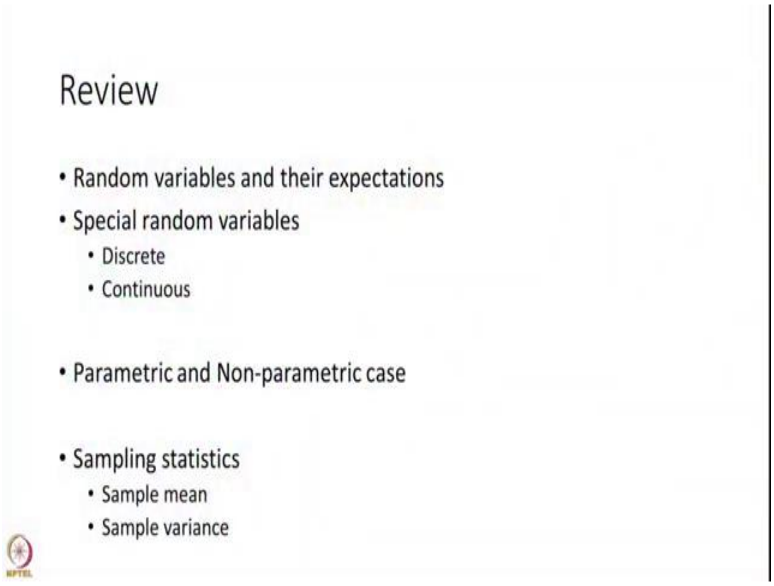
Dealing with Materials Data
Professor M P Gururanjan
Professor Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture 50
Parameter Estimation 1

Hello and welcome to the course on Dealing with Materials Data. We have come a long way in understanding the statistics while we want to understand the materials data and its component both in the theoretical part that I am trying to cover and what has been covered in the hour sessions by professor Gururanjan. In next few sessions we are going to consider the case of parametric estimation.

(Refer Slide Time: 00:56)



Let us recall what we have done so far. We introduced random variables and their expectations. Then we also talked about certain discrete and continuous special random variables. Then in the last few sessions we discussed what is a parametric and non-parametric case, that is if you now we are in the realm of the population. There is a population for which we are trying to make an estimate or to judge or to understand what that population is like. And this we can do through a small sample that we draw from the population.

This sample we expected it is a fully random sample and it gives us on a small number all information that we may require from the population, so that we can estimate or we can judge what the population is like. So, if we are aware that the population is coming from a specific distribution function, then we need to estimate the parameters of the distribution. We need to talk about the parameters of, unknown parameters of the distribution. But, suppose the form of the distribution itself is not known then it is called a non-parametric case.

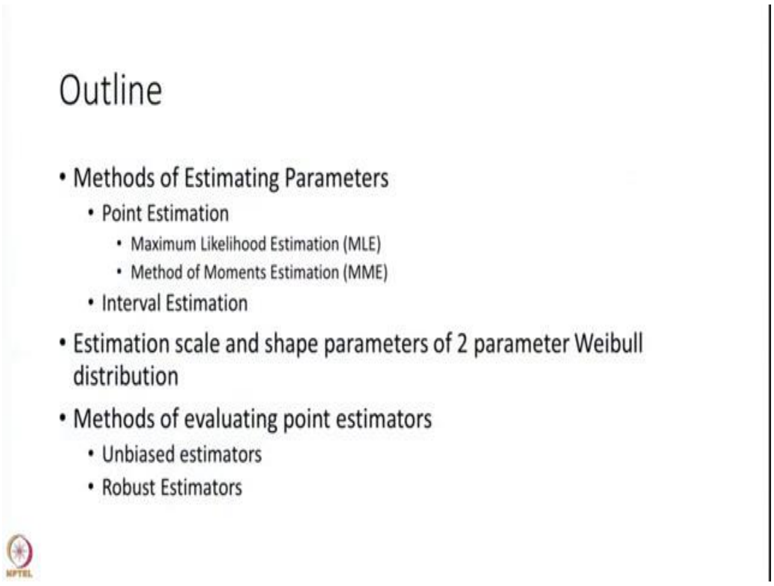
As we mentioned in the previous sessions, in this particular course we are going to concentrate on the parametric session that is all the cases which are referring to parametric cases. Then we also studied under parametric cases the sampling statistics in which we talked about sample mean and sample variance. And we showed that the sample mean expected value is actually mean of the population and we also showed that sample variance is its expected value is also sample, sorry, population variance for the population.

Now, we want to get into what is call estimation, the parameters? See in sampling statistics what we did is that, every distribution or every population has its mean value and its standard deviation. Which is it is a tendency of, central tendency where the data will be generally located as we said in the descriptive statistics it talks something like, the centre of gravity of the data.

And the dispersion actually says how the data is spreaded. But, suppose I say that a particular data is coming from a log normal distribution or I say that, particular data is coming from a Weibull distribution. I also specify that it comes from the two parameter log normal distribution. Then we would like to know, what these parameters are? Because once you define these parameters the whole distribution is known, so in a way you know the whole population.

Now, when you want to find the parameter you need to estimate it again from the small sample that we have drawn from the population. And therefore in this few sessions we are going to talk about parametric estimation. We will be assuming that a population has a certain distribution form. We say that, population distribution has a certain form and it has certain unknown parameters and our endeavour in the coming few sessions is to estimate these parameters in various ways. So, we would like to consider here, the case of point estimator.

(Refer Slide Time: 05:10)

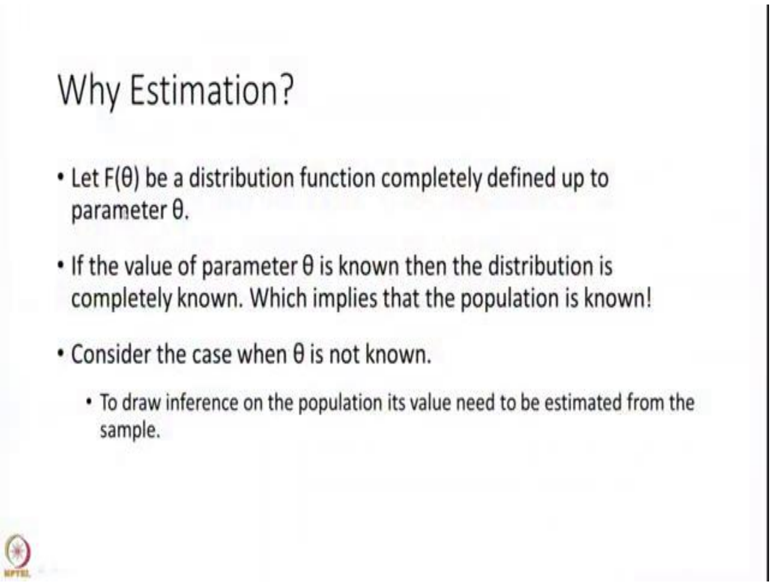


There are other methods of estimation for example, there is called an interval estimation also that will also follow. But in this particular session, today session we are going to consider the methods of estimating parameters through point estimation. In which we will cover maximum likelihood estimator and method of moments estimation.

Then we will also talk about interval estimation in the future sessions then we will have estimation of scale and shape parameters of two parameter Weibull distribution. We are going to give examples of Bernoulli as well as normal but they are very straight forward. So, we would like to give an example of two parameters, Weibull distribution which is bit involved and we will also talk about methods of evaluating point estimators that is unbiased estimator and robust estimator.

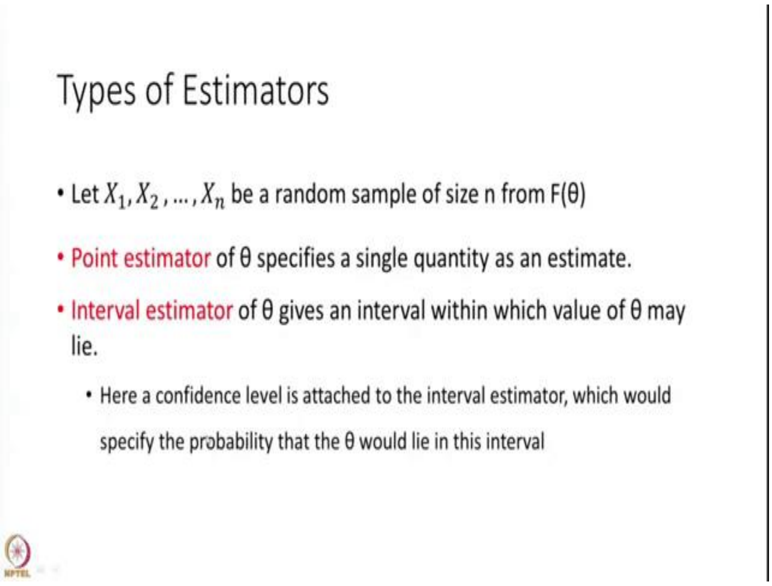
Though I must mention that this interval estimation we will cover in subsequent sessions and similarly the methods of evaluating point estimators we will also cover in the subsequent session. In this particular session we planned to cover maximum likelihood estimation for parameters, so let us start. We already have talked about it let just repeat it.

(Refer Slide Time: 06:45)



Why estimation? as I said F theta be a distribution function completely defining the population. Only the parameter theta is unknown. So, if we know the parameter theta then we know the whole population and therefore we want to consider the case of theta not known. And we would like to estimate theta in order to draw inference on the population.

(Refer Slide Time: 07:13)

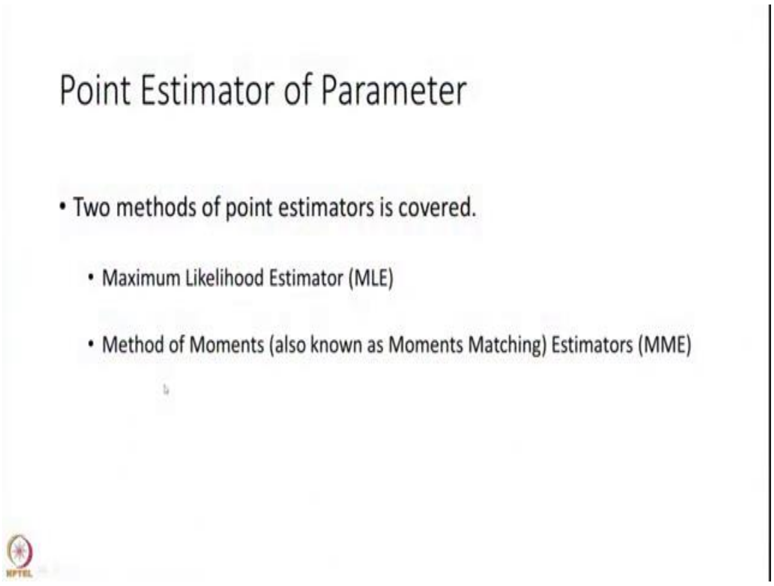


There are two types of estimators, one is called a point estimator in which it finds a single quantity as an estimate for unknown parameter theta. While, the interval estimator gives an interval within which a value theta may lie with a probability attached to it. This probability which is attached to it is either called a confidence level or it is called a significance level I am sorry it comes in hypothesis testing.

But actually here it is called a confidence level of the interval estimator. So, for example, we can say that a theta will lie in the interval between say minus 3 and plus 3 with a probability of 95 percent. Or I may say that a theta will lie between minus 4 and 4 with the probability of say 68 percent. I mean you can give a variety of probabilities.

But this probability is to be derived. So, you can have an interval estimator with the different level of confidence. As I said in this session we are going to talk about maximum likelihood estimator but in completeness we are going to talk about 2 methods.

(Refer Slide Time: 08:38)



Maximum likelihood estimator along with method of moments or moments matching estimator which is also known as MME. I have particularly chosen this because it has become very common with many software's and also with respect to the R-programming.

So, it is better to know its merits and its demerits with respect to methods of moments. Of course, when we study, when we talk about evaluating the point estimator we will go in further details, so let us start. We start this session we are going to talk about maximum likelihood estimator, so let us begin.

(Refer Slide Time: 09:26) and 12.00

- Let $X_1,\ X_2, \dots, X_n$ be a Random sample of size n from F(θ)
- Thus $X_1,\ X_2, \dots, X_n$ are independently and identically distributed as F(θ).
- Let $f_{X_i}(x_i; \theta)$ be pdf of X_i , $i = 1,\ 2,\ \cdots .\ n$

- Hence, the joint pdf of $X_1,\ X_2, \dots, X_n$ can be given by

$$f(x_1, x_2, \cdots, x_n|\theta) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta)$$

Here $x_1, x_2, \cdots, x_n$ are the realisation of random sample $X_1,\ X_2, \dots, X_n$.

You please note the difference between the capital Xi and the small xi and please recall.

Capital Xi are the random variables, small xi is are the realisation of respective capital Xi. So, it is a small x1, x2, xn are the realisation of a random sample of X1, X2, X3, Xn all X capital here. It is important to realise that joint PDF that is this function has x1, x2, x3, xn known. Only unknown is theta and as far as the population is concerned we can recognize the population only through its realisation of the random sample.

And therefore whatever information that you wish to have with respect to theta is all contained in this joint distribution function. And therefore this joint distribution function is given a special name and it is called a likelihood function of parameter theta. Because it contains the likelihood what the parameter theta can take a value because again let us repeat.

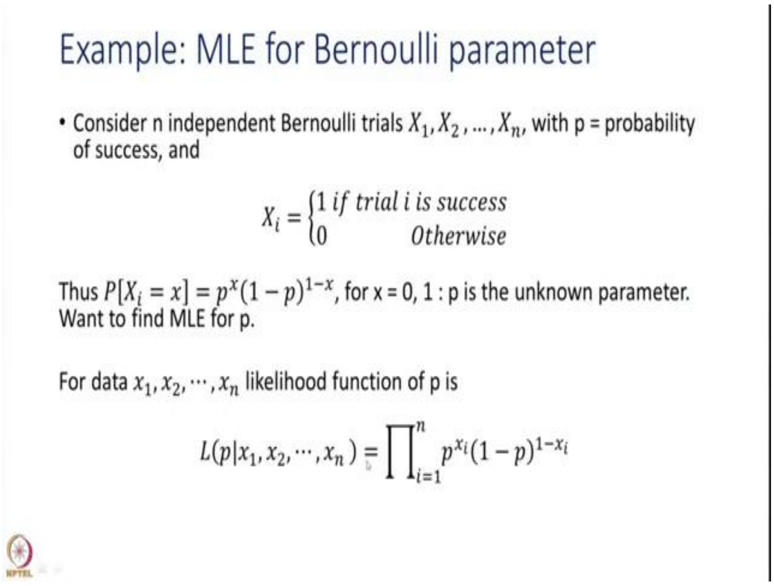
What I am trying to say is that population is unknown only thing we know is a random sample that we have drawn and the realisation that we have made. So, it is an abstract phenomenon which we say that, random sample of size n is drawn in reality what we see is the actual realisation. The small x1, x2, x3, xn that we observed this is our data, small x1, x2, x3, xn is our data.

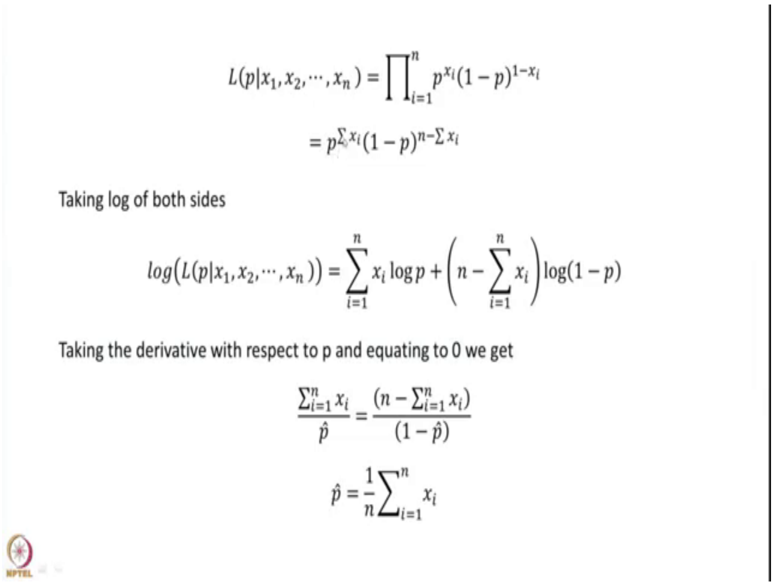
So, data has all information possible with respect to the unknown parameter theta and therefore this joint density function is called a likelihood function of theta. Now, we would like to find out theta and one of the philosophy is that, take the value of theta which maximizes this likelihood. Take the value of theta which maximizes this likelihood this sounds very logical argument.

So, here we wish to find out the maximum likelihood function and at which point it will become maximum that theta we would like to call a maximum likelihood estimator. So, we have that so we would like to maximize the likelihood function of theta and that estimator we are going to call a maximum likelihood estimator. Now, this is a purely little math that if, likelihood function is maximized then log likelihood function is also maximized.

And we will see in future this log likelihood function makes life easy in order to find a maximum mathematically. Generally, the maximum likelihood estimator of theta is denoted by, what is called theta hat as it is shown here. Let us see how do we actually do it. So I am going to consider 3 examples in this session.

(Refer Slide Time: 14:23)





The first example is that of a Bernoulli parameter. So, consider n independent Bernoulli trials X1, X2, X3, Xn with a p as a probability of success and p is unknown okay.

$$X_i = \begin{cases} 1 \; if \; trial \; i \; is \; success \\ 0 \qquad\qquad Otherwise \end{cases}$$

So, we can say that the probability mass function we do not have probability density function, this is a discrete case and therefore we have a probability mass function.

So, the probability mass function that xi takes a value x, where x is either 0 or 1,

$$P[X_i = x] = p^x(1-p)^{1-x}$$, for x = 0, 1 : p is the unknown parameter

And now we want to find a maximum likelihood estimator of unknown parameter p. So, we must find first the likelihood function of p which is nothing but the joint density of x1, x2, x3, xn and that we can find out by simply multiplying the densities. Because they are all independent and identically distributed so we get the likelihood function of p as a

$$L(p|x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

Take log on both sides, remember we said that maximizing likelihood or maximizing log likelihood it is going to give us the same answer with respect to the parameter theta.

$$log\big(L(p|x_1, x_2, \cdots, x_n)\big) = \sum_{i=1}^{n} x_i \log p + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-p)$$

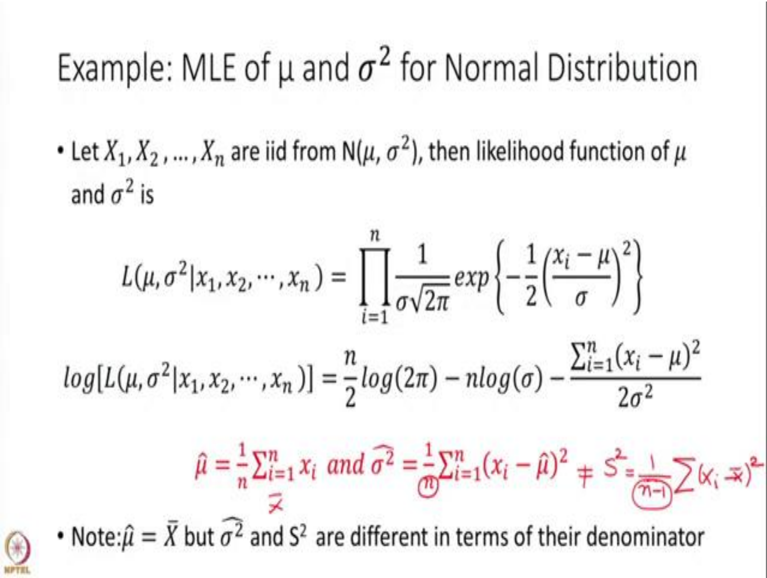
So, here we take a log function of the likelihood and you see how it nicely simplifies to take a derivative of this is a very tough, this is very simple. So, we get the

$$\frac{\sum_{i=1}^{n} x_i}{\hat{p}} = \frac{(n - \sum_{i=1}^{n} x_i)}{(1-\hat{p})}$$

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

It is MLE, it is maximum likelihood estimator and that value p hat is 1 over n summation of xi, in other word it is x bar, it is a mean value of the total realization of sample of size n.

(Refer Slide Time: 17:41)



Okay let us consider another example, this is a continuous distribution. We take a normal distribution with two unknown parameters, mean of the distribution and variance of the distribution Mu and sigma square respectively.

So, here again the likelihood function of

Let $X_1, X_2, \ldots, X_n$ are iid from N(μ, σ^2), then likelihood function of μ and σ^2 is

$$L(\mu, \sigma^2|x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right\}$$

$$log[L(\mu, \sigma^2|x_1, x_2, \cdots, x_n)] = \frac{n}{2}log(2\pi) - nlog(\sigma) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

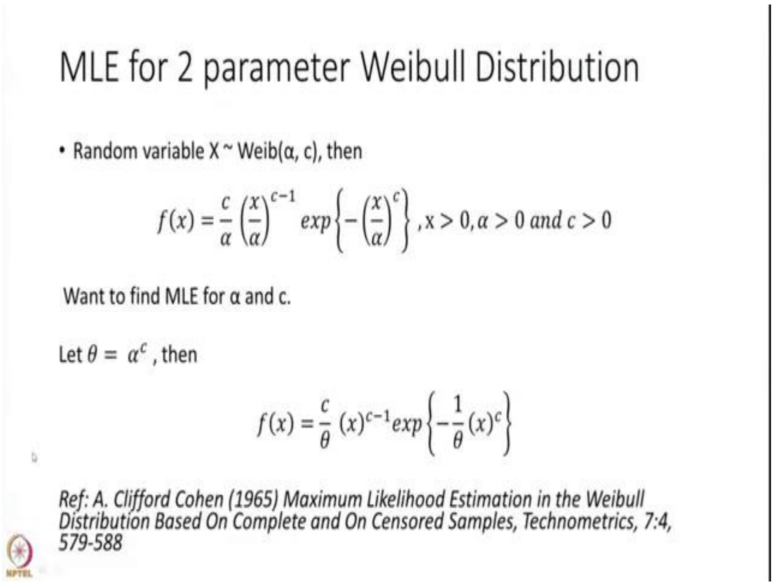
$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \ and \ \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

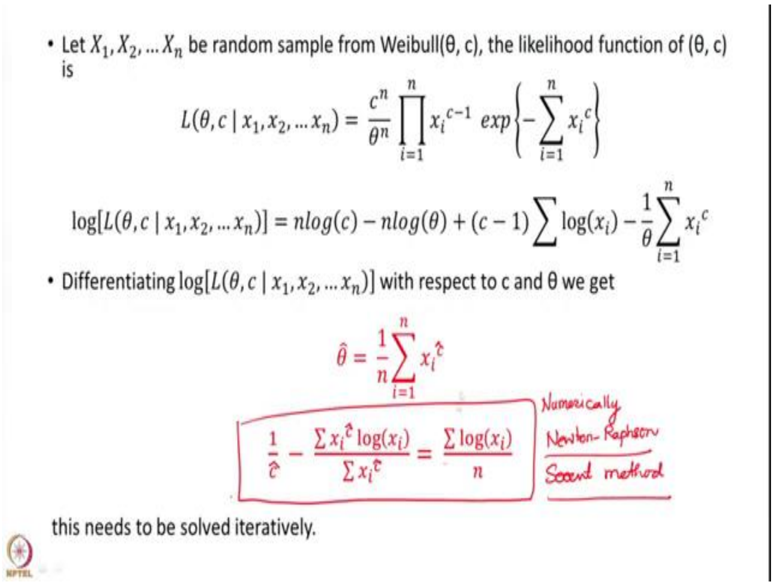
- Note: $\hat{\mu} = \bar{X}$ but $\widehat{\sigma^2}$ and S^2 are different in terms of their denominator

Please remember that this is not capital S square, please recall this is very important to note that this is not equal to capital S square. Remember that capital S square is 1 over n minus 1 summation xi minus x bar whole square. Now this is x bar, so Mu hat and x bar are same but they are different in the denominator.

So please note this that maximum likelihood estimator is not the sample mean it differs in the denominator, sample variance has denominator as n minus 1. Maximum likelihood estimator of population variance has a denominator of n.

(Refer Slide Time: 20:43)





Let us take the next example, now you see so far I have given you examples in very beautifully you get a closed form solution for maximum likelihood estimator. Well these are the lucky cases this does not happen all the time. Here I would like to show you the case of two parameter Weibull distribution in which we will at the end will get the two equations or one equation that we need to solve iteratively to find the value of maximum likelihood estimator. Let us go through it.

Please note that I have given a reference here, it is good to read this is freely available on internet for downloading. You can download this particular PDF and go through it, you will get a test as to how people tend to derive this kind of estimator. Anyway, briefly I am going to discuss it here. So, let the random variable X be a Weibull distribution with two parameters the scale and shapes.

X ~ Weib(α, c), then

Scale is α, and the shape is c then it is denoted as

$$f(x) = \frac{c}{\alpha} \left(\frac{x}{\alpha}\right)^{c-1} exp\left\{-\left(\frac{x}{\alpha}\right)^{c}\right\}, \text{ x} > 0,\ \alpha > 0 \text{ and } c > 0$$

Now, we would like to find MLE of alpha and c. Here it is easier to make one transformation, so that the calculations become easier. So, if you take another parameter theta which is a function of alpha and c, Let $\theta = \alpha^c$, then

$$f(x) = \frac{c}{\theta} (x)^{c-1} exp\left\{-\frac{1}{\theta}(x)^{c}\right\}$$

Then this density function very beautifully simplifies to this and this is easier to work with to find the maximum likelihood estimator. And once you find it you can always revert back and find out the actual estimator of alpha and c. So, here c is as it is, so only alpha is replaced by theta.

So, Let $X_1, X_2, \ldots X_n$ be random sample from Weibull(θ, c), the likelihood function of (θ, c) is

$$L(\theta, c \mid x_1, x_2, \ldots x_n) = \frac{c^n}{\theta^n} \prod_{i=1}^{n} {x_i}^{c-1}\ exp\left\{-\sum_{i=1}^{n} {x_i}^{c}\right\}$$

$$\log[L(\theta, c \mid x_1, x_2, \ldots x_n)] = nlog(c) - nlog(\theta) + (c-1)\sum \log(x_i) - \frac{1}{\theta}\sum_{i=1}^{n} {x_i}^{c}$$

Now, if you Differentiating $\log[L(\theta, c \mid x_1, x_2, \ldots x_n)]$ with respect to c and θ we get

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} {x_i}^{c}$$

$$\frac{1}{c} - \frac{\sum {x_i}^{c} \log(x_i)}{\sum {x_i}^{c}} = \frac{\sum \log(x_i)}{n}$$

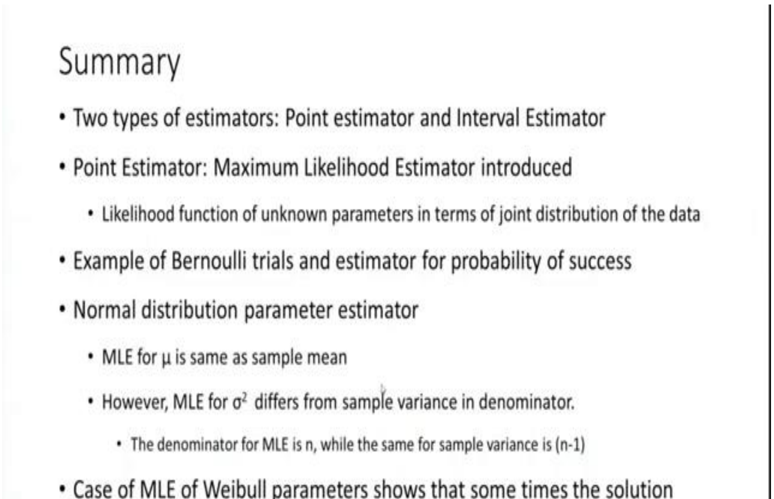
this needs to be solved iteratively.

And therefore you please note that this equation based on the data has to be solved iteratively. You can use Newton-Raphson method, you can do secant method and I mean you can choose any method. And solve this particular equation to find its zero that is going to be your maximum likelihood estimator of c which is c hat and then you replace in the equation of theta, the c hat here.

Actually in this equation I should have everywhere written c hat, so let me make that correction. So, here it should be c hat, this is c hat, this is c hat, this is c hat and therefore if you solve this equation numerically. You can use Newton-Raphson's method or you can use the Secant method.

Newton-Raphson works out very well but you can use either method and find a zero of this function. You have to take this j on the other side and you find to solve this iteratively. Once you find the c hat you plugged that c hat into the expression of theta hat and what you get in as an estimator, maximum likelihood estimator of theta hat.

(Refer Slide Time: 26:39)



So, with this 2-3 examples let us summarized, what we learnt about maximum likelihood estimator and what we talked in this session. There are two types of estimators, we say there is a point estimator and there is a interval estimator. In the point estimator we introduced a maximum likelihood estimator, where the likelihood function of unknown parameter in terms of joint distribution of the data is taken and it is maximized.

And the place where it takes the maximum value, the parametric value which makes it maximum is the called the maximum likelihood estimator. The examples of Bernoulli trials and we estimate we found the estimator of probability of success. Please note that, that estimator is also the sampled mean which is same as the, it is also the sample mean is the estimator, which expected value of sample mean turns out to be the probability of success.

Remember that MLE may not have all the time this property. Then it with respect to normal distribution we talked about the maximum likelihood estimator Mu which is same as sample mean. However, we found that the maximum likelihood estimator of variance differs from the sample variance in the denominator.

We then showed that very time finding maximum likelihood estimator is not an easy task at times you end up with a set of equations or a single equation which you need to solve iteratively. Now in the next session we will move on to discuss the method of moments estimator. Thank you.