Statistical Inference

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

Lecture No. # 05
Finding Estimators – II

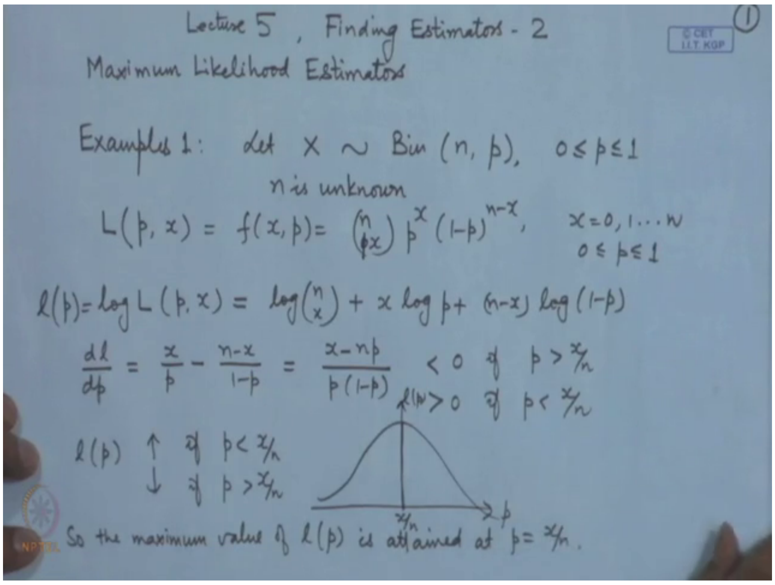
So, friends in my earlier lecture I have told the some criteria for judging the goodness of estimators. For example, unbiaseness is 1 criteria, consistency is 1 criteria. That means, if an estimator is unbiased it is in general preferable to an estimator which is not unbiased. Similarly, an estimator which is consistent is preferable to an estimator which is inconsistent. So, there are many other criteria's which we will be discussing in the further discussion, then we dwelt upon how to find out the new estimators or how to propose the estimators.

We have mainly discussed 2 methods. one is the method of moments, the method of moments concentrated on equating the sample moments with the population moments and thereby obtaining the estimates of the parameters. The method is quite simple, and in general it provides good estimators, but then there are certain criteria which it does not satisfy; for example, in many cases we saw that the method of moment estimators were not unbiased although in many cases they were consistent.

Another popular method which was introduced in 1920s by R A Fisher is the well known method of maximum likelihood estimation. Here the idea is that whenever a sample is observed we look at the probability of that sample being observed, and what is the parameter value for which these probabilities or likelihood is maximized. So, we define what is known as a likelihood function. In the previous class I have given an example illustrating that and the general form of a likelihood function.

Today, we start with various applications that is in many probability models, what are the method of maximum likelihood estimators. So, we call in general M L E that is the maximum likelihood estimators.

(Refer Slide Time: 02:38)



Lecture 5 , Finding Estimators - 2

Maximum Likelihood Estimators

Examples 1: Let $X \sim Bin(n, p)$, $0 \le p \le 1$

n is unknown

$L(p, x) = f(x, p) = \binom{n}{px} p^{x} (1-p)^{n-x}$, $x = 0, 1 \dots n$, $0 \le p \le 1$

$l(p) = \log L(p, x) = \log\binom{n}{x} + x \log p + (n-x) \log(1-p)$

$\frac{dl}{dp} = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}$ < 0 if $p > x/n$

> 0 if $p < x/n$

$l(p)$ ↑ if $p < x/n$

↓ if $p > x/n$

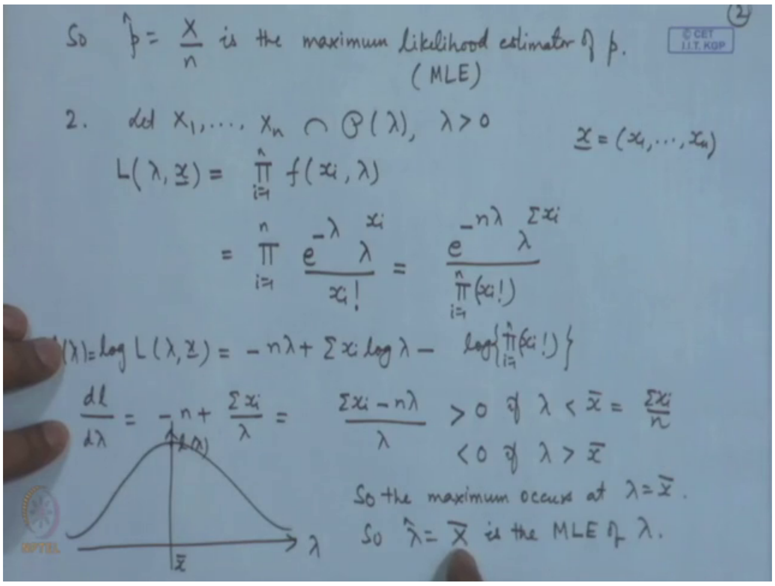
So the maximum value of $l(p)$ is attained at $p = x/n$.

So, let me discuss some applications which are applicable to popular distributional models, maximum likelihood estimators. So, let me start with some familiar examples. Let x follow a binomial n p distribution, now if we say that n is known then p is a parameter, let us consider the likelihood function. So, the likelihood function is written as a function of the parameter which is actually the density function and in this particular case it is n c x p to the power X 1 minus p to the power n minus x. Here x takes value 0 1 to n and p is a number between 0 to 1. Our objective is to maximize this likelihood function with respect to p.

A usual practice is to take the log of likelihood function which we call log likelihood and we use a another notation small l for this. So, small L p is equal to log of likelihood that is equal to log of n c x plus x log p plus n minus x log of 1 minus p. Now, if you look at this function we can apply the usual method of the calculus for finding out the maximum with respect to p. So, we can consider for example, derivative of this with respect to p. So, this vanishes you get x by p minus n minus x by 1 minus p which we can write as x minus n p divided by p into 1 minus p.

Now, if you notice this thing this is less than 0 if p is greater than x by n and it is greater than 0 if p is less than x by n. So, we can see from here that L p this will be increasing if p is less than x by n and it is decreasing if p is greater than x by n therefore, the shape of the likelihood function is something like this if you are plotting L p then it is attaining the maximum at the point x by n. So, the maximum value of L p is attained at p is equal to x by n.

(Refer Slide Time: 05:59)



So $\hat{p} = \frac{X}{n}$ is the maximum likelihood estimator of p. (MLE)

2. Let $X_1, \ldots, X_n \sim \mathcal{P}(\lambda)$, $\lambda > 0$ $\qquad \underline{x} = (x_1, \ldots, x_n)$

$$L(\lambda, \underline{x}) = \prod_{i=1}^{n} f(x_i, \lambda)$$

$$= \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\Sigma x_i}}{\prod_{i=1}^{n} (x_i!)}$$

$$l(\lambda) = \log L(\lambda, \underline{x}) = -n\lambda + \Sigma x_i \log \lambda - \log\left\{\prod_{i=1}^{n} (x_i!)\right\}$$

$$\frac{dl}{d\lambda} = -n + \frac{\Sigma x_i}{\lambda} = \frac{\Sigma x_i - n\lambda}{\lambda} \begin{array}{l} > 0 \text{ if } \lambda < \bar{x} = \frac{\Sigma x_i}{n} \\ < 0 \text{ if } \lambda > \bar{x} \end{array}$$

So the maximum occurs at $\lambda = \bar{x}$.

So $\hat{\lambda} = \bar{X}$ is the MLE of λ.

So, we say that p hat is equal to x by n is the maximum likelihood estimator of of p. Now, you notice this thing x by n is actually the sample proportion. So, we are getting that the sample proportion is the maximum likelihood estimator of the population proportion p. So, this is the natural estimator and from the method of maximum likelihood estimator we are actually getting that as an estimator. Let me take some more examples for the popular distributional models, suppose I have a random sample X 1, X 2, X n from a Poisson distribution with parameter say lambda.

Our interest is to find out the maximum likelihood estimator for the parameter lambda as you recall the parameter lambda in the Poisson distribution represents the average arrival rate or the mean of the process in which the Poisson distribution is generated. So, if you write down the likelihood function L lambda and let me use the notation x for the sampled observation X 1, X 2, X n this is nothing, but the joint probability mass function of X 1, X 2, X n written at the points X 1, X 2, X n.

So, this is nothing, but the product i is equal to 1 to n. Now, this is for 1 particular X I, if we write it is e to the power minus lambda, lambda to the power X i divided by X i factorial. So, this can be further simplified e to the power minus n lambda, lambda to the power sigma X i divided by product of X i factorial. Now, as you notice we have to maximize this function with respect to lambda and this function here lambda is occurring in the exponent as well as lambda has an exponent. Therefore, it will be convenient if once again in place of the likelihood function we consider log likelihood function.

So, we take log of this, we call it log likelihood that is equal to minus n lambda plus sigma X i log of lambda minus log of product X i factorial. Once again if you observe this is a non-linear function of lambda. We can apply the usual method of analysis for finding out the maximum with respect to lambda. So, let us consider the simple derivation of this with respect to lambda. So, that is equal to minus n plus sigma X i by lambda that is equal to sigma X i minus n lambda divided by lambda. Easily you can see that it is greater than 0 if lambda is less than X bar where X bar is actually sigma X i by n and it is less than 0, if lambda is greater than X bar.

So, naturally if you plot the behavior of the L function. So, suppose this is my x axis represents lambda and on the y axis, I represent l of lambda then for lambda less than X bar the value is positive of the derivative therefore, the L lambda function will be increasing and for lambda greater than X bar this d l by d lambda is negative therefore, this L lambda will be a decreasing function. Therefore, the maximum occurs at lambda is equal to X bar.

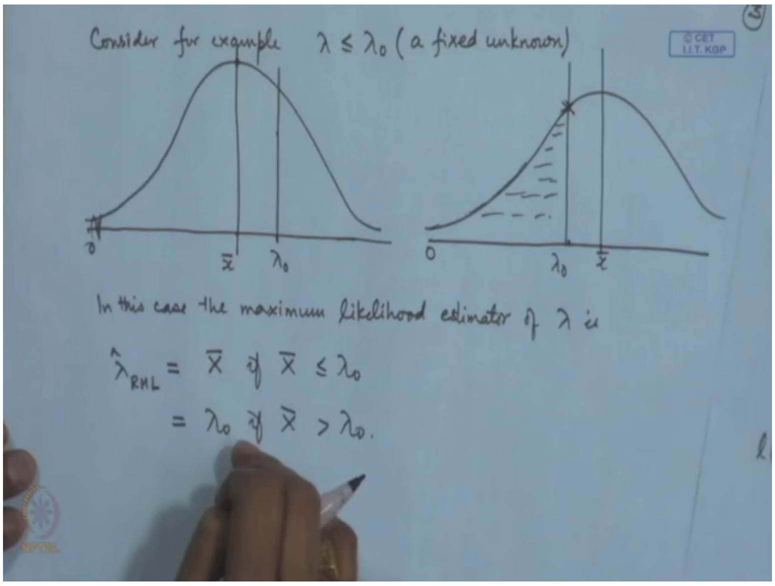
So, the maximum occurs at lambda is equal to X bar. So, we say that lambda hat is equal to X bar is the maximum likelihood estimator of lambda. Once again you observe here this is the sample mean and in this particular case it turns out that the sample mean is the maximum likelihood estimator of lambda. In the method of moments also we would have got the estimator because expectation of X bar would have been equal to because the first moment is lambda and first sample moment is X bar. So, this would have also been the method of moments estimator for lambda in the case of Poisson distribution.

However, in the case of maximum likelihood estimator we have a restriction. Restriction means that whatever be the required parameter space the maximization is over only in that region that thing is not necessarily satisfied suppose we are considering the method of moments, because there we simply equate the sample moments with the population moments. We do not bother about what is the region of the parameter, that means, the region where the parameter can vary.

Similarly, when we apply the concept of unbiasedness or consistency we do not look at the parameter space. In that sense the maximum likelihood estimation is more powerful and all encompassing procedure because it takes into account what is the sampled observations as well as what is the required parameter space where you are actually considering the estimation. In that sense this has more applicability and acceptance for the user point of view.

To give an example in this case I have taken lambda to be greater than 0. That means, the arrival rate is positive which is true in general for a Poisson process, but suppose your physical constraints restrict the parameter space for example, it could be a service queue where if the number of required persons exceeds a certain number then the service, that means, then no more persons are allowed then you may have a situation of this nature.

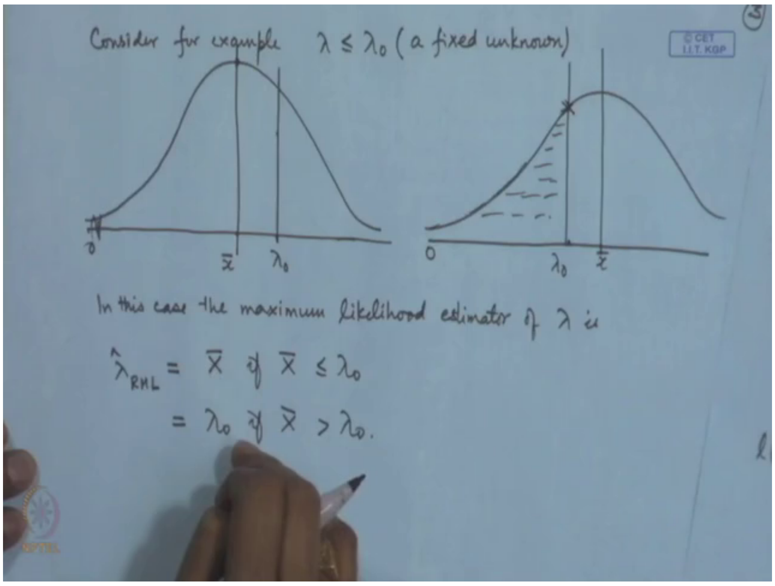
(Refer Slide Time: 13:02)



Consider for example $\lambda \leq \lambda_0$ (a fixed unknown)

In this case the maximum likelihood estimator of λ is

$$\hat{\lambda}_{RML} = \bar{X} \text{ if } \bar{X} \leq \lambda_0$$
$$= \lambda_0 \text{ if } \bar{X} > \lambda_0.$$



Consider for example $\lambda \leq \lambda_0$ (a fixed unknown)

In this case the maximum likelihood estimator of λ is

$$\hat{\lambda}_{RML} = \bar{X} \text{ if } \bar{X} \leq \lambda_0$$
$$= \lambda_0 \text{ if } \bar{X} > \lambda_0.$$

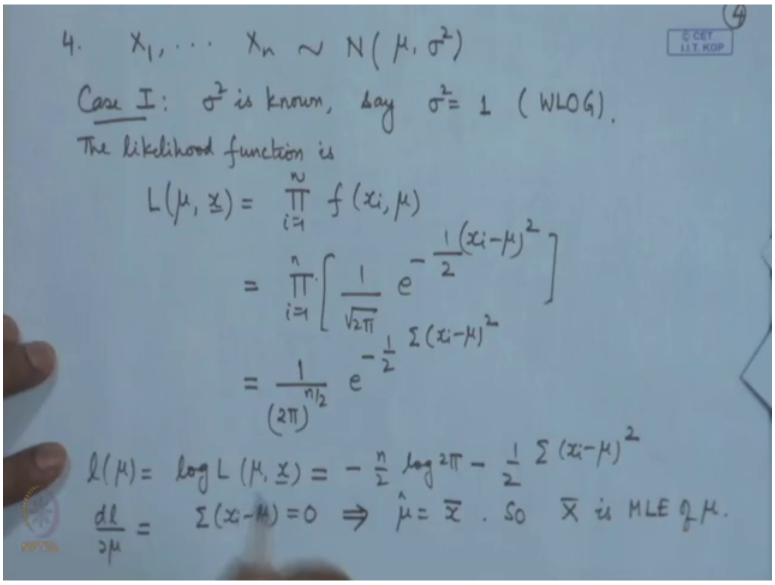
Consider for example, lambda is less than r equal to naught where lambda naught is a fixed unknown. Now, in this case if you see we have here looked at the maximum value lambda is equal to X bar. Now, you may have 2 cases let me give the plot here. So, see this is X bar. Now, there may be 2 cases. It could happen that lambda naught value is here. If lambda naught is here, then the maximum of likelihood function is in the region 0 to suppose this is starting point is your 0.

So, 0 to lambda naught the maximum value is still occurring at lambda naught. It is still occurring at X bar. Whereas, you may have another situation where your lambda naught may be on this side, your X bar is here. Now, if you look at the likelihood function we are concerned only for this portion and therefore, if you see the maximum value that is occurring here that is at lambda naught.

So, we cannot say here that the maximum likelihood estimator is X bar. It is actually lambda naught. So, in this case the maximum likelihood estimator of lambda is let me call it lambda hat R M L restricted M L E. So, this is equal to X bar if X bar is less than or equal to lambda naught, it is equal to lambda naught if X bar is greater than lambda naught. So, you note here that this estimator is certainly different from the method of moments estimator for this problem.

Because the methods of moments estimator does not take care of this fact that lambda is bounded by lambda naught. So, the answer would have been still X bar for the methods of moment estimator. Let me explain the situation with some other examples also.

(Refer Slide Time: 15:45)



4. $X_1, \cdots X_n \sim N(\mu, \sigma^2)$

Case I: σ^2 is known, say $\sigma^2 = 1$ (WLOG).

The likelihood function is

$$L(\mu, \underline{x}) = \prod_{i=1}^{n} f(x_i, \mu)$$

$$= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2} \right]$$

$$= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\Sigma(x_i-\mu)^2}$$

$$l(\mu) = \log L(\mu, \underline{x}) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\Sigma(x_i-\mu)^2$$

$$\frac{dl}{\partial\mu} = \Sigma(x_i-\mu) = 0 \Rightarrow \hat{\mu} = \bar{x}. \text{ So } \bar{x} \text{ is MLE of } \mu.$$

Let us for example, take X 1, X 2, X n following normal mu sigma square estimation. Now, I consider different cases because when we are dealing with the 2 parameter problem then there may be some information regarding 1 parameter or there may be information regarding both the parameters. I will consider all these cases.

Let us take say case 1, say sigma square is known. So, in that case without loss of generality we can take sigma square to be 1 without loss of generality. So, if we write down the likelihood function the likelihood function is L mu x because when sigma square is known only 1 parameter is occurring here. So, it is the joint density function of X 1, X 2, X n at the observed values small X 1, small X 2, small X n that is equal to product i is equal to 1 to n, 1 by root 2 pi e to the power minus 1 by 2 X i minus mu square.

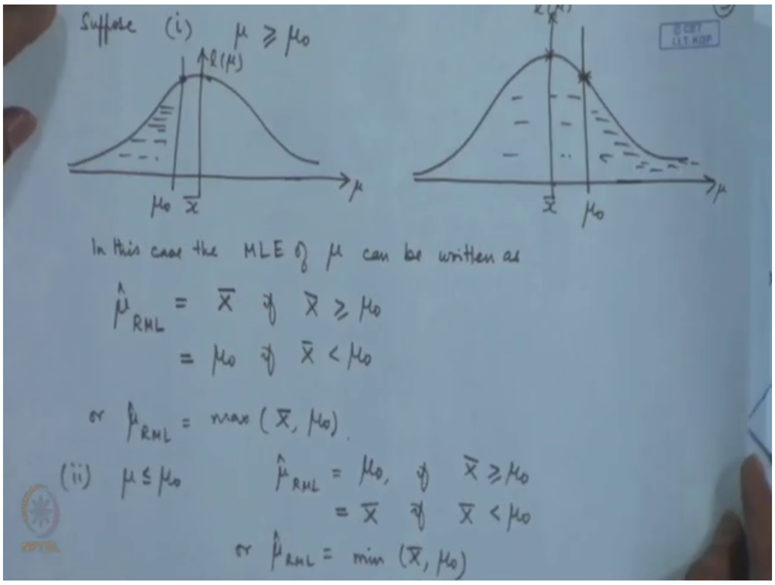
Now, we try to write it in a slightly compact fashion. So, you get 2 pi e to the power n by 2 e to the power minus 1 by 2 sigma X i minus mu square. So, as before you can see here mu is occurring in the exponent therefore, it is beneficial if we consider the log likelihood. So, we consider log likelihood function as minus m by 2 log 2 pi minus half sigma X i minus mu square.

In order to maximize this with respect to mu we consider simple derivative respect to mu which gives us sigma X i minus mu is equal to 0 which are the extremely simple solution mu hat is equal to X bar. So, X bar is the maximum likelihood estimator of mu. Now, if you look at here the parameter space for mu is minus infinity to infinity, for square it is 0 to infinity.

So, when we took sigma square is equal to 1 the parameter space is simply 0 to minus infinity to infinity and if you look at the X bar, X bar is likely to be n e value because in the normal distribution case the variable lies on the real line and therefore, the average value will also lie on the real line.

Now, if we had considered the methods of moments estimator in this problem then for mu the methods of moment estimator also would have been X bar; however, let us consider say a slightly different situation in the same case

(Refer Slide Time: 19:24)



$l(\mu)$
μ_0
$\bar{x}$
μ
$l(\mu)$
$\bar{x}$
μ_0
μ

Suppose (i) $\mu \geqslant \mu_0$

In this case the MLE of μ can be written as

$$\hat{\mu}_{RML} = \bar{X} \text{ if } \bar{X} \geqslant \mu_0$$
$$= \mu_0 \text{ if } \bar{X} < \mu_0$$

or $\hat{\mu}_{RML} = \max(\bar{X}, \mu_0)$.

(ii) $\mu \leq \mu_0$ $\hat{\mu}_{RML} = \mu_0, \text{ if } \bar{X} \geq \mu_0$

$$= \bar{X} \text{ if } \bar{X} < \mu_0$$

or $\hat{\mu}_{RML} = \min(\bar{X}, \mu_0)$

Suppose, we know from the physical considerations that mean mu is either greater than or equal to mu naught, less than or equal to mu naught or it lies in a interval say mu 1 to mu 2.

So, let me take 1 case say mu is greater than or equal to mu. Now, you look at the behavior of the likelihood function. So, we have observed here d l by d mu is equal to sigma X i minus mu. Now, this you can write as n times X bar minus mu. Now, once again you notice this. This is less than 0 if mu is greater than X bar, it is greater than 0 if mu is less than X bar. So, the nature of the likelihood function would have been of this nature that if this is mu on the x axis on the y axis we plot l mu, then for mu less than X bar the likelihood function is, the log likelihood function is increasing and it is decreasing thereafter.

Therefore a maximum is occurring at X bar. Now, if I use the restriction mu is greater than or equal to mu naught then there are 2 cases. Let us make the plot of the likelihood function. On this side we show mu and on this side we show L mu. So, we may have a situation that say mu naught is here. Now, our parameter expresses mu greater than r equal to mu naught. So, if you see it carefully our region of consideration is on the right side of this X is equal to mu naught, this mu is equal to mu naught. Now, the maximum value that mu is equal to X bar that is occurring within this region.

So, the maximum likelihood estimator for mu is still remains X bar. Let us look at the other case. Suppose, mu naught is on the right side here. Now, there is a problem, mu is greater than or equal to mu naught. So, our region of maximization is only this. Now, in this region if you see the likelihood function is decreasing, the maximum value is attained at mu naught therefore, your formal maximum likelihood estimator has got modified. So, we in this case the maximum likelihood estimator of mu can be written as mu hat. Let me put R M L just to denote a restriction that is equal to X bar if X bar is greater than or equal to mu naught and it is equal to mu naught if X bar is less than mu naught.

Or we can also express it in this fashion that mu hat R M L is equal to maximum of X bar and mu naught-. So, immediately you can notice that it has got changed from the original maximum likelihood estimator and therefore, it is certainly different from the method of moments estimator also because this procedure takes care of the exact parameter space where the maximization problem is solved, which is not true in the method of moments estimator. I will consider other type of restrictions for this problem.

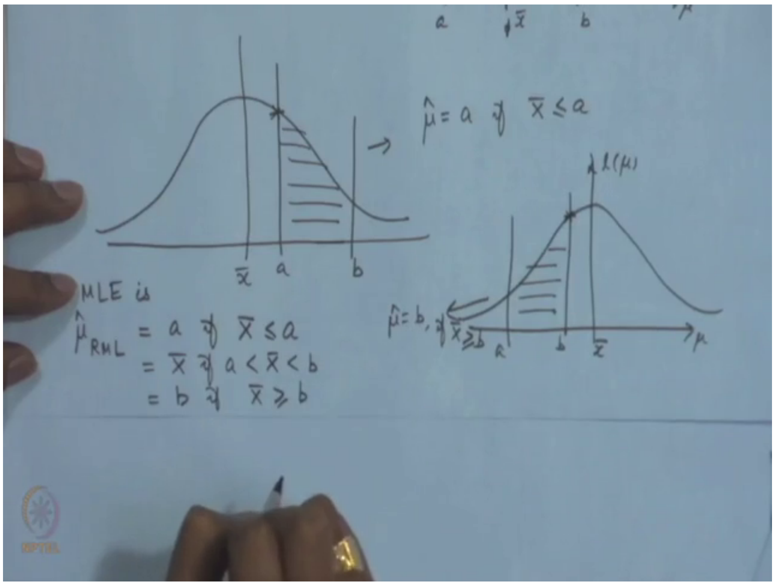
So, let us take say mu less than or equal to mu naught. Now, if you take mu less than or equal to mu naught, we can go back to the same graph and see this. If mu is less than or equal to mu naught and mu naught is in this position then our region of maximization is here therefore, the maximum value is occurring at mu naught. That means, I will say that mu had R M L it is equal to mu naught if X bar is greater than or equal to mu naught and it is equal to.

Now, in this case if you see if mu naught is on this side then our region of maximization is this full thing and here the maximum is occurring at X bar. So, it is equal to X bar if X bar is less than mu naught. So, this you can also say in other words as mu had R M L is equal to minimum of X bar and mu naught. So, notice here if we have the full region we get X bar as the maximum likelihood estimator for mu in the case of estimating the mean of a normal distribution when the variance is known.

Then there are certain restrictions like a lower bound placed or an upper bound placed for the parameter mu then accordingly the maximum likelihood estimator gets modified. In this case it is becoming maximum of X bar and mu naught and in this case it is becoming minimum of X bar and mu naught. Let me take another kind of restriction in many of the practical problems. It may happen that the mean mu lies between 2 values for example, you look at the average income levels, you look at the average rainfall, you look at the average weight, average height.

So, there are various parameters which occur in the practical situations which are actually bounded in nature, they are not unbounded; that means, we cannot say that they take values from minus infinity to infinity. So, when that information is available to us in that case we should utilize that and our estimator should reflect that.

(Refer Slide Time: 26:05)



MLE is

$$\hat{\mu}_{RML} = a \text{ if } \bar{X} \le a$$
$$= \bar{X} \text{ if } a < \bar{X} < b$$
$$= b \text{ if } \bar{X} \geq b$$

$\hat{\mu} = a$ if $\bar{X} \le a$

$\hat{\mu} = b$, if $\bar{X} \geq b$

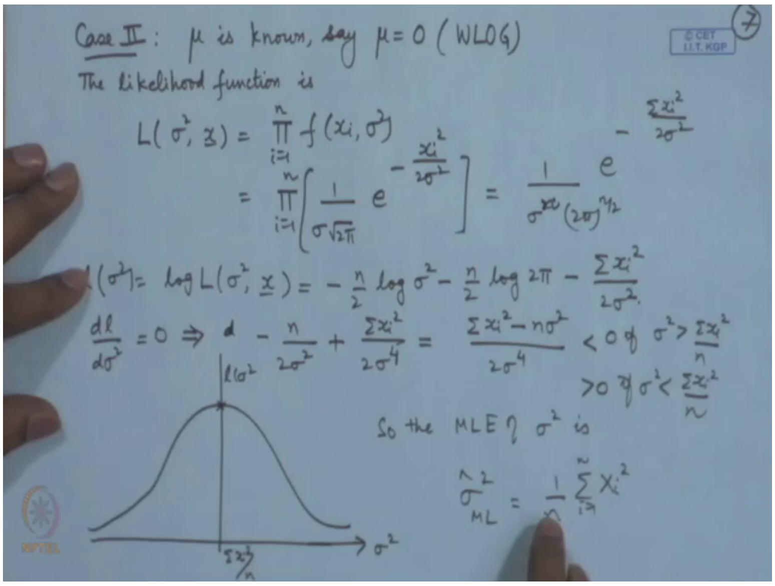
That means, let me take the third restriction of this nature that say a is less than or equal to mu is less than or equal to b. Now, this is even more interesting.

We look at the likelihood function as we have plotted in this particular case. So, so if your a and b is for example, containing X bar, that means, X bar lies between a to b then the maximum occurs as usual at X bar. However, you could have had other kind of situations. So, in this case in this case u hat is equal to X bar, that means, when X bar is lying between a to b. You consider another situation for example, a and b are here. If a and b are here, then we have to look at the maximum of likelihood function within this region alone and obviously, the maximum occurs at a.

So, in this particular case then the maximum likelihood estimator is becoming a if X bar is less than a. And a similar situation would occur if we consider say say a and b are to the left to the X bar. In this case our maximization problem is restricted to this region and if you see the maximum is occurring at b. So, in this particular case then mu hat will become equal to b, that means, if X bar is greater than or equal to b. Therefore, our solution for the full problem of mu lying between a to b is that mu hat R M L, it is equal to a if X bar is less than or equal to a it is equal to X bar if a is less than X bar less than b and it is equal to b if X bar is greater than or equal to b.

So, if there is any prior information about the parameter the method of maximum likelihood estimation takes care of that. Now, let me take additional cases in the case of normal distribution. See, here we have taken the case for estimating mu because sigma square was known. Now, you may have another identical situation where mu may be known and we may be interested in the estimation of sigma square. So, let us look at this situation then say mu is known.

(Refer Slide Time: 29:28)



Case II: μ is known, say $\mu = 0$ (WLOG)

The likelihood function is

$$L(\sigma^2, \underset{\sim}{x}) = \prod_{i=1}^{n} f(x_i, \sigma^2)$$

$$= \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x_i^2}{2\sigma^2}} \right] = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{\Sigma x_i^2}{2\sigma^2}}$$

$$l(\sigma^2) = \log L(\sigma^2, \underline{x}) = -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{\Sigma x_i^2}{2\sigma^2}.$$

$$\frac{dl}{d\sigma^2} = 0 \Rightarrow d \quad -\frac{n}{2\sigma^2} + \frac{\Sigma x_i^2}{2\sigma^4} = \frac{\Sigma x_i^2 - n\sigma^2}{2\sigma^4} \begin{array}{l} < 0 \text{ if } \sigma^2 > \frac{\Sigma x_i^2}{n} \\ > 0 \text{ if } \sigma^2 < \frac{\Sigma x_i^2}{n} \end{array}$$

So the MLE of σ^2 is

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

If mu is known then without loss of generality, we may put mu is equal to 0 because you can always shift all the observations by mu naught for example, if I say mu is equal to mu naught then we may put it is equal to 0. So, now, you look at the likelihood function notice here the problem gets modified in the maximum likelihood estimation as soon as the the information about the parameters is changed.

So, the likelihood function is the product of the density functions of X 1, X 2, X n that is equal to 1 by sigma root 2 pi e to the power minus X i square by 2 sigma square; i is equal to 1 to n. So, we can write it in a more compact fashion. It becomes equal to 1 by sigma square, sigma to the power n, 2 pi to the power n by 2, e to the power minus sigma X i square by 2 sigma square. Notice here that sigma is occurring in the denominator as well as it is occurring in the denominator of the exponent therefore, it is beneficial to consider the log likelihood function that is equal to minus n by 2 log of sigma square minus n by 2 log of 2 pi minus sigma X i square by 2 sigma square.

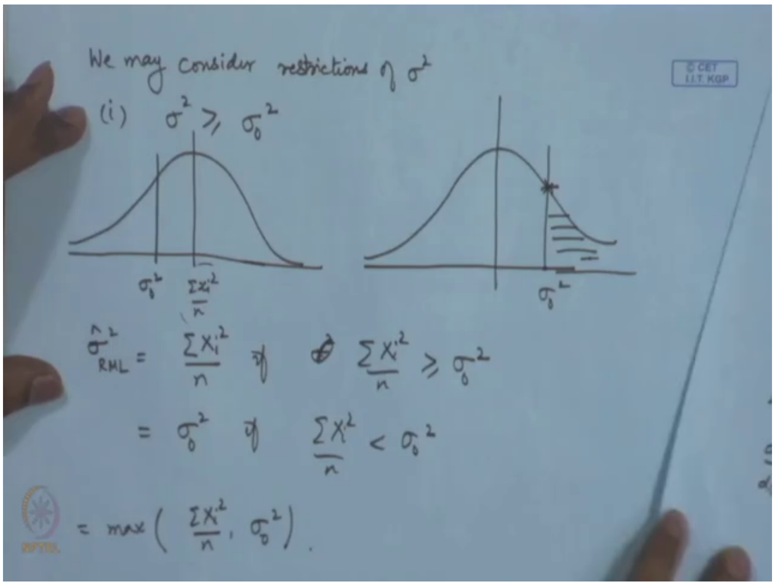
So, we consider the likelihood equation that is d l by d sigma square is equal to 0. So, so when you differentiate this you will get minus n by 2 sigma square plus sigma X i square by twice sigma to the power 4. Notice here that I am considering sigma square as a parameter. One we misled by considering sigma as the parameter and then you may be getting a slightly different derivative here. So, later on we will show that the 2 procedures will lead to the same answer identical answer; that means, whether you are considering estimation of sigma or you are considering estimation of sigma square, it should not lead to contradictory statements.

Now, we write it in a slightly modified fashion sigma X i square minus n sigma square by twice sigma to the power 4. So, notice here this will be less than 0 if sigma square is greater than sigma X i square by n and it is greater than 0, if sigma square is less than sigma X i square by n. So, if we look at the plot of the likelihood function then naturally the likelihood function is increasing up to sigma X i square by n because the derivative is positive for sigma square less than sigma X i square by n.

So, it is increasing up to this and thereafter it is decreasing. So, the maximum occurs at sigma X i square by n. So, the maximum likelihood estimator of sigma square is we will write m l just to denote that it is the maximum likelihood estimator that is standing out to be 1 by n sigma X i square. Now, you can look at the variation in place of mu is equal to 0 if we had put mu is equal to mu naught then what would have been the modification. Here, we would have got X i minus mu naught whole square therefore, when we considered the derivative here we would have got an increasing and decreasing nature for sigma X i minus mu naught whole square by n.

Thereby the answer would have been 1 by n sigma X i minus mu naught whole square. So, now, once again let me show you the effect of the prior information in this. Suppose, on sigma square we have certain information because as you know sigma square is a variance. Now, the variance are a reciprocal of that is known as the precision. So, the variability may be known in advance or it may have certain restrictions for example,

(Refer Slide Time: 34:45)



We may consider restrictions of σ^2

(i) $\sigma^2 \geqslant \sigma_0^2$

σ_0^2
$\frac{\Sigma x_i^2}{n}$
σ_0^2

$$\hat{\sigma}^2_{RML} = \frac{\Sigma X_i^2}{n} \text{ if } \frac{\Sigma X_i^2}{n} \geqslant \sigma_0^2$$

$$= \sigma_0^2 \text{ if } \frac{\Sigma X^2}{n} < \sigma_0^2$$

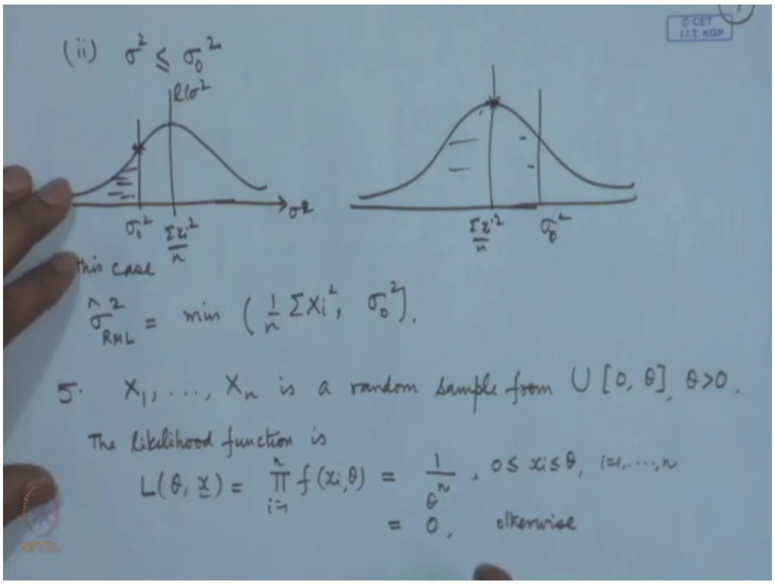
$$= \max\left(\frac{\Sigma X_i^2}{n}, \sigma_0^2\right).$$

we may consider say restrictions on sigma square say for example, sigma square may be greater than or equal to sigma naught square.

Now, if you consider sigma square greater than or equal to sigma naught square then in this case there will be 2 cases because sigma naught square may occur here or sigma naught square may occur here. So, let us see. This is sigma X i square by n and it may happen that sigma naught square is here. So, in this case the maximum occurs at this point whereas, if sigma naught square occurs here in that case our region of maximization is here because sigma square is greater than or equal to sigma naught square.

In that case the maximum will occur at sigma naught square. So, we conclude that sigma hat square restricted M L E is equal to sigma X i square by n if sigma square, if sigma X i square by n is greater than or equal to sigma naught square, it is equal to sigma naught square if sigma X i square by n is less than sigma naught square. That means, you can write it as maximum of sigma X i square by n and sigma naught square. In a similar way one may consider the case of an upper bound on sigma square

(Refer Slide Time: 36:46)



(ii) $\sigma^2 \leq \sigma_0^2$

this case

$$\hat{\sigma}^2_{RML} = \min\left(\frac{1}{n}\Sigma X_i^2, \sigma_0^2\right).$$

5. $X_1, \ldots, X_n$ is a random sample from $U[0, \theta]$, $\theta > 0$.

The likelihood function is

$$L(\theta, \underline{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{\theta^n}, \quad 0 \leq x_i \leq \theta, \; i=1, \ldots, n$$
$$= 0, \quad \text{otherwise}$$

Let me take sigma square less than or equal to sigma naught square So, once again if we look at the plot of the likelihood function in that case if sigma naught square is occurring here now this is our region of maximization. So, the maximum will occur at sigma naught square whereas, if sigma naught square occurs here then this is our region of maximization and we get the maximum here. So, in this case the maximum likelihood estimator of sigma square will be minimum of 1 by n sigma X i square sigma naught square.

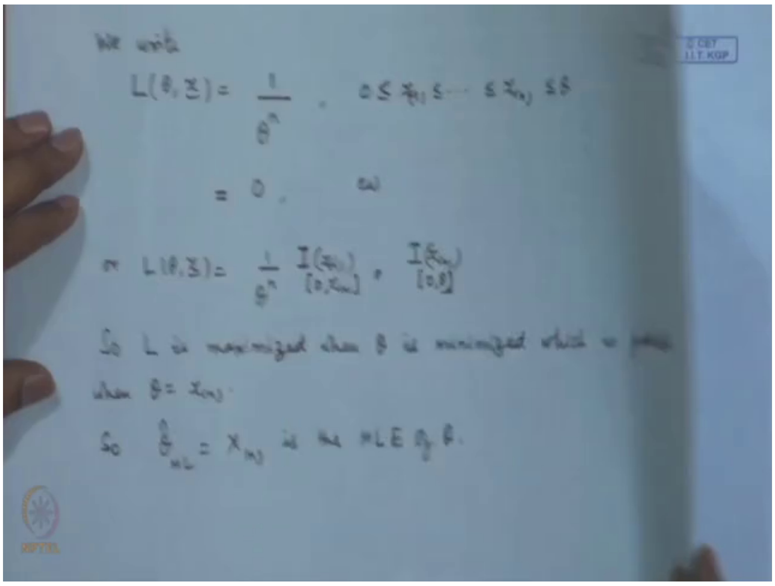
So, the effect of the information or the prior information about the parameter plays a role in the maximum likelihood estimation and that is 1 important feature which distinguishes the method of maximum likelihood estimation from various other methods. The examples that I have discussed take into account that the likelihood function are the log of the likelihood function is a nice or you can say a smooth function, because we are able to differentiate and carry out the usual arguments of the analysis.

Now, in certain situations that may not be possible. Let me take up another case say X 1, X 2, X n is a random sample from uniform 0 theta distribution where theta is the unknown parameter which is certainly positive. We are interested in the maximum likelihood estimation for theta. If you recollect the method of moments estimator for theta was 2 X bar because the mean of the uniform distribution is theta by 2. So, the first sample moment that is X bar would be the moments estimator for theta by 2, that means, 2 X bar will be the method of moments estimator for theta.

Let us look at the maximum likelihood estimator here. So, the likelihood function is l theta x that is equal to product of f X i theta, i is equal to 1 to n. Now, this we write as 1by theta to the power n because the density function of the uniform distribution on the interval 0 to theta it is 1 by theta. So, it will become 1 by theta to the power n, but at the same time let us not forget that each of the X i is lies between 0 and theta this is for i is equal to 1 to n.

Now, we should also write that it is 0 at other places. Now, a common thing which we have been applying earlier that you take the log of this and differentiate with respect to theta and put equal to 0. Now, in this case what it would lead to? You will get minus n log theta and if you differentiate you will get minus n by theta which you put equal to 0 will give you an absurd answer. The reason for this absurdity is that we have not taken care of the full likelihood function. The full likelihood function takes into account this portion also.

(Refer Slide Time: 41:01)



We write

$$L(\theta, \underset{\sim}{x}) = \frac{1}{\theta^n}, \quad 0 \le x_{(1)} \le \cdots \le x_{(n)} \le \theta$$

$$= 0, \quad \text{ow}$$

or $L(\theta, \underset{\sim}{x}) = \frac{1}{\theta^n} I_{[0, x_{(n)}]}(x_{(1)}) \cdot I_{[0,\theta]}(x_{(n)})$

So L is maximized when θ is minimized which is possible when $\theta = x_{(n)}$.

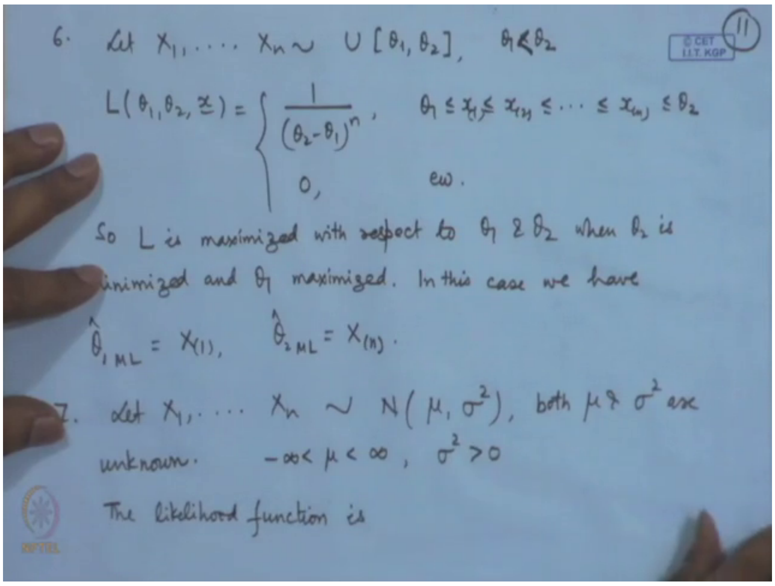
So $\hat{\theta}_{ML} = X_{(n)}$ is the MLE of θ.

So, we write it in a slightly more compact fashion as follows. We may write the likelihood function as 1 by theta to the power n, 0 less than or equal to X 1, less than or equal to X n, less than or equal to theta or we can also write theta as 1 by theta to the power n i. Here, we can say that all the X i's are from 0 to X n and multiplied by X n itself lies between 0 to theta. Now, if you look at the maximization of this with respect to theta, now the theta is occurring in the denominator.

So, that means, what is the minimum value of theta, the minimum possible value of theta is X n, theta cannot be below X n, because of the observations. Each of the observations lies between 0 to theta. So, l is maximized when theta is minimized which is possible when theta is equal to X n. So, theta hat m l is equal to X n is the maximum likelihood estimator of theta, that is the maximum of the observations. So, you can see here the result is quite different from the method of moments estimation here

Because in the method of moments we would have got 2 X bar. So, this is certainly different and later on we will study the criteria that which 1 should be preferred here. That means, whether m m e is better here or M L E is better here which one one should prefer. So, we will discuss about those criteria later on. This example shows that one should not blindly use the differentiation and put equal to 0 because this will not be the answer in this particular situation. Similar, thing would occur for example, if I consider 2 parameter uniform distribution.

(Refer Slide Time: 43:45)



11

© CET I.I.T. KGP

6. Let $X_1, \ldots X_n \sim U[\theta_1, \theta_2]$, $\theta_1 < \theta_2$

$$L(\theta_1, \theta_2, \underset{\sim}{x}) = \begin{cases} \frac{1}{(\theta_2-\theta_1)^n}, & \theta_1 \le x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)} \le \theta_2 \\ 0, & \text{ew}. \end{cases}$$

So L is maximized with respect to θ_1 & θ_2 when θ_2 is minimized and θ_1 maximized. In this case we have

$\hat{\theta}_{1ML} = X_{(1)}$, $\hat{\theta}_{2ML} = X_{(n)}$.

7. Let $X_1, \ldots X_n \sim N(\mu, \sigma^2)$, both μ & σ^2 are unknown. $-\infty < \mu < \infty$, $\sigma^2 > 0$

The likelihood function is

Suppose, I take a random sample from uniform theta 1 to theta 2, here theta 1 is certainly less than or equal to theta 2. So, in this particular case we have 2 unknown parameters here and we consider the maximum likelihood estimation. So, as before we consider the likelihood function and this will be and it is equal to 0 ensure. Now, you notice the likelihood function here. The likelihood function has theta 2 minus theta 1 in the denominator which is positive quantity and we are looking at the maximization. That means, theta 2 minus theta 1 should be minimum. That means, theta 2 should be minimum and theta 1 should be maximum.

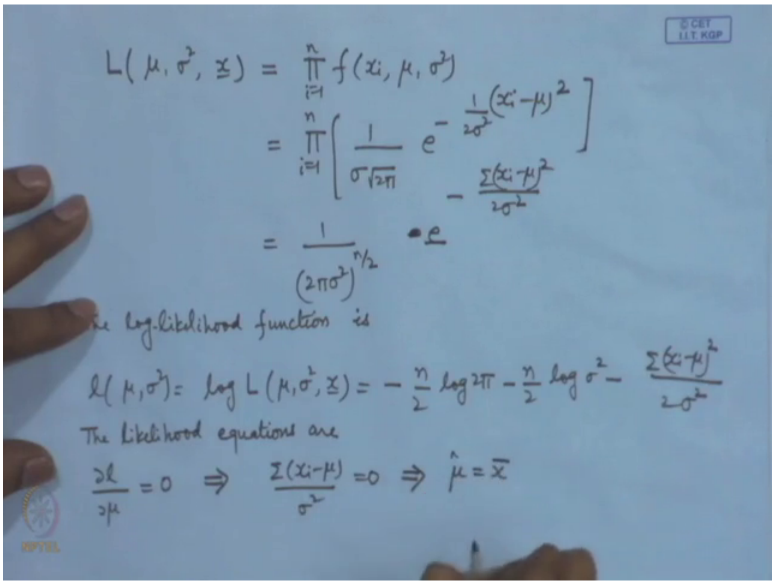
Now, if you look at the nature of the observations all the observations lie between theta 1 to theta 2 therefore, the minimum of the observations is certainly greater than or equal to theta 1 and the maximum of the observations is certainly less than or equal to theta 2. So, l is maximized with respect to theta 1 and theta 2 when theta 2 is minimized and theta 1 maximized. So, in this case we have theta 1 hat maximum likelihood estimator is equal to the minimum of the observations and theta 2 hat M L Equal to the maximum of the observations.

And all this is an example where I have considered 2 parameter problem. So, the method of maximum likelihood estimator can be used for the maximization of the likelihood function when there can be more than 1 parameter and in that case the maximization should be considered a with respect to all the parameters. So, in this case you can see the simultaneous maximum is occurring.

Now, let us go back to the case of normal distribution that I discussed earlier. Here, I had taken special cases. If you see carefully if we consider normal mu sigma square here, I have taken sigma square to be known. So, in effect I have reduced it to 1 parameter problem. Similarly, if you look at mu is known then once again the parameter has been reduced to sigma square alone. So, in effect this problem also reduced to 1 parameter problem. However, in general both the parameters in a normal distribution may be unknown and in that case let us look at the solution.

So, let me discuss in detail. So, we have X 1, X 2, X n a random sample from normal mu sigma square as before. However, both mu and sigma square are unknown. So, in general you remember that in the normal distribution the mean parameter may vary from minus infinity to plus infinity and the variance parameter will be from 0 to infinity. Now, in this case when we want to find out the maximum likelihood estimator, we will like to find out for both mu and sigma square. So, let us write down the likelihood function.

(Refer Slide Time: 48:31)



$$L(\mu, \sigma^2, \underline{x}) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2)$$

$$= \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \right]$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\Sigma(x_i-\mu)^2}{2\sigma^2}}$$

e log-likelihood function is

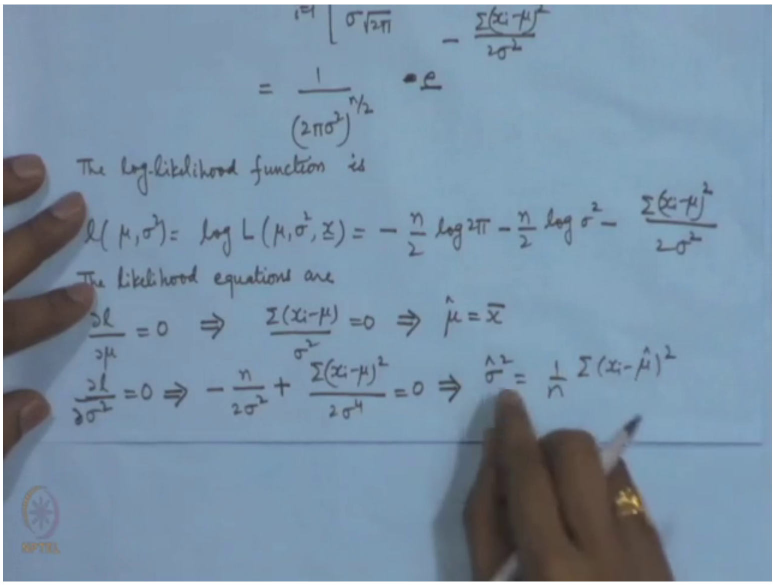
$$l(\mu,\sigma^2) = \log L(\mu,\sigma^2,\underline{x}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{\Sigma(x_i-\mu)^2}{2\sigma^2}$$

The likelihood equations are

$$\frac{\partial l}{\partial \mu} = 0 \Rightarrow \frac{\Sigma(x_i-\mu)}{\sigma^2} = 0 \Rightarrow \hat{\mu} = \bar{x}$$

So, the likelihood function is L mu sigma square X. Notice here that this has become function of both mu and sigma square now. So, this is a joint density function as before. In the earlier cases I had substituted special values of mu or sigma square as the case was. In this case we will have to write down the full form of the density function of a normal distribution that is 1 by sigma root 2 pi e to the power minus 1 by 2 sigma square X i minus mu whole square.

(Refer Slide Time: 48:31)



$$i=1 \Big[\sigma\sqrt{2\pi} \qquad -\frac{\Sigma(x_i-\mu)^2}{2\sigma^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e$$

The log-likelihood function is

$$l(\mu,\sigma^2) = \log L(\mu,\sigma^2,\underline{x}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{\Sigma(x_i-\mu)^2}{2\sigma^2}$$

The likelihood equations are

$$\frac{\partial l}{\partial \mu} = 0 \Rightarrow \frac{\Sigma(x_i-\mu)}{\sigma^2} = 0 \Rightarrow \hat{\mu} = \bar{x}$$

$$\frac{\partial l}{\partial \sigma^2} = 0 \Rightarrow -\frac{n}{2\sigma^2} + \frac{\Sigma(x_i-\mu)^2}{2\sigma^4} = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\Sigma(x_i-\hat{\mu})^2$$

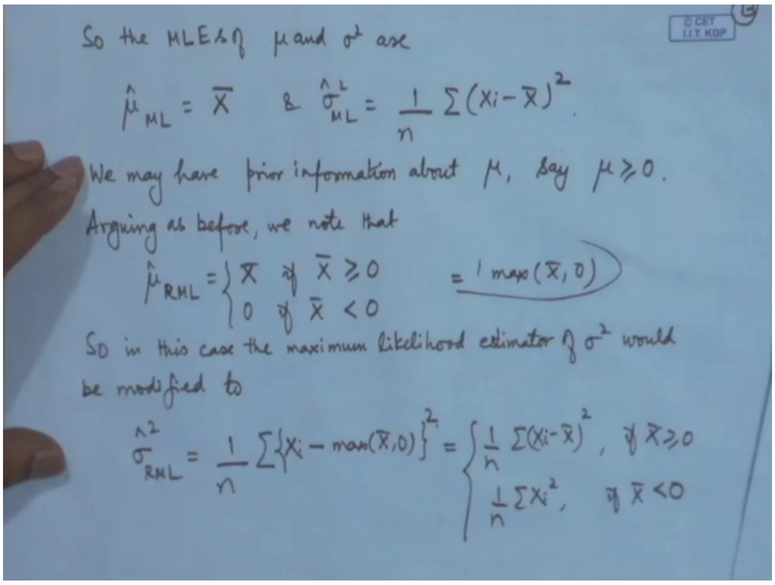
So, we write it in a slightly more compact fashion. This becomes 1 by 2 pi sigma square to the power n by 2 e to the power that when you take the. So, it will become e to the power minus sigma X i minus mu square by twice sigma square. Again, you observe the parameters for which we need the estimators they are occurring in the exponent as well as they are occurring in the main form here. So, it will be beneficial if we consider the log likelihood as before. So, the log likelihood L mu sigma square, log of L mu sigma square x that is equal to minus n by 2 log of 2 pie minus n by 2 log of sigma square minus sigma X i minus mu square divided by twice sigma square.

This equation this function involves mu and sigma square 2 variables. We need to maximize this with respect to both mu and sigma square. So, since this function is still a very nice smooth function. So, we can still use the direct calculus methods for example, by taking the first order derivatives putting them equal to 0 they are giving us the likelihood equation. The solutions of that will be the points of minimum or maximum which we can check separately that they would be actually leading to the maximization points. They will not be the points of minimum.

So, in this case for example, we write down the the likelihood equations. The likelihood equations are del l by del mu is equal to 0 that is sigma X i minus mu by sigma square is equal to 0 which we can further write because this can be easily simplified. Sigma square is in the denominator that would give mu hat is equal to X bar. The other equation is del l by del sigma square is equal to 0 which will give me minus n by y sigma square plus sigma X i minus mu square by twice sigma to the power 4 equal to 0

Which will give me sigma square is equal to 1 by n, sigma X i minus mu hat square. Actually, the equation is sigma square is equal to 1 by n sigma X i minus mu square. We substitute the value of mu from the first equation and substitute here. So, the maximum likelihood estimators then turn out to be.

(Refer Slide Time: 52:50)



So the MLEs of μ and σ^2 are

$$\hat{\mu}_{ML} = \overline{X} \quad \& \quad \hat{\sigma}^2_{ML} = \frac{1}{n} \sum (X_i - \overline{X})^2.$$

We may have prior information about μ, say $\mu \geq 0$.

Arguing as before, we note that

$$\hat{\mu}_{RML} = \begin{cases} \overline{X} & \text{if } \overline{X} \geq 0 \\ 0 & \text{if } \overline{X} < 0 \end{cases} \quad = \max(\overline{X}, 0)$$

So in this case the maximum likelihood estimator of σ^2 would be modified to

$$\hat{\sigma}^2_{RML} = \frac{1}{n} \sum \{X_i - \max(\overline{X}, 0)\}^2 = \begin{cases} \frac{1}{n} \sum (X_i - \overline{X})^2, & \text{if } \overline{X} \geq 0 \\ \frac{1}{n} \sum X_i^2, & \text{if } \overline{X} < 0 \end{cases}$$

So, the maximum likelihood estimators of mu and sigma square are mu hat m l is equal to X bar and sigma hat square m l is equal to one by n sigma X i minus X bar whole square.

In this case you may notice that these are the same as the method of moments estimator for this particular problem, but once again as I mentioned earlier the method of maximum likelihood can take care of many other possibilities also. For example, we may have say prior information about mu say mu is greater than or equal to 0. In that case once again we look at the likelihood function here we are getting n X bar minus mu. So, if we plot the behavior with respect to mu then the maximum is occurring at X bar, but if X bar is greater than 0, I will consider 0 here and this region is coming.

So, the maximum likelihood estimator will be X bar. However, if 0 occurs on this side and then we have this portion then the maximum will occur at 0. So, arguing as before we note that mu hat restricted m l will be equal to X bar, if X bar is greater than or equal to 0 it will be equal to 0 if X bar is less than 0, which we can actually write as maximum of X bar and 0. Now, if we use this in that case the second equation the solution will get modified because for sigma square the estimator was 1 by n sigma X i minus the estimator of mu.

And if the estimator for mu gets modified immediately the estimator for sigma square will also get modified. So, in this case the maximum likelihood estimator of sigma square would be modified to sigma hat square or m l is equal to 1 by n sigma X i minus maximum of X bar

0 which we can write as 1 by n sigma X i minus X bar whole square if X bar is greater than or equal to 0 and it will become 1 by n sigma X i square if X bar is less than 0.

So, the placing of additional information about the parameter changes the maximum likelihood estimators. I will consider a few more examples in the next class and also then we will see there are certain desirable properties, which are basically called the large sample properties that the maximum likelihood estimators satisfy and because of this the method has wide applicability among statisticians. So, in the tomorrow's class we will consider various properties of the maximum likelihood estimators and then we will proceed to determining the criteria for judging the goodness of the estimators. So, thank you.