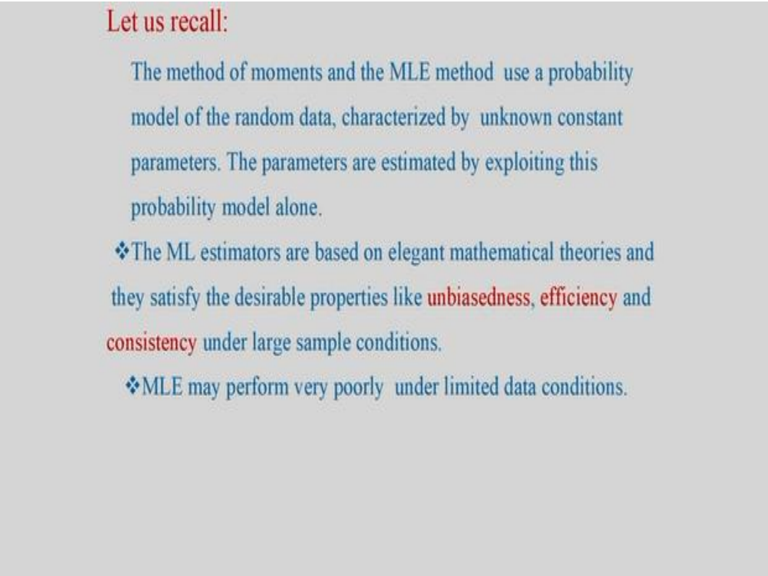
Statistical Signal Processing
Prof. Prabin Kumar Bora
Department of Electronics & Electrical Engineering
Indian Institute of Technology, Guwahati

Lecture 15
Bayesian Estimator

Hello students will come to the lecture, in this lecture we will introduce a powerful class of estimators known as Bayesian estimators

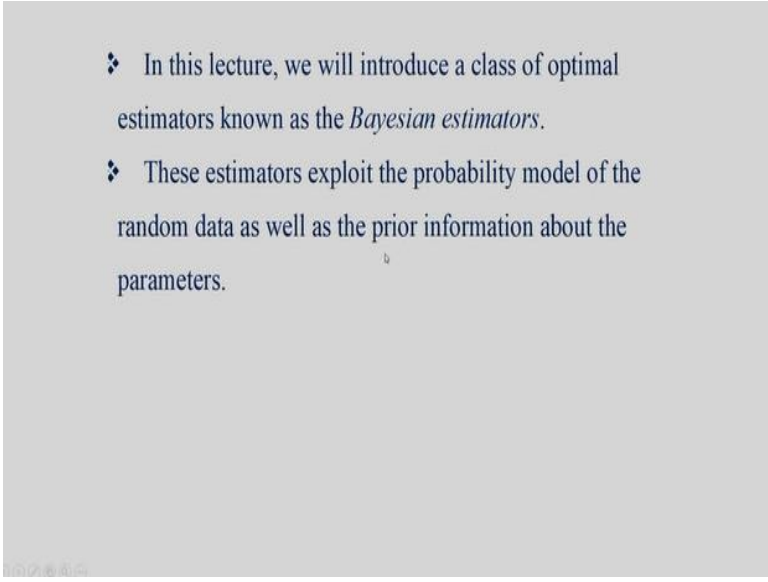
(Refer Slide Time: 00:49)



Let us recall; the method of moments and the maximum likelihood estimation method use a probability model of the random data, this model is characterized by unknown constant parameters. The parameters are estimated by exploiting this probability model alone. The ML estimators are based on the elegant mathematical theories and they satisfy the desirable properties like unbiasedness, efficiency and consistency under large sample conditions.

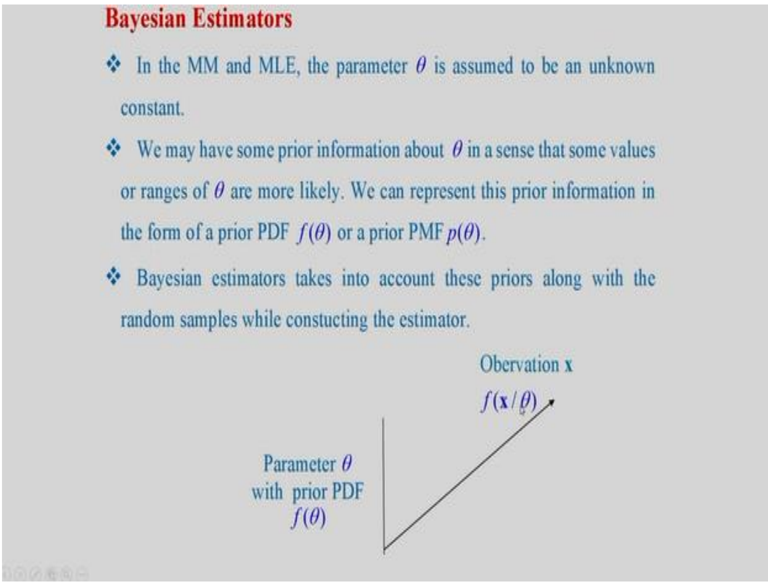
So to be a good estimators ML estimators required lot of data. MLE may perform very poorly under limited data a condition that is one drawback of MLE.

(Refer Slide Time: 01:47)



In this lecture we will introduce a class of optimal estimators known as the Bayesian estimators. These estimators exploit the probability model of the random data as well as the prior information about the parameters.

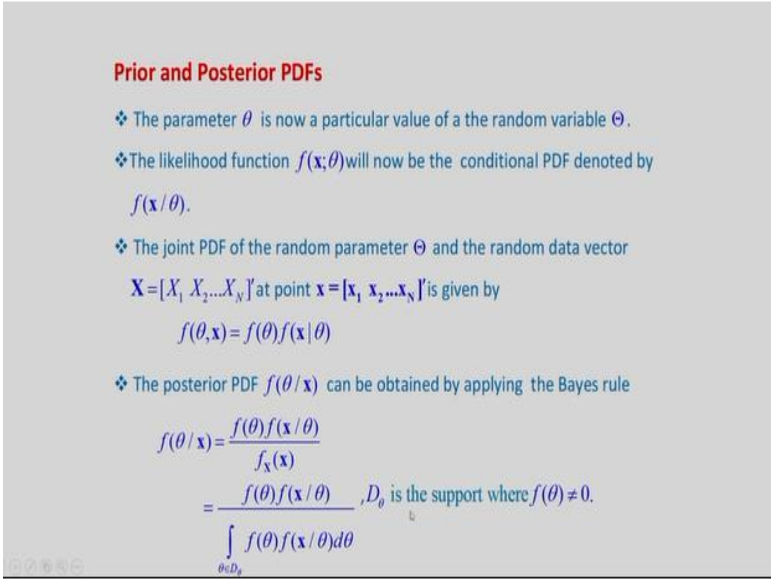
(Refer Slide Time: 02:10)



In the MM and MLE, the parameter theta is assumed to be an unknown constant. We may have some prior information about theta in the sense that some values or ranges of theta are more likely. We can represent this prior information in the form of a prior PDF prior density function f theta or a prior mass function p theta. So this is important, now whatever prior information about theta is available that is expressed in the form of a PDF or a PMF.

A Bayesian estimator takes into account of these priors along with the random samples while constructing the estimator. That is an important aspect. Now we will use the random data as well as the prior f theta or p theta. So that way parameter theta is now random and it has a prior PDF f theta and data is generated according to this model f of x given theta. Suppose here theta is there which is a random variable and for a given theta we will have this f of x given theta which will generate the observed data.

(Refer Slide Time: 03:39)



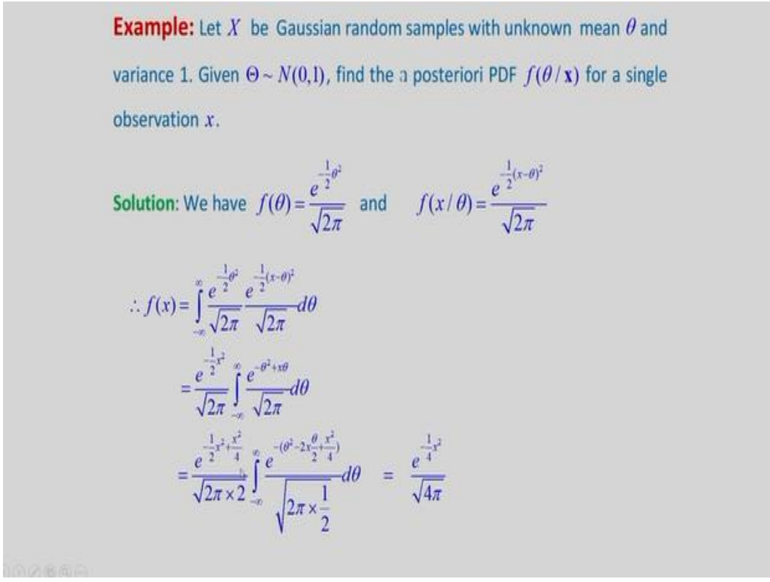
Let us explain what are prior and posterior PDFs; the parameter theta is now a particular value of random variable under this condition that theta is random therefore this theta whatever parameter we get that is a particular value of a random variable big theta. So big theta is the random variable and small theta is a particular value the likelihood function f x, theta.

Now with the conditional PDF denoted by f of x given theta earlier we used the likelihood function as a function of theta but now it will be a conditional PDF f of x given theta. the joint PDF of the random variable theta and the random data vector X this is the random data vector at the point small x vector that is x1, x2 up to xN transpose is given by this expression, this is the expression for a joint PDF f of theta N, x is equal to f theta multiplied by f of x given theta.

The posterior PDF f of theta given x can be obtained by applying the Bayes rule f of theta given x that will be equal to f of theta into f of x given theta that is designed PDF of data N x divided by f x of X, this is the marginal PDF of X. now this marginal PDF F x of X can be written in terms of the prior and the conditional PDF, that is integration of f theta into f of x given theta D theta over theta belonging to D theta where D theta is the support of f theta where f theta is not equal to 0.

So that way we have expressed the posterior PDF in terms of the prior PDFs and conditional PDF. So f of theta given X is given as f theta into f of x given theta divided by integration f theta f of x given theta d theta, theta belongs to D theta where D theta is the support of f theta.

(Refer Slide Time: 06:16)



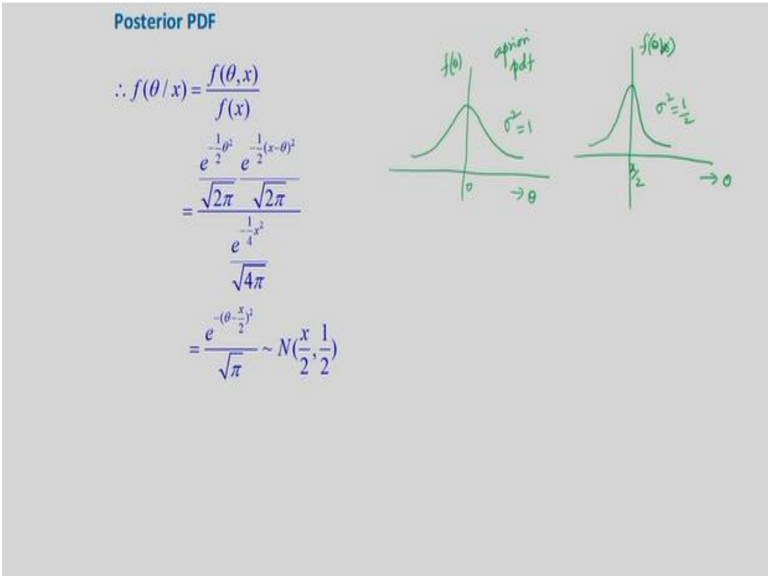
Let us consider one example; let X be Gaussian random samples with unknown mean theta and variance 1. Given theta distributed as normal 0, 1 find a posteriori PDF f theta given x for a single observation x. In this case only one observation is there and given that theta that parameter is distributed as a normal random variable with mean 0 and variance 1. So let us solve this problem we have to find the a posteriori PDF f of theta given x.

Now we have f theta this is normal with mean 0 and variance 1, so it will be f x, theta will be e to the power - half theta square divided by root over 2 pi and the likelihood function or conditional PDF is given by f of x given theta is equal to e to the power - half of x - theta whole square divided by root over 2 pi. So we have to find out the marginal PDF f x, so that will be a joint PDF and if we integrate with respect to theta.

So that way it will be integration from - infinity to infinity e to the power - half theta square divided by root 2 pi into e to the power - half of x - theta whole square divided by root 2 pi into d theta. So this integration now we will carry out so we are integrating with respect to theta so we can take out all the terms involving x only, so that way this e to the power - half x square will come out and then we can make it a whole square, so that it may be expressed as a Gaussian.

So that we have we write x square by 4 and subtract a square by 4 here, so that way this part will be a Gaussian now and integration of the Gaussian is equal to 1, therefore what will be left to it, e to the power - ¼ th of x square divided by root 4 pi.

(Refer Slide Time: 08:38)



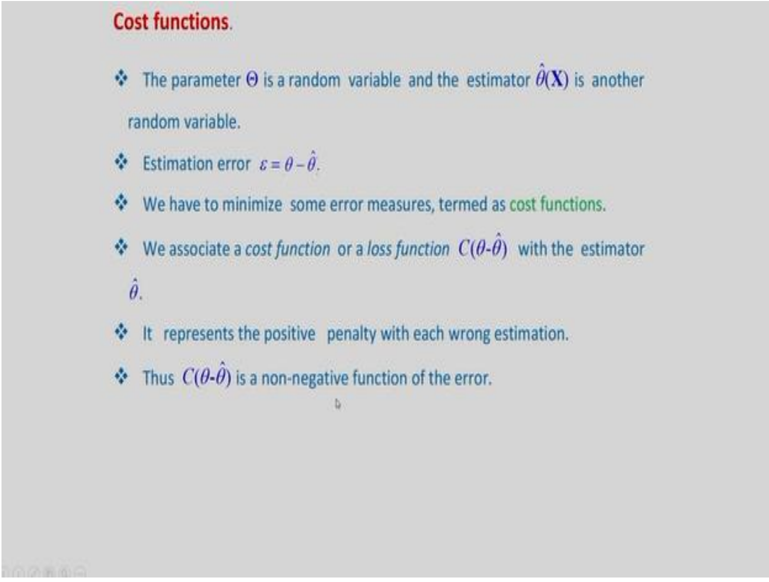
We have to find out the posterior PDF f of theta given x, it is given by f of theta x divided by f of x and if we substitute f of theta x already we have seen that e to the power - 1/2 of theta square into e to the power - half of x - theta square divided by 2 pi into root 2 pi and f of x will be equal to this quantity e to the power - x square by 4 divided by root over 4 pi

And this if we simplify, we will get e to the power - theta - x by 2 whole square divided by root pi, this is a normal distribution with mean x by 2 and variance ½. So what we have got that this posterior distribution is a normal distribution with mean x by 2 and variance half, earlier we have f theta that is a priori theta, f theta and this is 0 and earlier it was it like this normal distribution, this is the a priori PDF it has variants 1.

Now if we consider f of theta given x a posteriori PDF, so here theta now it will be a distribution this point is X by 2 and variance is now earlier variance was 1 now variance is half. So, that way what we see that given the information about the data the a posteriori PDF is now governed by the data the data will determine the a posteriori PDF and also we see that earlier whatever uncertainty was there that variant was 1.

Now it has billions has reduced, so this once we have the data the uncertainty about the parameter is reduced earlier this PDF was how its standard deviation Sigma squared it is equal to 1 and here Sigma square is equal to half. So uncertainty is reduced, as well as the mean is now shifted towards the data so that way from 0 we have x by 2 that way this distribution is now biased towards the actual data. So that way we have the idea of prior PDF and posteriori PDF.

(Refer Slide Time: 11:55)

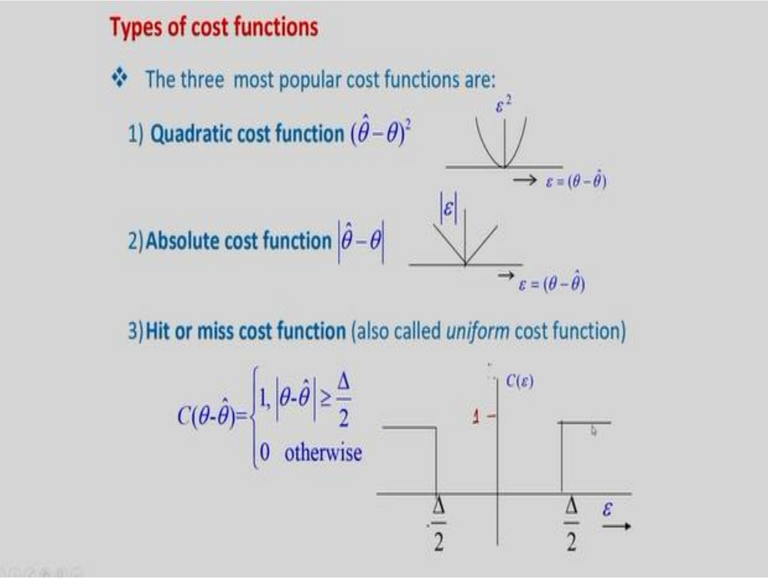


Let us define another term what is known as cost function. The parameter theta is a random variable and the estimator theta hat X is another random variable. So we have two random variable and estimation error at a particular point is given by theta hat - theta. In Bayesian framework, we have to minimize some error measures termed as cost function. These error measures are called cost functions.

We associate a cost function or a loss function of these forms C of theta - theta hat with the estimator theta hat. So if theta hat is the estimator this cost function is a function of the error theta is estimated value. So it is a function of error. It represents the positive penalty with each wrong estimation. So we have a this function will be a positive function if we commit more error than this function should be large.

Thus C of theta - theta hat is a non-negative function of the error theta - theta hat is the error, therefore C of theta - theta hat is a non-negative function of the error. So the cost function is a non-negative function of the error.

(Refer Slide Time: 13:22)

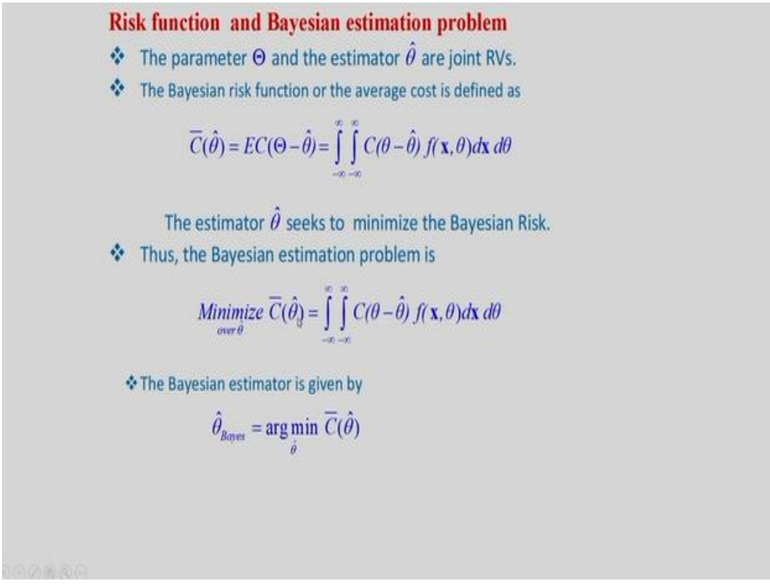


The three most popular cost functions are; number one quadratic cost function or it is also called square error cost function because this is the square of the error, squared error cost function. So if we plot it suppose this x is epsilon is equal to theta - theta hat error and this side is the cost function. So if we put epsilon squared that will be a parabola, this is the quadric or square error cost function.

Absolute cost function, instead of considering the square here we can take the absolute value of theta hat - theta, so then we get this cost function is plotted here actually it is a plot of mode of e. So that way there will be two lines like this. The third cost function is hit-or-miss cost function and it is also called the uniform cost function. it is given by C of theta - theta hat is equal to 1, that is uniformly 1 for absolute value of theta - theta hat to get our than equal to delta by 2 and 0 otherwise.

So this is the plot of the hit-or-miss cost function here, this delta is a small quantity. So it is a very small quantity within this error is and that is cost to the error is 0 and if we cross delta by 2 this side already said that error will be uniformly 1. So that is a very strict condition on the cost function

(Refer Slide Time: 15:07)

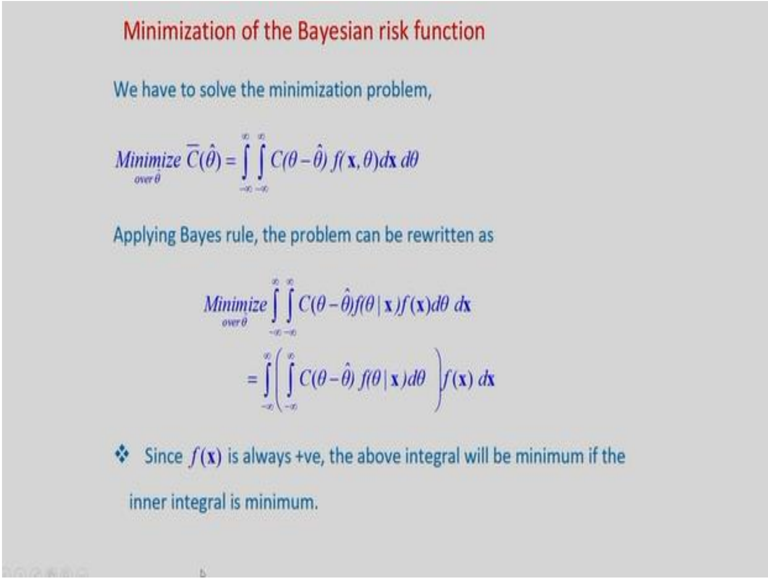


And now, we will define this function and the Bayesian estimation problem the parameter theta and the estimator theta hat are joint random variables. The Bayesian risk function or the average cost function is defined as this, actually we have defined the cost function expected value of the cost function is the least function that we denote by C bar theta hat it is a function of theta hat and it is given by expected value of C of theta - theta hat.

And if we expand it, it will be integration from - infinity to infinity - infinity to infinity C of theta - theta hat into f x, theta dx, d theta. The estimator theta hat seeks to minimize the Bayesian risk. So a Bayesian estimation problem is to minimize this risk function and thus the Bayesian estimator problem is given by minimize over theta hat C bar theta hat that is equal to integration from - infinity to infinity - infinity to infinity of C of theta - theta hat into f of x theta dx, d theta.

The Bayesian estimator is now given by, theta hat Bayes is equal to arg min theta hat of C bar theta hat. So the value of theta hat corresponding to the minimum is the base estimator.

(Refer Slide Time: 16:41)

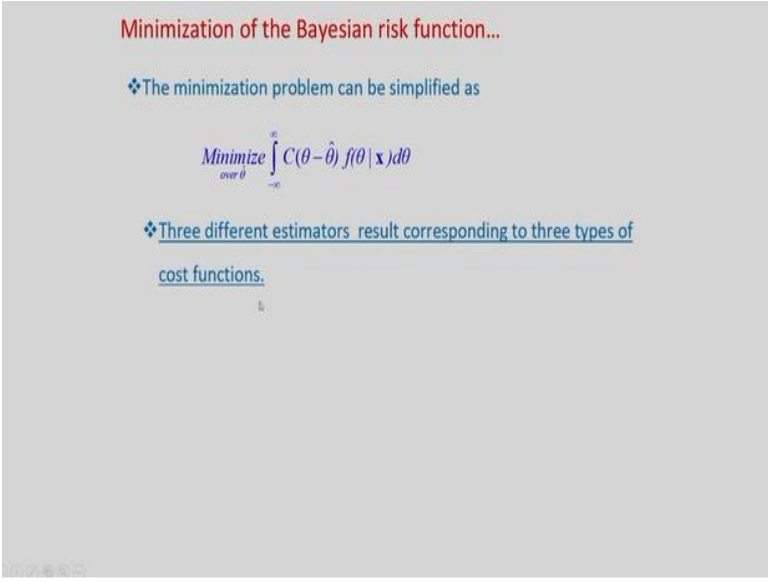


Now let us see, how to minimize the Bayesian risk function; we have to solve the minimization problem, that is minimize over theta hat C bar of theta hat is equal to integration - infinity to infinity, - infinity to infinity C of theta - theta hat into f of x theta dx, d theta. Now we will apply the Bayes rule and express this joint PDF as a product of the prior PDF and the conditional PDF.

So that way now minimizes some problem will become minimize over theta hat integration - infinity to infinity, - infinity to infinity of C theta - theta hat into f of theta given x that is the conditional PDF into f x d theta dx. Now we can express this integration as two integration like this one is involving theta only and other part involving x. So that way the inner integral involves theta, theta hat and x but the outer integral involves only x.

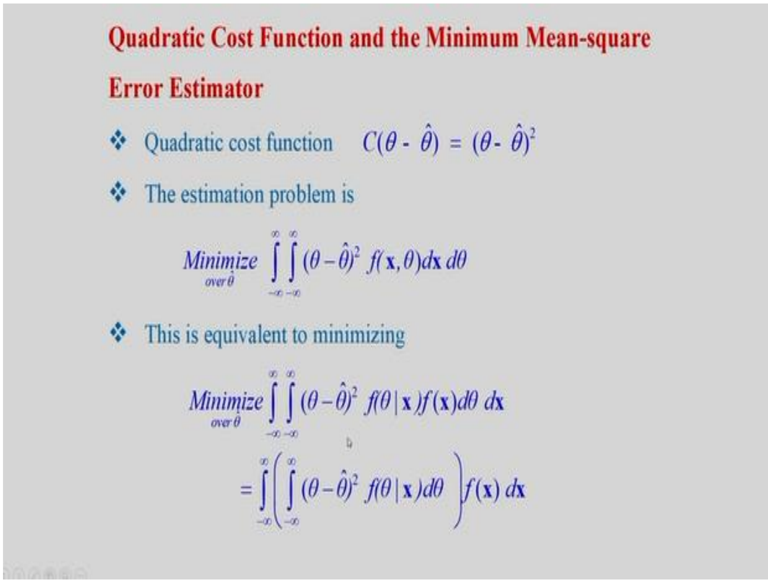
So that way this minimization problem we have written like this. Now, we know that this f x is always a positive quantity, therefore this entire double integral will be minimum whenever the inner integral is minimum. So since f x is always positive the above integral will be minimum, if the inner integral is minimum.

(Refer Slide Time: 18:30)



The minimization problem is now simplified and it can be written as minimize that inner integral basically over theta hat integration - infinity to infinity C of theta - theta hat into f of theta given x, d theta. Now three different estimators result corresponding to three types of cost function. We know that we have three cost function; squared error cost function, absolute error cost function and uniform error cost function.

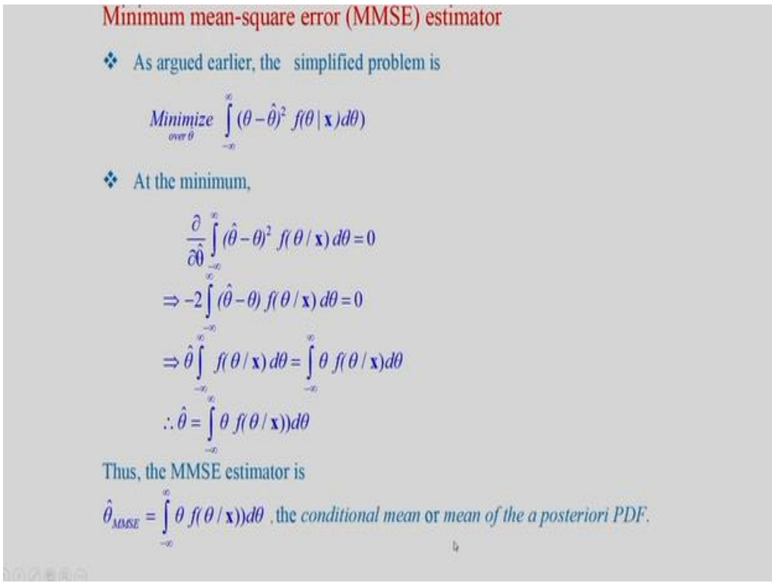
(Refer Slide Time: 19:03)



So first we will be considering quadratic cost function or square error cost function and the the corresponding minimum mean square error estimator. Quadratic cost function or square error cost function that is given by C of theta - theta hat is equal to theta - theta hat all square. The estimation problem is this minimize over theta hat integration - infinity to infinity - infinity to infinity of theta - theta hat whole square f of x theta dx, d theta

And then as I told this we will write in terms of the prior and conditional PDF, so this is the expression and as I did earlier so I will have an inner integral like this and outer integral involving f x like this.

(Refer Slide Time: 20:06)



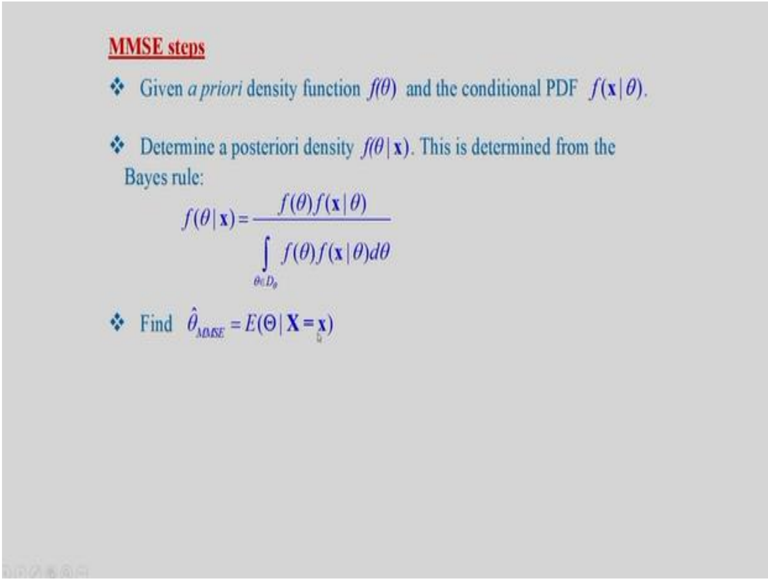
So this simplified problem is now minimize integral from - infinity to infinity of theta - theta hat whole square into f of theta given x, d theta. So this is the mean square error we are trying to minimize. At the minimum, now we will take the partial derivative with respect to theta hat del L theta hat of this expression must be equal to 0 and we can take this differentiation inside because these limits do not involve in any theta hat time.

Therefore we can take this partial differentiation inside and carrying out the partial differentiation we will get that - 2 times integral from - infinity to infinity of theta hat - theta into f of theta given x, d theta is equal to 0 and after simplifying we will get that corresponding to this term theta hat integration - infinity to infinity f of theta x, d theta is equal to integral - infinity to infinity theta f of theta given x, d theta.

Now this part I know because this is a PDF, if you integrate from - infinity to infinity you will get 1, so left side is simply theta hat and right side this quantity is the conditional expectation theta f of theta given x, d theta integral of from - infinity to infinity. So this is the conditional expectation, therefore theta hat will be equal to integration from - infinity to infinity of theta f theta given x, d theta, this is the conditional expectation.

So therefore what we get is that the MMSE, minimum mean square error estimator is given by theta hat MMSE is equal to integral - infinity to infinity of theta f of theta given x, d theta, this is the conditional mean or mean of the a posterior a PDF. Therefore we have to find out the a posteriori PDF f of theta given x and mean of this will be the theta hat MMSE.

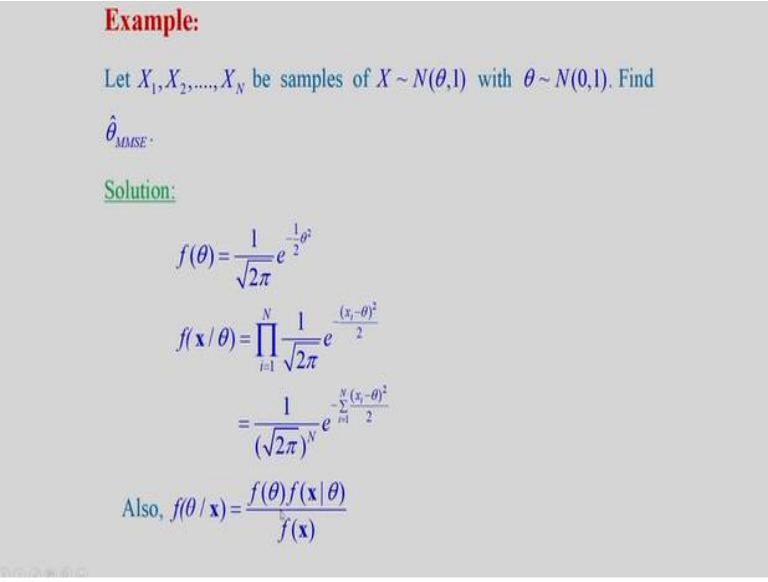
(Refer Slide Time: 22:39)



So therefore we can find out the MMSE, minimum mean square error estimate or in the following steps given a priori density function f theta and the conditional PDF of x given theta determined a posteriori PDF f of theta given x, this is determined using this Bayes rule; f of theta given x equal to f of theta, this is the prior PDF into f of x given theta that is the likelihood function divided by integral theta belonging to d theta, d theta is the support of f theta, f theta into f of x given theta d theta.

So this integral is the expression for f of x marginal density of x and once we have this conditional PDF, now we can find out the conditional mean that is theta hat MMSE will be equal to E of theta that is the parameter vector given that data vector X is equal to x, that is the observed data. So that way we can find out the conditional mean and that is d theta hat MMSE.

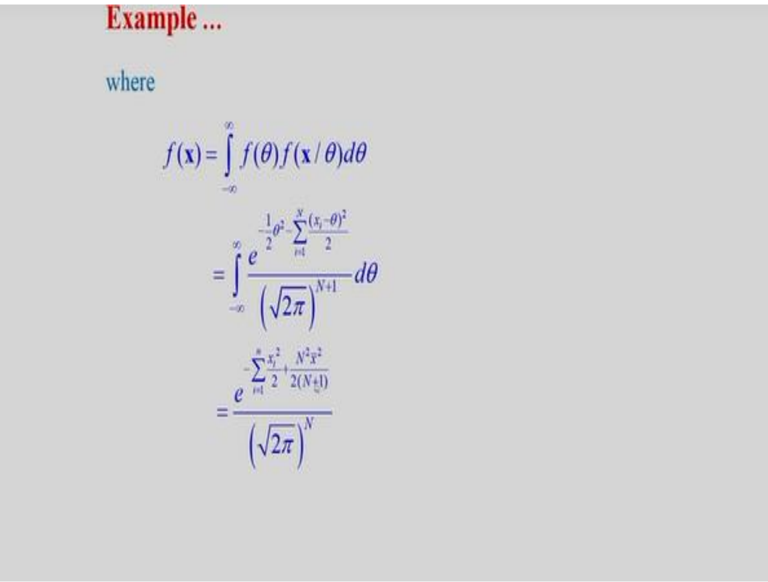
(Refer Slide Time: 23:59)



Let us consider one example; let x1, x2 up to xN be samples of normal distribution X is normally distributed with mean theta and variance 1 and theta is also a random quantity with distribution normal 0, 1. That is the PDF of theta, find theta hat MMSE. So let us solve this problem f theta is given by this is the normal distribution with mean 0 and variance 1 and conditional PDF f of x given theta there are n observations are there, therefore there iid.

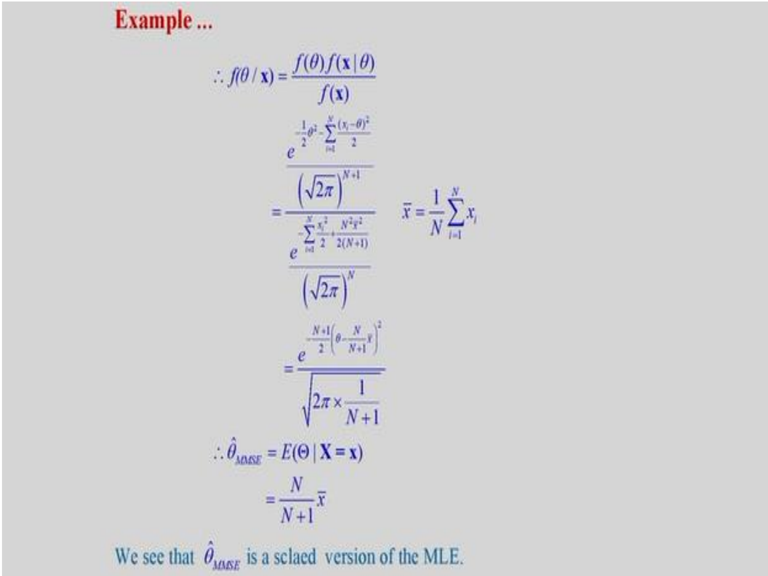
So it will be product i going from 1 to N of 1 by root over 2 pi e to the power - half of x i - theta whole square, okay? And this, if I carry out the product and then I will get in terms of summation like this, so it will be 1 by root over 2 pi to the power N into e to the power - 1/2 of x i - theta whole square i going from 1 to N. Now the a posteriori PDF f of theta given x is given by f theta into f of x given theta divided by f of x, f x is the marginal PDF.

(Refer Slide Time: 25:36)



So we can calculate f x like this that is join density integrated over theta, so we can put the expression for f theta and f of x given theta and then we are all carry out the integration and this will give us like this, e to the power - summation i going from 1 to N of x i squared divided by 2 + N into x bar square divided by 2 into N+1. So this we can simplify it like this,

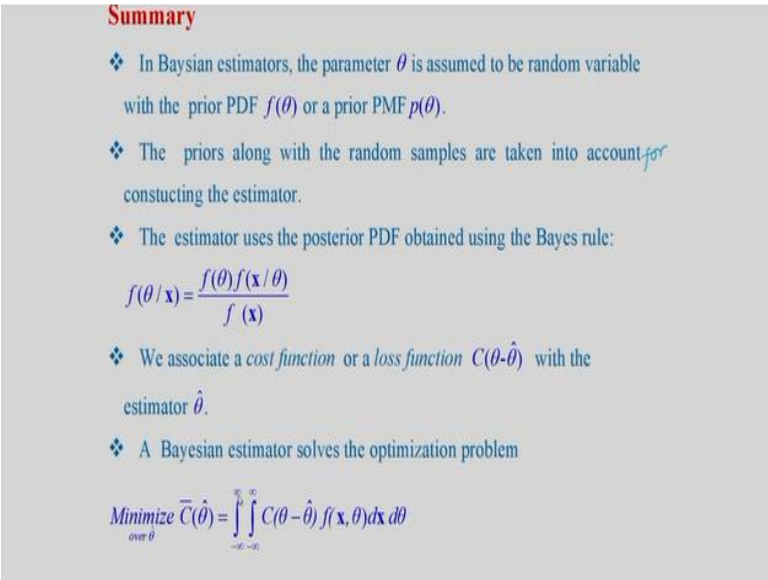
(Refer Slide Time: 26:19)



This x bar is and this quantity x bar is 1 by N into summation xi, i going from 1 to N this is the sample mean. Therefore the posterior PDF is given by f of theta given x is equal to f theta into f of x given theta divided by f x and we will substitute all this quantity now, we have the expression for f x also this is the expression for f x. So if we substitute and simplify we will get this expression, e to the power - half into N+ 1 into theta - N by N+ 1, x bar whole square.

So this is again a Gaussian mean is given by this therefore this conditional mean or theta hat MMSE will be equal to e of theta given x is equal to small x vector and that will come out to be this N divided by N+1 into x bar, where x bar is this sample mean this is the expression for sample mean. We see that theta hat MMSE is a scaled version of the MLE. So with the prior information about theta we are able to find out theta hat MMSE given by this N divided by N+1 into x bar, whereas what is this sample mean and we hope it would do better than MLE.

(Refer Slide Time: 27:51)

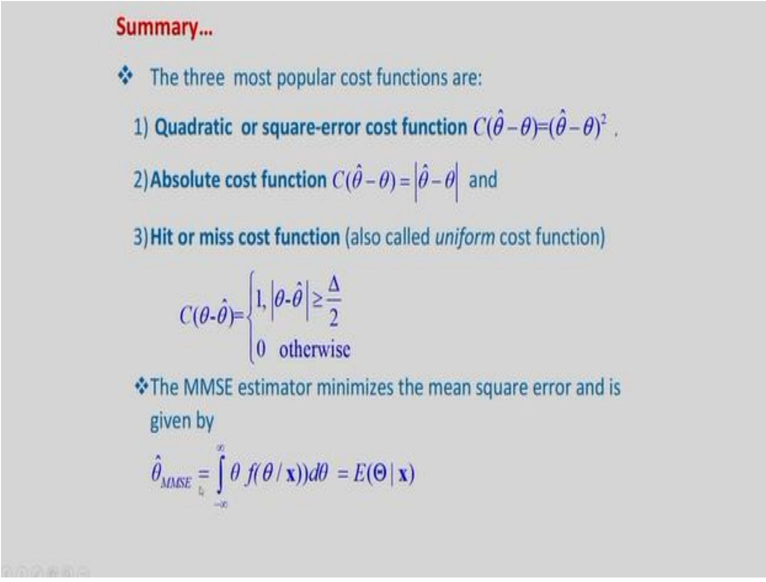


Let us summarize the lecture; in Bayesian estimators, the parameter theta is assumed to be random variable with the prior PDF f theta or a prior PMF P theta. The priors along with the random samples are taken into account for constructing the estimator. So we have to consider both the prior and model of this random samples. The estimator uses the posterior PDF obtained using the Bayes rule; that is f of theta given x equal to f of theta into f of x given theta divided by f x.

So this is the marginal PDF, this is the conditional PDF of x given theta and this is the prior PDF. In Bayesian estimator, we associate a cost function or a loss function C of theta - theta hat with the estimator theta hat. So this is a positive error measure and this cost function we want to minimize. A Bayesian estimator solve the optimization problem, this is the optimization problem here, minimize over theta hat C bar up theta hat that is the expected cost function or average cost function or least function

And that is equal to integral from - infinity to infinity - infinity to infinity C of theta - theta hat into f of x theta dx, d theta, this is the Bayesian estimation problem as an optimization problem.

(Refer Slide Time: 29:38)



The three most popular cost functions are; quadratic or square error cost function that is C of theta hat - theta is equal to theta hat - theta whole square. So this is this square error. Absolute cost function that is C of theta hat - theta will be equal to mode of theta hat - theta; we take the absolute value of error. Third cost function is hit or miss cost function also called uniform cost function.

And it is given by C of theta - theta hat is equal to 1 if mode of theta hat - theta that is greater than equal to delta by 2 and equal to 0 otherwise, where delta is a small positive number the MMSE estimator minimizes the mean square error. This mean square error is minimized and is given by theta hat MMSE is equal to this expression integral from - infinity to infinity theta f of theta given x, d theta and that is the conditional expectation of theta given x, this is the MMSE. Thank you.