CIS 2033, Spring 2017

Instructor: David Dobor

April 22, 2017

We will now go through an example that brings together all of the concepts that we have introduced in this part of the course – marginal PDFs, joint PDFs, conditional PDFs and the relations between them. The example also offers you a chance to practice with the calculations of expectations and conditional expectations, as well as with the total probability theorem.

Stick-breaking example

We have a stick of length l. We break the stick at some random location, which corresponds to a random variable, X.

We assume that this random variable X is uniform over the length of the stick, so its PDF has the shape shown in Figure 1. Of course, for the PDF to integrate to 1, the height of this PDF must be equal to $1/l$.

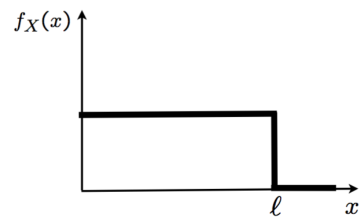

Figure 1: The PDF of X, the break point of the stick. X is assumed to be uniformly distributed over the length of the stick and therefore $f_X(x) = 1/l$.

Then we take the piece of the stick that we are left with, which has length X, and we break it at a random location, which we call Y. We assume that this location Y is uniformly distributed over the length of the stick *that we were left with*, and we are ultimately interested in $\mathbf{E}[Y]$, the expected length of the stick that we are left with.

What does this assumption mean?

It means that if the first break was at some particular value, x, then the random variable Y has a conditional distribution, which is uniform over the interval from 0 to x. So the *conditional* PDF is uniform. A conditional PDF, like any other PDF, must integrate to 1. So the height of this conditional PDF, shown in Figure 2, is constant and is equal to $1/x$.

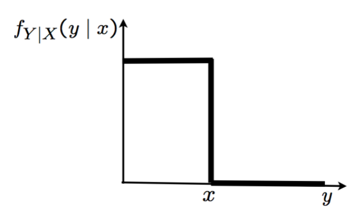

Figure 2: The conditional PDF of the second break point of the stick. This second break point, denoted by Y, is assumed to be uniformly distributed over the length of the remaining part of the stick, which is x, and therefore $f_{Y|X}(y) = 1/x$.

Are X and Y independent? No. One way to see it is that if you change little x, the conditional PDF of Y would have been something different. Whereas if we have independence, all the conditional PDFs have to be the same when you change the value of little x.

Another way to see it is that if somebody tells you that x is 0.5, this gives you lots of information about Y. It tells you that Y has to be less than or equal to 0.5. So the value of the random variable X gives you plenty of information about the other random variable Y, which means that we do not have independence.

Notice that in this example, instead of starting with a full description of the random variables in terms of a joint PDF, we use a marginal PDF and then a conditional PDF to construct our model. Of course, with these two pieces of information, we can reconstruct the joint PDF using the multiplication rule:

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(x \mid y) = \frac{1}{lx}$$

But for which values of x and y is this the correct expression? It is correct only for those values that are possible. So y must be to the left of x, because of the way we broke the stick in two parts, and of course both x and y are between 0 and l:

$$0 \leq y \leq x \leq l$$

We can visualize these values. These values correspond to the shaded triangle in Figure 3.

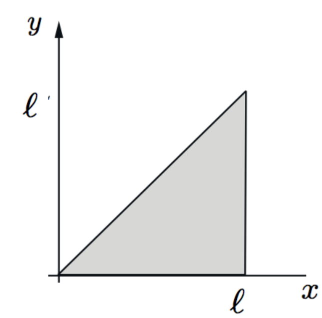

Figure 3: The joint distribution of X and Y is non-zero over the shaded region shown here.

Now take a look at Figure 4 and try to visualize the joint PDF. Consider any one of the "slices" along the red lines shown in this figure. Notice that along any one of these slices, since the PDF only depends on x and not on y, the value of the joint PDF is going to be a constant.

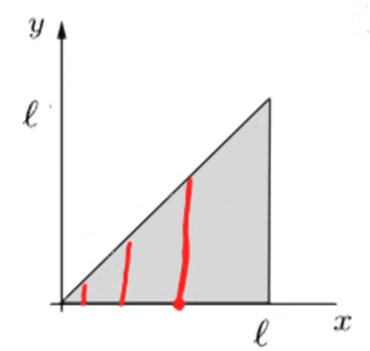

Figure 4: The joint PDF of X and Y is constant over the red "slices" shown here, and the smaller the length of the slice, the larger that constant becomes.

Notice also that along the longer slice (the rightmost red line) that constant (the value of the PDF) is going to be smaller than along the shorter slices. On the shortest slice that constant is going to be the largest. And actually, this constant is bigger and bigger and goes to infinity as the length of the slice approaches 0. Of course, this is just a reflection of the fact that the conditional PDF is constant over the range of values that the random variables can take.

LET US NOW continue with some calculations. Let us find the marginal PDF of Y. How do we do it? Since we have in our hands the joint PDF, we can find the marginal by integrating the joint. And in our case, the joint is equal to $1/lx$, and so we integrate it over all x's:

$$f_Y(y) = \int f_X(x) f_{Y|X}(x \mid y)\, dx = \int \frac{1}{lx} dx. \tag{1}$$

Now, what is the range of the integration? Take a look at Figure 5. Let's fix a certain value of y. The joint PDF is 0 in the shaded region, so we should only integrate over x's that correspond to this interval. What is that interval?

It's the interval that ends at l. And because the longest side of the shaded triangle has slope 1, the x coordinate of the left end of the interval is also y. So we integrate over an interval where x ranges from y to l. We rewrite (1), then, as

$$f_Y(y) = \int_y^l \frac{1}{lx} dx. \tag{2}$$

Now, the integral of $1/x$ is a logarithm. Using this fact, we can evaluate this integral.

$$f_Y(y) = \int \frac{1}{lx} dx = \frac{1}{l} \ln\left(\frac{l}{y}\right). \tag{3}$$

For what y's is this a correct expression? Well, it makes sense only for those y's that are possible in this experiment, and that's the range from 0 to l. The plot of the marginal PDF of Y, of $f_Y(y)$ is shown in Figure 6.

Is it a problem having a PDF that blows up to infinity? Not really. As long as the area under this PDF is equal to 1, it's still a legitimate PDF, and blowing up to infinity is not an issue.

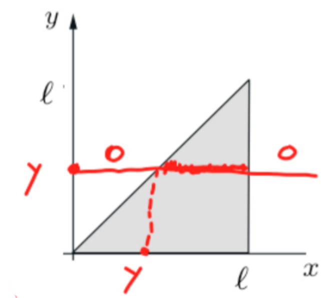

Figure 5: The region of integration where the marginal PDF of Y, $f_Y(y)$, is non-zero is shown as the red solid line over shaded triangle.

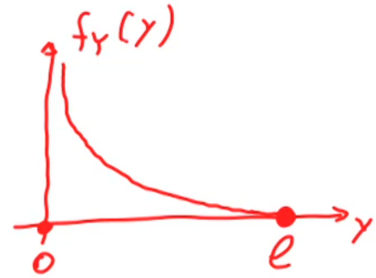

Figure 6: A plot of the marginal PDF of Y, $f_Y(y)$. Note that it "blows up" to infinity as y approaches 0. Of course this is not a problem as long as the area under the curve integrates to 1, making $f_Y(y)$ a valid PDF.

LET US NOW calculate the expected value of Y. One way of doing this is by using the definition of the expectation. By definition, it is the following integral over those values for which we have a non-zero density:

$$\mathbf{E}[Y] = \int_0^l y \frac{1}{l} \ln\left(\frac{l}{y}\right) dy.$$

This integral is pretty messy. One can actually integrate it using integration by parts. But the calculation is a bit tedious. So let us look for an alternative and more clever approach.

The idea is to divide and conquer. We're going to use the *total expectation theorem*, where we're going to condition on X. The total expectation theorem tells us that the expected value of Y is the integral over all possible values of the random variable X, which is from 0 to l:

$$\mathbf{E}[Y] = \int_0^l \frac{1}{l}\, \mathbf{E}[Y \mid X = x]\, dx. \tag{4}$$

Why is this simpler? When we condition on X taking a specific value, Y has a uniform distribution between 0 and x. And therefore, this conditional expectation is the expectation of a uniform, which is 1/2 the range of that uniform.

So we can continue with the calculation in (4) as follows:

$$\mathbf{E}[Y] = \int_0^l \frac{1}{l}\,\mathbf{E}[Y \mid X = x]\,dx \tag{5}$$

$$= \int_0^l \frac{1}{l} \cdot \frac{x}{2}\,dx \tag{6}$$

Finally, this is an integral that we can evaluate easily.

Or we can think even in a simpler way. Inside the last integral, the expression $1/l$ is the density of X. So we can calculate (6) as follows:

$$\begin{aligned}
\mathbf{E}[Y] &= \int_0^l \frac{1}{l} \cdot \frac{x}{2}\,dx \\
&= \frac{1}{2}\int_0^l \frac{1}{l} \cdot x\,dx \\
&= \frac{1}{2}\int_0^l x \cdot f_X(x)\,dx \\
&= \frac{1}{2} \cdot \mathbf{E}[X] \\
&= \frac{1}{2} \cdot \frac{l}{2} \\
&= \frac{l}{4}
\end{aligned}$$

And so this is our final answer, $\mathbf{E}[Y] = l/4$. This answer makes intuitive sense. If we break a stick once, the expected value or what we're left with is half of what we started with. But if we break it once more, then we expect it on the average to be cut by a factor again of 1/2. And so we expect to be left with a stick that has length 1/4 of what we started with.

This example is a particularly nice one, because we used all of the concepts that we have introduced– marginal PDFs, joint PDFs, conditional PDFs, the relations between them, as well as expectations, calculations of expectations, and conditional expectations, as well as the total probability theorem.