We now continue with our example and turn to the performance evaluation question. As you recall, we have a Theta that has a certain prior distribution. We're given a model for the observations. We came up with the joint distribution for X and Theta, which was uniform on this particular shape, and we found that the least mean squares estimator, namely the conditional expectation of Theta given any particular value of X, is given by this particular piecewise linear function.

Now, let us look at the performance of this estimator. We judge the performance given any particular value of X by looking at the corresponding mean squared error, which is the square of the distance between the unknown parameter and the estimate with which we came up. And as we have discussed, this is the same as the variance of Theta but in the conditional universe where X has been observed. It's the variance of the conditional distribution of Theta.

As we have discussed, if I tell you that X takes on this particular value, Theta is uniform on this interval. Therefore, the conditional variance of Theta is the variance of a uniform on an interval of this particular length. Now, we know that the variance of a uniform on an interval from a to b is equal to b minus a squared divided by 12. In this particular instance, the interval has length 2. Therefore, we have 2 squared divided by 12. So the variance is equal to 1/3.

This is what we get when the picture is of this type. On the other hand, if X falls in this range, then this interval on which Theta is now constrained to live, has a smaller length and we're going to get a different variance. So in order to keep track, let us come up with a plot. When X is between 5 and 9, Theta has a conditional distribution which is uniform on an interval of length 2 and a variance of 1/3. And therefore, the variance is constant, takes this value.

In the extreme case, when X is equal to 3, then this interval has 0 length. In fact, we have perfect certainty about the value of Theta. If X is equal to 3, then we know that Theta is equal to 4. There's no uncertainty. There's zero variance. What happens in between? As we increase x moving away from 3, the length of this interval increases linearly with x. And this means that the variance increases quadratically with x, so we have a quadratic that starts at 0 and rises to this value. And by a symmetric argument, on the other side, we also get function, which is 0 at 11, and which rises quadratically as x gets reduced from 11 to 9.

So this is a complete plot of the conditional variance of Theta as a function of the particular observation that we have obtained. We notice that some x's are more favorable than others. An observation equal to 3 is extremely favorable because it tells us unambiguously the value of Theta. But other choices of x, other possible observations, will lead to more uncertainty in Theta, and this is reflected in this diagram.

In case we are now interested in the overall mean squared error, then we have to calculate the average of this conditional variance where the average is taken over all values of X. This is going to be an integral of the conditional variance of Theta integrated over all possibilities for x. But, of course, each possibility of x has to be weighted according to the corresponding probability, or in this case, probability density function of X.

What is the PDF of X? It is not a given to us, but it is something that can be easily determined from what we have already done. We know the joint distribution of Theta and X, and whenever we know the joint we can also find the marginal. So once we find the marginal PDF of X, then we can plug it in, multiply by this function that we have already obtained, carry out the integration, and we will end up with a numerical value. Since it is an average of what we have here, it's going to end up being some number between 0 and 1/3. It's the average of these values, and closer to 1/3 rather than 0. And this would complete the way that a performance of a particular estimator gets evaluated.