MITx Video | 6.041x | Spring 2014 | Video transcript | L17-8: Choices in representing the observations

In this final segment, we want to discuss an interesting point about linear estimators. Here's what the issue is. You obtain an observation, X, on the basis of which you want to estimate Theta. But perhaps you measure X on a different scale, let's say on a cubic scale, so that what you record actually is X cubed.

So you're faced with two possible estimation problems. One estimation problem is to use X to estimate Theta. Another estimation problem is to use X cubed to estimate Theta. Does it make a difference?

Let's consider the case of least mean squares estimation, without any linearity constraint. If you use X to estimate Theta, your estimator is going to be this conditional expectation. If you use X cubed to estimate Theta, your estimator will be this conditional expectation. Are they different?

Well, X and X cubed carry the same information about Theta. In particular, the posterior distribution of Theta given X is going to be the same as the posterior distribution of Theta given X cubed. You will be getting the same information, the same knowledge about X.

And in particular, if you calculate conditional expectations, these will also be the same. What about the linear case? If we restrict to linear estimators, then on the basis of X, you would form a linear estimation of this kind. But if your observation is in the form of X cubed, then a linear estimator would form a linear function of X cubed.

So this would be a different kind of estimator. We have seen a formula on how to obtain the best estimator, the best choices of a and b for estimators of this kind. We can use that same formula to obtain the best estimator of that kind. It's going to be, of course, a different estimator.

Here, we're optimizing within a different class. Which one of the two is better? Well, this depends on what you know about the particular problem at hand. If you have some reason to believe, or if you know that Theta and X are roughly related by some kind of cubic relation, then perhaps estimators in this class are going to perform better than estimators in that class.

Let me also point out a related issue that would come here. To find the right choice of a, you need to know the covariance between X and Theta. That's why the formula tells us about the optimal linear estimator. Here you would you need to know the covariance between Theta and X cubed.

In addition, the formula requires the variance of X. But here, instead of X, we're using X cubed. So in this case, we would need the variance of X cubed. Now, this could be more challenging.

In general, the higher the powers that you have, the more difficult these quantities are to calculate or to know what they are. But leaving that issue aside, what we have here is two alternative choices for the structure of the estimator that we're using. Now, we can push this story further. Instead of considering just estimators of this kind, we might consider as well estimators of this kind.

Is this a linear estimator? We still call it a linear estimator, because it is linear in the coefficients that we have to choose on how to optimize. That's the more important part. It's the linearity in these coefficients that's important, rather than the linearity in the X's.

So as a function of X, this is non-linear. On the other hand, we can think of this X as one observation, X squared as another observation, X cubed as a third observation, and what we've got here is a linear function of three different observations. So we can still pose a least squares problem in which we try to find the best choices for the coefficients a1, a2, and a3, as well as the coefficient b, find those choices that they're going to give us the smallest possible mean squared error. So we can optimize within this class.

Within this class of estimators, we certainly have more flexibility. This is a more general class of estimators than either of this one or that one. So within this class, we should be able to do even better.

On the other hand, we would have to pay a price that this is a more complex structure. It would be more difficult to find the optimal coefficients. And also, we're going to need higher order moments or expectations related to the X's and the Thetas.

Finally, there's nothing special in us using powers of X and using a polynomial. We could also look at estimators that have some other type of structure. For example, we might want to mix an exponential function in X and a logarithmic function of X, look at estimators of this form, and try to choose the best one. Find the best choice of the coefficients.

Again, this is something that is possible. And again, it's going to boil down to solving a system of linear equations in the coefficients. On the other hand, we need to know various expectations about X that might be difficult to obtain. How do we choose which structure to use should it be this one, this one, this one, or that one?

There's a trade-off, that more complicated structures introduce more complexity and make the problem more difficult. But there's also another issue. It has to do with what do we know about the particular problem at hand. If we know or have reason to believe that third order polynomials are going to give us excellent estimates of theta, then we may want to work within this class.

In any case, the moral of this story is that if we are to use the linear estimation methodology, we do have some choices. Linear in what? And different choices will give us different performance. But this now gets somewhat away from the subject of a mathematical methodology, and it gets closer to the art that you need to exercise in any particular problem domain.