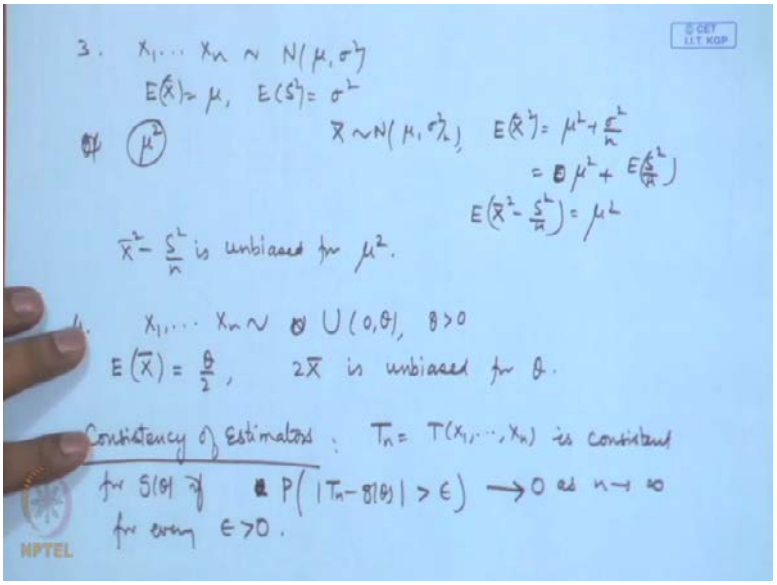
So, you can say here, that the sample proportions X by n; sample proportion is unbiased for population proportion which is unknown to us. Let us take another case say X 1, X 2, X n following Poisson distribution. Then we may define, say T 1 is equal to say X 1, let me define say T 2 is equal to say X 1 plus X 2 by 2, let may define T n that is equal to say X 1 plus X 2 plus X n by n. Let me also define, say S square that is 1 by n minus 1 sigma X i minus X bar whole square. Let us check, expectation of T 1 is lambda, what is expectation of T 2 that will be lambda plus lambda by 2, which is equal to lambda. If I consider expectation of T n that is also lambda, if I consider expectation of S square then that is also equal to lambda.

So, we have several unbiased estimators so T 1, T 2, T n, S square; these are all unbiased for lambda of course, then we will introduce some other criteria to check which one is

preferable among these, we introduce the concept of minimum variance unbiased estimation. So we say T is minimum variance unbiased estimator let us say MVUE.

So we say uniformly, because over the whole parameter is based should be UMVUE; UMVUE of say g theta, if T is unbiased, and if T 1 is also unbiased. Then variance of T is less than or equal to variance of T 1, for all theta. So, there are methods for deriving the unbiased estimators, for example, there are methods of minimum variance unbiased estimator, for example, there are method of lower bounds there is a method using the completeness un sufficiency of the statistics etcetera. However, we will not get too much into detail here; I will end up with two more applications of the unbiasness here. Let us consider, say X 1, X 2, X n following normal mu sigma square, then expectation of X bar is mu and expectation S square is sigma square.

(Refer Slide Time: 14:43)

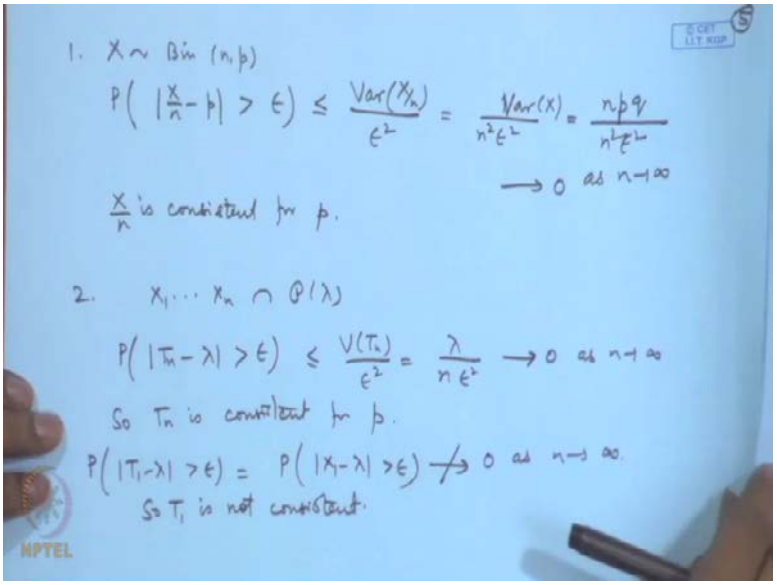


So, unbiased estimator of mu and sigma square exist here suppose, I want to calculate for say mu square, suppose my g function is mu square. Now, then let us look at this, X bar follows normal mu, sigma square by n so expectation of X bar square is equal to mu square plus sigma square by n, and this we can write as mu square plus expectation of X square by n. This you bring to the left hand side so you get expectation of X bar square minus S square by n that is equal to mu square. So, X bar square minus S square by n is unbiased for mu squared. In fact one can show that, this is your UMVUE; but that will require some additional arguments. Let us take X 1, X 2, X n following normal say

uniform (0, theta) distribution. If I consider X bar, then expectation of X bar is equal to theta by 2.

So, 2 X bar is unbiased for theta, we have another concept; that is called consistency of estimators so T n that is equal to T of X 1, X 2, X n so we are showing exact dependence that there are n observations taken. So, I am writing here T n; this is consistent for g theta. If probability that modulus T n minus g theta greater than epsilon goes to 0 as n tense to infinity; for every epsilon greater than 0. Let us take the case of unbiased estimation, I have considered several examples; I will take each of these here.

(Refer Slide Time: 17:22)

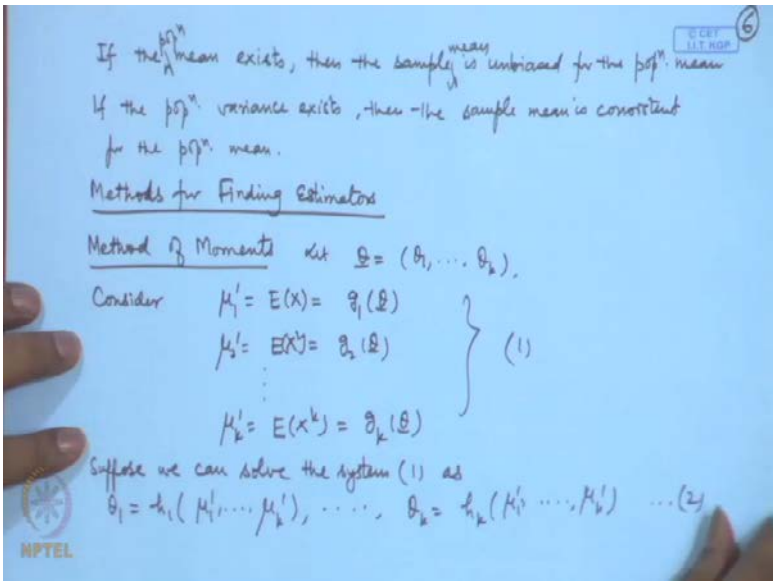


Let us consider say binomial (n, p) here, x follows binomial (n, p). Let us consider probability of modulus X by n minus p greater than epsilon, we can use semi shapes equal to here then this is less than or equal to variance of X by n divided by epsilon square; that is equal to 1 by n square, epsilon square in two variance of X; variance of X in the binomial distribution is n p q divided by n square epsilon square. Now, in the denominator have n squared here; so this goes to 0 as n tense to infinity.

So, X by n is consistent for p let us take the second example X 1, X 2, X n following Poisson (lambda). Here, I have introduced several estimators T 1, T 2, T n etcetera. Let us take for example T n, then T n minus lambda greater than epsilon once again it is less than or equal to variance of T n by epsilon square, that is equal to lambda by n epsilon square so this goes to 0 as n tense to infinity. So, T n is consistent for p however, if I

consider T 1 then that is equal to probability of modulus X 1 minus lambda greater than epsilon; now, this does not depend upon n. So, these cannot go to 0 as n tense to infinity.

(Refer Slide Time: 19:31)



So, T 1 is not consistent. I will stay end this criteria unbiasedness and consistency by stating two results. If the mean exists than the sample mean is unbiased; sample mean is unbiased for the population mean. So, if the population means exist; that means, if expectation of X is defined for example, in the case of quasi distribution expectation X is does not exist.

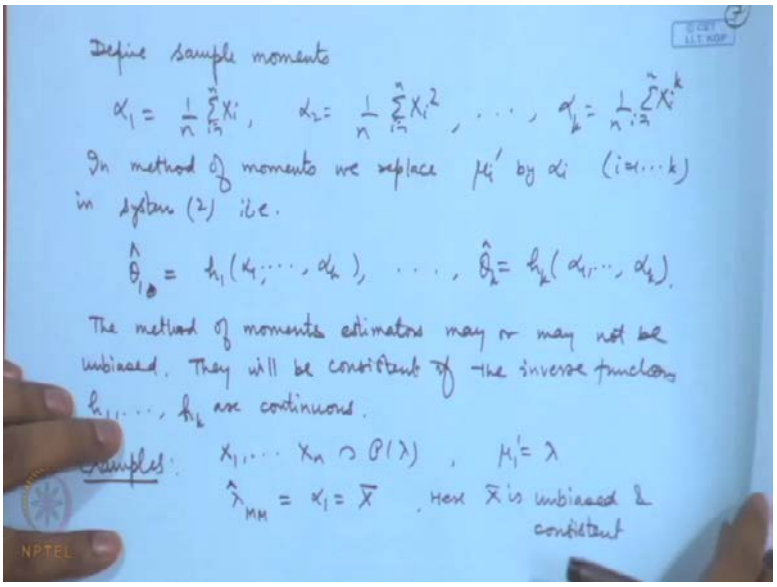
So, in that case exist statement will not be true, if the populations mean exist then the sample mean is unbiased for the population mean. If the population variance exists then the sample mean is consistent for the population mean; this is statement I am saying, because of using the semi shapes and equality in the previous two examples, because here variance is being used in this case also variance is being used. However, if we use the legal of large numbers etcetera than we may not also require this condition, and we can say only that the sample mean is always consistent for the population in that means sample mean must exist. Now, we discuss certain methods for finding out the estimators methods.

For finding estimators, the finding out of unbiased estimator are consistent estimator is relatively easy, because we can guess about the function here, but there can be other cases for example, I may consider a log normal distribution, I may consider a uniform

distribution with two parameters, and I may consider exponential distribution with two parameters. So, in these cases it is not so easy to guess the form of the unbiased estimator or the consistent estimator.

So, in that case firstly, letters have proper method for deriving the estimator, and then we can proceed to check their desirable properties. So, one of the first or you can say elementary methods are the method of moments let theta be k dimensional parameter and consider mu 1 prime that is expectation X that is the first moment. So, it will be some function of theta 1, theta 2, and theta k. mu 2 prime that is expectation of X square that is equal to say g 2 of theta and so on. Consider k Th moment that is a g k of theta. Suppose we can solve, let us call it system one; suppose we can solve the system one, as theta 1 is equal to say h 1 of mu 1 prime, mu 2 prime, mu k prime and so on theta k is equal to h k of mu 1 prime, mu 2 prime, mu k prime.

(Refer Slide Time: 23:47)

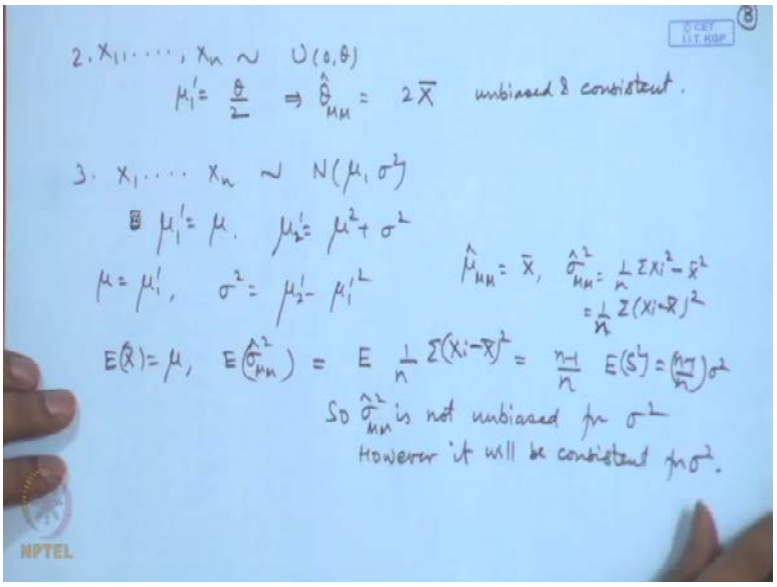


Let us define sample moments, say alpha 1 is equal to 1 by n sigma X i, i is equal to 1 to n, alpha 2 is equal to 1 by n sigma X i square; i is equal to 1 to n and so on, alpha k is equal to 1 by n sigma X i to the power k is equal to 1 to n. In method of moments, we replace mu I prime by alpha I, for i is equal to 1 to k in system two; that is theta 1 method of moment's estimator; let me call it theta 1 ahead; that is equal to h 1 of alpha 1, alpha 2 and alpha k and so on. Theta k ahead is equal to h k of alpha 1, alpha 2, alpha k .now, this is the general guide line, if I have a two dimensional parameter then I will

consider in general two moments, but sometimes we may have to take three also or sometimes we may have to take only one, because there may be entire relationship between those parameters.

So, but this is general guideline, that if I have a k dimensional parameter then I will consider k moments. Let me explain through the examples here, the method of moments estimators may or may not be unbiased. They will be consistent, if the inverse functions h 1, h 2, h k are continuous. Let us consider the Poisson lambda case, and then here mu 1 prime is equal to lambda. So, lambda ahead method of moments estimator is equal to simply alpha 1 that is equal to X bar. So, if we see here, our several proposed estimators for the lambda in the Poisson distribution case that T 1, T 2, T 3 and T n, and S square etcetera, among them X bar is the one; which is obtained through the method of moments and here X bar is unbiased as well as consistent.

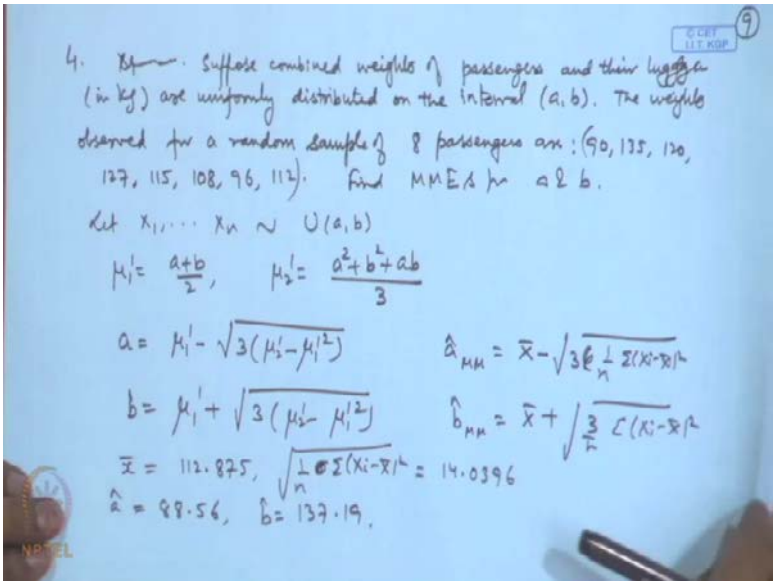
(Refer Slide Time: 27:16)



Let us consider say X 1, X 2, X n following uniform (0, theta) here, the first moment is theta by 2 and therefore we get theta ahead method of moments estimator as 2 mu 1 prime; so, mu 1 prime will be replace by alpha 1 that is X bar so that is 2 X bar and once again this is unbiased and consistent. Let us consider say x 1, x 2, x n following normal mu sigma squared here, mu 1 prime is equal to mu, and mu 2 prime is equal to mu square plus sigma square. So, when we write the solution, mu is equal to mu 1 prime and sigma square is equal to mu 2 prime minus mu 1 prime square.

So the method moments estimators, for mu it will be X bar and sigma it will be equal to 1 by n sigma X i square minus X bar square; that is 1 by n sigma X i minus X bar square. Here, if you look at expectation of X bar; that is mu, but expectation of sigma had square mm; that is equal to, because 1 by n minus 1 is unbiased. So, 1 by n will be expectation of 1 by n sigma X i minus X bar square; that is equal to n minus 1 by n expectation of S square; that is n minus 1 by n sigma square. So, sigma ahead square mm is not unbiased; however, it will be consistent for sigma square.

(Refer Slide Time: 29:30)

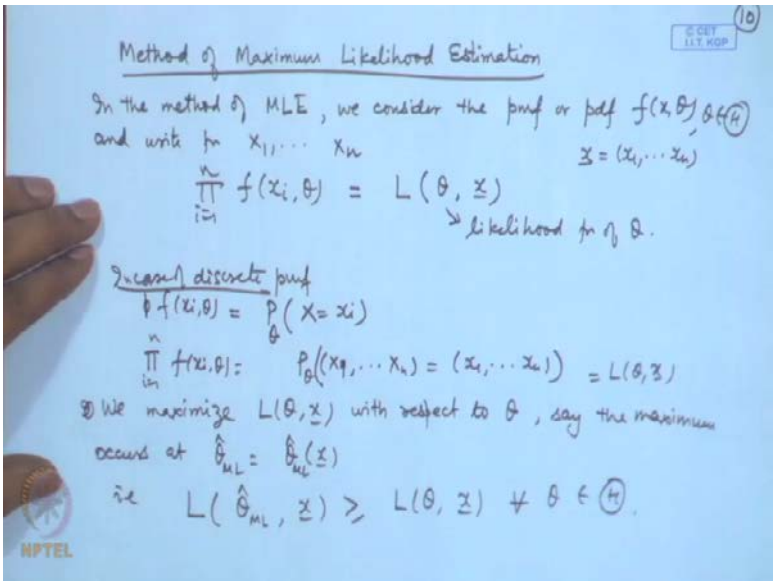


Let us consider, say suppose combined weights of passengers and their luggage in kilograms are uniformly distributed on the interval a to b. The weights observed for random samples of 8 passengers are; 90, 135, 120, 127, 115, 108, 96, 112. We want to find out the method of moments estimators for a and b. Now, if I assume say X 1, x 2, X n following uniform a to b. Then mu 1 prime is equal to a plus b by 2, and mu 2 prime is equal to a square plus b square plus ab by 3. So, if we solve for a and b, here I will get a is equal to mu 1 prime minus square root 3 mu 2 prime, minus mu 1 prime square, and b is equal to mu 1 prime plus square root 3 mu 2 prime, minus mu 1 prime square.

So, a head mm will be equal to X bar minus square root thrice 1 by n sigma X i minus X bar square, and b head mm will be equal to X bar plus square root 3 by n sigma X i minus X bar n whole square. So, if I consider this particular sample here, then X bar in this problem turns out to be 112.875 and this value 1 by n sigma X i minus X bar whole

square, that value turns out to be 14.0396. Therefore ahead turns out to be 88.56 and b ahead turns out to be 137.19.

(Refer Slide Time: 33:06)

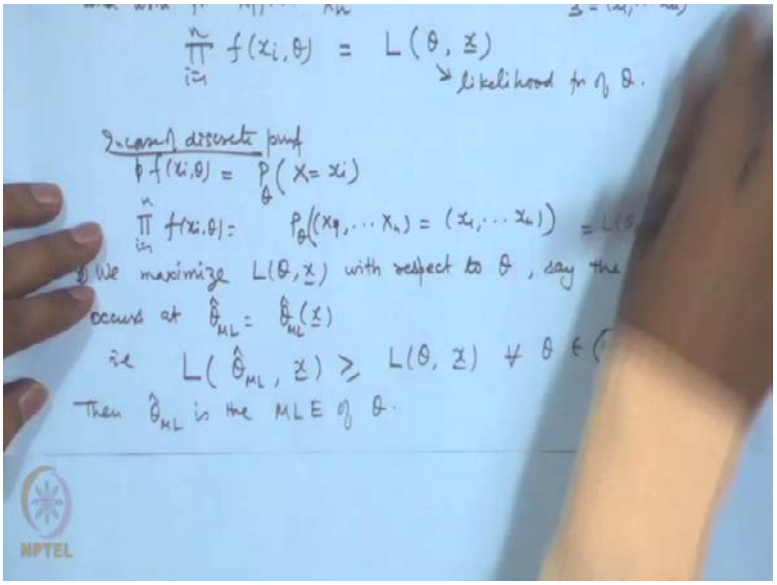


So, the method of moments estimators for a and b are 88.56 and 137.19 another more popularly used and more powerful method of estimation is for finding the estimators is called the method of maximum likelihood. In method of maximum likelihood; so, I will use the word MLE or maximum likelihood estimation or maximum likelihood estimators. We consider the probability mass function or probability density function f (x, theta) and write for X 1, X 2, X n. So, this will become equal to f (x i, theta) product i is equal to 1 to n. And we denote it to be likelihood function, and I will change the nomenclature from f (x i, theta) 2 f, L(theta, x), where x is denoting the sample values here x 1, x 2, x n; this is called the likelihood function of theta.

What is it mean, suppose I am considering probability mass function than f (x i, theta) in case of discrete PMF. You will have f (x i) is probability of X is equal to x i; when theta is the true value of the parameter. So, when we consider product of f (x i, theta) we can write it as probability of X equal to X 1, X 2, X n is equal to x small x 2 small x 1; that means, this is the probability of observing the sample X 1, X 2, X n; when theta is the true parameter. In method of maximum likelihood, we interpret in a different way we call it the likelihood of the sample, when theta is the true parameter value. And we maximize this L (theta, x); this L (theta, x) as a function of theta. Naturally, when you

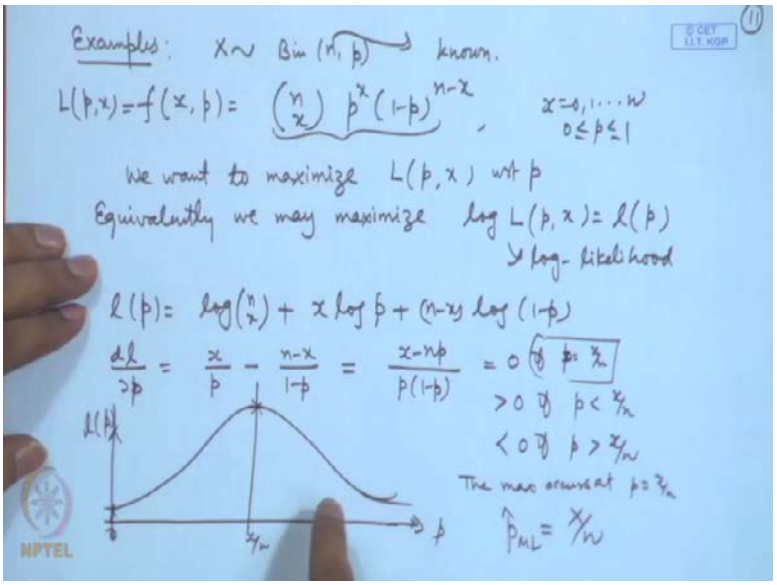
maximize over theta, you have to they consider all values of the parameter over the parameter space here, where you may have theta belonging to a parameter space theta, we maximize L (theta, x) with respect to theta, say the maximum occurs at theta head ML; that is equal to theta head ML X 1, X 2, X; that is you will have L (theta, head ML) is greater than or equal to L (theta, x), for all theta belonging to theta.

(Refer Slide Time: 36:45)



Then we say that, theta head ML is the maximum likelihood estimator of theta, let me explain through certain examples here.

(Refer Slide Time: 36:59)

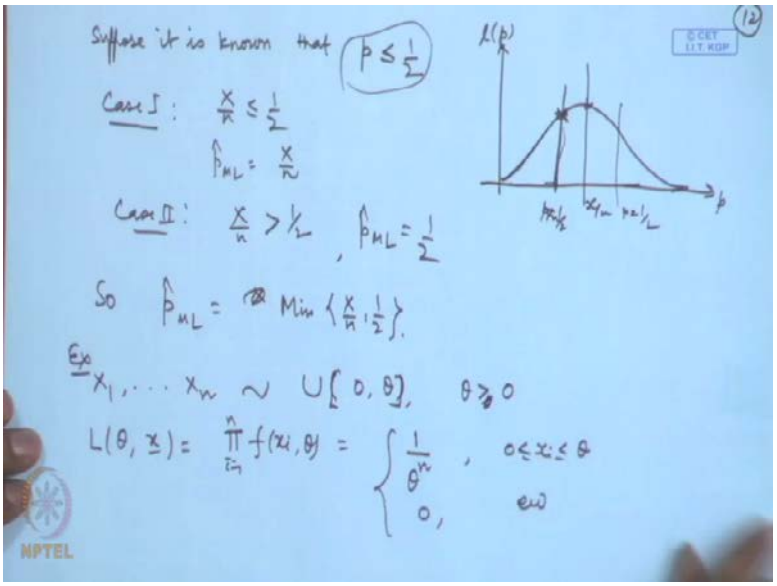


Let us consider binomial (n, p); where n is known here the probability mass function n c x p to the power x 1 minus p to the power n minus x; where x can take values 0, 1 to n and p lies between 0 to 1. Then we are considering these us now the likelihood function. So, I will call it as a function of this, now we want to maximize L of p, x with respect to p. Now, p is ranging over an interval, so we can apply the usual methods of the calculus like we can try to see that, whether L is an increasing function or decreasing function or the range where it is increasing and then where it is decreasing at sector. We can simplify the situation, by looking at equivalently we may maximize log of L.

So, we use different notation log likelihood; so this is called log likelihood. So, l p here turns out to be log of n c x plus x log of p plus n minus x log of 1 minus p. Let us consider the derivative of log likelihood respect to p. So, here x is not a variable here, we are considering it as a function of p, that is why I am not putting x here; that is equal to x by p minus n, minus x divided by 1 minus p; that is equal to x, minus n p divided by p into 1 minus p. Now, note here that this is equal to 0, actually if p is equal to X by n. So, this is suddenly greater than 0, if p is less than X by n; it is less than 0, if p is greater than X by n. Say if we plot the L function, and this side we have p, and on this side you have l p then starting from 0 on words. The value of this as p goes to X by n; this is increasing and when p is greater than X by n; this is decreasing. Naturally, the maximum is occurring at X by n.

So, the maximum occurs at p is equal to X by n so p head ML the maximum likelihood estimator for the population proportions are the probability of success in a binomial distribution terms out to be x by n. Now, notice here that, here p was in the interval 0 to 1 and you can see this x, y and also lies between 0 to 1. There may be a situation, where we have some prayer information about this parameter p. For example, we may know that suppose it is a related to certain success failure experiment, where we may know that say p is less than or equal to half or p is greater than or equal to half. In that case, we will not write the answer X by n; because X by n then may cross the region, in the method of maximum likelihood estimation we have to consider the maximization over the given parametric space only.

(Refer Slide Time: 41:33)

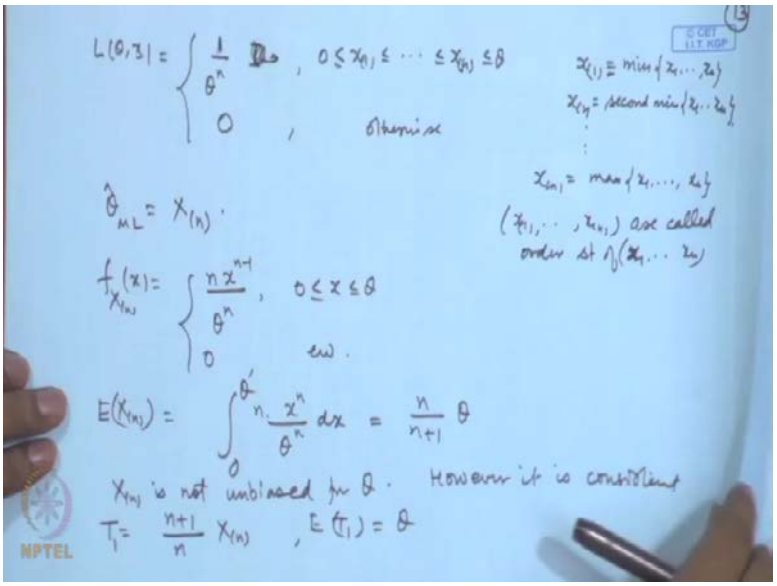


So, let us consider that analysis here, suppose it is known that say p is less than or equal to half in this problem; if p is less than or equal to half. Let us consider there are two possible t is them. See we had this has X by n, on this side you have p; this is l (p). If p is less than or equal to half given to us, then there may be two cases; case one, X by n is less than or equal to half; that means, p is somewhere here, p is equal to half is somewhere here; in that case the maximization occurs at X by n but there may be another case that p is equal to half is here, in that case, if you see this x by n goes out of the region of the parameter. Therefore, we cannot consider X by n as the maximum likelihood estimator you will notice the likelihood function in the region 0 to half itself; now in these region the maximum value is attained at half. So, in these cases when x by n is say greater than half, then p head ML you have take to be half. So, the answer is p head ML it is equal to X by n that is it is equal to minimum of X by n and half.

So, you can see here, that there is a direct effect of the parametric space on the maximization problem or the optimization problem as we may call. Let us take some more problem say X 1, X 2, X n follows say uniform distribution on the interval 0 to theta; where theta is of course, greater than 0. Now, in this case the likelihood function; this is equal to the joint density function. Now, in the case of uniform distribution the density is 1 by theta, over the region that each X i is between 0 to theta; it is equal to 0, elsewhere. Now, if you look at this thing directly and try to maximize with respect to theta, then you will get an absolute result, because this is theta in the denominators of the

theta should go to 0, but that will give us the value as infinite so this will give observed T here and of course, when you are considering the parametric space from 0 to infinity, and then saying that theta is equal to 0; which is not dependent upon the observations is an absorbed result. So, where we are missing is that we are ignoring the region.

(Refer Slide Time: 44:40)

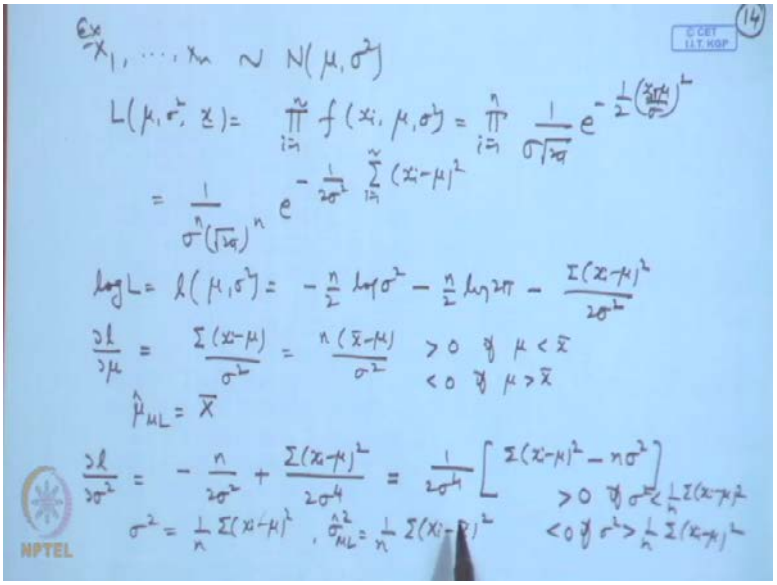


So, if we look at the region properly, we can write this likelihood function in a more appropriate fashion as 1 by theta to the power n indicator function of the set. So, firstly let me say, we can write it as, if I consider the order is statistics; that is X 1 is equal to minimum of X 1, X 2 X n. X 2 as the second minimum of X 1, X 2, X n and so on. X n is equal to the maximum of X 1, X 2, X n; these are known as X 1, X 2, X n are called order a statistics of X 1, X 2, X n. So, we can write the region as, now you see from here, if I am looking at 1 by theta to the power n the minimum value of theta; that is possibly is X n. So, theta had ML is equal to the largest orderly statistics you compare it with the method of moments estimator for uniform distribution; the method of moments estimator for the uniform distribution was 2 X bar.

So, here the two things are quite different and we may also consider, whether it is unbiased or not. For example, we may consider the distribution of X n; that is n x to the power n minus 1 divided by theta to the power n. So, if we consider expectation of X n that is equal to integral power n x to the power n by theta to the power n d x, 0 to theta; that is equal ton by n plus 1 theta.

So, X n is not unbiased for theta; however, it remains consistent and we may also consider alternatively like T 1 is equal to n plus 1 by n x n, then expectation of T 1 will be equal to theta. So, from the maximum likelihood estimator we can consider little bit of adjustment to make it unbiased, in fact it can be shown that this is minimum variance and biased estimator of theta.

(Refer Slide Time: 47:28)



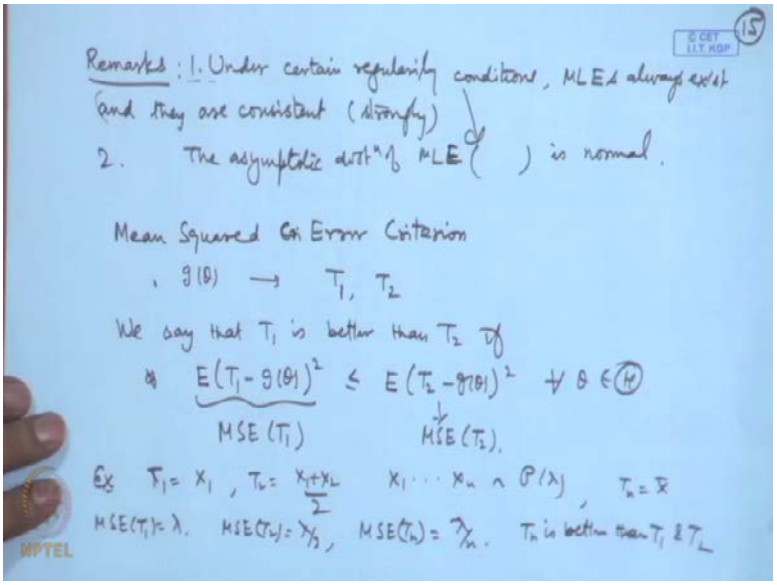
Let us consider X 1, X 2, X n a random sample from say normal mu, sigma square distribution; this is a two parameter problem. So, here the likelihood function will depend upon mu and sigma square, that is the joint density of X 1, X 2, X n that is equal to product i is equal to 1 to n, 1 by sigma root 2 pi e to the power minus 1 by 2 x minus mu by sigma x i minus mu by sigma square.

So, that is equal to 1 by sigma to the power n root 2 phi to the power n e to the power minus 1 by 2 sigma square sigma x i minus mu square i is equal to 1 to n so log likelihood function, which I get denote by small l; it is equal to minus n by 2 log sigma square minus n by 2 log 2 phi, minus sigma X i minus mu square by 2 sigma square. If we consider say del l by del mu, I get sigma X i minus mu by sigma square and this is nothing but n times X bar minus mu by sigma square.

So easily we can see, it is greater than 0, if mu is less than X bar, it is less than less than 0, if mu is greater than X bar. So, as a function of mu if you see, it is increasing up to X bar and decreasing after X bar. Therefore, the maximum likelihood estimator of mu will

become equal to X bar. On the other hand, if I want to consider with respect to sigma square then I differentiate with respect to sigma square, I get minus n by 2 sigma square plus sigma X i minus mu square by 2 sigma to the power 4; which we can consider as 1 by 2 sigma to the power 4 sigma X i minus mu square minus n sigma square square is less than 1 by n sigma X i minus mu square and it is less than 0 if sigma square is greater than 1 by n sigma X i minus mu square. Therefore, the maximization of sigma square occurs at 1 by n sigma X i minus mu square, but this involves mu which is unknown, but maximization with respect to mu, we have already consider that is obtain that X bar. So, sigma head square ML that turns out to be 1 by n sigma X i minus X bar square; in this particular case these estimators are same as the method of moments estimators for this problem and naturally, here X bar is n biased for mu, but 1 by n sigma X i minus X bar whole square is not unbiased.

(Refer Slide Time: 51:01)



We have certain remarks here regarding the maximum likelihood estimators, under certain regularity conditions maximum likelihood estimators always exist and they are consistent. In fact, they are strongly consistent; under these conditions the asymptotic distribution of MLE; under these regularity conditions is normal.

So, these maximum likelihood estimators are quite useful. In general, they are dependent upon the sufficient statistics is not introduce a concept of sufficiently statistics in this code still now. They are strongly consistent they always exist under regularity conditions

and the asymptotic distribution is normal; so, these are all large sample properties which are satisfied by the maximum likelihood estimators, and that is why they are quit preferred in a statistical theory of course, there are certain situations we are one can find better estimators then maximum likelihood estimators also, but by for these are most commonly used, I will just end these lecture by introducing the concept of better.

So, we can define the concept of mean squared error criteria so for estimating a parametric function g theta we may have estimator say T 1 and T 2, then we say that T 1 is better than T 2. If expectation of T 1 minus g (theta) whole square is less than or equal to expectation of T 2 minus g (theta) whole square for all theta, this term; this is called the mean squared error estimate of the estimator T 1, this is called the mean squared error of the estimator T 2. Let us consider the estimators T 1 and T 2 in the Poisson example, T 1 was X 1 and T 2 was X 1 plus X 2 by 2, where X 1, X 2, X n; where Poisson (lambda) and also we have consider T n as equal to X bar. Now consider here, mean squared error of T 1; that one is lambda, mean square error of T 2; that will be equal to lambda by 2. If consider mean squared error of T n; that will be equal to lambda by n.

So, certainly T n is better than T 1 and T 2. There are general concepts of loss functions, in place of this is squared error; one may consider some other loss functions are change absolute error power 4 linear loss function log squared error or in trophy losses. That general concept of loss gives rise to expression known as risk function, and than one prefers the estimator; which as the smaller risk function this topic comes under the general concept of decision theory so we do not intend to cover in this particular course; in this particular course we will also consider interval estimation and testing of hypothesis. So, I will plan to cover it in the following lectures here.