

Let us now illustrate the linear least mean squares estimation methodology in the context of an example. And we're going to revisit our familiar example that we considered earlier in the context of general least mean squares estimation. Let us remind ourselves what were the assumptions behind this example. There is an unknown random variable that we wish to estimate, and that random variable happens to be uniform on the range from 4 to 10.

What we get to observe is a random variable X , which is equal to Θ plus or minus something. So X is Θ plus a noise term. And that noise term can be anything in the range from minus 1 to plus 1. Furthermore, the distribution of U is this particular uniform no matter what θ is, so Θ and U are independent.

These modeling assumptions correspond to this picture. This is the range of X and Θ . And the joint distribution of X and Θ happens to be a uniform distribution on this particular shape. However, we're not really going to use this picture other than for illustration purposes. You could take just this as the formulation of the problem that we're interested in.

So now, to develop the form of the optimal linear estimator, all we need to do is to determine the various constants that show up. So let's start with expectations. Θ is uniform from 4 to 10. Therefore, the expected value is the midpoint, which is equal to 7.

U has a symmetric distribution around 0, so its expected value is going to be equal to 0. X is the sum of Θ and U . Therefore, its expected value is the sum of these two expected values and is equal to 7.

Let us now look at variances. The variance of a uniform is equal to the square of the length of the interval on which the uniform is distributed-- in this case, it's 6 squared-- divided always by a coefficient of 12. 6 squared is 36, so we obtain three.

The variance of U is determined by a similar formula, except that now we have an interval of length 2, so we obtain 2 squared over 12. And this is $1/3$. Now let us look at the variance of X . Since X is the sum of Θ and U , and since the two of them are independent, the variance of X is going to be the sum of these two variances, which is $10/3$.

Now let us try to calculate the covariance term. The covariance of Θ with X is this expression,

because X is Θ plus U . And then, using linearity properties of covariances, this is the covariance of Θ with itself, plus the covariance of Θ with U . Now, Θ and U are independent, so this covariance is equal to 0. The covariance of Θ with itself is just the same as the variance, so here we obtain an answer of 3.

And so now we have all the pieces of information that we need, and we can proceed and write down the form of the linear estimator. The expected value of Θ is 7. Then, the covariance of Θ with X is 3, divided by the variance, which is 10 over 3. So this ratio becomes $9/10$. And then X minus the expected value of X gives us this term.

So this is the form of the optimal linear estimator, and if you wish to plot it, it is a curve of this kind. It is actually interesting to contrast this solution to the shape of the optimal estimator, the least mean squares estimator, or the conditional expectation estimator, which we had found earlier, and which corresponds to this particular blue curve. So in some sense, the linear estimator is a pretty good approximation of the optimal non-linear one. It does the best job that it can do, subject to the constraint that it has to be a linear function.

Notice also that this coefficient here is, of course, positive. This reflects the fact that the two random variables, X and Θ , are positively correlated. This should be clear from this diagram. When X increases, Θ tends to also increase, and vice versa. It's also reflected in the fact that the covariance is a positive number. On the other hand, because this coefficient is $9/10$ and not equal to 1, this red line is somewhat slanted in comparison with the orientation of the diagram that we have.

The calculations that we went through in this particular example are pretty generic. This is what you need to do in general. You just look at the random variables involved, you calculate their means, you calculate their variances. Then you may have to do some extra work to calculate the covariance of interest. And once you're done, you plug in the numerical values that you have found, and you obtain the form of the linear estimator.