# 6.041SC Probabilistic Systems Analysis and Applied Probability, Fall 2013
# Transcript – Recitation: Inferring a Parameter of Uniform Part 2

Welcome back. So now we're going to finish the rest of this problem. For part e, we've calculated what the map and LMS estimators are. And now we're going to calculate what the conditional mean squared error is. So it's a way to measure how good these estimators are.

So let's start out generically. For any estimator theta hat, the conditional MSE is-- conditional mean squared error-- is equal to this. It's the estimator minus the actual value squared conditioned on X being equal to some little x. So the mean squared error. So you take the error, which is the difference between your estimator and the true value, square it, and then take the mean. And it's conditioned on the actual value of what x is. Or, conditioned on the data that you get is.

So to calculate this, we use our standard definition of what conditional expectation would be. So it's theta hat minus theta squared. And we weight that by the appropriate conditional PDF, which in this case would be the posterior. And we integrate this from x-- from theta equals x to theta equals 1.

Now, we can go through some algebra and this will tell us that this is theta hat squared minus 2 theta hat theta minus plus theta squared. And this posterior we know from before is 1 over theta times absolute value of log x d theta.

And when we do out this integral, it's going to be-- we can split up in to three different terms. So there's theta hat squared times this and you integrate it. But in fact, this is just a conditional density. When you integrate it from x to 1, this will just integrate up to 1 because it is a valid density. So the first term is just theta hat squared.

Now, the second term is you can pull out of 2 theta hat and integrate theta times 1 over theta times absolute value of log of x d theta from x to 1.

And then the last one is integral of theta squared 1 over theta absolute value of log x d theta from x to 1. OK, so we can do some more-- with some more calculus, we get a final answer is this. So this will integrate to 1 minus x over absolute value of log x. And this will integrate to 1 minus x squared over 2 times absolute value of log x.

So this tells us for any generic estimate theta hat, this would be what the conditional mean squared error would be. Now, let's calculate what it actually is for the specific estimates that we actually came up with.

So for the MAP rule, the estimate of theta hat is just equal to x. So when we plug that into this, we get that the conditional MSE is just equal to x squared minus 2x 1 minus x absolute value of log x plus 1 minus x squared over 2 times absolute value of log of x.

And for the LMS estimate, remember this was equal to-- theta hat was 1 minus x over absolute value of log x. And so when you plug this particular theta hat into this formula, what you get is that the conditional mean squared error is equal to 1 minus x squared over 2 times absolute value of log of x minus 1 minus x over log of x quantity squared.

So these two expressions tells us what the mean squared error is for the MAP rule and the LMS rule. And it's kind of hard to actually interpret exactly which one is better based on just these expressions. So it's helpful to plot out what the conditional mean squared error is.

So we're plotting for x. For each possible actual data that we observe-- data point that we observe, what is the mean squared error? So let's do the MAP rule first. The MAP rule would look something like this.

And it turns out that the LMS rule is better, and it will look like this dotted line here on the bottom. And so it turns out that if your metric for how good your estimate is is the conditional mean squared error, then LMS is better than MAP. And this is true because LMS is actually designed to actually minimize what this mean squared error is. And so in this case, the LMS estimator should have a better mean squared error than the map estimator.

OK, now the last part of the problem, we calculate one more type of estimator, which is the linear LMS estimator. So notice that the LMS estimator was this one. It was 1 minus x over absolute value of log of x. And this is not linear in x, which means sometimes it's difficult to calculate. And so what we do is we tried to come up with a linear form of this, something that is like ax plus b, where a and b are some constant numbers. But that also does well in terms of having a small mean squared error.

And so we know from the class that in order to calculate the linear LMS, the linear LMS we know we just need to calculate a few different parts. So it's equal to the expectation of the parameter plus the covariance of theta and x over the variance of x times x minus expectation of x.

Now, in order to do this, we just need to calculate four things. We need the expectation of theta, the covariance, the variance, and the expectation of x. OK, so let's calculate what these things are.

Expectation of theta. We know that theta is uniformly distributed between 0 and 1. And so the expectation of theta is the easiest one to calculate. It's just 1/2. What about the expectation of x?

Well, expectation of x is a little bit more complicated. But remember, like in previous problems, it's helpful when you have a hierarchy of randomness to try to use the law of iterated expectations. So the delay, which is x, is random. But it's randomness depends on the actual distribution, which is theta. Which itself is random. And so let's try to condition on theta and see if that helps us. OK, so if we knew what theta was, then what is the expectation of x?

Well, we know that given theta, x is uniformly distributed between 0 and theta. And so the mean would be just theta over 2. And so this would just be expectation of theta over 2. And we know this is just 1/2 times the expectation of theta, which is 1/2. So this is just 1/4.

Now, let's calculate the variance of x. The variance of x takes some more work because we need to use the law of total variance, which is this. That the variance of theta is equal to the expectation of the conditional variance plus the variance of the conditional expectation.

Let's see if we can figure out what these different parts are. What is the conditional variance of x given theta?

Well, given theta, x we know is uniformly distributed between 0 and theta. And remember for uniform distribution of width c, the variance of that uniform distribution is just c squared over 12. And so in this case, what is the width of this uniform distribution?

Well, it's uniformly distributed between 0 and theta, so the width is theta. So this variance should be theta squared over 12.

OK, what about the expectation of x given theta? Well, we already argued earlier that the expectation of x given theta is just theta over 2. So now let's fill in the rest.

What's the expectation of theta squared over 12? Well, that takes a little bit more work because this is just-- you can think of it as 1/12. You could pull the 1/12 out times the expectation of theta squared. Well, the expectation of theta squared we can calculate from the variance of theta plus the expectation of theta quantity squared. Because that is just the definition of variance. Variance is equal to expectation of theta squared minus expectation of theta quantity squared. So we've just reversed the formula.

Now, the second half is the variance of theta over 2. Well, remember when you pull out a constant from a variance, you have to square it. So this is just equal to 1/4 times the variance of theta.

Well, what is the variance of theta? The variance of theta is the variance of uniform between 0 and 1. So the width is 1. So you get 1 squared over 12. And the variance is 1/12.

What is the mean of theta? It's 1/2 when you square that, you get 1/4. Finally for here, the variance of theta like we said, is 1/12. So you get 1/12.

And now, when you combine all these, you get that the variance ends up being 7/144. Now we have almost everything. The last thing we need to calculate is this covariance term.

What is the covariance of theta and x? Well, the covariance we know is just the expectation of the product of theta and x minus the product of the expectations. So the expectation of x times the expectation of theta.

All right, so we already know what expectation of theta is. That's 1/2. And expectation of x was 1/4. So the only thing that we don't know is expectation of the product of the two. So once again, let's try to use iterated expectations. So let's calculate this as the expectation of this conditional expectation. So we, again, condition on theta.

And minus the expectation of theta is 1/2. Times 1/4, which is the expectation of x. Now, what is this conditional expectation?

Well, the expectation of theta-- if you know what theta is, then the expectation of theta is just theta. You already know what it is, so you know for sure that the expectation is just equal to theta. And what is the expectation of x given theta?

Well, the expectation of x given theta we already said was theta over 2. So what you get is this entire expression is just going to be equal to theta times theta over 2, or expectation of theta squared over 2 minus 1/8.

Now, what is the expectation of theta squared over 2? Well, we know that-- we already calculated out what expectation of theta squared is. So we know that expectation of theta squared is 1/12 plus 1/4. So what we get is we need a 1/2 times 1/12 plus 1/4, which is 1/3 minus 1/8. So the answer is 1/6 minus 1/8, which is 1/24.

Now, let's actually plug this in and figure out what this value is. So when you get everything-- when you combine everything, you get that the LMS estimator is-- the linear LMS estimator is going to be-- expectation of theta is 1/2. The covariance is 1/24. The variance is 7/144. And when you divide that, it's equal to 6/7 times x minus 1/4 because expectation of x is 1/4.

And you can simplify this a little bit and get that this is equal to 6/7 times x plus 2/7. So now we have three different types of estimators. The map estimator, which is this. Notice that it's kind of complicated. You have x squared terms. You have more x squared terms. And you have absolute value of log of x. And then you have the LMS, which is, again, nonlinear.

And now you have something that looks very simple-- much simpler. It's just 6/7 x plus 2/7. And that is the linear LMS estimator.

And it turns out that you can, again, plot these to see what this one looks like. So here is our original plot of x and theta hat. So the map estimator-- sorry, the map estimator was just theta hat equals x. This was the mean squared error of the map estimator. So the map estimator is just this diagonal straight line.

The LMS estimator looked like this. And it turns out that the linear LMS estimator will look something like this. So it is fairly close to the LMS estimator, but not quite the same.

And note, especially that depending on what x is, if x is fairly close to the 1, you might actually get an estimate of theta that's greater than 1. So for example, if you observe that Julian is actually an hour late, then x is 1 and your estimate of theta from the linear LMS estimator would be 8/7, which is greater than 1.

That doesn't quite make sense because we know that theta is bounded to be only between 0 and 1. So you shouldn't get an estimate of theta that's greater than 1. And that's one of the side effects of having the linear LMS estimator. So that sometimes you will have an estimator that doesn't quite make sense.

But what you get instead when sacrificing that is you get a simple form of the estimator that's linear. And now, let's actually consider what the performance is.

And it turns out that the performance in terms of the conditional mean squared error is actually fairly close to the LMS estimator. So it looks like this. Pretty close, pretty close, until you get close to 1. In which case, it does worse. And it does worse precisely because it will come up with estimates of theta which are greater than 1, which are too large. But otherwise, it does pretty well with a estimator that is much simpler in form than the LMS estimator.

So in this problem, which had several parts, we actually went through, basically, all the different concepts and tools within Chapter Eight for Bayesian inference. We talked about the prior, the posterior, calculating the posterior using the Bayes' rule. We calculated the MAP estimator. We calculated the LMS estimator. From those, we calculated what the mean squared error for each one of those and compared the two.

And then, we looked at the linear LMS estimator as another example and calculated what that estimator is, along with the mean squared error for that and compared all three of these. So I hope that was a good review problem for Chapter Eight, and we'll see you next time.