

After our warm-up, we can now come to the real problem. We have, again, a random variable  $\Theta$  with a known prior distribution. And we're interested in a point estimate. What will be different this time, however, is that we now have an observation. And we also have a model of that observation as a conditional distribution given the value of the true parameter.

We observe a value of that random variable. That value is little  $x$ . And on the basis of that value, we would like to now come up with a point estimate of the unknown random variable  $\Theta$ .

How do we proceed? We can, of course, use the Bayes rule. And the Bayes rule is going to give us a distribution for the unknown random variable given the observation that we have obtained. And that distribution could be discrete or continuous. Let me just plot something as if it's continuous.

And now that we have the posterior distribution of  $\Theta$ , we would like to come up with a point estimate. How do we do it? Remember our earlier conclusion.

If we do not have any observations, if we live in a universe where we have a distribution of  $\Theta$  and we want a point estimate, the optimal, the one that minimizes the mean squared error, is the expected value of our random variable. But now we live in a different universe, in a universe where we have a conditional distribution of  $\Theta$ . We want to minimize the conditional mean squared error, because this is the mean squared error that applies to this conditional universe in which we have obtained a particular observation.

What is going to be the result of this minimization? Well, this is a problem that's identical to the problem of minimizing this quantity, except that now this problem is posed in a conditional universe. So we just follow the same steps. And obtain the same solution, the solution is going to be the expected value of the unknown random variable, except that now we live in a conditional universe. And therefore, we should take the relevant expected value which is the conditional expectation given the information that we have available in our hands.

So to summarize, what we obtain is that the optimal estimate is the conditional expectation. And this is a relation between numbers. But if we want to think about it more abstractly, we have designed an estimator which is based on a random variable, capital  $X$ , and calculate the expected value of our

random variable that we're trying to estimate, namely  $\Theta$ , on the basis of  $X$ .

Let us now continue with some observations. Remember that the expected value of  $\Theta$  minimizes this quantity. And we can write this more explicitly in terms of the following inequality-- that if we use the expected value as an estimate, the resulting mean squared error is less than or equal to the mean squared error that we would have obtained if we had used any other estimate,  $c$ . So this is a relation that's true for all  $c$ .

Now, let us take this inequality and translate it into our more interesting context where we have an observation available. Once more, the conditional expectation minimizes the mean squared error. Let us write out explicitly what this means in a form analogous to what we wrote down earlier.

What it means is that the expected value of  $\Theta$  minus the estimate, namely the conditional expectation, squared. In this conditional universe in which we live, this is less than or equal to the mean squared error that we would have obtained if we had used any other estimate in the place of the conditional expectation. So for any value  $g$  of  $x$  that we might have used, the error would have been at least as large.

Why am I using this notation  $g$  here? Let us go back to the bigger picture. What we have is that we are obtaining a numerical value  $x$ .

We do some processing to it which corresponds to some function  $g$ . And we come up with an estimate which is a function of the little  $x$  that we have observed. So no matter what estimate we use, the mean squared error is going to be at least as large as the mean squared error that we obtain if we use the conditional expectation.

Now, let us take this inequality here and write it in a more abstract form. Suppose that we have settled on some particular estimator and we want to compare this estimator with the expected value estimator. Then we're going to get the following inequality.

If we use the conditional expectation as an estimator in a conditional universe where we know the value of the random variable  $X$ , the corresponding mean squared error is going to be less than or equal to the mean squared error obtained by the alternative estimator  $g$ . What does this inequality say? This inequality is simply an abstract version of the previous inequality. The previous inequality is true for all little  $x$ .

Here we have an inequality between random variables. This random variable here is a random variable that takes this specific numerical value, whenever capital  $X$  takes the value little  $x$ . When  $X$  takes the value little  $x$ , we're conditioning on this event. And when  $X$  is equal to little  $x$ , this quantity takes on this particular numerical value. And similarly, on the other side, this is a random variable that takes this particular numerical value whenever capital  $X$  is equal to little  $x$ .

Now that we have an inequality between random variables, actually between conditional expectations, we can take expectations of both sides. And we use the law of iterated expectations. The expectation of a conditional expectation is an unconditional expectation.

So we obtain this as the expected value of this quantity. And it's less than or equal to the expected value of this quantity using, again, the law of iterated expectations. We obtain this relation here. And what this inequality is saying is that the overall mean squared error, if we use the conditional expectation, now as an estimator, is less than or equal to the mean squared error that we would obtain if we had used any other estimator.

So this inequality refers to the following picture. We obtain an observation which is a random variable. We process that random variable to come up with an estimator which is a function of the random variable that we observe and so is itself a random variable.

So when we use this random variable to estimate  $\Theta$ , we obtain a certain mean squared error. This is going to be at least as large as the mean squared error that we obtain if we use the conditional expectation as our estimator. So to summarize, the conditional expectation of  $\Theta$ , viewed as a random variable, as an estimator, what we call the LMS estimator of  $\Theta$ , has the property that it minimizes the mean squared error over all possible alternative estimators. So if you want to design this box using some other function  $g$ , you're going to obtain a mean squared error that's going to be no better than what you obtain if you were to use the conditional expectation.