6.041SC Probabilistic Systems Analysis and Applied Probability, Fall 2013 Transcript – Tutorial: Using the Conditional Expectation and Variance

Hey guys. Welcome back. Today we're going to do a fun problem that will test your knowledge of the law of total variance. And in the process, we'll also get more practice dealing with joint PDFs and computing conditional expectations and conditional variances.

So in this problem, we are given a joint PDF for x and y. So we're told that x and y can take on the following values in the shape of this parallelogram, which I've drawn. And moreover, that x and y are uniformly distributed. So the joint PDF is just flat over this parallelogram. And because the parallelogram has an area of 1, the height of the PDF must also be 1 so that the PDF integrates to 1. OK.

And then we are asked to compute the variance of x plus y. So you can think of x plus y as a new random variable whose variance we want to compute. And moreover, we're told we should compute this variance by using something called the law of total variance. So from lecture, you should remember or you should recall that the law of total variance can be written in these two ways.

And the reason why there's two different forms for this case is because the formula always has you conditioning on something. Here we condition on x, here we condition on y. And for this problem, the logical choice you have for what to condition on is x or y. So again, we have this option. And my claim is that we should condition on x. And the reason has to do with the geometry of this diagram.

So notice that if you freeze an x and then you sort of vary x, the width of this parallelogram stays constant. However, if you condition on y and look at the width this way, you see that the width of the slices you get by conditioning vary with y. So to make our lives easier, we're going to condition on x. And I'm going to erase this bottom one, because we're not using it.

So this really can seem quite intimidating, because we have nested variances and expectations going on, but we'll just take it slowly step by step. So first, I want to focus on this term-- the conditional expectation of x plus y conditioned on x. So coming back over to this picture, if you fix an arbitrary x in the interval, 0 to 1, we're restricting ourselves to this universe. So y can only vary between this point and this point.

Now, I've already written down here that the formula for this line is given by y is equal to x. And the formula for this line is given by y is equal to x plus 1. So in particular, when we condition on x, we know that y varies between x and x plus 1. But we actually know more than that. We know that in the unconditional universe, x and y were uniformly distributed. So it follows that in the conditional universe, y should also be uniformly distributed, because conditioning doesn't change the relative frequency of outcomes.

So that reasoning means that we can draw the conditional PDF of y conditioned on x as this. We said it varies between x and x plus 1. And we also said that it's uniform, which means that it must have a height of 1. So this is py given x, y given x.

Now, you might be concerned, because, well, we're trying to compute the expectation of x plus y and this is the conditional PDF of y, not of the random variable, x plus y. But I claim that we're OK, this is still useful, because if we're conditioning on x, this x just acts as a constant. It's not really going to change anything except shift the expectation of y by an amount of x. So what I'm saying in math terms is that this is actually just x plus the expectation of y given x.

And now our conditional PDF comes into play. Conditioned on x, this is the PDF of y. And because it's uniformly distributed and because expectation acts like center of mass, we know that the expectation should be the midpoint, right? And so to compute this point, we simply take the average of the endpoints, x plus 1 plus x over 2, which gives us 2x plus 1 over 2. So plugging this back up here, we get 2x/2 plus 2x plus 1 over 2, which is 4x plus 1 over 2, or 2x plus 1/2. OK.

So now I want to look at the next term, the next inner term, which is this guy. So this computation is going to be very similar in nature, actually. So we already discussed that the joint-- sorry, not the joint, the conditional PDF of y given x is this guy. So the variance of x plus y conditioned on x, we sort of have a similar phenomenon occurring. x now in this conditional world just acts like a constant that shifts the PDF but doesn't change the width of the distribution at all. So this is actually just equal to the variance of y given x, because constants don't affect the variance.

And now we can look at this conditional PDF to figure out what this is. So we're going to take a quick tangent over here, and I'm just going to remind you guys that we have a formula for computing the variance of a random variable when it's uniformly distributed between two endpoints. So say we have a random variable whose PDF looks something like this.

Let's call it, let's say, w. This is pww. We have a formula that says variance of w is equal to b minus a squared over 12. So we can apply that formula over here. b is x plus 1, a is x. So b minus a squared over 12 is just 1/12. So we get 1/12. So we're making good progress, because we have this inner quantity and this inner quantity.

So now all we need to do is take the outer variance and the outer expectation. So writing this all down, we get variance of x plus y is equal to variance of this guy, 2x plus 1/2 plus the expectation of 1/12. So this term is quite simple. We know that the expectation of a constant or of a scalar is simply that scalar. So this evaluates to 1/12.

And this one is not bad either. So similar to our discussion up here, we know constants do not affect variance. You know they shift your distribution, they don't change the variance. So we can ignore the 1/2. This scaling factor of 2, however, will change the variance. But we know how to handle this already from previous lectures. We know that you can just take out this scalar scaling factor as long as we square it. So this becomes 2 squared, or 4 times the variance of x plus 1/12.

And now to compute the variance of x, we're going to use that formula again, and we're going to use this picture. So here we have the joint PDF of x and y, but really we want now the PDF of x, so we can figure out what the variance is. So hopefully you remember a trick we taught you called marginalization. To get the PDF of x given a joint PDF, you simply marginalize over the values of y.

So if you freeze x is equal to 0, you get the probability density line over x by integrating over this interval, over y. So if you integrate over this strip, you get 1. If you move x over a little bit and you integrate over this strip, you get 1. This is the argument I was making earlier that the width of this interval stays the same, and hence, the variance stays the same. So based on that argument, which was slightly hand wavy, let's come over here and draw it.

We're claiming that the PDF of x, px of x, looks like this. It's just uniformly distributed between 0 and 1. And if you buy that, then we're done, we're home free, because we can apply this formula, b minus a squared over 12, gives us the variance. So b is 1, a is 0, which gives variance of x is equal to 1/12. So coming back over here, we get 4 times 1/12 plus 1/12, which is 5/12. And that is our answer.

So this problem was straightforward in the sense that our task was very clear. We had to compute this, and we had to do so by using the law of total variance. But we sort of reviewed a lot of concepts along the way. We saw how, given a joint PDF, you marginalize to get the PDF of x. We saw how constants don't change variance. We got a lot of practice finding conditional distributions and computing conditional expectations and variances.

And we also saw this trick. And it might seem like cheating to memorize formulas, but there's a few important ones you should know. And it will help you sort of become faster at doing computations. And that's important, especially if you guys take the exams. So that's it. See you next time.

6.041SC Probabilistic Systems Analysis and Applied Probability
Fall 2013