

Least mean squares estimation is remarkable because it has such a simple answer.

The way to come up with estimates, if what you care about is to keep the mean squared error small, the way to come up with estimates is to just report the conditional expectation, which is going to be a number, once you have obtained some values of the data.

Or more abstractly, you can think of it as a random variable, if you do not know ahead of time what data you're going to obtain.

Because this estimator is so important, it is worth writing down what the performance of that estimator is.

So suppose that you have obtained a particular measurement, a particular value of the observation, then the resulting mean squared error within that conditional universe where you have already obtained that value, is just this quantity.

It's the mean square of the error between the variable that you're trying to estimate and your estimate.

And everything gets computed within this conditional universe.

Now this is a very familiar quantity, however.

It's the expected value of a random variable difference from its mean, squared.

This is just the variance, except that because all quantities are calculated in a conditional universe, this is the conditional variance.

So the conditional variance is the optimal mean squared error, the mean squared error that you obtain when you use this particular estimate.

And it's the value that you would report to your boss if you were asked how good is the estimate that you're giving me.

But suppose that you have not yet obtained a measurement, but you're going to your boss and you're proposing this particular estimator as your design.

What are you going to report to your boss as the performance of your design?

Since you have not yet obtained the value of X , X is a random variable, you do not know what the value of this conditional expectation is going to be.

It's a random variable.

But no matter what it is, this is going to be the error that you're going to be obtaining.

And this is the overall value of the mean squared error.

So this is the quantity that you would report to your boss as your overall mean squared error, the value that you report before obtaining any specific measurement.

Now, what is this quantity?

This quantity is just the average of this quantity up here, averaged over all the possible values of X .

And in our more abstract notation, it is just the expectation of the conditional variance.

The conditional variance, the abstract conditional variance, is a random variable that takes this value whenever capital X happens to be equal to little x .

And when we average it over all possible values of X , we just have the expectation of this random variable.

Let me continue now with a few more comments on LMS estimation.

First, something that should be pretty clear at this point is that LMS estimation is only relevant to estimation problems.

This is because in hypothesis testing problems we typically care about the probability of error, not the mean squared error.

A second important comment is that in some cases the LMS estimates and the MAP estimates turn out to be the same.

When is that the case?

If the posterior distribution of Θ happens to have a single peak, and it is also symmetric around a certain point, so that the peak also occurs at that particular point, then clearly the peak occurs here.

But the conditional expectation is also that same point, because it is the center of symmetry.

So in those cases, the two types of estimates, or estimators, coincide.

When does this happen?

This happens in one particular important special case.

We have seen that in linear-normal models the posterior distribution is normal.

And so this is one case where the MAP estimate and the LMS estimate are going to coincide.