Statistical Methods of Scientists and Engineers

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

Lecture - 10
Parametric Methods - II

In the previous lecture, I have introduced the concept of point estimation, what is the problem and we are considering the parametric methods. That means we are assuming that the unknown populations distribution is known. However, it may depend upon unknown parameter. We have considered certain criteria for judging the goodness of estimators. For example, we have considered the criteria of unbiasedness then consistency.

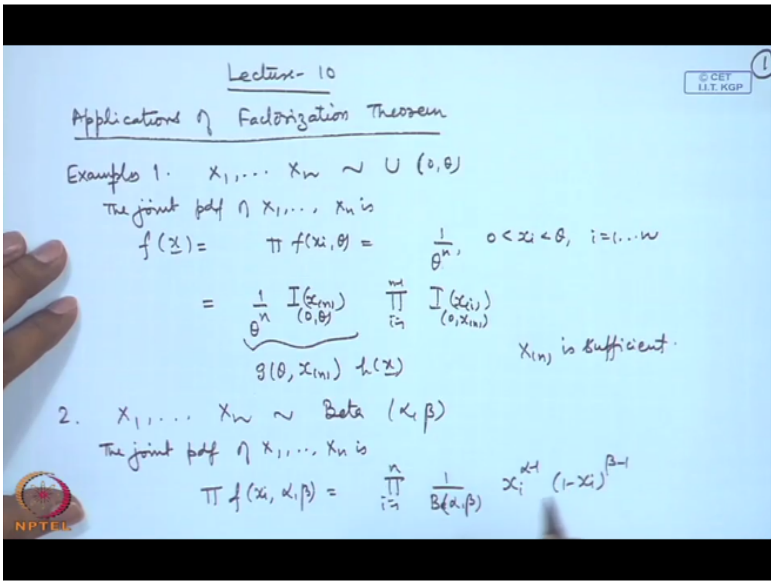
I also introduced the concept of mean squared error criterion, that means an estimator which has a smaller mean squared error over the parameter space will be considered better than the one which is having slightly larger mean squared error. If the estimator is unbiased, then the mean squared error reduces to the variance of an estimator. So, therefore we have the concept of uniformly minimum variance and biased estimator, which we call shortly UMVUE.

I mentioned that in order to obtain the UMVUE, we have broadly speaking 2 methods. One is the method of lower bounds. So under certain conditions or sometimes without conditions, one can obtain a lower bound for the variance of an unbiased estimator. Therefore, an estimator which will achieve that lower bound will be called the minimum variance or it will be the minimum variance unbiased estimator.

In this particular course, we will not be discussing those methods. However, let me briefly introduce another method which is based on the concept of completeness and sufficiency. So, I introduced a sufficient statistics and I gave a consequence of that, which is called Rao-Blackwell theorem that if there is an unbiased estimator, which may not depend upon the sufficient statistics, then we can construct another unbiased estimator.

Which will be simply a function of the complete sufficient statistics and whose variance will be less than or equal to the variance of the original estimator and this also be unbiased, now coupled with another concept of completeness.

(Refer Slide Time: 02:40)



Let me introduce that, and firstly let me consider the applications of the factorization theorem, which basically produces the sufficient statistics in given problem. Of course, one may see that from the definition, if conditional distribution of X1, X2, Xn given T is independent of the parameter, if T itself is a function of say U, then U will also be sufficient. However, we can consider something called minimal sufficiency that means maximum reduction of the data.

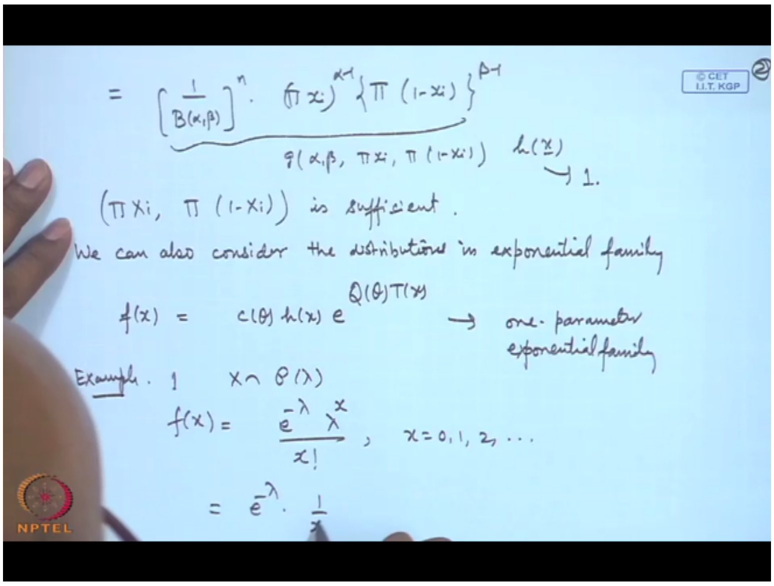
I will not get to much into technical details here, rather we will look at the direct application. So, let us consider say X1, X2, Xn follow say uniform distribution on the interval 0 to theta. Now, how do you write down the join density? The joint probability density function of X1, X2, Xn is so I will just write Fx that = product of Fxi theta that = 1/theta to the power n, for 0 < xi < theta, for I = 1 to n. now in order to apply the factorization theorem, we need to represent in a slightly compact form, because here this range is coming separately.

So, we write it as 1/theta to the power n indicator function of xn over the interval 0 to theta * the product of xi, i = 1 to n-1 and all of them will be from 0 to xn. If we look at this, this can be considered as g theta and xn and this is a function of observations alone. So, here xn is sufficient, that is the maximum of the observations. If we remember one exercise, which I did for the consistency. In this one I proved that xn is consistent for theta.

Now, here I am observing that xn is also sufficient. Now, here in the uniform distribution theta/2 is the mean, that means x bar will be unbiased. But x bar is not based on xn. Therefore, I can construct another estimator which will be based on xn and whose variance will be smaller than x bar than 2 x bar. For theta, it will be 2 x bar. So, we will show it later. Now, let us consider some more examples.

Say consider X1, X2, Xn follow say beta distribution with parameters alpha, beta. That means I am considering the joint pdf. So, that is product of Fxi alpha, beta that = product of i = 1 to n, 1/beta function alpha, beta x to the power alpha-1 1-xi to the power beta-1. So, this you can see it will be.

(Refer Slide Time: 06:52)

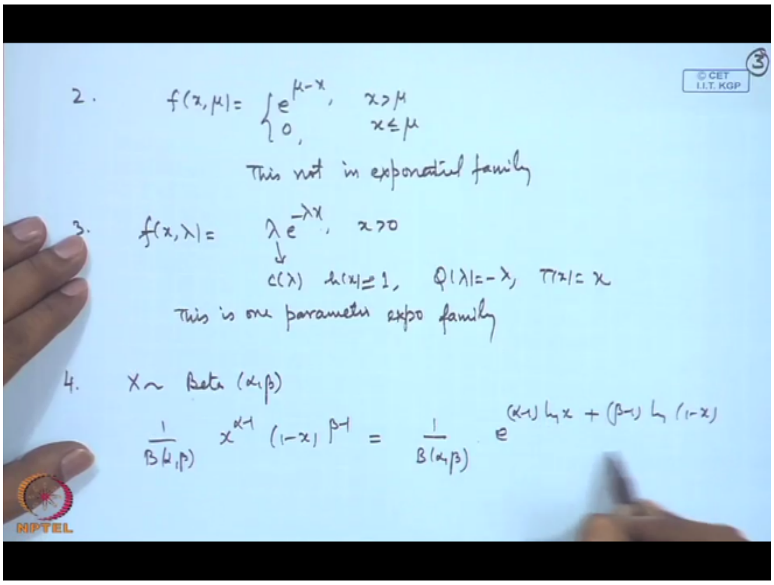


This can be written as 1/beta alpha, beta to the power n product of xi to the power alpha-1 product of 1-xi to the power beta-1. Here, this entire thing can be considered as a function of parameters alpha, beta product xi and product 1-xi and then hx, you can consider to be 1 itself. So, here product xi and product of 1-xi that is sufficient. Another way of looking at this concept of sufficiency is in the form of we can consider the distributions in exponential family.

Let me define one parameter exponential family and multi parameter exponential family. So, we consider c theta hx e to the power Q theta Tx, this is called one parameter exponential family. To give an example, say you consider X following Poisson lambda. How do you write down the distribution? E to the power -lambda, lambda to the power x/x factorial, for x = 0, 1, 2. This we can write as e to the power -lambda 1/x factorial e to the power x log lambda.

So, if I define Q lambda = lambda, Tx = x, c lambda = e to the power -lambda and hx = 1/x factorial. Then this an example of one parameter exponential family. That means the Poisson distribution belongs to one parameter exponential family. Note that this exponential family is different from exponential density that we discussed earlier. This is exponential family.

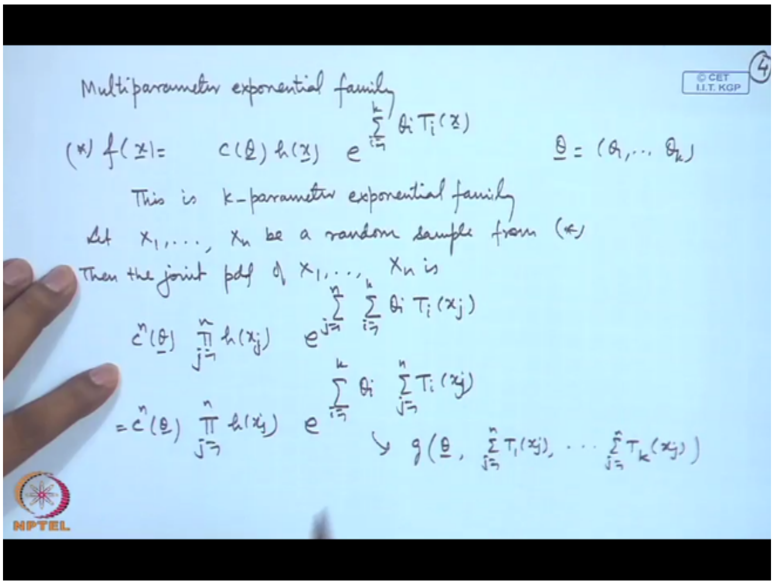
(Refer Slide Time: 10:05)



Let us take say exponential distribution itself, say fx mu that = e to the power mu-x, for x > mu 0 for x<=mu. Then, this is not in exponential family. Let us consider say fx lambda = lambda e to the power -lambda x, then here this can be considered as c lambda hx is 1, Q lambda = -lambda, Tx = x. So this is again one parameter exponential family. Let us consider this beta distribution that I wrote beta alpha, beta.

This is 1/beta alpha, beta x to the power alpha-1 1-x to the power beta-1. Now, this we can write as 1/beta alpha, beta. This is e to the power alpha-1 log x+beta-1 log 1-x. now, that gives rise to multi parameter exponential family. So, let me introduce that here. Because here we are having 2 terms coming here.

(Refer Slide Time: 11:55)

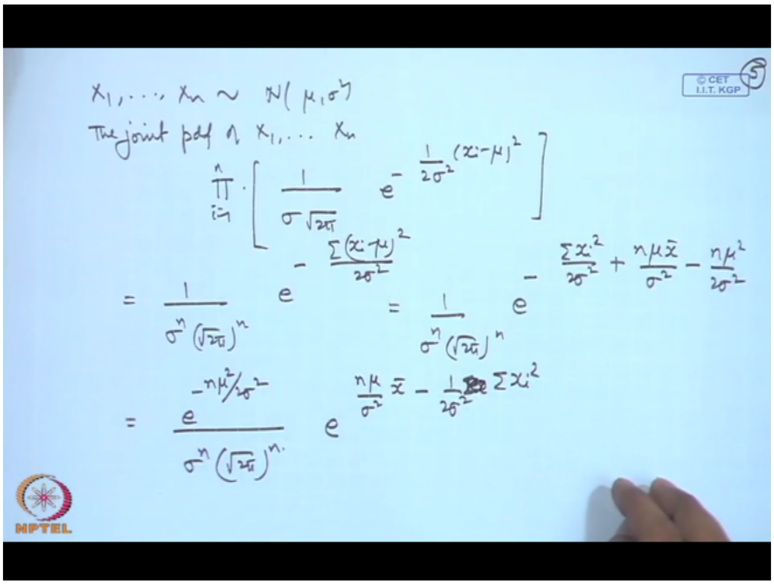


So in general we can define multi parameter exponential family. So, let us consider fx as c theta hx e to the power sigma theta i Ti x, for i = 1 to k. So, here theta is a vector parameter, theta 1, theta 2, theta k. this is k parameter exponential family. So, if you look at the distribution that I introduced here the beta this one then we can write as c of alpha beta and this is then theta 1, this is theta 2, this is T1x, this is T2x. So, this is an example of 2-parameter exponential family.

Now, if we look at distributions in the k parameter exponential family and let us apply the factorization theorem and see what is the effect. Let X1, X2, Xn be a random sample from say this distribution star. Then the joint pdf of X1, X2, Xn is c to the power n theta product hxi, i = 1 to n e to the power sigma let me put here j because i is being used here so j = 1 to n, i = 1 to k theta I Ti xj. So, this we can write as c to the power n theta product hxj, j = 1 to n theta i sigma Ti xj j = 1 to n, i = 1 to k.

So, if I consider factorization theorem, then by factorization theorem, I am able to express this as a function of so this is a function of theta and sigma T1xj, sigma T2xj, sigma Tkxj j = 1 to n. therefore, we can say that sigma T1xj and so on, sigma Tkxj is sufficient by factorization theorem. To give an example here, if we consider this beta distribution, in this case sigma log xi and sigma log 1-xi that will be sufficient.

(Refer Slide Time: 16:09)



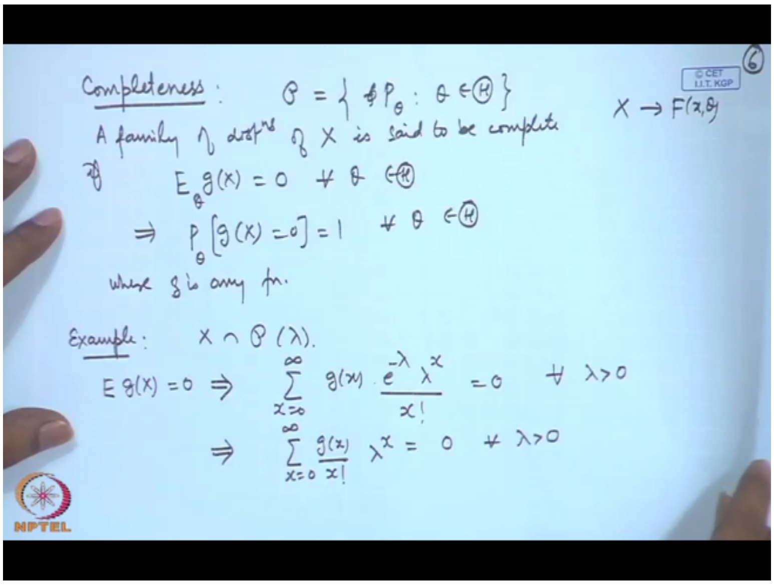
Let us take the more popular normal distribution say X1, X2, Xn follow normal mu sigma square. So, if I write down the joint pdf of X1, X2, Xn, then that = product i = 1 to n 1/sigma root 2 pi e to the power-1/2 sigma square xi-mu square. So, that = 1/sigma to the power n root 2 pi to the power n e to the power -sigma xi-mu square/2 sigma square. Now, this term we can expand and you can write it as e to the power -sigma xi square/2 sigma square+ n mu x bar/sigma square-n mu square/2 sigma square.

So, this is becoming e to the power -n mu square by 2 sigma square divided by sigma to the power n root 2 pi to the power n e to the power n mu/sigma square x bar-1/2 sigma square x bar-1/2 sigma xi square. Now, we can put it in the form of 2 parameters exponential family by defining so this term is simply the function of parameters. So, this is some function of mu and sigma square. Now, this we can call theta 1, that is n mu/sigma square and T1x is x bar, then we can call theta 2 = -1/2 sigma square T2x = sigma xi square.

So naturally you can see that this is a 2 parameter exponential family. This is a 2 parameter exponential family at the same time, we conclude that x bar and sigma xi square is sufficient. We can also write sigma xi and sigma xi square is sufficient, because this is a one to one function. We can also write x bar and sigma xi-x bar whole square is sufficient. Because these are all one to one functions of each other. So we can write down in this any of these forms.

Now after this concept of sufficiency is introduced, let me introduce the concept of completeness and that will help in obtaining a form for the or a methodology to obtain the UMVUE.

(Refer Slide Time: 19:30)



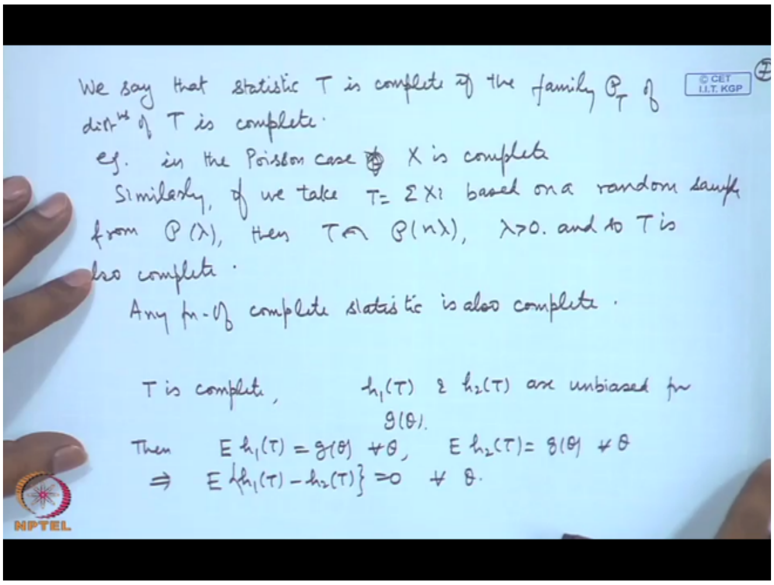
Let us use a notation of P, so if we are considering the distributions P theta. So a family of distributions so X. so we are actually using the notation that x has cdf fx theta. So, in general we can use some abstract notation P theta just to not to mention x there. So family of distributions of x is said to be complete if expectation of gx = 0 for all theta implies probability of gx = 0 = 1 for all theta belonging to theta, where g is any function.

Now to look at some simple application, first of all what is the meaning of this thing. Let us consider say X following Poisson lambda distribution, let us consider expectation of gx = 0. Now, this is equivalent to sigma gx e to the power - lambda, lambda to the power x/x factorial = 0. Now, we can multiply by e to the power +lambda on both the sides, then that is giving us gx/x factorial lambda to the power x.

Now, if you look at the left hand side is a power series in x and we are saying it is vanishing identically over the entire positive real line. The only possibility is that the coefficients must be all 0. That means we are having that gx = 0 for all x, which implies that the probability that gx is o is 1 for all lambda. So, the family of Poisson distributions that is P lambda, lambda > 0 is complete.

Now, we extend this concept of completeness of a family of distributions to a statistic.

(Refer Slide Time: 22:35)

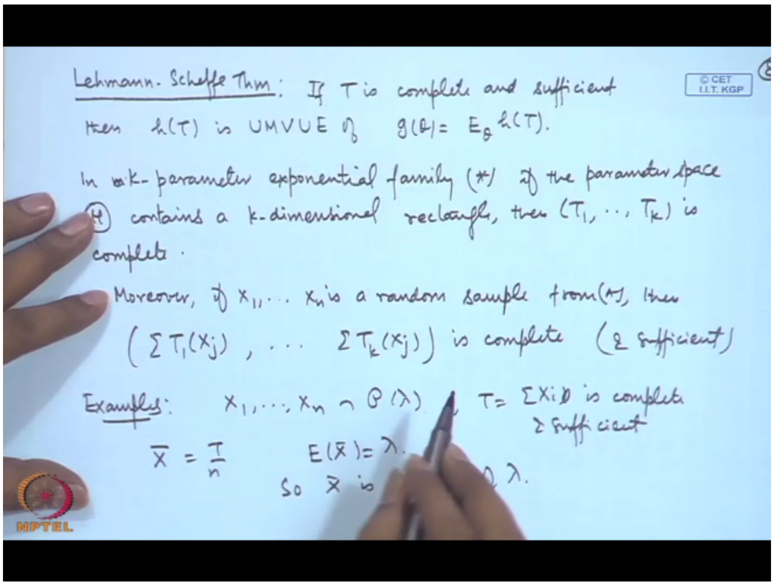


So, we say that a statistic T is complete if the family let me say P T of distributions of T is complete. For example, in the Poisson case X is complete, similarly if we take T = sigma Xi based on a random sample from Poisson lambda then T will follow Poisson n lambda and so T is also complete and of course a consequence that function of complete statistics is also complete. Now, this completeness concept is extremely useful in the sense basically it says that if I am having an unbiased estimator of 0, then that estimator must be 0.

Now that yields to some interesting thing for example, if I say T is complete and I say 2 estimators say h1 T and h2 T are unbiased for say g theta, then expectation of h1 T = g theta and also you have expectation of h2 T = g theta. If I take the difference, then I will get expectation of h1 T- h2 T that = 0 for all theta. Now, h1 T-h2 T is a function of T and if T is complete, then this will imply that probability that h1 T-h2 T = 0, that will be = 1 for all theta.

Basically this means that h1 T = h2 T almost everywhere. That is unbiased estimator based on complete statistic is unique almost everywhere. Therefore, you can say that uniformly minimum variance unbiased estimator can be obtained.

(Refer Slide Time: 26:22)



So, there is a result called Leymann-Scheffe Theorem. In fact, you have a slightly relaxed version of this completeness. That is called bounded completeness. That means if I consider here g to be any bounded function then I can change this to boundedly complete. So, that means, in place of any function if I put only bounded function if for only bounded function this is true then we call boundedly complete. However, this is not required here.

So, if T is complete and sufficient then, hT is UMVUE of g theta that = expectation of hT. Now, once again one can prove actually a completeness for various families for example normal distribution, binomial distribution, Poisson distribution etc. but, in exponential distribution, we have a result which can straight away give the completeness property. I introduce the multi parameter exponential family that is of this form Fx = c theta hx e to the power -sigma theta i Ti x.

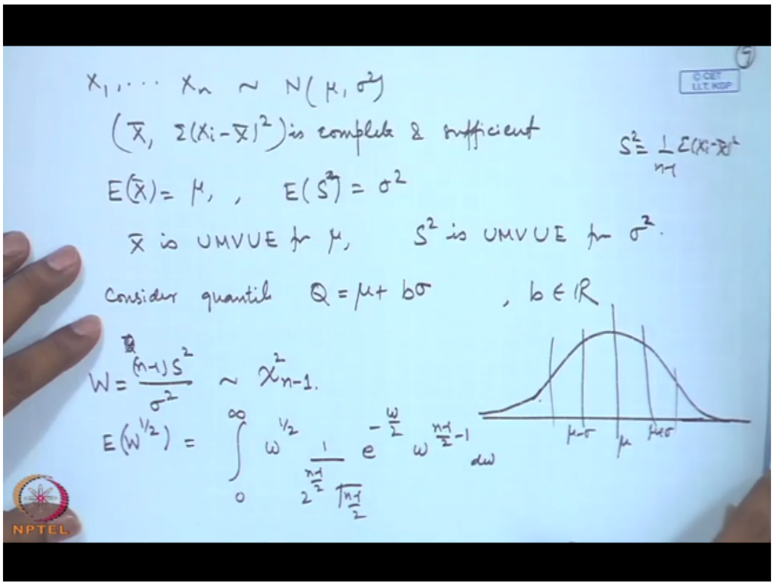
So, if we have distribution of this nature and we have the parameter space say theta, if it is a k parameter exponential family and if the space theta contains a k dimensional rectangle, then T1, T2, Tk will be complete and this result is very useful in proving completeness in various distributions. In k-parameter exponential family star (*) if the parameter space theta contains a k dimensional rectangle then, T1, T2, Tk is complete.

Moreover, if X1, X2, Xn is a random sample from star (*) then, sigma T1 Xj and so on sigma Tk Xj that will be complete and of course sufficient. That means the problem of obtaining the UMVUE reduces to actually determination of complete sufficient statistics and then by making use of that we can simply consider functions of that which are unbiased for the required parametric functions and then you will have UMVUE.

So, let me give you example here so X1, X2, Xn follow Poisson lambda then, T = sigma Xi this is complete and sufficient. So, if I consider X bar, which is simply T/n so, expectation of X bar = lambda so, X bar is UMVUE of lambda. Now, this resolves the problem that for example based on this sample, I could have considered any number of unbiased estimators for lambda.

For example, in Poisson distribution, 1/n-1 sigma Xi-X bar whole square, let me call it U, this is also unbiased for lambda but, since this is not dependent upon X bar alone, because it is using other observations also. So, you will have variance of X bar <= variance of U.

(Refer Slide Time: 31:45)



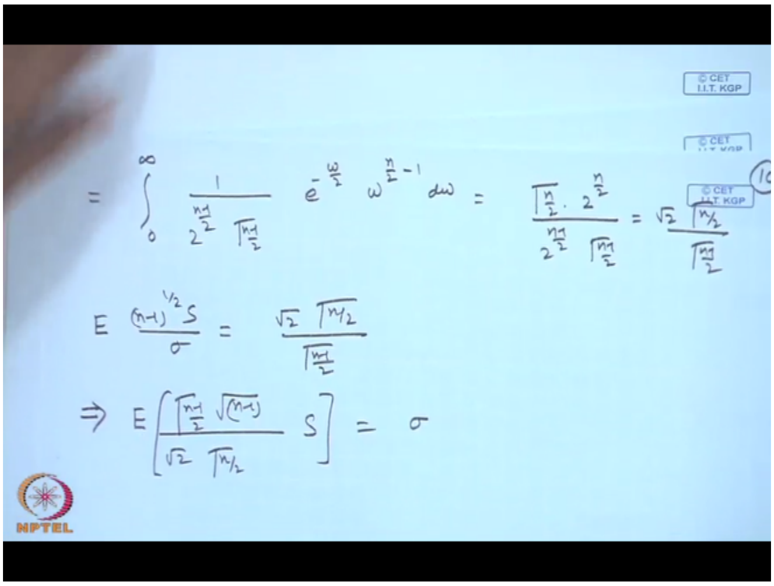
Let us consider say X1, X2, Xn from normal distribution, the popular one so, we have already seen that it is 2-parameter exponential distribution. I showed here in the form X bar and sigma Xi square or X bar and sigma Xi-X bar whole square. So, here X bar and sigma Xi-X bar whole square, this is complete and sufficient. So let us look at expectation X bar that is mu, if I look at let me call this S square is 1/n-1 sigma Xi- X bar whole square.

So, expectation of S square is sigma square. So, X bar is UMVUE for mu, S square is UMVUE for sigma square. Not only that, we can also consider unbiased estimator for other parametric functions for example, in this problem a popular thing could be considered say quantile of the form mu+say b sigma, where b is an arial number. Basically in the normal distribution as I have explained, this is mu, you may have mu-sigma, mu+sigma and so on.

So, in general mu+b sigma is any position on the curve here. So, if we consider this as a function, let me call it Q then, for mu we have X bar, now let us consider estimation of sigma also so, we can make use of n-1 S square/sigma square this follows X square distribution on n-1, these are as I mentioned yesterday in the discussion of the sampling distribution. Now if I make use of this, I can consider expectation of say W to the power half.

So, that = integral 0 to infinity W to the power half 1/2 to the power n-1/2 gamma n-1/2 e to the power -W/2 W to the power n-1/2-1dw. This is the density of the Chi-squared distribution on n-1 degrees of freedom.

(Refer Slide Time: 34:56)



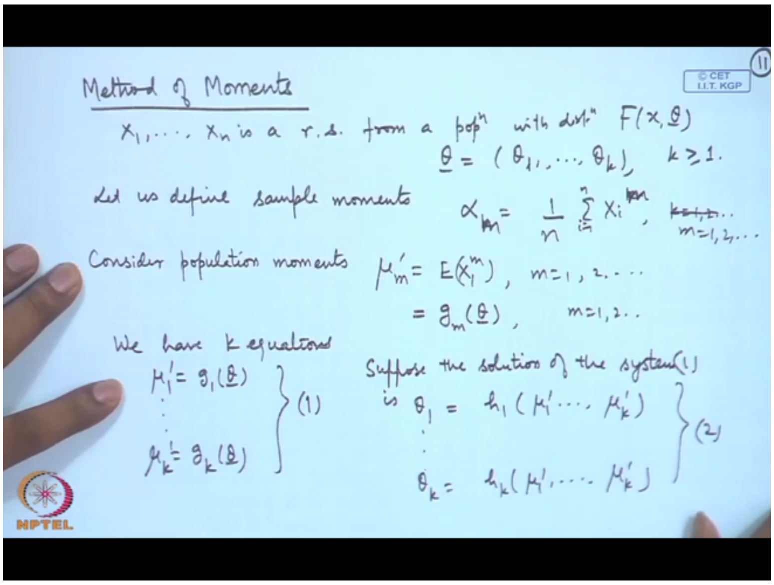
So, let us simplify this terms, this we can write as integral 0 to infinity and these constants will remain as it is and here I can adjust the power n/2-1 dw so, this is nothing but gamma n/2 and 2 to the power n/2 divided by 2 to the power n-1/2 gamma n-1/2. So, that is giving us square root to gamma n/2/gamma n-1/2. So, what we have proved expectation of w to the power half that is n-1 to the power half S/sigma that = root 2 gamma n/2 divided by gamma n-1/2.

That means we can write expectation of gamma n-1/2*n-1 square root divided by square root 2 gamma n/2 S = sigma. So, we are able to obtain. So first of all, since X bar and S square is complete and sufficient this gives this is the UMVUE for standard deviation. Another thing is that if I plug in Q so I get X bar + this root n-1/2 gamma n-1/2/ gamma n/2 S, this is UMVUE for quantile.

So, you can see this concept of complete sufficient statistics is extremely helpful in deriving the uniformly minimum variance unbiased estimators and not only that see if we had not considered the complete sufficient statistics, then for the estimation of sigma perhaps we would have simply used 1/square root 1/n sigma X i-X bar whole square as for sigma square we were using 1/n sigma Xi-X bar square or 1/n-1 sigma Xi-X bar square.

But if you see this one, we are not using that this is slightly different. If we use the concept of minimum mean squared error, then some other estimator is also possible but that I will delay here I will not be considering right now. Now let us consider the method of obtaining estimators. Right now we have discussed the criteria for obtaining estimator and we have shown that, there are estimators which will fulfill those criteria. But, for any population, we can also give some general methods for obtaining estimators.

(Refer Slide Time: 38:02)

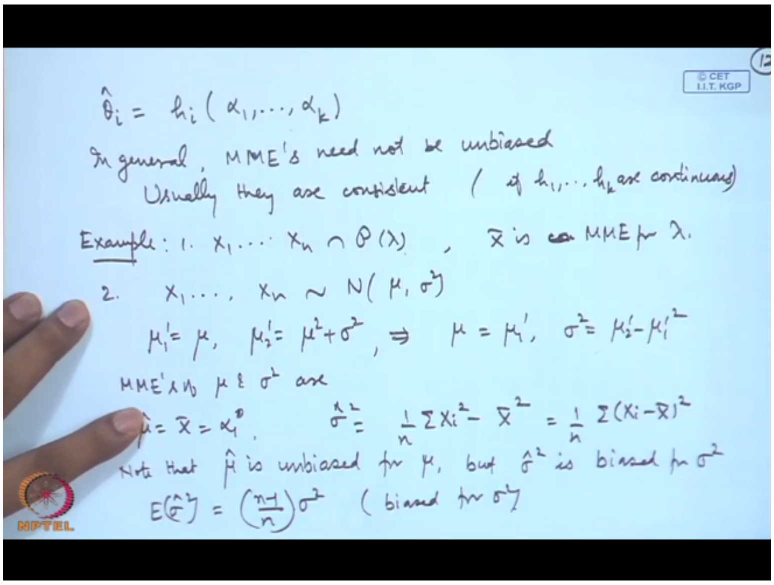


So, first of such methods is method of moments. This was introduced by Karl Pearson, one of the founders of the subject of statistics. So, if we are considering that X1, X2, Xn is a random sample from a population with distribution say Fx theta, I am putting it in the vector form. In general I am assuming it is a k parameter distribution for k >= 1.

Suppose we want to estimate theta 1, theta 2, theta k so, let us define sample moments that is alpha k that = 1/n sigma Xi to the power ki = 1 to n, for k = 1, 2 and so on. Let me change it, I put alpha m here because k is used here. Consider population moments so, mu prime that = expectation of say X1 to the power m for m, for m = 1, 2 and so on. Now, naturally this mu m prime, this will be some function of the parameter. So, let me call it this function as gm theta.

So, for m = 1,2. So, we have k equations that is we write mu 1 prime = g1 theta and so on mu k prime = g k theta, let me call this system 1. Suppose the solution of the system 1 is theta 1 = say h1 of mu 1 prime and so on mu k prime and so on theta k = say hk of mu 1 prime and so on mu k prime. In method of moments, we plug in for mu 1, mu 2, mu k prime alpha 1, alpha 2, alpha k.

(Refer Slide Time: 41:35)



In method of moments, estimators of theta 1, theta 2, theta k are obtained as theta I head = hi of alpha 1, alpha 2, alpha k. so you can say that the basic method is that they estimate the population moment by the corresponding sample moment. Of course, when we write this equations, this must exist, that is this must exist, if they do not exist then you cannot write the equation here. So this is the basic method of moments here.

In general, method of moments estimators need not be unbiased that means sometimes they may be biased and sometimes they may be unbiased. Usually, they are consistent. Now, in fact you can write the conditions, if this functions h1, h2, hk are continuous functions, if they are continuous, then we have already done the weak law of large numbers. So, from there this alpha m will be actually consistent for mu m prime.

If alpha m is consistent for mu m prime and hi are continuous functions, then this theta heads will be consistent for h i's. so, you can consider here this is following say Poisson lambda, then X bar is consistent and this is an MMV, method of moments estimator for lambda. If I consider say X1, X2, Xn following normal mu sigma square then, what are the moments here? Mu 1 prime = mu, mu 2 prime = mu square + sigma square. So, if we solve the equation, you get mu = mu 1 prime, and sigma square = mu 2 prime - mu 1 prime square.

So if I substitute here, so method of moments estimators of mu and sigma square, they will be mu head = X bar that is alpha 1 prime, alpha 1 and sigma head square, that will be = 1/n sigma Xi square - X bar square, that is 1/n sigma Xi - X bar square. Note that, this is not unbiased, note that mu head is unbiased for mu, but, sigma head square is biased for sigma square. Because we have seen actually that 1/n-1 sigma Xi- X bar whole square is unbiased for sigma square.

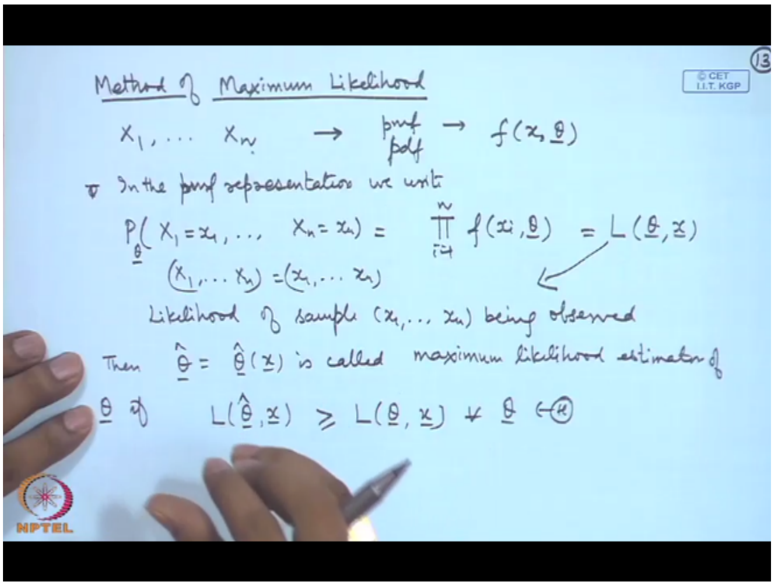
So, if I consider expectation of sigma head square, then that will be = n-1/n sigma square. So this is biased for sigma square. So this is a simple and heuristic method for obtaining the estimators for parameters in any given problem. Now there may be some times some sort of discrepancies for example, here if I am writing 2 parameters, then I am writing 2 equations here. If I have 1 parameter I write 1 equations.

Sometimes, it may happen that due to peculiarity of the distribution, that the required number of equations may be more. For example, if I consider uniform distribution on the interval say -theta to +theta, then the mean is 0, then the first moment is not useful. So, you can consider the second moment that will be theta square/3 and then you can use second sample moment to estimate theta. Another thing that was observed in the method of moments estimator is that, we have to actually solve the equations.

In examples that I constructed here, it is simple, but sometimes you may end up with some very complicated functions. For example, if I consider gamma distribution, or I consider 2 parameter uniform distribution or if I consider beta distribution, where the mean is somewhat complicated function of the parameter. In that case, the solution of the equations will give rise to some complicated functions.

So, certainly unbiasedness will be ruled out, not only that, sometimes continuity of the function may also be in question. A more practical and also you can say theoretically sound procedure was proposed in 1925 by RA Fisher, which is known as the method of maximum likelihood.

(Refer Slide Time: 47:11)



So, in the method of moments, we are making use of the moment structure of the distribution where as in the maximum likelyhood estimation, we make use of the probability structure or the density structure of the distribution. So roughly speaking, let me give the interpretation here, suppose X1, X2, Xn is a random sample from a distribution, either pmf or say pdf of course, you may have somewhat different situation in which you may have a mixture also, that means partly pmf and partly pdf but, for the time being, let me write in a simpler form.

So suppose, it is written as Fx theta okay. So, let me consider the pmf representation, in the pmf representation, we write probability of X1 = say x1 and so on, Xn = xn that will = product of f xi theta, i = 1 to n. now, let me put this in a different form. Here what we are saying? If theta is the 2 parameter value, the probability that capital X1 = x1, Xn = xn is given by this expression. Now, depending upon different values of theta, this value will change.

So, if I am considering that means a sample this has been observed, we can actually consider it as X1, X2, Xn = x1, x2, xn. That means, what is the probability of this sample being observed? Now, we can call it likelihood of sample x1, x2, xn being observed. So I give a new name and I call it L theta, x. This is called the likelihood function. That value of theta, we consider as that means we maximize this with respect to theta.

Then that value of theta, theta head = say theta head x is called maximum likelihood estimator of theta, if L theta head x >= L theta x for all theta. That means, we are considering maximization of the probability of observing or likelihood of observing that particular sample. We can consider some typical example, suppose I take say Poisson lambda and I specify say lambda = either 1 or lambda = 2, that means 2 values are possible.

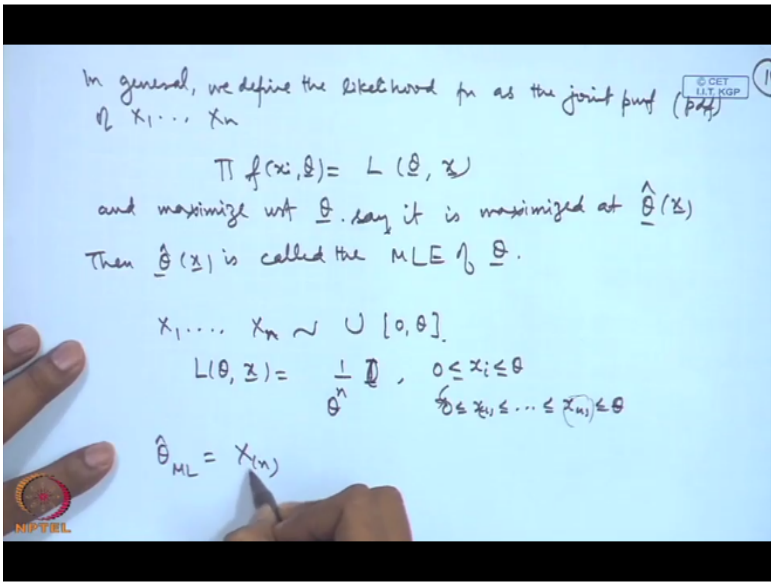
2 values in the parameter space okay and we observe say X = 2 for example or let us take X = 1. If I observe x = 1, let us write down this probability of X = 1 that = e to the power - lambda, lambda to the power x that is 1, so this is simply divided by X factorial. Now, if lambda = 1, then this is e to the power -1. If I observe lambda = 2 then this = 2 e to the power -2. So, we look at the comparison of this values, which value is larger, that is 1/e or 2/e square, so we compare let us just write down so I multiply by e square, so this is e square < 2e. Or if I consider e < 2.

So that means this is actually larger. We are getting e > 2, which is true. So, this number is larger, that means likelihood of observing X=1 is more when lambda = 1, so we say lambda head = 1 is the maximum likelihood estimate. Since it is observed already, so we call it estimate of lambda. So look at this, I am telling here that 2 values lambda = 1 and lambda = 2 are allowed here. We do not know which one is the correct value. Now we observe the sample in this particular case, one observation I take, and it = 1.

Now, I calculate the probability of this X = 1 under this lambda so I am getting e to the power - lambda lambda. I look at under both the conditions, for lambda = 1, this = e to the power -1 for lambda = 2 this is 2 e to the power -2. Now, I compare these 2 and I just write a simple inequality 1/e > 2/e square, which is equivalent to e > 2, which is true. Therefore, we conclude that this probability is higher, therefore, lambda = 1 will be called the maximum likelihood estimate of lambda here.

So, you can say this is the fundamental principle of the maximum likelihood estimation that we consider the likelihood function. We look at that value of the parameter, which is actually maximizing. That means we are basically maximizing the likelihood function which is actually nothing but, I have given the probability mass function interpretation. So, now we generalize this in place of this one suppose I consider pdf, then we maximize that.

(Refer Slide Time: 54:23)

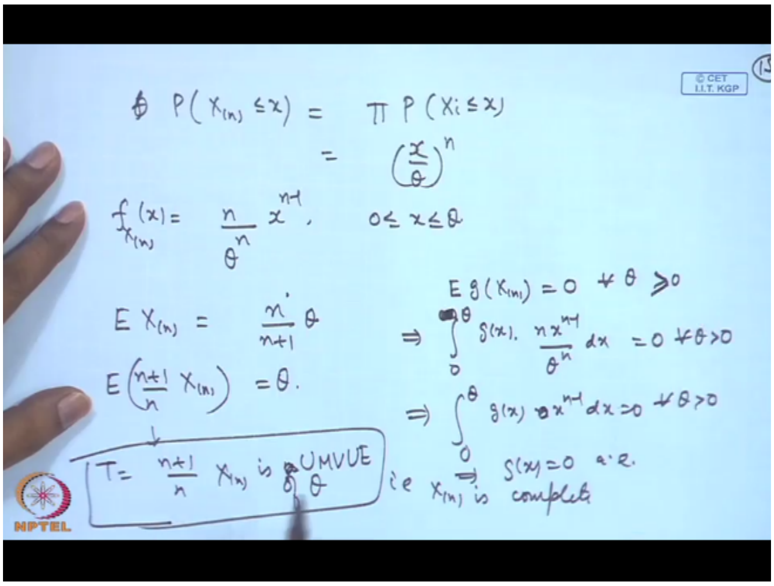


So in general, we define the likelihood function as the joint pmf or pdf of X1, X2, Xn. So that is product of Fxi theta and we call it L theta X and maximize with respect to theta. So, say it is maximized at theta head x. Then, theta head x is called the maximum likelihood estimator of theta. So, I will be showing through various example of this, let me consider a simple application, which we have been considering earlier for the discussion of consistency and sufficiency at sector.

So, now let us consider this for this purpose. Now, you can see that the likelihood function will be 1/theta to the power n indicator function of, so let me just write this and this we can actually write as 0 < = x1 <= x2 < = xn. Now, to maximize this we see the maximum value will be attained when theta is minimum but the minimum value of theta will be xn. So theta head mL = Xn.

In fact, we already proved that this is sufficient, we can also show it is complete. This was already shown to be consistent, it was sufficient. We can also show it to be complete. We can also show that Xn is complete. Just briefly I will obtain actually the UMVUE based on this to complete this to complete this discussion.

(Refer Slide Time: 57:03)



See, we had obtained the probability of Xn <= x = product of probabilities Xi <= x, that = x/theta to the power n. so, the density function of xn is actually = n/theta to the power n x to the power n-1. If I consider the expectation of this, what I get here? This = n/n+1 theta. So, that means expectation of n+1/n Xn = theta. Also, let us consider say expectation of g Xn = 0 for all theta. Then this will imply integral gx n x to the power n-1 theta to the power n dx 0 to theta that = 0 for all theta positive.

Now, you are saying that integral over intervals of the form 0 to theta for all such intervals. Then, you can consider say (()) (58:30) result by differentiation etc. You can prove that actually that gx = 0 almost everywhere. That means Xn is actually complete. Now, Xn is complete sufficient and this is an unbiased estimator based on Xn. So, T = n+1/n Xn is UMVUE of theta.

In tomorrow's class, I will discuss few more examples of maximum likelihood estimation and the method of moments and what is the comparison between them and then we will move over to the concept of interval estimation also. So, we stop today's lecture at this point.