

Let's now go through another example, which will be a little more challenging.

We're going to revisit an old problem.

We have a coin that has an unknown bias,  $\Theta$ .

And we have a prior distribution on this  $\Theta$ .

We fix some positive integer,  $n$ , we flip a coin  $n$  times, that has this unknown bias.

And we record the number of heads.

On the basis of the number of heads that have been observed, we wish to estimate the bias,  $\Theta$ , of the coin.

To make things more concrete, we're going to assume a prior distribution on  $\Theta$  that is uniform on the unit interval.

Now, this is a problem we have considered before.

We have calculated the expected value of  $\Theta$  given  $X$ .

And we did find that the expected value takes this particular form.

Now, notice that this is a linear function of  $X$ .

And if it turns out the least mean squares estimator is a linear function of  $X$ , then we're guaranteed, since this is the best, that this is also the best within the class of linear estimators.

So we immediately have the conclusion that the linear least mean squares estimator is this particular function of  $X$ .

So there's not much left to do.

On the other hand, just for practice, let us derive this answer directly from the formulas that we have for the linear least mean squares estimator, and see whether we're going to get the same answer.

So we want to use this formula.

And in order to apply this formula, all that we have to do is to calculate these expected values, this variance, and this covariance.

So now let's move on to this particular calculational exercise.

Let's start by writing down what we know about the random variables involved in this problem.

About  $\Theta$ , we know that it is uniform.

And so it has a mean of  $1/2$  and a variance of  $1/12$ .

About  $X$ , what we know is the following.

If you fix the bias of the coin, then the number of heads you're going to obtain in  $n$  flips has a binomial distribution, with parameters  $n$  and  $\Theta$ .

But of course,  $\Theta$  itself is a random variable.

So for this reason, this is a conditional distribution.

But within the conditional universe, we know the mean and the variance of a binomial, and they are as follows.

The mean of a binomial is  $n$  times the bias of the coin.

But because we're talking about the conditional universe, this is a conditional expectation.

And it's a random variable, because it's affected by the value of the random variable  $\Theta$ .

And similarly, for the variance, it's the usual formula for the variance of a binomial, except that now the bias itself is a random variable.

So now let's continue with the calculation of the quantities that we need for the formula for our estimator.

Let's start with the expected value of  $X$ .

Since we know the conditional expectation of  $X$ , we can use the law of iterated expectations.

The unconditional expectation is the expected value of the conditional expectation, which is  $n$  times  $\Theta$ .

And since the mean of  $\Theta$  is  $1/2$ , we obtain  $n/2$ .

Let us now continue with the calculation of the variance.

There are different ways that we can calculate it.

One could be the law of total variance.

But we will take the alternative approach, which is to use the general formula for the variance, that the variance is

equal to the expected value of the square of a random variable, minus the square of the expected value.

We know the expected value of  $X$ .

So all that's left is to calculate the expected value of  $X$  squared.

How are we going to calculate it?

Well, we know the conditional distribution of  $X$ .

So it should be easy to calculate the conditional expectation of  $X$  squared in the conditional universe, and then use the law of iterated expectations to obtain the unconditional expectation.

So now, we need to calculate this conditional expectation here.

How do we do it?

The expected value of a square of a random variable is always equal to the variance of that random variable plus the square of the expected value.

We're going to use this property, but we're going to use it in the conditional universe.

So in the conditional universe, this is going to be equal to the variance, in the conditional universe, which is  $n$  times  $\theta$  times  $1 - \theta$ , plus the square of the expected value of  $X$ , but the expected value in the conditional universe, which is this quantity--  $n$  times  $\theta$ .

So we obtain another term--  $n$  squared,  $\theta$  squared.

So now we can go back to our previous calculation, and right here, the expected value of this expression, which is  $n$  times  $\theta$ .

And then we have some  $\theta$  squared terms.

One is  $n$  squared.

The other is a minus  $n$ .

So we obtain plus  $n$  squared minus  $n$   $\theta$  squared.

The expected value of  $n$  times  $\theta$  is  $n$  times the expected value of  $\theta$ , which is  $1/2$ .

So we obtain a factor of  $n/2$ .

But then we have this additional term here.

We need the expected value of Theta squared.

What is it?

Well, since we know the mean and the variance of Theta, we can calculate the expected value of Theta squared.

It is equal to the variance plus the square of the mean.

And this evaluates to  $1/3$ .

So from here, we're going to obtain  $n^2$  minus  $n$  divided by 3.

And you can rewrite this term in a different way by collecting first the  $n$  terms,  $n/2$  minus  $n/3$ .

This gives us  $n/6$ .

And then there's the  $n^2$  term, which is  $n^2$  over 3.

Now that we found the expected value of  $X^2$ , we can go back to this calculation.

We have  $n/6$  plus  $n^2$  over 3 minus the square of the expected value of  $X$ , which is this expression here.

So we obtain  $n^2$  over 4.

$1/3$  minus  $1/4$ -- that makes  $1/12$ .

So we obtain  $n/6$  plus  $n^2$  over 12.

Or another way of writing this is  $n$  times  $n + 2$  over 12.

And this completes the calculation of the variance of  $X$ .

The last quantity that's left for us to calculate is the covariance of Theta with  $X$ .

We're going to calculate it using the alternative formula for the covariance, which is the expectation of the product minus the product of the expectations.

We have the expectations, but we do not have the expectation of the product.

So we need to calculate it.

Once more, it's going to be the same trick.

We're going to condition on Theta, and then use the law of iterated expectations.

So the law of iterated expectations, when you condition on Theta, takes this form.

And to continue here, we need to find this conditional expectation in the inside.

This conditional expectation-- what is it?

If I give you Theta, then you know Theta.

It is becoming now a constant.

There's nothing random to it, so it can be pulled outside the expectation.

And we obtain Theta times the conditional expectation of X.

We know what the conditional expectation of X is.

It's n times Theta.

So from here, we obtain overall, a term n times Theta squared.

So now we can go back here.

We have the expected value.

And the term in the inside-- we just found it.

It's n times Theta squared.

And since the expected value of Theta squared is  $1/3$ , from here, we obtain  $n/3$ .

And now we can go back, finally, to the calculation of the covariance.

It's going to be  $n/3$ , minus the expected value of Theta, which is  $1/2$ , times the expected value of X, which is  $n/2$ .

So it's minus n over four.

And this evaluates to  $n/12$ .

So we have succeeded in calculating all the quantities that are needed in the formula for the linear least mean squares estimator.

We can now take those values that we have just found and substitute them into this formula.

And after a little bit of algebra and moving terms around, everything simplifies to this expression.

Just to verify that this makes sense, what is the coefficient next to  $X$ ?

It's the covariance divided by the variance.

$n/12$  divided by this expression-- this  $n$  cancels that  $n$ .

This 12 cancels that 12.

We're left with an  $n + 2$  in the denominator.

And indeed, the coefficient that multiplies  $X$  is the term  $n + 2$  in the denominator.

And you can similarly verify that the constant term as well is the correct one.

So of course, this answer is what we had found in the past to be the optimal, the least mean squares estimator of  $X$ . As we discussed earlier, when this is linear in  $X$ , it has to be the same as the linear least mean squares estimator.

So this answer is not a surprise, but it was an interesting and perhaps useful exercise to go through the details of this calculation to see what it takes to figure out the different terms in this formula.