As we discussed earlier-- even if we have multiple observations we can still find the structure of the best linear estimator in a fairly simple, computational way-- by solving a system of linear equations.

But usually, we do not get nice and simple formulas.

But here is a nice example, in which we will get a simple formula.

The example is something that we have seen before-- in various guises.

We're trying to estimate a certain quantity-- Theta.

And what we obtain is multiple, noisy observations of Theta.

That is-- at each observation we see Theta plus a noise term.

The assumptions that we make is that Theta has a prior distribution with a certain mean and a certain variance.

And the noise terms are zero mean-- but they have some variance.

And the additional assumption that we make is that all of these random variables are independent of each other.

So the noise terms are independent between themselves-- and also, the noise terms are independent of Theta.

This is the usual assumption-- but actually-- in the linear estimation problem, we do not need to make an independence assumption.

It's enough for our purposes to just assume that they are uncorrelated.

So we will assume that the correlation coefficient between any two of these random variables is equal to zero.

Now we can write down the form of the mean squared estimation error criterion that we have-- and try to find good choices for the coefficients to be attached to each one of the observations.

However-- we're going to find the solution to this problem using a shortcut that's going to bypass all kinds of computations.

Here's the trick.

Let us suppose-- in addition-- that these random variable were not just uncorrelated, but independent.

And that they happen to be normal random variables.

This is a problem that we did study before-- and we did find the maximum a posteriori probability estimate of Theta.

Because the posterior was normal, and we found that this was also the conditional expectation estimator of Theta.

And we did find a formula for it-- which took this form.

This was the form of the optimal estimate of Theta.

If you obtain values for the different observations-- little xi.

On the other hand-- if you want to translate this into random variable notation-- then notice that this is going to be a random variable, this is our estimator.

it's a conditional expectation of Theta given X-- and it's random because it depends on the values of the data that we see-- which are themselves random variables.

On the other hand-- this x0-- is actually the prior mean of Theta.

So this is a constant-- it's not random, and that's why we keep it with a lowercase notation.

Now-- what is interesting about this form?

It is a linear function of the observations.

And as we have discussed earlier-- if it turns out that the conditional expectation is linear in the observations, then this is also the best possible linear estimator.

So for the special case-- where our random variables are independent and normal-- we have a formula for the best linear estimator.

Now what if they are not normal?

Suppose that they're not normal.

But that they have the same means and variances as here-- as in the normal example.

Since these means and variances are the same-- and since these random variables are uncorrelated-- it follows that also all kinds of covariances between the random variable is involved here-- are going to be the same as for the normal example.

Now, the optimal solution to the linear estimation problem-- as we discussed earlier-- only cares about the means, variances, and covariances of the random variables involved.

The details of the distribution do not matter.

So whether we have normal distributions-- or non-normal distributions-- as long as we're making enough assumptions that fix all the means, variances, and covariances of interest-- we should be getting exactly the same solution.

Therefore-- this solution remains valid for the case of general random variables, as well.