MITx Video | 6.041x | Spring 2014 | Video transcript | L16-7: Properties of the LMS estimation error

In this segment, we're going to go over a few theoretical properties of the estimation error in least mean squares estimation. Recall that our least mean squares estimator is the conditional expectation of the unknown random variable, given our observations. Let us define the error, which is the difference between the estimator and the random variable that we are trying to estimate.

Let us start with some observations. What is the expected value of our estimator? Well, using the law of iterated expectations, the expectation of a conditional expectation is the same as the unconditional expectation. And using this property, by moving this Theta to the other side, what we obtain is that the estimation error has an expectation of 0. So this tells us that the estimation error, on the average, is equal to 0, which is good news.

In fact, something stronger is true. Not just the overall average of the estimation error is 0, but even if you condition on a particular measurement, still the conditional expectation of your estimation error is going to be equal to 0.

Let us derive this relation. We're looking at the expected value of Theta tilde, which is Theta hat minus Theta, conditional on a value of X. Now, if I tell you the value of X, then the estimator is completely determined-- there's no uncertainty about it-- so the expectation of Theta hat, in this conditional universe, is just Theta hat itself. And we're left with the second term, but the second term is also Theta hat, and therefore we obtain a difference of 0.

Let us now move to a slightly more complicated question. What is the covariance between the estimation error and the estimate? We will calculate the covariance as follows. It is the expected value of the product of the two random variables that we are interested in, minus the product of their expectations.

Now, we already calculated that the expected value of the estimation error is equal to 0, and therefore, this term here disappears. This term is equal to 0. So we now need to calculate the first term. This may seem difficult, but conditioning is always a great trick, so let's do that.

Let us start by calculating the conditional expectation of this product. As before, in the conditional universe, where we're told the value of X, the value of Theta hat is known. It is becoming a constant, so

it can be pulled outside the expectation. But then we can apply the fact that we established earlier that this term is 0, and therefore, we obtain a 0 here.

Now, the expected value of a random variable is the same as the expected value of the conditional expectation. This is, again, the law of iterated expectations. Since the conditional expectation is 0, when we apply the law of iterated expectations to this quantity, we also obtain a 0. Therefore, this term is 0 as well, and we have established what we wanted to show.

Using this fact, now we can figure out that the following is true. We write the random variable Theta as the sum of Theta hat minus Theta tilde. This comes simply from this definition here, by just moving Theta to this side, and Theta tilde to the other side. So Theta is the difference of two random variables, and these two random variables have 0 covariance.

When two random variables have 0 covariance, then the variance of their sum, or of their difference, is the sum of the variances. And this leads us to this relation-- that the variance of our random variable can be decomposed into two pieces. One of them is the variance of the estimator, and the other is the variance of the estimation error. This is an interesting fact. It can actually be derived in a different way, as well. It is just a manifestation of the law of total variances, but hidden in somewhat different notation.

And this concludes our discussion of theoretical properties of the estimation error. Unfortunately we will not have the opportunity to use them in any interesting ways. On the other hand, they are a foundational piece for the more general theory of least-squares estimation. If you try to develop it in a more sophisticated and more deep way, it turns out that these properties are cornerstones of that theory.