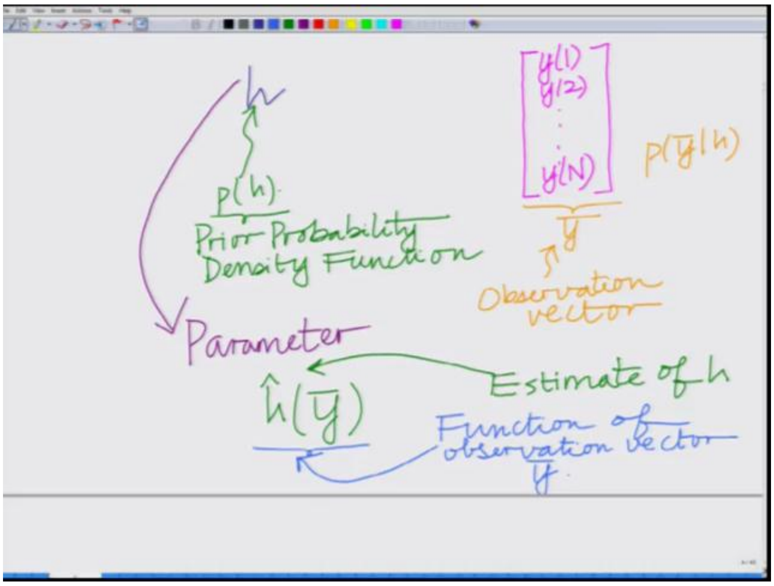
Bayesian/MMSE Estimation for MIMO/OFDM Wireless Communications
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture - 02
Optimal Bayesian Minimum Mean Squared Error (MMSE) Estimate

Hello. Welcome to another module, in this massive open online course in Bayesian, minimum mean squared error estimation, for wireless network. So, yesterday we motivated, or in the previous module, we basically motivated the need for estimation of a parameter h, more specifically a Bayesian parameter h.

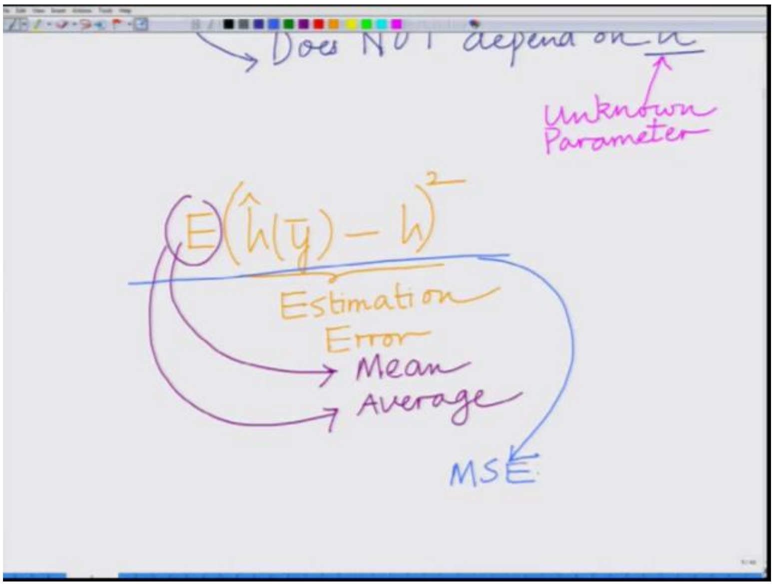
(Refer Slide Time: 00:33)



By Bayesian, we mean, a parameter about which there is some prior information that is available, in terms of the probability density function p of h correct? This is the prior probability density function. This is the prior probability density function. Also we have a set of measurements, or observations, which is also called the Observation Vector. We have a set of n observations. We can represent them, using the Vector y1, y2 up to yn. This is y bar, which is, we also said, this is our observation, this is our Observation Vector, all right? And we also have the conditional probability, density of the Observation Vector py bar, given the parameter h, remember we also termed this h; this quantity h is also termed as a parameter, which has to estimated, all right?

So, this parameter h has to be estimated, given the observations y1, y2, yn, the n observations, or the Observation Vector y bar, all right? And now, let us denote this estimate of h, by h hat, all right? So, h hat, which is estimate of h, is naturally going to be a function of the observation, Vector, all right? So, h hat, this is the standard notation used to denote the estimate, h hat, is a function of the Observation Vector. So, h hat is the estimate, estimate of the parameter h, and observe that, and it is important to realize that, although it is a simple point, h hat is a function of the Observation Vector. This is a function of the Observation Vector, y bar. And another important to be realized, although it is also very simple point, is that, the estimate h hat, can only be a function of the Observation Vector y bar, and not a function of the unknown parameter h, because h, the underline parameter itself is unknown, and that is a reason we are trying estimate h.

(Refer Slide Time: 03:37)



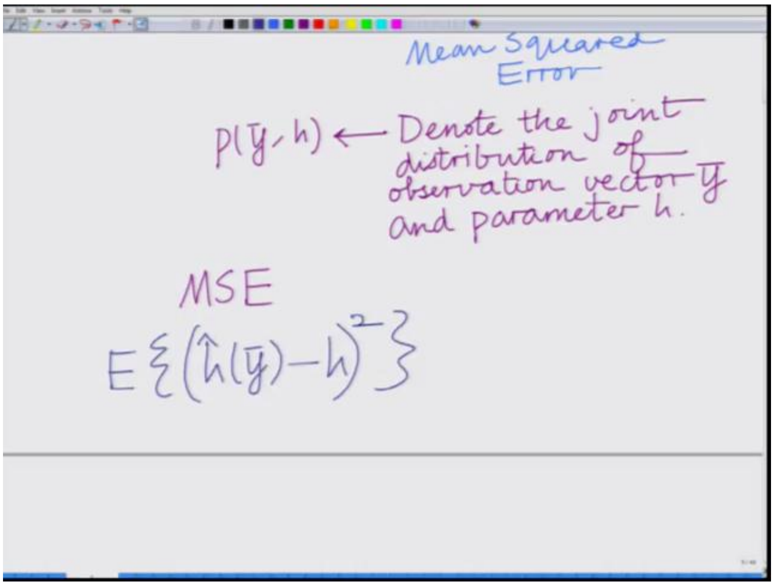
Therefore, h hat cannot depend on the unknown parameter h, and that is an important parameter, important aspect, although simple, it is an important, it is an important property, that as to kept in mind.

So, h hat cannot depend on h. So, it does not, it does not depend on h, or rather cannot depend on h, which is basically your unknown parameters, because h is the unknown parameter. This is the unknown, because h is the unknown parameter. Now what we would like to do, is that, we would like to come with an estimate, which minimizes the estimation error. Now what is the estimation error? Estimation error is naturally, the difference between the estimate and the parameter, all right? So, we have the error, this is the estimate, correct? So, this is, I have my estimate, h hat of y bar, minus h, this is the estimation error. And what I am going to do, is am going to consider the square, of the estimation error, because I can have the scenario in which the estimation error is positive, and negative, and cancels each other, all right? So, I am interested in minimizing the absolute value of the error, so I am going to considering the square, of the estimation error.

And more importantly, I am not interested in minimizing, a particular instantiation of the estimation error, rather, I am interested minimizing, the estimation error on an average, that is where I repeat this experiment several times, I would like to come up with an estimator, which minimizes the average estimation error. That is also termed as the mean squared error of the estimation.

So, I am going to denote this average, by the expectation operator, which gives me the, basically the mean. So, basically this mean, also represents, this quantity also represents, basically your averaging, this also represents the average, and therefore, now, this total quantity that you have over here, this is known as the MSE, which is the mean squared error.

(Refer Slide Time: 06:06)

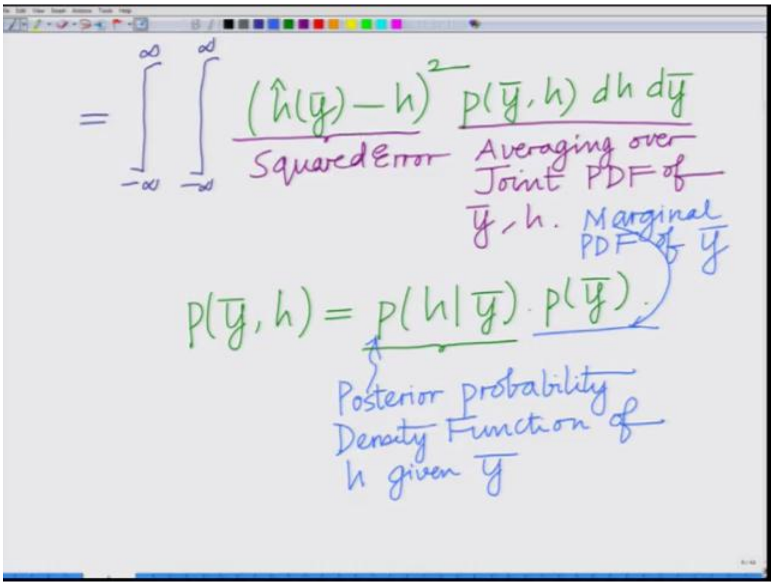


And this is an important quantity, that we will be consider considering, in the rest of this course, that is the, this is the, MSE or the mean squared error, and that is from which we get the term MMSE, that is we want to find the minimum, of this MSE, that is the minimum, of this mean square error.

Now, how do we come up with an estimator that yields the minimum of this mean squared error? That is the procedure, we are going to, which we are going to discuss, that is the general principle, to obtain the minimum mean squared error estimate, that is the procedure, which we are going to describe now all right. So, now, let p y bar comma h, all right? Let this quantity, denote the, denote the, joint distribution of the Observation Vector y bar, and your parameter.

This denote the joint observation, the joint distribution, or the joint probability density function, of the Observation Vector y bar, and parameter h. Therefore, what I would like to do, is to basically average, so what now, the mean squared error, the MSE, which we have represented as the expected value, of h hat, y bar, that is the, let me just write this clearly, that is h hat, of y bar, minus h, whole square. Now this quantity can be written as, given the joint probability density function, I am going to integrate it, I am going to write this expression, and I am going to, and I am going to write this expression.

(Refer Slide Time: 08:07)



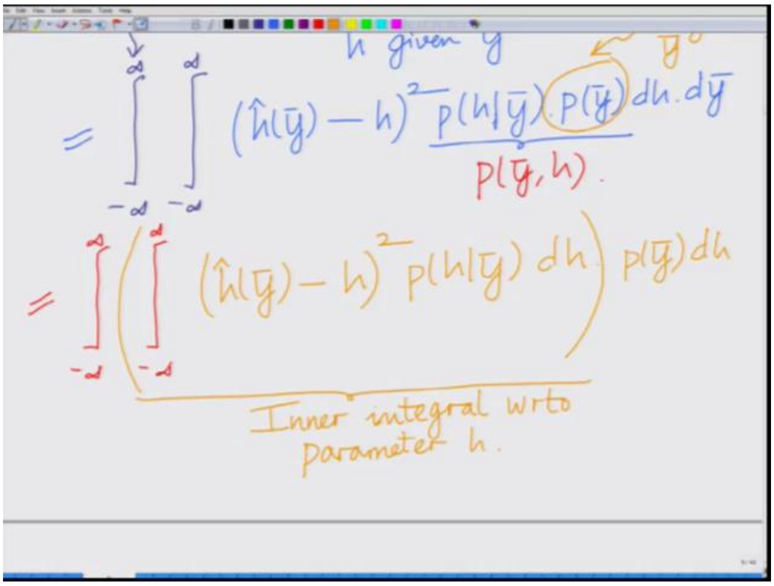
So, this expression is basically, h hat, of y bar, that is the error, all right? So, this is the squared error, what is this? This is the squared error, that is, h hat, of y bar, minus h, whole square. I have to average this over the joint probability density function, of; I have to average this over the joint probability density function of, the parameter and the observation error, all right? So, this is an important expression, all right? So, let me take some time to explain it. So, h hat, y bar minus h, whole square, is the squared error, all right? But it depends on a particular value, of the parameter h, and the particular value of the Observation Vector y bar.

So, now to get the average, or this mean squared error, I have to average this random quantity, right? Over the joint probability density function of the both, the Observation Vector y bar, and the parameter h, all right? Because as we have said in the Bayesian framework, the observations y bar, as well as the parameter h, is random in nature, therefore, to compute the mean squared error, I have to average, this instantaneous quantity, over the joint probability density function, of the random, of the Observation Vector y bar, and the parameter h. So, that is what we are doing over here, this is your squared error, this is the squared error which is the function of y bar, and h, and what we are doing over here, is averaging, over the joint, PDF of y bar, comma h. And now, I am going to use important property, of this joint probability density function, to simplify this expression. Now we are going to use the following property, of the joint property, probability density function, that is, p of y bar, comma h, sorry, this is not h bar, this is h, can be expressed as, p of h, given y bar, times p of y bar.

So, the joint probability density function p of y bar, given h, can be expressed as p of h, that given y bar, the conditional probability density function, of the parameter h, given the Observation Vector y bar, times, p of y bar, that is the marginal probability density function, of the Observation Vector y bar. So, this is an important quantity, p of h given y bar, this is also the posterior probability density function, or the a posteriori probability density function, of, h, given your Observation Vector y bar, and what is this? This is the marginal probability density function of, where PDF basically represents, probability density function, the marginal probability density function, and marginal probability density function of y bar, so p h given y bar.

This is an important quantity; this is the a posteriori probability density function, of the parameter h, right? So, remember we said, we have a prior probability density function of h, which gives information of the parameter h. Now the a posteriori probability density function, p of h, given y bar, basically gives the modified probability density function of h, on observing, the Vector y bar right. So, now, we have the Observation Vector y bar, having observed, this Vector y bar, what information does that convey about the parameter h, that is captured in this a posteriori probability density function, p of h, given y bar, all right? I am now going to use this, plug this simplification, or plug this property of this probability, joint probability density function of y bar, comma h, in the expression for the mean squared error above, and simplify the mean squared error as follows.

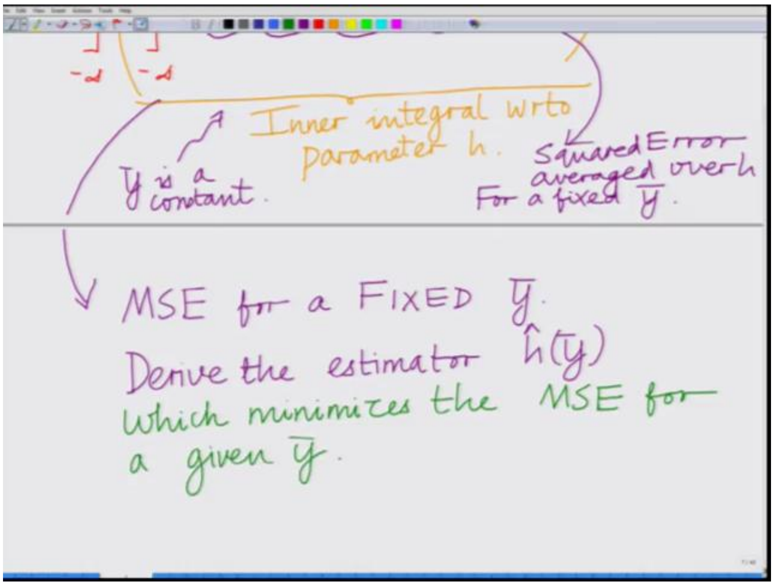
(Refer Slide Time: 13:05)



So, now I can right the same expression, for the mean squared error. This is integral minus infinity, integral minus infinity. Remember there are 2 integrals, one with respect to y bar, the outer integral with respect to I bar, y bar, the inner integral with respect to h, and we have well, h hat, y bar, minus h, whole square, times, instead of the joint probability density function, I am going to write it as, p of h, given y bar, into, p of y bar, d of h, d of y bar. And what is this? This is basically, were we had your joint probability density function, p of y bar, comma h, earlier. I am substituting for p of y bar, comma h, as p of h, given y bar, times, p of y bar. And now I am going to simplify this, as basically first I am going to write the outer integral, with respect to y bar. and I am going to write the inner integral, now with respect to h. I am going to right it, as follows, that is going to be, h hat, of y bar, minus h, whole square, p of h, given y bar, times, d of h, correct?

Now, I am going to take all the quantities, with respect to y bar, because p of y bar, this depends only on y bar. I can remove it, from the integral, inner integral, which depends only on h. So, p of y bar, depends this, depends, this depends, only on y bar. So, it is a probability density function of h, given by y bar. So, this depends only on y bar. So, I can move this into the outer integral, alright. So, this is basically your inner integral with respect to h. Let me just write that clearly. This is your inner integral, with respect to h. And it depends only on the parameter h, with respect to your, because in this inner integral, with respect to h, basically y bar (Refer Time: 15:58), because y bar is in the outer integral. So, we are carrying out the double integration, in 2 steps. First, an inner integration, with respect to h, we get a function of y bar, and then integrated with respect to y bar.

(Refer Slide Time: 16:16)

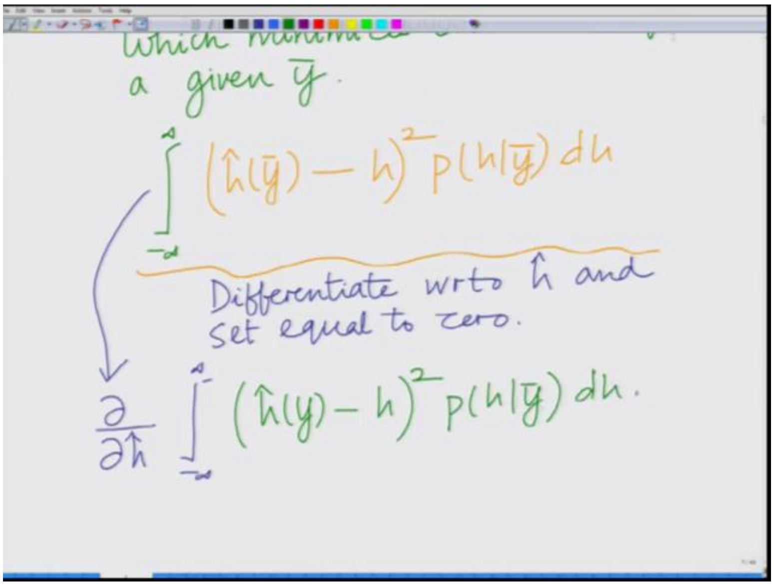


So, therefore, as far as the inner integral is concerned, y bar is a constant. So, that is an important to realize. For this inner integral, y bar is a, y bar is a constant for this inner integral. Now, therefore, what I am going to do, all right. Now let us take a look on this inner integral. What is this inner integral? This inner integral is the mean squared error, averaged over, what is this, this is the mean squared error, error or a squared error, average over the parameter h, for a fixed y bar. That is, for a given Observation Vector y bar, you were to ask the question, what is the mean, what is the average of the squared error corresponding to a given Observation Vector y bar, that is given by this inner integral, because we are averaging only with respect to the parameter h, while fixing y bar in the inner integral. So, the inner integral basically is the mean squared error, average, with respect to h, that is it is the mean squared error, for a fixed y bar, that is important.

So, this inner integral is the MSE for a fixed Observation Vector y bar. That is the first that is the first important point to keep in mind. Now, the second important point is, basically, that if you were to come up with an estimator h hat of y bar, remember h hat of y bar, depends only on y bar. So, if I obtain the minimum, obtain an estimator h hat of y bar, which yields a minimum squared error, or which is the minimum mean squared error, for each y bar, at, for a given y bar, all right? So, for a given y bar, if I obtain the minimum mean squared, obtain the estimator h hat a y bar, which is the minimum error, I do that for every value of y bar, then naturally, it is mean squared error, with respect to averaging over y bar, is going to be it is going to be the minimum mean squared error, with respect to y bar.

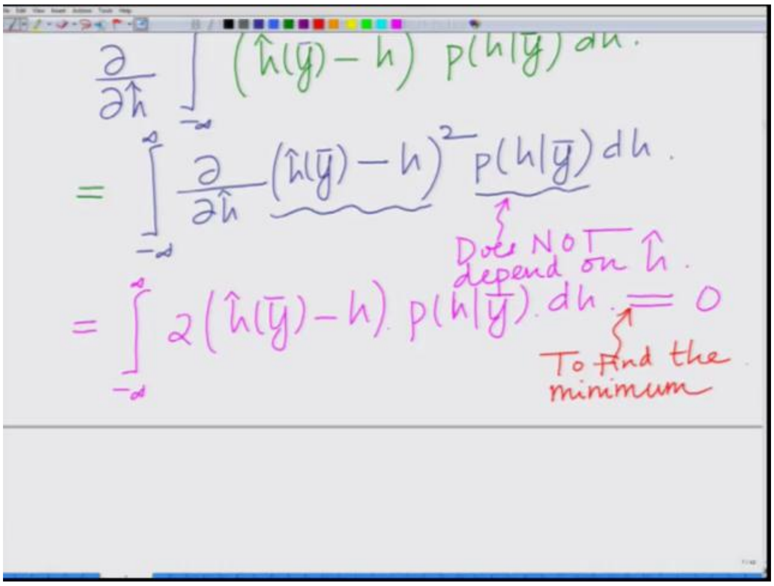
So, what I am going to do, I am going to now, derive, the estimator, h hat y bar, which yields, which minimizes the MSE. So, for every given y bar Observation Vector y bar, we derive the error, we derive the estimator h hat of y bar, which minimizes the, which minimizes the squared error, or which minimizes the mean squared error, averaged, with respect of the parameter h. And therefore, now, since this is the best estimator, or this is the best, or this yields the minimum error, with respect to every y bar, naturally, when you average the performance, over y bar, it is going to yield the best performance, in that sense, it is going to be the minimum mean squared error estimator.

(Refer Slide Time: 20:14)



So, now, how do we derive the estimate, which yields the minimum error, for a given y bar? So, to do that naturally, look at this, we have this expression, for the mean squared error, for a given y bar. This is h hat y bar, minus h, whole square, p of h, given y bar, d of h. This is the expression of the mean squared error for a given y bar, to find the best, to find the expressed y, in the h hat, y bar, that is the estimator which minimizes the mean squared error, we simply differentiate this with respect to h, differentiate with respect to h hat, and set this equal to, and set this equal to, 0. So, this is, this mean squared error, is the function of h hat. So, naturally to find the minimum, we differentiate with respect to h hat, and set this equal to 0, all right? So, now, we differentiate this, with respect to h hat, h hat of y, minus h whole square, p of h, given y bar, d of h.

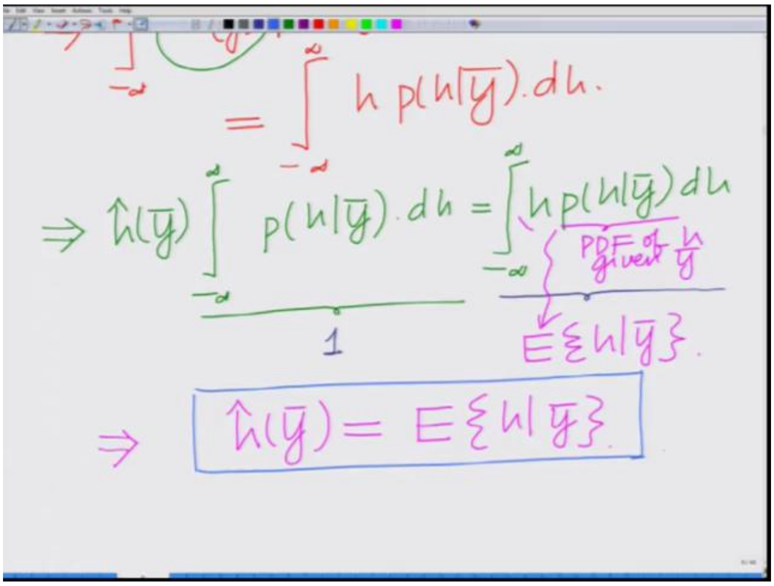
(Refer Slide Time: 21:52)



Now take the derivative inside, inter change the derivative, and the integral, we have, minus infinite to infinity, duo by duo h hat, h hat of y bar. In fact, this has to be the Observation Vector y bar, minus h, whole square, p of h, given y bar, d of h.

Now, look at this, only this quantity here, depends on h hat, this does not depend on h hat, does not depend on, this does not depend on h hat. Therefore, now if I differentiate this with respect to h hat, what I am I going to have? It is integral, this is going to be integral, minus infinity to infinity, twice, h hat y bar, minus h, this is differentiating with respect to h hat, I have twice h hat y bar minus h, p of h, given y bar, of times d h. Now this has to be equated to 0. To find the minimum, to find the h hat, which is the minimum, basically I want to differentiate this, and equal to 0, this we equate into 0, basically to find the, to find the, to find the minimum.

(Refer Slide Time: 23:35)



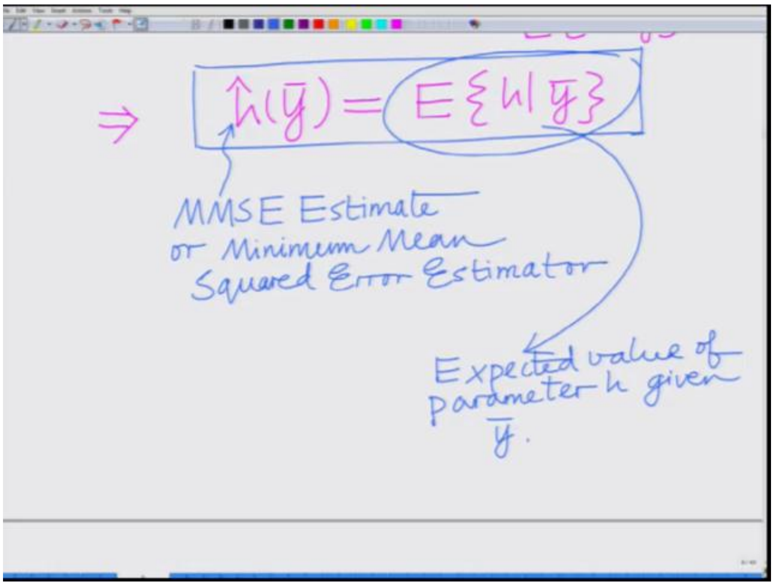
And now what we get is basically, what this implies is basically, this implies, now this 2 cancels, because we have 0. So, I can cancel the 2. This implies h hat, of y bar, p of h, given y bar, d of h, equals, integral minus infinity to infinity, h, p of h, given y bar, d h.

Now, observe these 2 quantities. Integral h hat, y bar, p of h, given y bar, d of h. Now remember the important property of the estimator, h hat, of y bar. The estimator h hat of y bar, does not depend on h. This integral is with respect to h; therefore, since h hat of y bar does not depend on h, it can come outside of the integral, that is important point. So, this does not depend on, this does not depend on h, which means I can take this and move it out, of the integral, which means, h hat, of y bar, integral, minus infinity to infinity, p of h, with respect to y bar, d of h, equals, this should be equal to, integral minus infinity to infinity, h p of h, given y bar, into d of h. Now look at these two, this integral first, this integral is an integration of the probability density function, p of h, given y bar. This is the a posteriori probability density function of h. We are integrating this, over the range, minus infinity to infinity.

So, will when we integrate any probability density function, over the range minus infinity to infinity, naturally the integral is going to be 1. Therefore, this integral is 1. Now look at this, this is basically, what is this? This is the probability density function of PDF of h, given y bar. I am multiplying this by h, and integrating from minus infinity to infinity. So, this is basically the average. So, what is this? This is the expected value of h, given y bar.

Remember, from the properties of the probability density function, if I take any random variable x. Probability density function f x, of x, multiplied by x, integrated from minus infinity to infinity, that is minus infinity to infinity, x times f x, of x, d x, that yields the expected value of x. Here, I have integral minus infinity to infinity, h times, of probability density function of h, given y bar, times d h integrated from minus infinity to infinity, that is nothing but, the expected value of h, but for a given y bar, that is condition on the Observation Vector y bar. And therefore, naturally finally, we have the estimator, which minimizes the mean squared error, is basically h hat, y bar, equals the conditional mean of h.

(Refer Slide Time: 27:30)



And this is the most important principle, of MMSE estimation, which basically says that, and that is, it this is basically the estimate of the parameter h, for a given y bar, what is this? This is the estimate which minimizes the mean squared error.

This is the MMSE estimate, or basically the minimum mean, the minimum mean squared error, estimator, all right? And that is given by, the expected value of the parameter, average of the parameter that is expected value of h, given by y bar that is the conditional mean of the parameter h that is expected value of h. What is this? This is your expected value of the parameter h, expected value of the parameter h, given y bar, and therefore, the simple and elegant result, basically gives us the minimum mean squared error estimator. What it says is, a very simple principle, that is basically, the estimate, which minimizes the mean squared error for a given y bar, for a given Observation Vector y bar, is basically the conditional mean, that is if you look at the a posteriori probability density function, p of h, given y bar, the probability density function of the parameter h, given the Observation Vector y bar, and look at its mean, look at the expected value, of this a posteriori probability density function, that itself gives me basically, the best estimator, or basically the estimator which minimizes the mean squared error, for a given value of h bar.

Now, naturally since it minimizes the mean squared error, which minimizes the squared error, for a given y bar, it minimizes the mean squared error, for every y bar. Therefore, it minimizes the mean squared error, on average, that is, when you average it over that probability density function of y bar. And this is the most important principle of minimum mean squared error estimate, estimation. That is, the h hat of y bar is basically expected value of h, expected value of the parameter h, given the Observation Vector y bar. And that basically captures, the sense of MMSE estimation, and we are going to use this principle of MMSE estimation, further, in wireless, in several wireless examples, to illustrate the application of this principle of MMSE estimation. So, we will stop this module, here.

Thank you very much.