

Our discussion of least mean squares estimation so far was based on the case where we have a single unknown random variable and a single observation. And we're interested in a point estimate of this single unknown random variable. What happens if we have multiple observations or parameters? For example, suppose that instead of a single observation, we have a whole vector of observations. And, of course, we assume that we have a model for these observations.

Once we observe our data, a numerical value for this vector, or what is the same numerical values for each one of these observation random variables. Then we're placed in the conditional universe where these values have been observed. Then, we notice that the arguments that we carried out did not rely in any way on the fact that  $X$  was one-dimensional. Exactly the same argument goes through for the multi-dimensional case, and simply, the answer is again, that the optimal estimate, the one that minimizes the mean squared error, is again, the conditional expectation of the unknown random variable given the values of the observations.

So this gives us a simple and much more general solution that also applies to the case of multiple observations. Now, what if we have multiple parameters? Once more, the argument is exactly the same, and we obtain that the optimal estimate of any particular parameter is going to be the conditional expectation of that parameter given the observations. So if our parameter vector is something of this form, consisting of several components, then the LMS estimate of the  $j$ th component of our parameter vector is going to be simply the conditional expectation of this parameter given the data that we have obtained.

And this gives us the most general solution to the program of least mean squares estimation when we have multiple parameters and multiple observations. One very simple concept that applies to all possible cases. Unfortunately, however, our worries are not over. Even though LMS estimation has such a simple and such a general solution, things are not always easy. Let us see what's happening.

No matter what, we have to first find out the posterior distribution of  $\Theta$  given the observations that we have obtained. And this is done using the Bayes rule, which we have written here, and this is how you evaluate the denominator in Bayes' rule. What are the difficulties that we may encounter? One first difficulty is that in many applications, we do not necessarily have a good model or we're not very

confident about our model of the observations.

If  $X$  and  $\Theta$  are multi-dimensional, such a model might be difficult to construct. Setting this issue aside, there's a further issue. The conditional expectation of  $\Theta$  given  $X$  may be a complicated non-linear function of the observations. This means that it may be difficult to analyze, but even more important, it may be very difficult to calculate even after you have obtained your data. Let us understand why this might be the case.

Suppose that  $\Theta$  is a multi-dimensional parameter. Then in order to calculate the denominator that's involved here in the Bayes rule, when you integrate with respect to  $\theta$ , you have to actually carry a multi-dimensional integral, and this can be very challenging or sometimes, practically impossible. Even if you had this denominator term in your hands, still, in order to calculate a conditional expectation, you would have to calculate once more an integral of  $\theta_j$  integrated against the posterior distribution of the vector  $\theta$ .

But this integral should be once more, over all the parameters. So it would be a multi-dimensional integral in the general case, and that's one additional source of difficulty. And this is the reason why we will also consider an alternative to least mean squares estimation, which is much simpler computationally and much less demanding in terms of the model that we need to have in our hands.