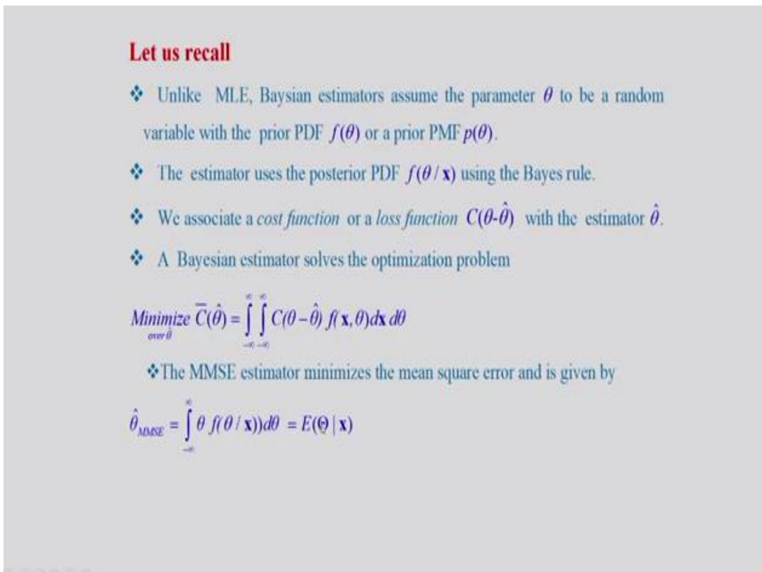
Statistical Signal Processing
Prof. Prabin Kumar Bora
Department of Electronics & Electrical Engineering
Indian Institute of Technology, Guwahati

Lecture No 16
Bayesian estimators 2

Hello students, welcome to the lecture on Bayesian estimators 2.

(Refer Slide Time: 00:45)

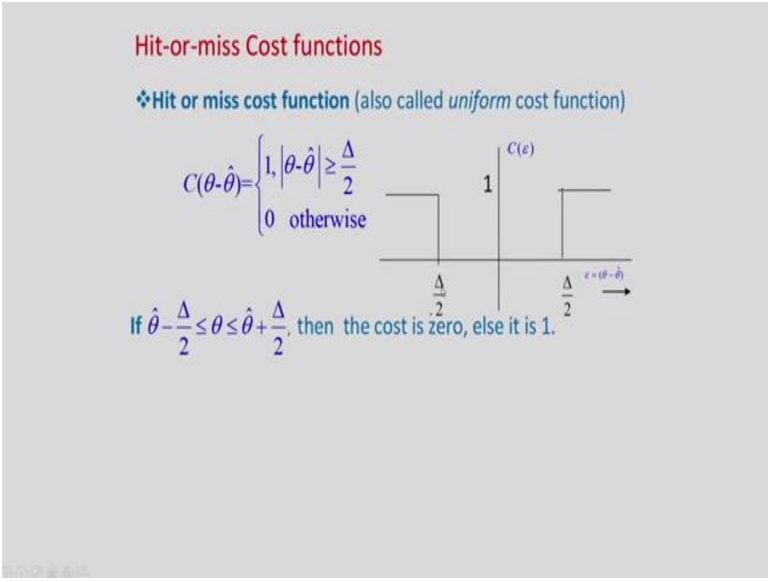


Let us recall unlike MLE Bayesian estimators assume the parameters theta to be a random variable with the prior PDF f theta or prior PMF p theta. The estimator uses the posterior PDF f of theta given X or posterior PMF p of theta given X using the Bayes rule. We associate a cost function or a loss function C of theta- theta hat with the estimator theta hat cost function is a function of error.

A Bayesian estimator solves the optimization problem, this is the optimization problem minimize C bar of theta hat over theta hat C bar of theta hat is equal to integration from minus infinity to infinity minus infinity to infinity C of theta minus theta hat into f of X theta dx d theta. This is the average cost function we want to minimize. The MMSE estimator minimum mean square error estimator minimizes the mean square error and is given by this expression theta hat remember C is equal to integration from minus infinity to infinity theta f of theta given X d theta.

That is the conditional expectation update and the variable theta given X so MMSE is given by the conditional expectation of theta given X. In this lecture I will focus on another Bayesian estimator known as the maximum a posteriori probability map estimator. I will also introduce another estimator known as the minimum absolute error MAE estimator.

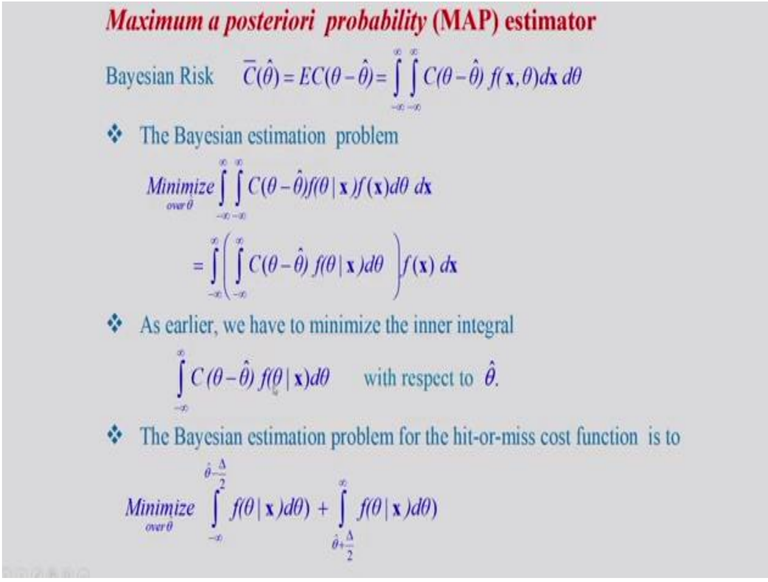
(Refer Slide Time: 02:40)



Let us start with hit or miss cost functions hit or miss cost function also known as uniform cost function is given by this G of theta - theta hat is equal 1, if the absolute value of theta - theta hat is greater than equal to Delta y2 is 0. Otherwise, so this is the plot of the hit or miss cost function if error is within plus minus Delta by 2 then cost is 0, otherwise cost this 1. In other words if parameter theta lies between theta hat minus Delta by 2.

And theta hat plus Delta by 2 then the cost is 0, else cost is 1 and here this Delta is very small quantity.

(Refer Slide Time: 03:34)

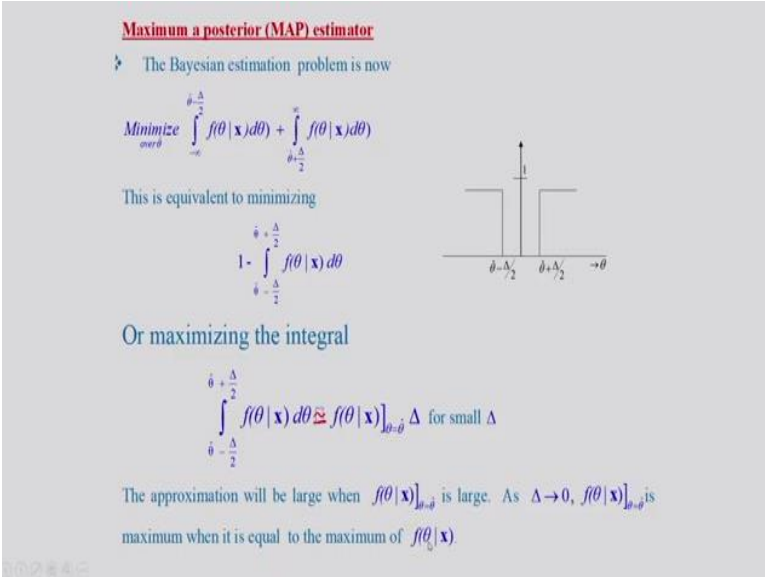


Now corresponding to this hit-or-miss cost function the Bayesian risk function is given by expected value of C theta minus theta hat, that is integration from minus infinity to infinity minus infinity to infinity of C of theta minus theta hat into f of X theta dx D theta. And now as earlier we will take this fx dx outside, so we will get the minimization problem at the minimization of this integral.

Integral from minus infinity to infinity integral from minus infinity to infinity C of theta minus theta hat into F of theta given X D theta into fx dx, so we have one inner integral and one outer integral. Since fx is always greater than equal to 0 therefore the double integral will be minimized, whenever the inner integral is minimized. So therefore our minimization problem is to find a minimum of this integral integral from minus infinity to infinity C of theta minus theta hat into f of theta given X D theta with respect to theta head.

Now using this cost function we have the cost function value at this interval is zero therefore we can write the minimizes some problem as minimize over theta hat integral from minus infinity to theta hat minus Delta by 2, f of theta given X D theta plus integral from theta hat plus Delta by 2 to infinity of f of theta given X D theta.

(Refer Slide Time: 05:38)

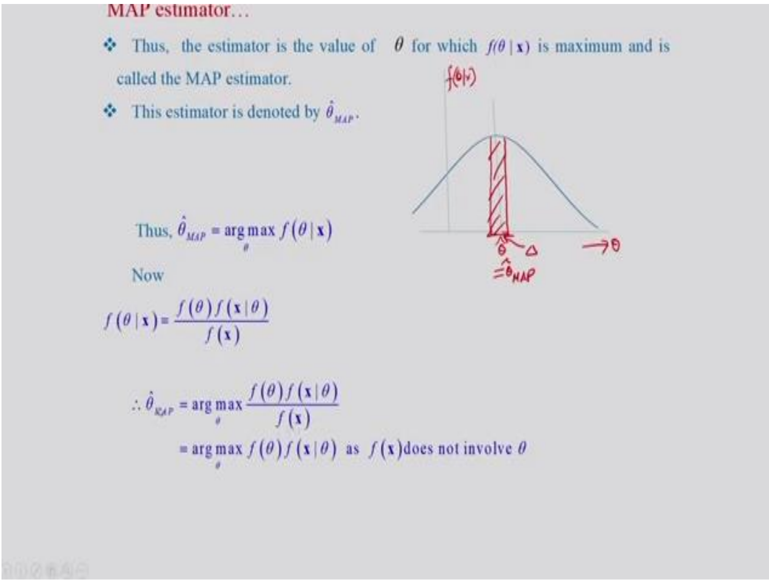


Thus the Bayesian estimation problem is now minimize over theta hat integral from minus infinity to theta hat minus Delta by 2 F of theta given X D theta plus integral from theta hat plus Delta by 2 to infinity f of theta given X D theta. Now this integral is equivalent to this expression because if we integral suppose this expression from minus infinity to infinity, will get 1, because this is a density function.

Therefore, the this minimization problem can be expressed as the equivalent minimization problem of this integral 1 minus integral theta hat minus Delta by 2 2 theta hat plus Delta by 2 F of theta given X D theta. So now we have to minimize this part since this is a difference, so this minimum will be same as maximizing the integral this integral if we maximize then this difference will be minimum therefore we have to maximize the integral theta hat minus Delta Y to 2 theta hat plus Delta by 2 F of theta given X D theta.

And this we can approximate as F of theta given X at point ax equal to theta hat into Delta for small Delta, because our Delta is very small therefore this integral we can approximate by this expression. Now consider this approximation Delta is a constant therefore this approximation will be large whenever this quantity is large and as delta tends to 0 f of theta given X at theta is equal to theta hat is maximum when it is equal to the maximum of F of theta given X.

(Refer Slide Time: 07:55)

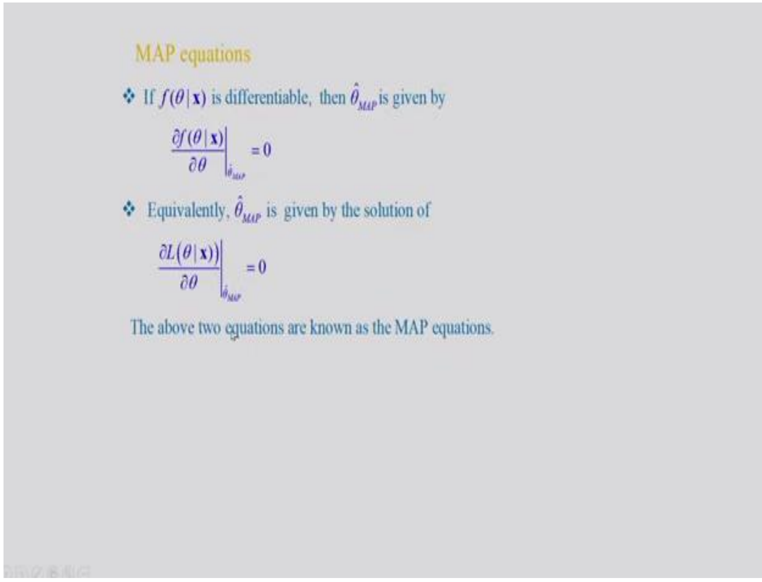


We will illustrate with this diagram, suppose this is my theta and this is my F of theta given X. And now this density function has a maximum at this point now if I put my theta hat here theta hat then this area for this area this interval is Delta so this area will be maximum if I put theta hat at this point corresponding to this maximum and therefore the estimator is the value of theta for which F of theta given X is maximum, and this called the MAP estimator.

Maximum is aposterior probability estimator, we will denote this estimator by theta hat MAP, because this is corresponding to the maximum a-posteriori PDF. So this point is theta hat is equal to theta hat MAP. So this point is equal to theta hat map the theta hat map is argument over theta of max f of theta given X. Now f of theta given X that is equal to f theta into f of X given theta divided by F X.

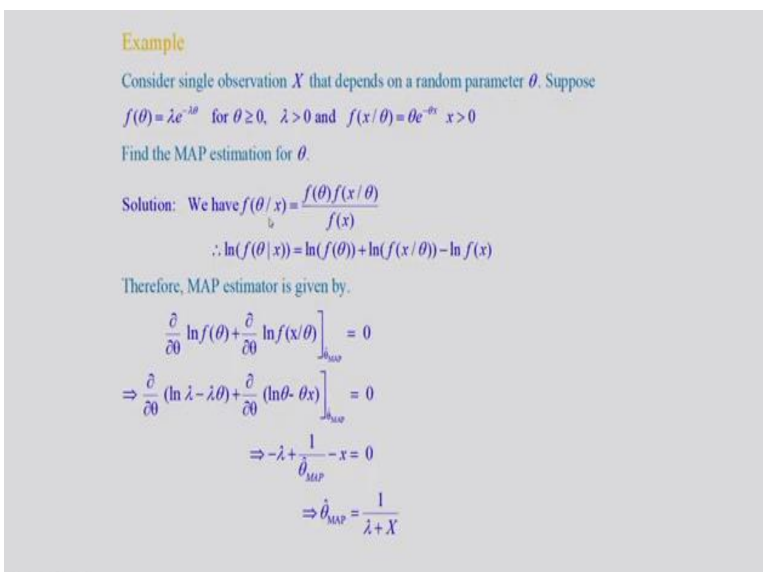
Therefore theta hat map will be arguing theta of max f theta into f of of X given theta divided by F X. Now this FS does not involve any theta they are prototype map is argmax theta of f theta into f of x given theta.

(Refer Slide Time: 09:49)



Now if F of theta given X is differentiable, the conditional PDF or a posterior PDF is differentiable then theta at map will be given by del F del theta at theta at map is equal to 0. So partial derivative of this a posterior PDF and theta hat map is equal to 0 or in terms of the log likelihood function we can write del L del theta at theta hat map is equal to 0, either we can solve this equation or this equation. The above two equations are known as the map equations.

(Refer Slide Time: 10:40)

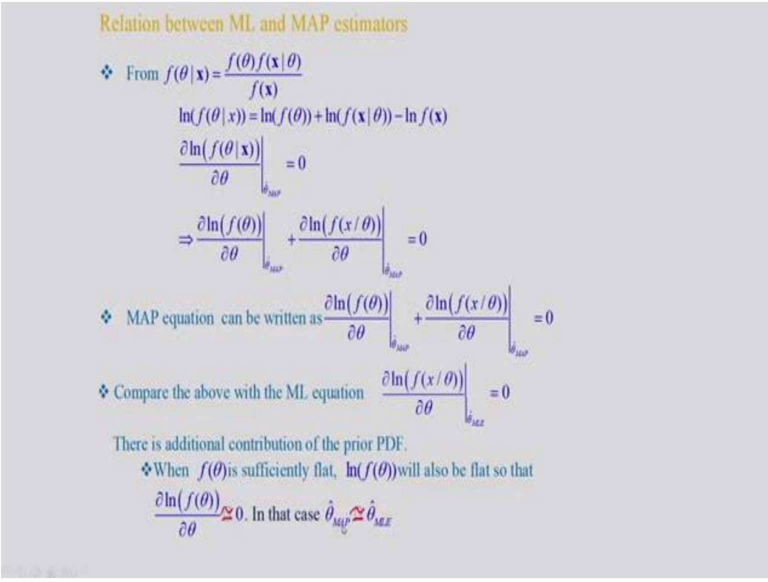


Example consider single observation X that depends on a random parameter theta, suppose F theta is equal to lambda into e to the power minus lambda theta for theta greater than equal to 0 and lambda greater than zero and the likelihood f x given theta is equal to theta into e to the power minus theta X, X greater than 0. find the map estimator for theta here lambda is assumed to be known we have the a posteriori PDF f theta given X that is equal to F theta into f of X given theta divided by F X.

And taking the logarithm log of F theta given X equal to log of F theta plus log of F x given theta minus log of F X so therefore map estimator is given by as Toller del del theta of log of f theta plus del L theta of log of x given theta at theta hat map must be equal to 0. So now we have to take the partial derivative or be given to both PDF is given here, so we can find out log of f theta. Similarly we can find out log of this expression.

And if we take the partial derivative we can write del del theta of log lambda minus minus lambda theta plus del Del theta of log theta minus theta X theta hat map is equal to 0. So we will take the partial derivative and then put theta is equal to theta hat map we will get minus lambda plus 1 y Tita at map minus X is equal to 0. So from this we can find out theta hat map and the estimator is given by 1 by lambda plus X. So this is this maximum a posteriori estimator.

(Refer Slide Time: 12:55)



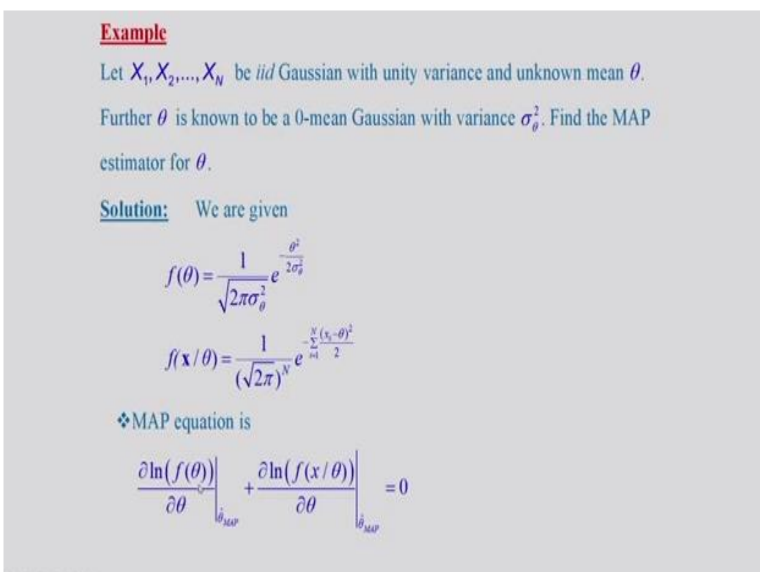
Now let us see the relation between ML and MAP estimators from f of theta given x equal to f theta into f of x given theta divided by f x, we can write log of F theta given X is equal to log of F theta plus log of F x given theta minus log of F X. Now this partial derivative of this with respect to theta is 0 that will imply that log of f theta del theta log of f x given theta del theta at theta hat map must be equal to 0.

So therefore we have this MAP equation del theta Delta of log of f theta plus del Del theta of log of fx given theta at theta at map must be equal to 0. And if we examine the ML equation this is simply del Del theta log likelihood function f theta MLE is equal to 0, so we see that this expression contains this term but it has one additional term. There is additional contribution from the prior PDF.

This is the contribution from the prior PDF when F theta is sufficiently flat in that case log of f theta will be also flat so that this quantity del Del theta log of F theta will be approximately equal to 0. So this will be approximately 0, so that theta hat map is approximately will be equal to theta hat MLE. So when f theta is sufficiently flat that means it's slope is very small in that case log of F theta will be also plate.

So that it's slope will be also very small nearly zero in that case theta hat map will be equal to theta hat MLE. That means if the parameter has more uncertainty in that case theta hat map will be closer to theta hat MLE.

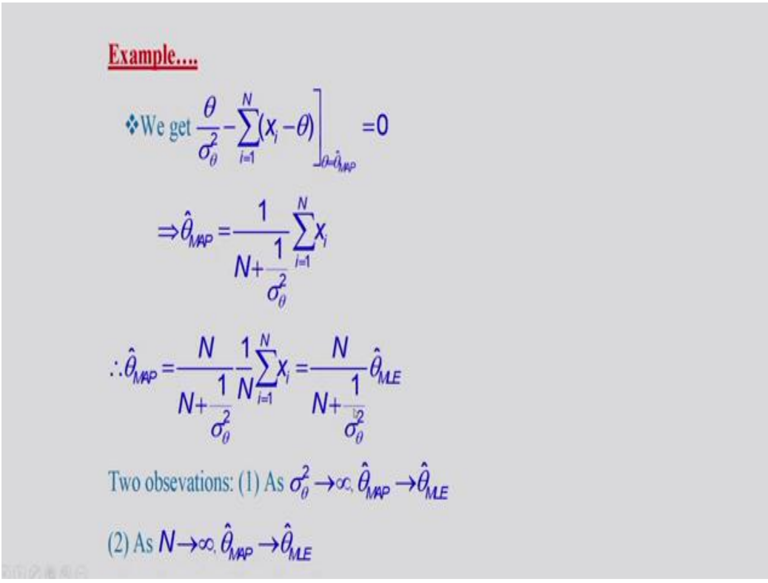
(Refer Slide Time: 15:17)



We will consider one example let X1,X2,...xn be iid Gaussian with unit variance and unknown min theta further theta is known to be a zero mean Gaussian with a variance sigma theta squared this is the prior PDF that is Gaussian with variance sigma theta square, find the map estimator for theta? We are given f theta this is the PDF prior PDF 1 by root over 2 pi sigma theta square into e to the power minus theta square divided by 2 Sigma Theta square.

So this is the 0 mean Gaussian with variance sigma theta square and the likelihood function f of x given theta that is also Gaussian with unknown min theta, so that way f of X given theta is equal to one world over two pi to the power n into e to the power minus summation, I going from 1 to N of X I minus theta whole square divided by 2. This is the likelihood function of random data. Now we will consider the map equation del Delta of log F theta plus del del theta log fx given theta at theta hat equal to theta hat map must be equal to 0.

(Refer Slide Time: 16:52)

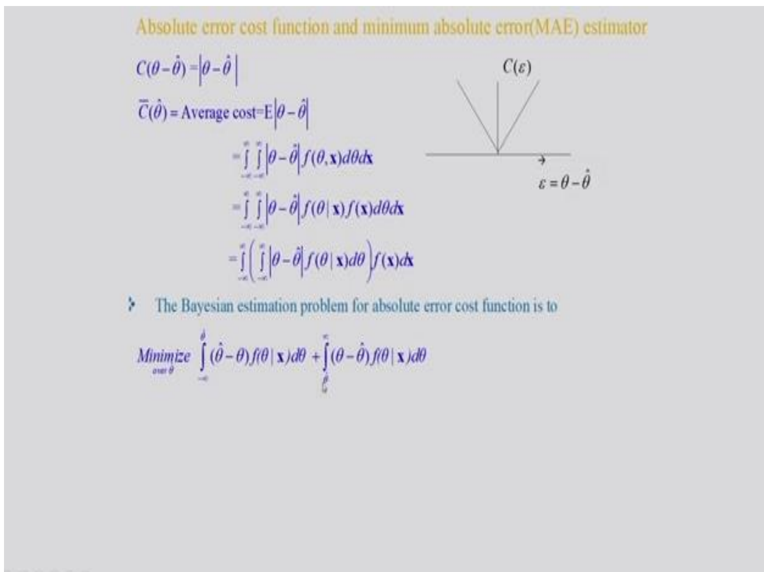


And this will give us this relationship theta by sigma theta square minus summation X I minus theta I going from 1 to N F when theta equal to theta hat map is equal to 0. So this will give theta hat map is equal to 1 by n plus 1 by sigma theta square into summation X i i going from 1 to n this is the expression for theta hat map. And now we can manipulate this expression if we divide by n then multiplied by n therefore n divided by n plus 1 by sigma theta squared into 1 by n into summation X i i going from 1 to n.

Now this quantity is theta hat MLE, so therefore theta hat map will be equal to n divided by n plus 1 by sigma theta square into theta hat MLE, it is a scaled version of theta hat MLE. Two observations are made number 1 as sigma theta square goes to infinity then this quantity will become 0 and n by n will be equal to 1, so in the case Teta hat map will approach to theta hat MLE.

If sigma theta square that is the variance of the prior PDF is very large in that case theta hat map will be approximately theta hat MLE. Now next is if number of data point is large, so as n tends to infinity, so in that case this ratio will approach 1 therefore theta hat map will be again I proceed to theta hat MLE. So as n becomes lost did I add map will tend to theta hat MLE and as sigma theta square become large then also theta hat map will tend to theta hat MLE.

(Refer Slide Time: 19:01)



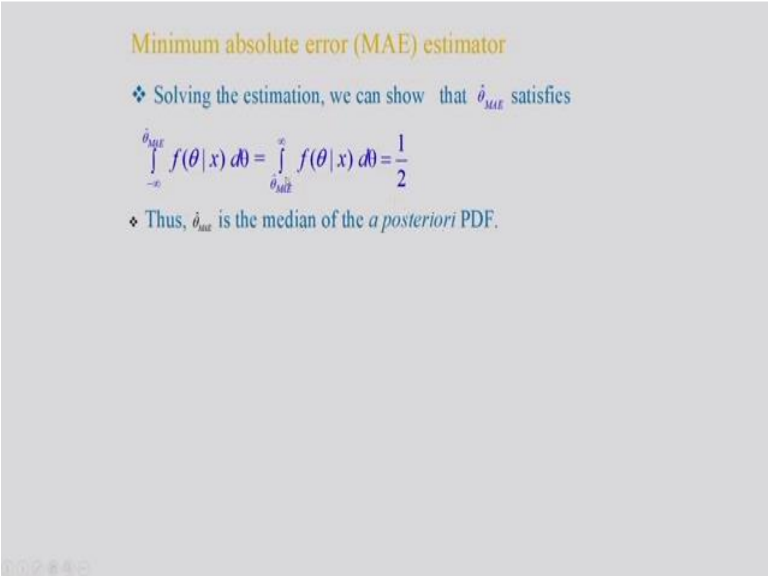
And now we will briefly introduce absolute error cost function and minimum absolute error MAE estimator. C of theta minus theta hat is equal to absolute value of theta minus theta hat this is the cost function, so as a function of error this is given by and positive a pair already positive this is a positive slope and if at all is negative then this is slope is negative. so average cost now C bar up theta hat, that is the Bayesian risk or ever average cost function.

That is equal to E of absolute value of theta minus theta hat and that is double integration from minus infinity to infinity minus infinity to infinity mod of theta minus theta hat into F of theta , X D theta D X. This is the joint PDF between theta and X. Now writing in terms of prior and conditional PDF we write this double integral as double integral of mod of theta minus theta hat into F of theta given X into F X D theta dx.

As earlier we can now take fx dx outside so that we have now two integral one outer integral and one inner integral and this quantity is always greater than equal to 0, there fore the minimum can be obtained by minimizing this expression only. Thus the Bayesian estimation problem for absolute error cost function is to minimize the word theta hat integration from minus infinity to theta hat theta hat minus theta f of theta given X D theta this is one part.

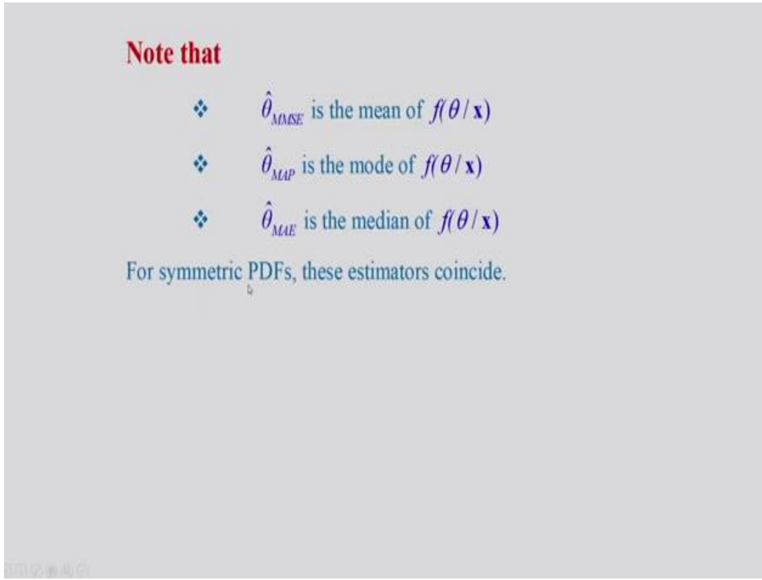
And the other part is theta hat to infinity theta minus theta hat into f of theta given s D theta in this interval theta hat to infinity D theta minus theta hat will be a positive quantity into F of theta given X D theta, so this is the integration and other side theta minus theta hat will be negative therefore we write here as theta hat minus theta into f of theta given X D theta. So that way we have integration from minus infinity to theta hat and theta hat to infinity.

(Refer Slide Time: 21:45)



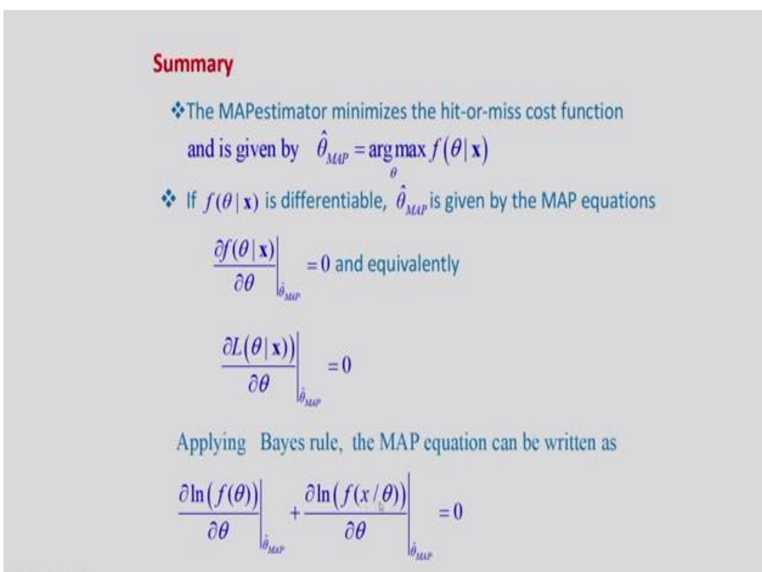
Solving the estimation problem we can show that theta hat MAE satisfy this equation that is integration from minus infinity to theta hat ma of f of theta given X D theta is equal to integration from theta hat ma to infinity of f of theta given X D theta and this is equal to ½. Thus theta MAE correspond to the median of the a-posteriori PDF. So we have to find out the median of the a posteriori PDF this is the approach to a PDF it's median value is theta hat MAE.

(Refer Slide Time: 22:32)



Now we know that theta hat MMSE is the conditional mean it is mean of f of theta given X to that map is the mode of f of theta given X, and finally theta hat ma is the median of f of theta given X. So MMSE is given by the mean of the a-posteriori PDF theta hat map is given by the mode of a posteriori PDF and that MAE is the median of a-posteriori PDF. for symmetric PDFs these estimators will coincide portion because porches has PDFs mean more than median coincide.

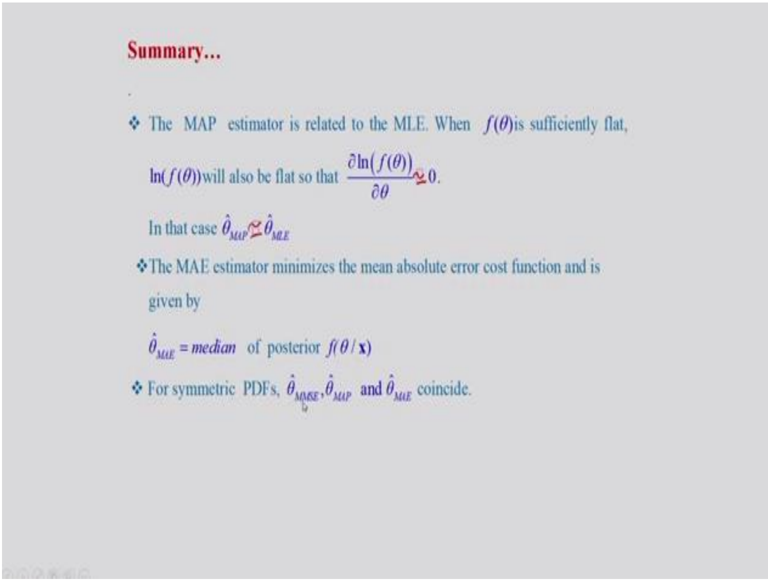
(Refer Slide Time: 23:21)



Let us summarize the map estimator minimizes the hit or miss cost function and is given by theta hat map is equal to argmax theta of f of theta given X. So a-posteriori PDF we have to consider and the mode of the a-posteriori distribution will give to that map theta hat map is the mode of the a posterity PDF. If f of theta given X is differentiable then theta hat map is given by the map equations that is either in terms of the PDF del Del theta of F theta given X map is equal to 0.

Or in terms of the log-likelihood function, del Del theta of L theta given X theta hat map is equal to 0. Applying Bayes rule the map equations can be written in this form del Del theta of log of f theta at theta ab plus del Del theta of log of f x given theta at theta at web must be equal to 0.

(Refer Slide Time: 24:39)



The map estimator is related to the MLE, when F theta is sufficiently flat log of F theta will also will flat, so that the partial derivative of the log likelihood function with respect to theta will be approximately equal to 0. So in that case theta hat map will be equal to theta hat MLE the minimum absolute error estimator MAE estimator minimizes this the mean absolute error cost function and is given by theta hat MAE is equal to median of the posterior PDF f of theta given X. For symmetric PDFs theta hat MMSC theta hat map and theta MAE coincide thank you.