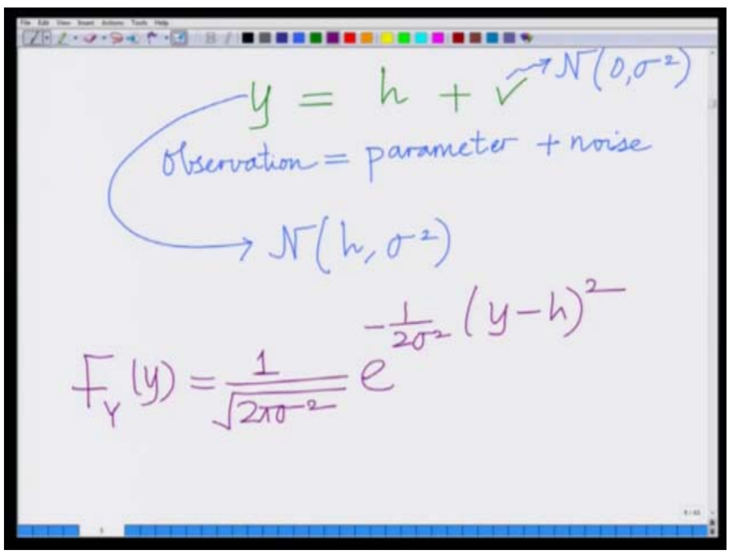
Estimation for Wireless Communication –MIMO/OFDM on Cellular and Sensor Networks
Professor A K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Lecture Number 02
Likelihood Function and Maximum Likelihood (ML) Estimate

Hello, welcome to another module in this massive open online course on Estimation or Estimation theory for Wireless Communication. And in the previous module we had started looking at the basic model for Estimation.

We had considered a sensor network and then noisy observation model for the Sensor network in which our observation.

(Refer Slide Time: 00:39)



Remember we have our observation y which is equal to the parameter h + the noise, so this is your observation at the Centre which is equal to the parameter that is denoted by h + the noise that is denoted by… And in addition we have seen then this v is Gaussian with mean 0 Sigma square.

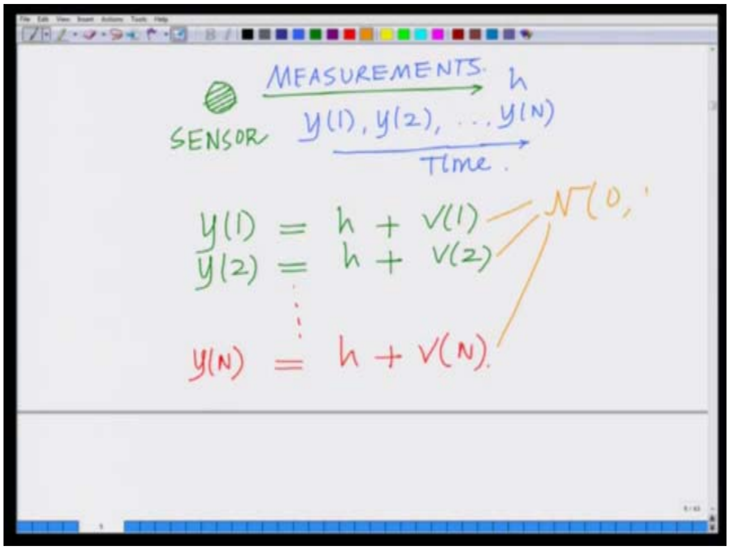
Then this y is also Gaussian with mean given by the unknown parameter h and variance Sigma square. Therefore, Probability Density Function of this observation f y of y that is the PDF of this observation y.

That is the random variable y is 1 over 2 pie square root Sigma square e raise to - 1 over 2 Sigma square y - h square, this is the Probability Density Function of the observation y at the

Sensor node corresponding to the parameter h and Gaussian noise v with mean 0 at variance or power Sigma square. So this is what we have seen in the previous module.

Now let us extend this module further, let us extend this to a scenario where a sensor node is making multiple measurements in time. For now I would like to consider our sensor node scenario. So let us go back to our sensor network, our sensor node which is now making multiple measurements. So our sensor node which is making multiple, so our sensor node is making…

(Refer Slide Time: 02:17)



Again remember measurements, but this time it is making multiple measurements of this parameter h in time. So we have measurements y 1, y 2, y N and these are the multiple measurements being made at different time instance 1, 2 up to and so we have the measurements y 1, y 2, y N capital N measurements being made a capital N time instance.

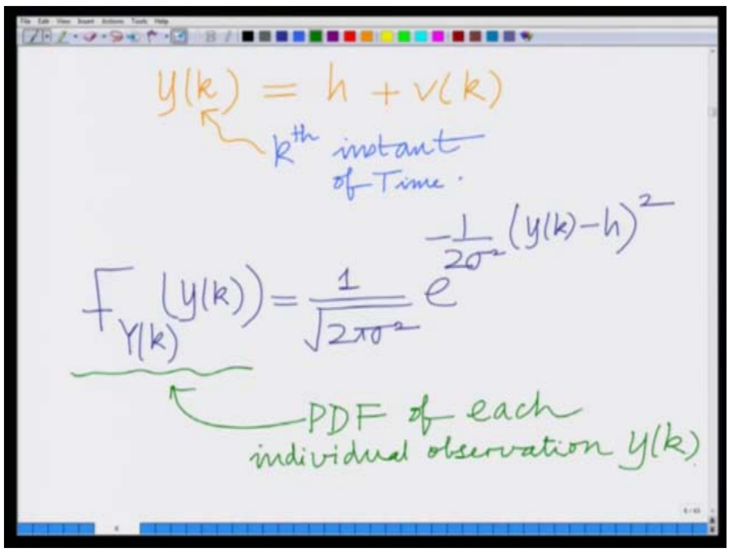
And interestingly now you can also consider this as a sensor network with multiple sensor nodes with capital N sensor nodes making capital N measurements, right. Instead of considering these such as single sensor making N measurements in time when can also consider this as N sensors making multiple measurements at a single instant of time.

So this so this basically captures both the single sensor and also a multiple sensor scenario for the measurement and subsequent estimation of the parameter h. Now therefore, these measurements can be modeled as each measurement is basically a noisy observation y 1 equals h + v 1, y 2 equals the parameter h + v 2.

So on and so forth y N, the Nth measurement is the parameter h + v N and similarly we also assume that all these noise elements v 1, v 2, v N are noise with 0 mean and variance Sigma square, so each measurement y K at the Kth instant of time is h + v K where K can range from 1 to N, we have N instant of time.

So in general I can represent this as the Kth measurement y K equals h + v K where K denotes the Kth instant the Kth instant of time.

(Refer Slide Time: 04:47)



And naturally the Probability Density Function corresponding to y K is f of y K of your observation y K is again the same. It is Gaussian with mean given by the parameter h and variance Sigma square.

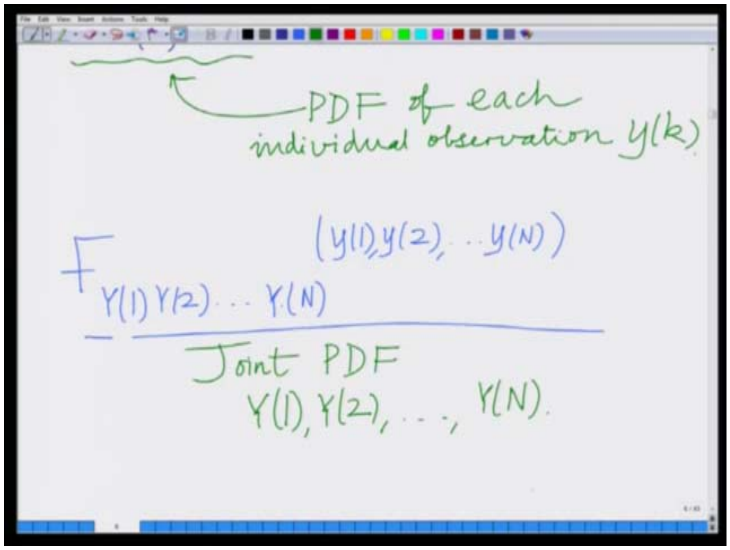
So this is given by the Probability Density Function which is 1 over square root of 25 Sigma pie Sigma square e raise to -1 over 2 Sigma square y K - h whole square, this is the Probability Density Function, this is the PDF of each individual of each individual observation y of K.

This is the PDF of each individual observation y of K, we are saying the same thing that is each individual observation made by the sensor at time instant K is a Gaussian, has a Gaussian Probability Density Function mean with given by the unknown parameter h and variance given by the variance of the noise v k which is Sigma square.

Now what we are interested is we want to come up with the joint density of the of the observation y 1, y 2 up to y N. We have to come up with a joint PDF or the joint Probability

Density Function of the observation y 1, y 2 up to y N which is represented by f subscript y 1, y 2, y N of y 1, y 2 up to y N.

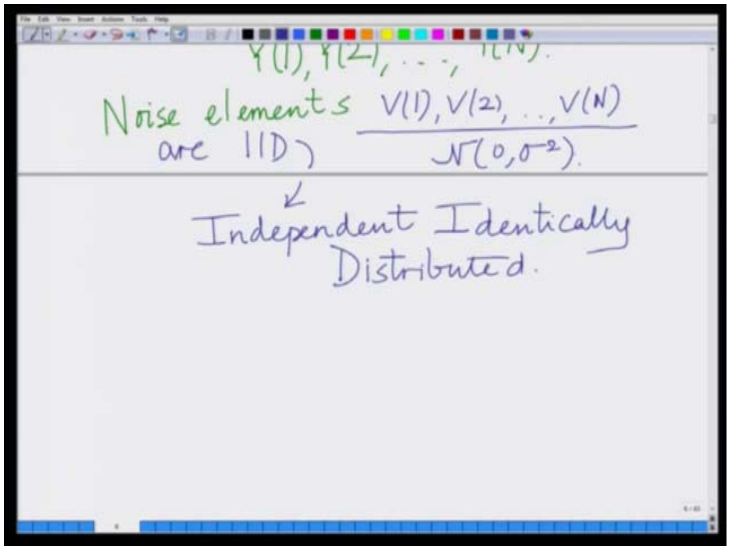
(Refer Slide Time: 06:39)



We have to come up with the joint Probability Density Function, this is a joint PDF of these observations y 1, y 2 up to y N and towards this and we will make a simplifying assumption to come up with a joint probability density function, we are going to assume initially that these noise elements v 1, v 2, v N are IID that is these are Independent Identically Distributed.

So the noise elements, the assumption that we are going to make is towards this end to develop the joint Probability Density Function. The noise elements or the noise samples v 1, v 2, v N are IID.

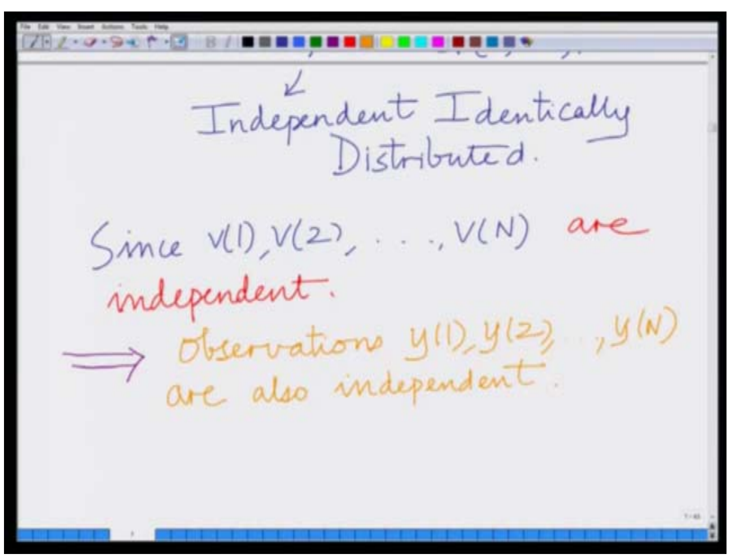
(Refer Slide Time: 07:55)



What do we mean by IID that is, they are independent and identically, they are Independent Identically Distributed. That is, all of these are identical distributed, that is, what we have already early on, that is Gaussian with mean 0 variance Sigma square.

And also they are Independent, right. These noise samples v 1, v 2, v N are Independent. Which means observations y 1, y 2, y N are also independent, so the observations y 1, y 2, y N of the noise sample v 1, v 2, v N are dependent. The observations y 1, y 2, y N are Independent and therefore what we have, since v 1, v 2 noise samples v 1, v 2 up to v N are Independent.

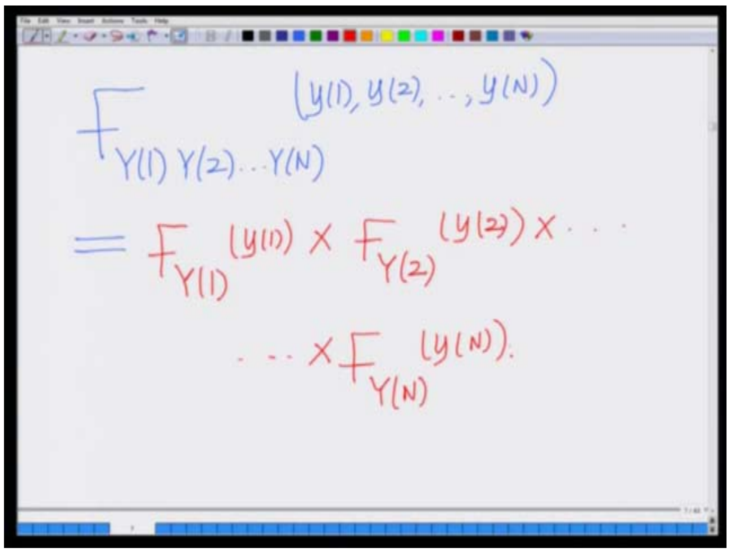
(Refer Slide Time: 08:56)



This implies that basically your observations y 1, y 2 up to y N are also independent. In fact, they are Independent identically distributed with mean h and variance Sigma square. Therefore, the probable joint Probability Density Function of the observation is of the observations y 1, y 2, y N is the product of the individual Probability Density Function corresponding to the observation because the observations are Independent.

And that gives us an important result.

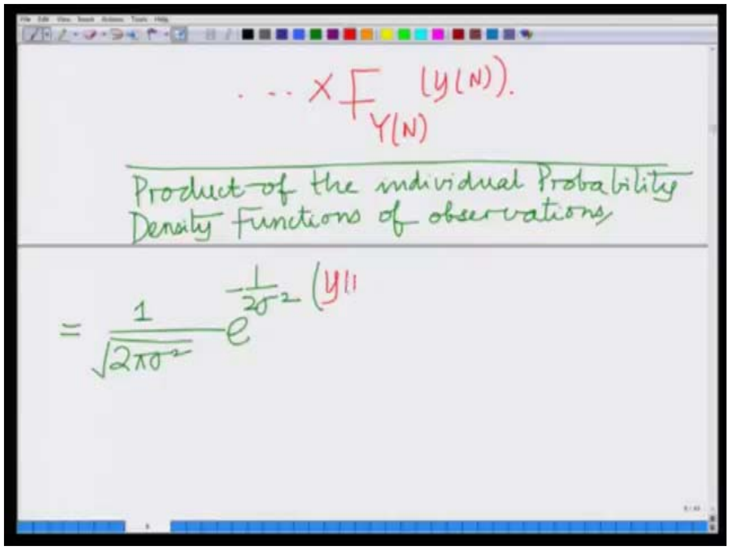
(Refer Slide Time: 10:11)



That is, the joint Probability Density Function which we are talking about before f of y 1 y 2 up to y N of y 1, y 2 up to y N. This is equal to the product of the individual Probability Density Function that is f of y 1 of y 1 times f of y 2 of y 2 times f of N of y N that is, this is the product of the individual Probability Density Function.

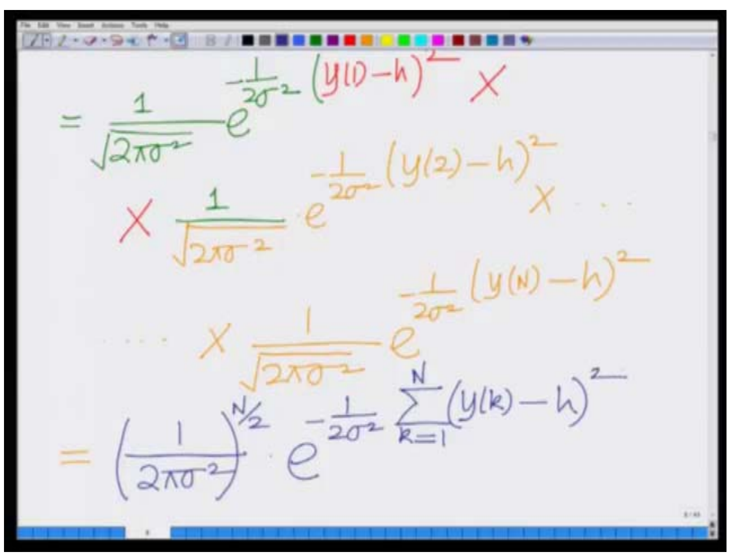
Product of individual Probability Density Function of the observation and we know what is the individual Probability Density Function of each observation that is…

(Refer Slide Time: 11:31)



1 over 2 pie Sigma square, square root of 2 pie Sigma square e raise to - 1 over 2 Sigma square times y 1 - h whole square…

(Refer Slide Time: 11:50)

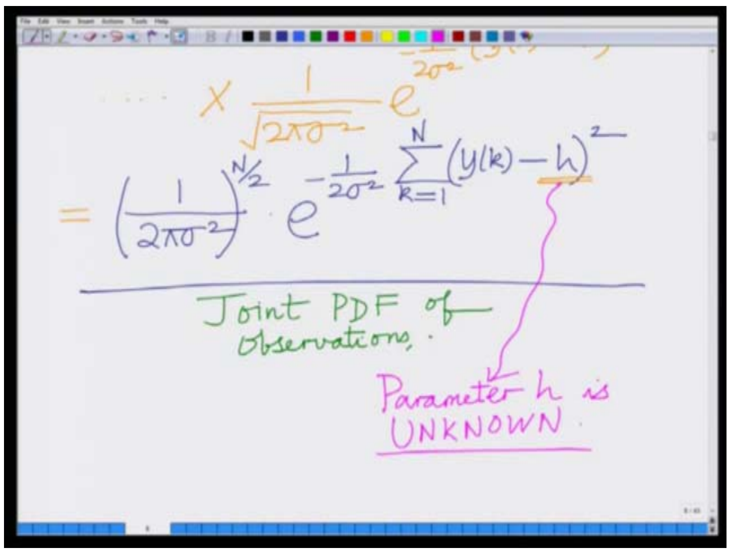


Times the product times the product 1 over square root 2 pie Sigma square e raise to - 1 over 2 Sigma square y 2 - h square times 1 over square root of 2 pie Sigma square e rise to - 1 over 2 Sigma square y N - h y N - h whole square. This is the product of the individual Probability Density Function corresponding to the observations y 1, y 2 up to y N.

And now simplifying this product of the individual Probability Density Function, you can clearly see that this is given as 1 over, that is collecting 1 over the square root of 2 pie Sigma square terms, I have 1 over 2 pie Sigma square raise to the power N by 2 e raise to - 1 over 2 Sigma square submission, all the expressions all the terms in the exponent will add up K equals 1 to N submission K equals 1 to N y K - h whole square.

And this is your joint Probability Density Function.

(Refer Slide Time: 13:48)



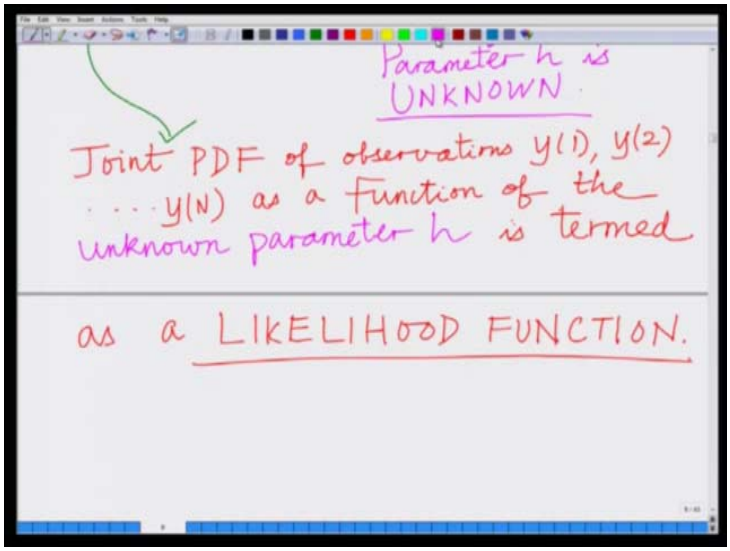
This is the joint PDF of this is the joint PDF of the observation y 1, y 2 up to y N. And now [we] I would like to bring to your attention an important point, note that while deriving this joint PDF we have ignored the facts so far that this parameter h is an unknown parameter. This is the parameter that we are trying to estimate at the sensor node or the fusion centre.

This parameter h as I pointed out to you also in the previous module that this parameter h is unknown. So to realize that in this, in our Probability Density Function, this parameter h is unknown. This parameter h is unknown and therefore one has to estimate this parameter h, this is known as Parameter Estimation.

This is what I pointed out to you in the previous module that is estimating this parameter h. That is finding on what this computing the value of this parameter h is termed as Parameter Estimation. And this probability joint Probability Density Function as a function of the unknown parameter h, because there is a parameter which is unknown, as a function of the unknown parameter h, this is known as a Likelihood Function.

And this is the important point, and one of the central points. I would like to point out in the entire framework of estimation theory. That is the Probability Density Function viewed as the function of the unknown parameter h is termed as Likelihood Function. That is the the joint Probability Density Function of the observation.

(Refer Slide Time: 15:40)



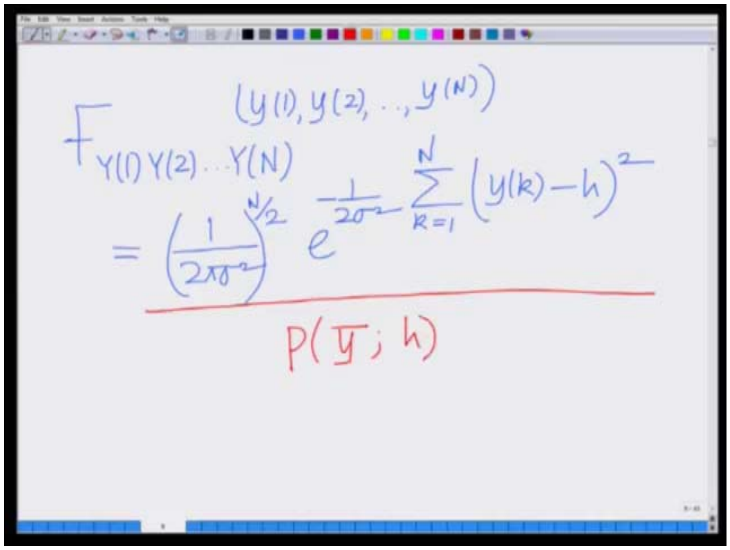
Let me put it this way, the joint PDF of your observations y 1 y 1, y 2 up to y N as a function of the unknown parameter H is termed the Likelihood Function. This is the Likelihood Function corresponding to the, this is the Likelihood Function which is basically the joint Probability Density Function of the observations y 1, y 2, y N which is viewed as a function of this unknown parameter h.

This is known as the Likelihood Function, Likelihood Function of the parameter h. The argument is the unknown parameter h in a certain sense, this represents for each value h this represents the likelihood corresponding to the unknown parameter h.

Right and we would like to use this Likelihood Function to find the estimate of the parameter h and therefore one natural framework to find an estimate of the parameter h is to find that value of h for which the likelihood is maximum and that is known as the Maximum Likelihood Estimate. And this is an important aspect of estimation or this is an important concept in estimation.

We have derived the likelihood function of the unknown parameter h. Now we would like to maximize this Likelihood Function and find out that value of h which maximizes this Likelihood Function and that is known as the Maximum Likelihood Estimate. So what we would like to do is first let us look at this likelihood function, right. What is this Likelihood Function?

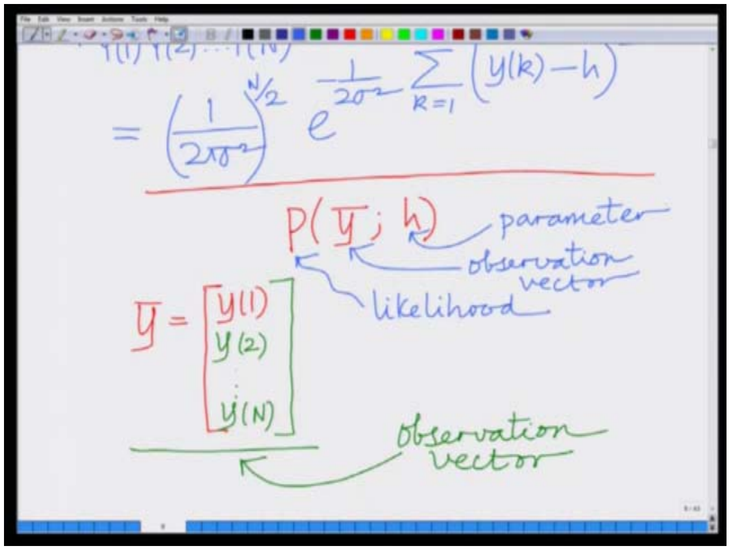
(Refer Slide Time: 18:30)



This Likelihood Function is your joint PDF that is f of y 1, y 2 up to y N y 1, y 2 up to y N which is equal to 1 over square root of 2 pie Sigma square e raise to - 1 over 2 Sigma. In fact, this is 1 over 2 pie Sigma square raise to the power of N by 2 e raise to - 1 over 2 Sigma square K equals 1 to N y K - h Whole Square. This is a function of the unknown parameter h.

Now what I am going to do, I am going to define this as the as the likelihood of the observations y bar parameterized by h where y bar is observation vector y bar is observation vector. Remember, we have multiple observations so y bar is the observation vector y 1, y 2 up to y N. So this is my basically my observation vector.

(Refer Slide Time: 19:40)

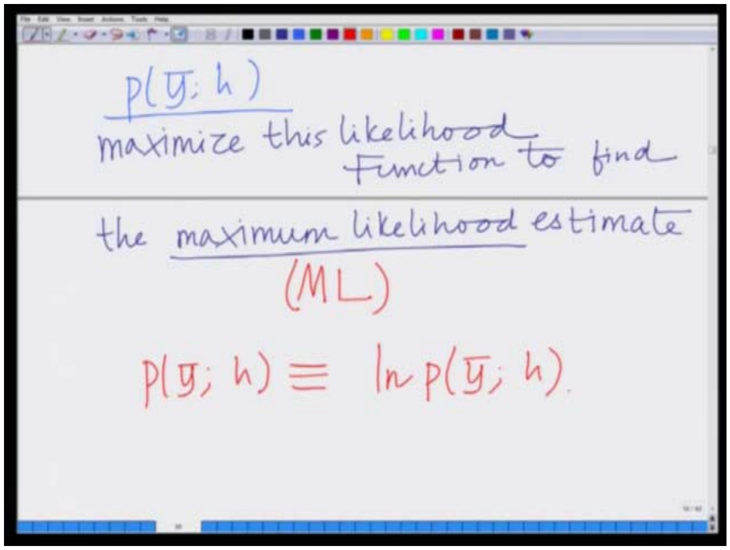


This contains all observation. This y bar is my observation vector and p of y bar h denotes this is the likelihood, what is this; this is the Likelihood Function of the observation vector and this is parameterized by h denotes, this is parameterized h, the terminology that is used is p. That is, we have this observation vector y bar which is of length N which contains observation y 1, y 2 up to y N.

And p of y bar semi colon h remember it is not a,, it is a semi-colon h which denotes or which signifies the Likelihood Function of the observation vector y bar parameterized by this unknown parameter h. That is, p of y bar h yeah, so this is the likelihood function of the observation vector y bar which is parameterized by h.

And now let us look at this thing, this like Likelihood Function; we have derived the expression for this Likelihood Function. Now as we said this basically, this is the Likelihood Function which corresponds to the likelihood of observation vector parameterized by h and now what we would like to do is we would like to maximize this Likelihood Function.

(Refer Slide Time: 21:24)



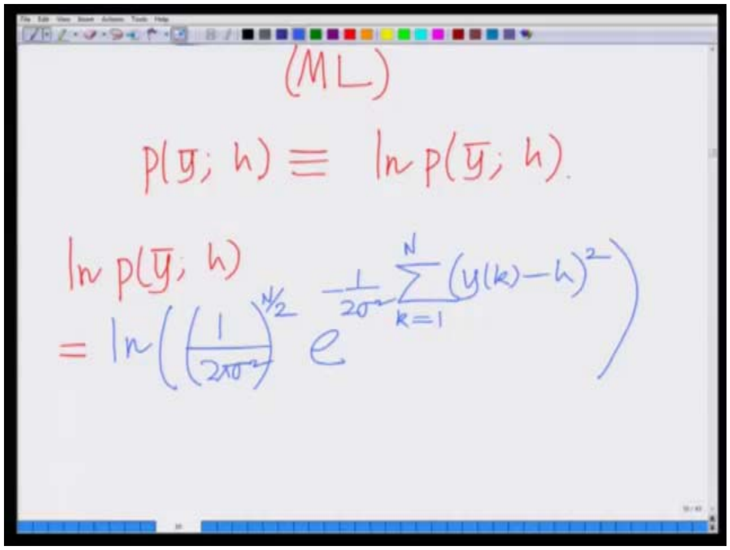
p of y bar y bar h, this is your Likelihood Function. What we would like to do is we would like to maximize this maximise this Likelihood Function to find the estimate which is termed as the Maximum Likelihood Estimate. In short, this is also denoted by m l. So we would like to maximise the Likelihood Function to find the Maximum Likelihood Estimate of the unknown parameter h.

This maximum likelihood, this concept of maximum likelihood is also abbreviated as m l which denotes the maximum likelihood. Now let us look at this, instead of maximizing p of y bar parameterized by h, I can also maximise that is, maximising this is equivalent to maximising the log of p of y bar parameterised by h. Because if the function maximum, the Logarithm is also maximum because the log is a monotonically increasing function.

We are talking about the natural logarithm. Let me denote this by l n that is log to the base e. When a function is maximum, its logarithm is also maximum. So function is maximise, instead of maximising the function, finding the point where the function is maximum, it is convenient in this problem since we have an exponential, we would like to find the parameter h for which the logarithm of the function maximum.

Now let us look at the logarithm of the joint density function.

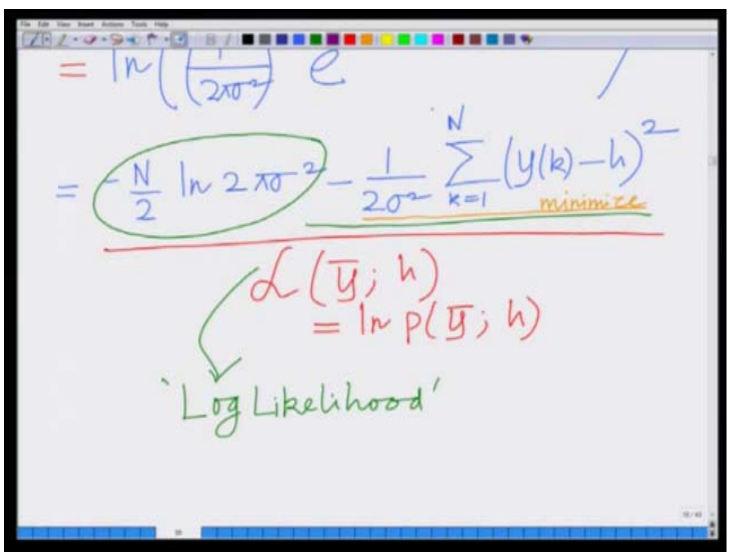
(Refer Slide Time: 23:30)



The logarithm of the [pro] of the likelihood function is basically your law logarithm of 1 over 2 pie Sigma square raise to the power of 2 over 2 e raise - 1 over 2 Sigma square submission K equal to 1 to N y K - h whole square.

And if I take the logarithm of this likelihood function, now can get, you can see

(Refer Slide Time: 24:13)



- N over 2 l n 2 pie Sigma square - 1 over 2 Sigma square submission K equals 1 to N of y K - h whole square and this is basically the log of the likelihood which is denoted by prescription l y bar parameterised by h. This is basically your log or the natural logarithm l n of p likelihood function p of observation vector y bar parameterised by h.

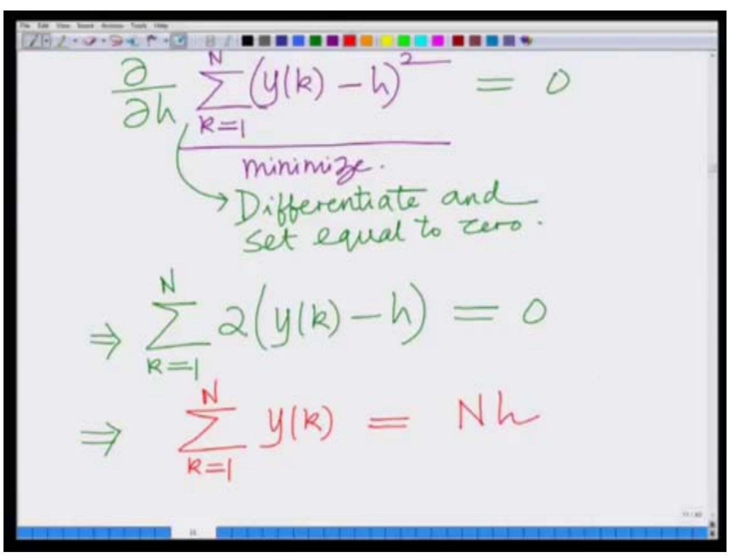
This is also termed as the log likelihood function. The logarithm of the Likelihood Function, this is also termed as the log, this is also termed as the log likelihood. And instead of maximising the likelihood, I can maximise the log likelihood of the observation vector y bar. And now we can observe something interesting, look at this part in the log likelihood, this is a constant.

So basically this is always a constant, so I did not worry about this. I would not worry about maximizing this, right; if the function is maximum, the function + constant is also a maximum. So I need to maximise this part, but this part has a negative sign, so I need to minimize this part, so I need to look at where I am, I mean this function has a negative sign, so to maximize the log likelihood function; I need to minimize this part.

That is, in order to maximise - 1 over 2 Sigma square submission K equal to 1 to N y K - h square. If I remove the negative sign, I can minimize the rest of the argument and now observe that this factor 1 over 2 Sigma square is a constant, so I all I am left with is minimising this component which is submission K equal to 1 to N y K - h square.

So what I am saying is that is equivalent this. Although this log likelihood function might seem complicated to begin with, it is equivalent to basically minimising this part. That is, minimise this part and now that minimisation is easier. I can simply differentiate this and said this equal to 0.

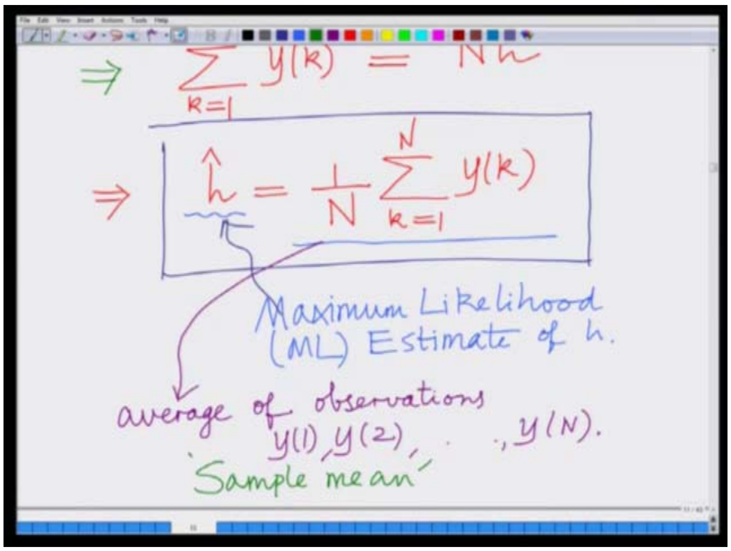
(Refer Slide Time: 26:50)



To minimise submission K equal to 1 to N y K - h square. To minimise this, differentiate and set equal to 0 and if I differentiate this, what am I going to get if I differentiate this with respect to h.

I am going to get submission K equal to 1 to N twice y K - h equal to 0 that is differentiating and setting equal to 0 imply submission K equal to 1 twice y K - h equal to 0 which basically implies that submission K equal to 1 to N of y K equal n times h which basically implies that this occurs at your h hat equals, what is the value of h that occurs?

(Refer Slide Time: 28:24)



This occurs at value of h equals, this occurs at value of h which is equal to 1 over N submission K equal to 1 to N y of K at this particular value of h is denoted by h hat. This is an estimate of the unknown parameter h that is the value of h at which the Likelihood Function is maximised is basically the estimate of the unknown parameter h.

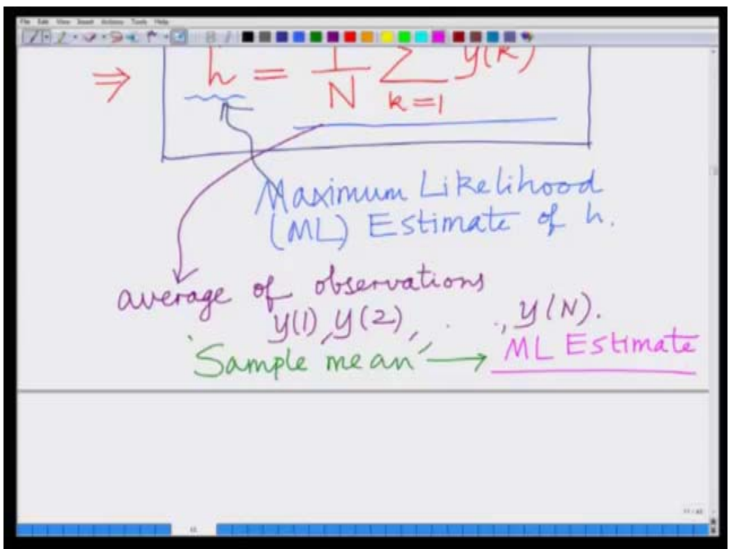
This is denoted by h hat, this is the Maximum Likelihood Estimate of h. So this is your h hat which we have calculated, this is the maximum likelihood of estimate; this is your Maximum Likelihood Estimate. Let me write this clearly, this is the maximum likelihood or ML estimate of h and one can also denote this by subscript ML to denote that this is the Maximum Likelihood Estimate.

So this h hat is the Maximum Likelihood Estimate and look at this, this is simply the average of the observations, what is this, this is the average, simple average or arithmetic mean of observations. Your observations y 1, y 2 up to y n. And this is also known as this average is also known as the sample mean. That is the Maximum Likelihood Estimate is given by the sample mean. What is the sample observation y 1, y 2, y N and the mean of this sample is 1 over N submission K equal to 1 to capital N y K.

This is the arithmetic mean or the sample mean of the observations and this is very intuitive, what we are saying is the Maximum Likelihood Estimate, they are Maximum Likelihood Estimate of the unknown parameter h is simply to take the average of the observations at the sensor load which is something very intuitive, but we have derived this analytically, rigorously and justified this that is we have formed the likelihood function and this value of h.

This is sample mean of the observations, maximises the likelihood function, therefore this is known as the Maximum Likelihood Estimate.

(Refer Slide Time: 31:05)



So this sample or turns out that in this case the sample mean is basically your ML, the sample mean is your ML estimate which is something very important which you have learnt in this module.

So what we have started in this module, we have extended our single observation model to a multiple observation or an observation vector model. We found out the joint Probability Density Function of these observations and is the function of the unknown parameter h, we said this is termed as a likelihood.

The value of h at which this likelihood is maximum which corresponds to the maximum likelihood of the parameter h is known as the Maximum Likelihood Estimate. And for this particular problem we have derived the Maximum Likelihood Estimate as h hat equals 1 over N submission K equal to 1 to N y K which is the sample mean.

And this is a very interesting aspect, this is a Maximum Likelihood Estimate which is a very interesting and very important and a very powerful estimation paradigm. This is the Maximum Likelihood Estimation, this example of sensor node making multiple measurements illustrates the the basics, the basics of this Maximum Likelihood Estimation principle.

And in the next module, in further module we will discover, we will explore this Maximum Likelihood Estimate further and discover other properties of this Maximum Likelihood Estimation procedure and the Maximum Likelihood Estimate that we have found. So let us stop here and we will continue in the subsequent model. Thank you very much.