6.041SC Probabilistic Systems Analysis and Applied Probability, Fall 2013 Transcript – Lecture 9

JOHN TSITSIKLIS: OK let's start. So we've had the quiz. And I guess there's both good and bad news in it. Yesterday, as you know, the bad news. The average was a little lower than what we would have wanted. On the other hand, the good news is that the distribution was nicely spread. And that's the main purpose of this quiz is basically for you to calibrate and see roughly where you are standing.

The other piece of the good news is that, as you know, this quiz doesn't count for very much in your final grade. So it's really a matter of calibration and to get your mind set appropriately to prepare for the second quiz, which counts a lot more. And it's more substantial. And we'll make sure that the second quiz will have a higher average.

All right. So let's go to our material. We're talking now these days about continuous random variables. And I'll remind you what we discussed last time. I'll remind you of the concept of the probability density function of a single random variable. And then we're going to rush through all the concepts that we covered for the case of discrete random variables and discuss their analogs for the continuous case. And talk about notions such as conditioning independence and so on.

So the big picture is here. We have all those concepts that we developed for the case of discrete random variables. And now we will just talk about their analogs in the continuous case. We already discussed this analog last week, the density of a single random variable.

Then there are certain concepts that show up both in the discrete and the continuous case. So we have the cumulative distribution function, which is a description of the probability distribution of a random variable and which applies whether you have a discrete or continuous random variable. Then there's the notion of the expected value.

And in the two cases, the expected value is calculated in a slightly different way, but not very different. We have sums in one case, integrals in the other. And this is the general pattern that we're going to have. Formulas for the discrete case translate to corresponding formulas or expressions in the continuous case. We generically replace sums by integrals, and we replace must functions with density functions.

Then the new pieces for today are going to be mostly the notion of a joint density function, which is how we describe the probability distribution of two random variables that are somehow related, in general, and then the notion of a conditional density function that tells us the distribution of one random variable X when you're told the value of another random variable Y. There's another concept, which is the conditional PDF given that the certain event has happened.

This is a concept that's in some ways simpler. You've already seen a little bit of that in last week's recitation and tutorial.

The idea is that we have a single random variable. It's described by a density. Then you're told that the certain event has occurred. Your model changes the universe that you are dealing with. In the new universe, you are dealing with a new density function, the one that applies given the knowledge that we have that the certain event has occurred.

All right. So what exactly did we say about continuous random variables? The first thing is the definition, that a random variable is said to be continuous if we are given a certain object that we call the probability density function and we can calculate interval probabilities given this density function. So the definition is that the random variable is continuous if you can calculate probabilities associated with that random variable given that formula. So this formula tells you that the probability that your random variable falls inside this interval is the area under the density curve.

OK. There's a few properties that a density function must satisfy. Since we're talking about probabilities, and probabilities are non-negative, we have that the density function is always a non-negative function. The total probability over the entire real line must be equal to 1. So the integral when you integrate over the entire real line has to be equal to 1. That's the second property.

Another property that you get is that if you let a equal to b, this integral becomes 0. And that tells you that the probability of a single point in the continuous case is always equal to 0. So these are formal properties.

When you want to think intuitively, the best way to think about what the density function is to think in terms of little intervals, the probability that my random variable falls inside the little interval. Well, inside that little interval, the density function here is roughly constant. So that integral becomes the value of the density times the length of the interval over which you are integrating, which is delta.

And so the density function basically gives us probabilities of little events, of small events. And the density is to be interpreted as probability per unit length at a certain place in the diagram. So in that place in the diagram, the probability per unit length around this neighborhood would be the height of the density function at that point.

What else? We have a formula for calculating expected values of functions of random variables. In the discrete case, we had the formula where here we had the sum, and instead of the density, we had the PMF. The same formula is also valid in the continuous case. And it's not too hard to derive, but we will not do it.

But let's think of the intuition of what this formula says. You're trying to figure out on the average how much g(X) is going to be. And then you reason, and you say, well, X may turn out to take a particular value or a small interval of values. This is the probability that X falls inside the small interval. And when that happens, g(X) takes that value.

So this fraction of the time, you fall in the little neighborhood of x, and you get so much. Then you average over all the possible x's that can happen. And that gives you the average value of the function g(X).

OK. So this is the easy stuff. Now let's get to the new material. We want to talk about multiple random variables simultaneously. So we want to talk now about two random variables that are continuous, and in some sense that they are jointly continuous.

And let's see what this means. The definition is similar to the definition we had for a single random variable, where I take this formula here as the definition of continuous random variables. Two random variables are said to be jointly continuous if we can calculate probabilities by integrating a certain function that we call the joint density function over the set of interest.

So we have our two-dimensional plane. This is the x-y plane. There's a certain event S that we're interested in. We want to calculate the probability. How do we do that?

We are given this function f_(X,Y), the joint density. It's a function of the two arguments x and y. So think of that function as being some kind of surface that sits on top of the two-dimensional plane. The probability of falling inside the set S, we calculate it by looking at the volume under the surface, that volume that sits on top of S. So the surface underneath it has a certain total volume.

What should that total volume be? Well, we think of these volumes as probabilities. So the total probability should be equal to 1. The total volume under this surface, should be equal to 1. So that's one property that we want our density function to have. So when you integrate over the entire space, this is of the volume under your surface. That should be equal to 1. Of course, since we're talking about probabilities, the joint density should be a non-negative function.

So think of the situation as having one pound of probability that's spread all over your space. And the height of this joint density function basically tells you how much probability tends to be accumulated in certain regions of space as opposed to other parts of the space. So wherever the density is big, that means that this is an area of the two-dimensional plane that's more likely to occur. Where the density is small, that means that those x-y's are less likely to occur.

You have already seen one example of continuous densities. That was the example we had in the very beginning of the class with a uniform distribution on the unit square. That was a special case of a density function that was constant. So all places in the unit square were roughly equally likely as any other places.

But in other models, some parts of the space may be more likely than others. And we describe those relative likelihoods using this density function. So if somebody gives us the density function, this determines for us probabilities of all the subsets of the two-dimensional plane.

Now for an intuitive interpretation, it's good to think about small events. So let's take a particular x here and then x plus delta. So this is a small interval. Take another small interval here that goes from y to y plus delta. And let's look at the event that x falls here and y falls right there.

What is this event? Well, this is the event that will fall inside this little rectangle. Using this rule for calculating probabilities, what is the probability of that rectangle going to be? Well, it should be the integral of the density over this rectangle. Or it's the volume under the surface that sits on top of that rectangle.

Now, if the rectangle is very small, the joint density is not going to change very much in that neighborhood. So we can treat it as a constant. So the volume is going to be the height times the area of the base. The height at that point is whatever the function happens to be around that point. And the area of the base is delta squared.

So this is the intuitive way to understand what a joint density function really tells you. It specifies for you probabilities of little squares, of little rectangles. And it allows you to think of the joint density function as probability per unit area. So these are the units of the density, its probability per unit area in the neighborhood of a certain point.

So what do we do with this density function once we have it in our hands? Well, we can use it to calculate expected values. Suppose that you have a function of two random variables described by a joint density. You can find, perhaps, the distribution of this random variable and then use the basic definition of the expectation. Or you can calculate expectations directly, using the distribution of the original random variables.

This is a formula that's again identical to the formula that we had for the discrete case. In the discrete case, we had a double sum here, and we had PMFs. So the intuition behind this formula is the same that one had for the discrete case. It's just that the mechanics are different.

Then something that we did in the discrete case was to find a way to go from the joint density of the two random variables taken together to the density of just one of the random variables. So we had a formula for the discrete case. Let's see how things are going to work out in the continuous case.

So in the continuous case, we have here our two random variables. And we have a density for them. And let's say that we want to calculate the probability that x falls inside this interval. So we're looking at the probability that our random variable X falls in the interval from little x to x plus delta.

Now, by the properties that we already have for interpreting the density function of a single random variable, the probability of a little interval is approximately the density of that single random variable times delta. And now we want to find a formula for this marginal density in terms of the joint density.

OK. So this is the probability that x falls inside this interval. In terms of the two-dimensional plane, this is the probability that (x,y) falls inside this strip. So to find that probability, we need to calculate the probability that (x,y) falls in here, which is going to be the double integral over the interval over this strip, of the joint density.

And what are we integrating over? y goes from minus infinity to plus infinity. And the dummy variable x goes from little x to x plus delta. So to integrate over this strip, what we do is for any given y, we integrate in this dimension. This is the x integral. And then we integrate over the y dimension.

Now what is this inner integral? Because x only varies very little, this is approximately constant in that range. So the integral with respect to x just becomes delta times f(x,y). And then we've got our dy.

So this is what the inner integral will evaluate to. We are integrating over the little interval. So we're keeping y fixed. Integrating over here, we take the value of the density times how much we're integrating over. And we get this formula.

OK. Now, this expression must be equal to that expression. So if we cancel the deltas, we see that the marginal density must be equal to the integral of the joint density, where we have integrated out the value of y. So this formula should come as no surprise at this point.

It's exactly the same as the formula that we had for discrete random variables. But now we are replacing the sum with an integral. And instead of using the joint PMF, we are using the joint PDF.

Then, continuing going down the list of things we did for discrete random variables, we can now introduce a definition of the notion of independence of two random variables. And by analogy with the discrete case, we define independence to be the following condition. Two random variables are independent if and only if their joint density function factors out as a product of their marginal densities. And this property needs to be true for all x and y. So this is the formal definition.

Operationally and intuitively, what does it mean? Well, intuitively it means the same thing as in the discrete case. Knowing anything about X shouldn't tell you anything about Y. That is, information about X is not going to change your beliefs about Y. We are going to come back to this statement in a second.

The other thing that it allows you to do-- I'm not going to derive this-- is it allows you to calculate probabilities by multiplying individual probabilities. So if you ask for the probability that x falls in a certain set A and y falls in a certain set B, then you can calculate that probability by multiplying individual probabilities. This takes just two lines of derivation, which I'm not going to do. But it comes back to the usual notion of independence of events. Basically, operationally independence means that you can multiply probabilities.

So now let's look at an example. There's a sort of pretty famous and classical one. It goes back a lot more than a 100 years. And it's the famous Needle of Buffon. Buffon was a French naturalist who, for some reason, also decided to play with probability. And look at the following problem.

So you have the two-dimensional plane. And on the plane we draw a bunch of parallel lines. And those parallel lines are separated by a length. And the lines are apart at distance d. And we throw

a needle at random, completely at random. And we'll have to give a meaning to what "completely at random" means.

And when we throw a needle, there's two possibilities. Either the needle is going to fall in a way that does not intersect any of the lines, or it's going to fall in a way that it intersects one of the lines. We're taking the needle to be shorter than this distance, so the needle cannot intersect two lines simultaneously. It either intersects 0, or it intersects one of the lines. The question is to find the probability that the needle is going to intersect a line.

What's the probability of this? OK. We are going to approach this problem by using our standard four-step procedure. Set up your sample space, describe a probability law on that sample space, identify the event of interest, and then calculate. These four steps basically correspond to these three bullets and then the last equation down here.

So first thing is to set up a sample space. We need some variables to describe what happened in the experiment. So what happens in the experiment is that the needle lands somewhere. And where it lands, we can describe this by specifying the location of the center of the needle.

And what do we mean by the location of the center? Well, we can take as our variable to be the distance from the center of the needle to the nearest line. So it tells us the vertical distance of the center of the needle from the nearest line.

The other thing that matters is the orientation of the needle. So we need one more variable, which we take to be the angle that the needle is forming with the lines. We can put the angle here, or you can put in there. Yes, it's still the same angle.

So we have these two variables that described what happened in the experiment. And we can take our sample space to be the set of all possible x's and theta's. What are the possible x's? The lines are d apart, so the nearest line is going to be anywhere between 0 and d/2 away. So that tells us what the possible x's will be.

As for theta, it really depends how you define your angle. We are going to define our theta to be the acute angle that's formed between the needle and a line, if you were to extend it. So theta is going to be something between 0 and pi/2. So I guess these red pieces really correspond to the part of setting up the sample space. OK. So that's part one.

Second part is we need a model. OK. Let's take our model to be that we basically know nothing about how the needle falls. It can fall in any possible way, and all possible ways are equally likely.

Now, if you have those parallel lines, and you close your eyes completely and throw a needle completely at random, any x should be equally likely. So we describe that situation by saying that X should have a uniform distribution. That is, it should have a constant density over the range of interest.

Similarly, if you kind of spin your needle completely at random, any angle should be as likely as any other angle. And we decide to model this situation by saying that theta also has a uniform distribution over the range of interest. And finally, where we put it should have nothing to do with how much we rotate it. And we capture this mathematically by saying that X is going to be independent of theta.

Now, this is going to be our model. I'm not deriving the model from anything. I'm only saying that this sounds like a model that does not assume any knowledge or preference for certain values of x rather than other values of theta. In the absence of any other particular information you might have in your hands, that's the most reasonable model to come up with. So you model the problem that way.

So what's the formula for the joint density? It's going to be the product of the densities of X and Theta. Why is it the product? This is because we assumed independence.

And the density of X, since it's uniform, and since it needs to integrate to 1, that density needs to be 2/d. That's the density of X. And the density of Theta needs to be 2/pi. That's the value for the density of Theta so that the overall probability over this interval ends up being 1.

So now we do have our joint density in our hands. The next thing to do is to identify the event of interest. And this is best done in a picture. And there's two possible situations that one could have. Either the needle falls this way, or it falls this way.

So how can we tell if one or the other is going to happen? It has to do with whether this interval here is smaller than that or bigger than that. So we are comparing the height of this interval to that interval. This interval here is capital X.

This interval here, what is it? This is half of the length of the needle, which is l/2. To find this height, we take l/2 and multiply it with the sine of the angle that we have. So the length of this interval up here is l/2 times sine theta.

If this is smaller than x, the needle does not intersect the line. If this is bigger than x, then the needle intersects the line. So the event of interest, that the needle intersects the line, is described this way in terms of x and theta.

And now that we have the event of interest described mathematically, all that we need to do is to find the probability of this event, we integrate the joint density over the part of (x, theta) space in which this inequality is true. So it's a double integral over the set of all x's and theta's where this is true. The way to do this integral is we fix theta, and we integrate for x's that go from 0 up to that number. And theta can be anything between 0 and pi/2. So the integral over this set is basically this double integral here.

We already have a formula for the joint density. It's 4 over pi d, so we put it here. And now, fortunately, this is a pretty easy integral to evaluate. The integral with respect to x -- there's nothing in here. So the integral is just the length of the interval over which we're integrating. It's l/2 sine theta.

And then we need to integrate this with respect to theta. We know that the integral of a sine is a negative cosine. You plug in the values for the negative cosine at the two end points. I'm sure you can do this integral . And we finally obtain the answer, which is amazingly simple for such a pretty complicated-looking problem. It's 2l over pi d.

So some people a long, long time ago, after they looked at this answer, they said that maybe that gives us an interesting way where one could estimate the value by pi, for example, experimentally. How do you do that? Fix l and d, the dimensions of the problem. Throw a million needles on your piece of paper. See how often your needless do intersect the line.

That gives you a number for this quantity. You know l and d, so you can use that to infer pi. And there's an apocryphal story about a wounded soldier in a hospital after the American Civil War who actually had heard about this and was spending his time in the hospital throwing needles on pieces of paper. I don't know if it's true or not. But let's do something similar here.

So let's look at this diagram. We fix the dimensions. This is supposed to be our little d. That's supposed to be our little l. We have the formula from the previous slide that p is 2l over pi d. In this instance, we choose d to be twice l. So this number is 1/pi. So the probability that the needle hits the line is 1/pi.

So I need needles that are 3.1 centimeters long. I couldn't find such needles. But I could find paper clips that are 3.1 centimeters long. So let's start throwing paper clips at random and see how many of them will end up intersecting the lines. Good.

OK. So out of eight paper clips, we have exactly four that intersected the line. So our estimate for the probability of intersecting the line is 1/2, which gives us an estimate for the value of pi, which is two. Well, I mean, within an engineering approximation, we're in the right ballpark, right?

So this might look like a silly way of trying to estimate pi. And it probably is. On the other hand, this kind of methodology is being used especially by physicists and also by statisticians. It's used a lot. When is it used?

If you have an integral to calculate, such as this integral, but you're not lucky, and your functions are not so simple where you can do your calculations by hand, and maybe the dimensions are larger-- instead of two random variables you have 100 random variables, so it's a 100-fold integral-- then there's no way to do that in the computer. But the way that you can actually do it is by generating random samples of your random variables, doing that simulation over and over many times. That is, by interpreting an integral as a probability, you can use simulation to estimate that probability. And that gives you a way of calculating integrals.

And physicists do actually use that a lot, as well as statisticians, computer scientists, and so on. It's a so-called Monte Carlo method for evaluating integrals. And it's a basic piece of the toolbox in science these days.

Finally, the harder concept of the day is the idea of conditioning. And here things become a little subtle when you deal with continuous random variables. OK. First, remember again our basic interpretation of what a density is. A density gives us probabilities of little intervals.

So how should we define conditional densities? Conditional densities should again give us probabilities of little intervals, but inside a conditional world where we have been told something about the other random variable. So what we would like to be true is the following. We would like to define a concept of a conditional density of a random variable X given the value of another random variable Y. And it should behave the following way, that the conditional density gives us the probability of little intervals-- same as here-- given that we are told the value of y.

And here's where the subtleties come. The main thing to notice is that here I didn't write "equal," I wrote "approximately equal." Why do we need that?

Well, the thing is that conditional probabilities are not defined when you condition on an event that has 0 probability. So we need the conditioning event here to have posed this probability. So instead of saying that Y is exactly equal to little y, we want to instead say we're in a new universe where capital Y is very close to little y.

And then this notion of "very close" kind of takes the limit and takes it to be infinitesimally close. So this is the way to interpret conditional probabilities. That's what they should mean.

Now, in practice, when you actually use probability, you forget about that subtlety. And you say, well, I've been told that Y is equal to 1.3. Give me the conditional distribution of X. But formally or rigorously, you should say I'm being told that Y is infinitesimally close to 1.3. Tell me the distribution of X.

Now, if this is what we want, what should this quantity be? It's a conditional probability, so it should be the probability of two things happening-- X being close to little x, Y being close to little y. And that's basically given to us by the joint density divided by the probability of the conditioning event, which has something to do with the density of Y itself. And if you do things carefully, you see that the only way to satisfy this relation is to define the conditional density by this particular formula.

OK. Big discussion to come down in the end to what you should have probably guessed by now. We just take any formulas and expressions from the discrete case and replace PMFs by PDFs. So the conditional PDF is defined by this formula where here we have joint PDF and marginal PDF, as opposed to the discrete case where we had the joint PMF and the marginal PMF. So in some sense, it's just a syntactic change.

In another sense, it's a little subtler on how you actually interpret it. Speaking about interpretation, what are some ways of thinking about the joint density? Well, the best way to think about it is that somebody has fixed little y for you. So little y is being fixed here. And we look at this density as a function of X.

I've told you what Y is. Tell me what you know about X. And you tell me that X has a certain distribution. What does that distribution look like? It has exactly the same shape as the joint density.

Remember, we fixed Y. So this is a constant. So the only thing that varies is X. So we get the function that behaves like the joint density when you fix y, which is really you take the joint density, and you take a slice of it.

You fix a y, and you see how it varies with x. So in that sense, the conditional PDF is just a slice of the joint PDF. But we need to divide by a certain number, which just scales it and changes its shape.

We're coming back to a picture in a second. But before going to the picture, lets go back to the interpretation of independence. If the two random the variables are independent, according to our definition in the previous slide, the joint density is going to factor as the product of the marginal densities. The density of Y in the numerator cancels the density in the denominator. And we're just left with the density of X.

So in the case of independence, what we get is that the conditional is the same as the marginal. And that solidifies our intuition that in the case of independence, being told something about the value of Y does not change our beliefs about how X is distributed. So whatever we expected about X is going to remain true even after we are told something about Y.

So let's look at some pictures. Here is what the joint PDF might look like. Here we've got our x and y-axis. And if you want to calculate the probability of a certain event, what you do is you look at that event and you see how much of that mass is sitting on top of that event.

Now let's start slicing. Let's fix a value of x and look along that slice where we obtain this function. Now what does that slice do? That slice tells us for that particular x what the possible values of y are going to be and how likely they are.

If we integrate over all y's, what do we get? Integrating over all y's just gives us the marginal density of X. It's the calculation that we did here.

By integrating over all y's, we find the marginal density of X. So the total area under that slice gives us the marginal density of X. And by looking at the different slices, we find how likely the different values of x are going to be.

How about the conditional? If we're interested in the conditional of Y given X, how would you think about it? This refers to a universe where we are told that capital X takes on a specific value.

So we put ourselves in the universe where this line has happened. There's still possible values of y that can happen. And this shape kind of tells us the relative likelihoods of the different y's. And this is indeed going to be the shape of the conditional distribution of Y given that X has occurred.

On the other hand, the conditional distribution must add up to 1. So the total probability over all of the different y's in this universe, that total probability should be equal to 1. Here it's not equal to 1. The total area is the marginal density. To make it equal to 1, we need to divide by the marginal density, which is basically to renormalize this shape so that the total area under that slice, under that shape, is equal to 1.

So we start with the joint. We take the slices. And then we adjust the slices so that every slice has an area underneath equal to 1. And this gives us the conditional.

So for example, down here-- you can not even see it in this diagram-- but after you renormalize it so that its total area is equal to 1, you get this sort of narrow spike that goes up. And so this is a plot of the conditional distributions that you get for the different values of x. Given a particular value of x, you're going to get this certain conditional distribution.

So this picture is worth about as much as anything else in this particular chapter. Make sure you kind of understand exactly all these pieces of the picture. And finally, let's go, in the remaining time, through an example where we're going to throw in the bucket all the concepts and notations that we have introduced so far. So the example is as follows.

We start with a stick that has a certain length. And we break it a completely random location. And-- yes, this 1 should be l. OK. So it has length l. And we're going to break it at the random place. And we call that random place where we break it, we call it X.

X can be anywhere, uniform distribution. So this means that X has a density that goes from 0 to l. I guess this capital L is supposed to be the same as the lower-case l. So that's the density of X. And since the density needs to integrate to 1, the height of that density has to be 1/l.

Now, having broken the stick and given that we are left with this piece of the stick, I'm now going to break it again at a completely random place, meaning I'm going to choose a point where I break it uniformly over the length of the stick. What does this mean? And let's call Y the location where I break it. So Y is going to range between 0 and x. x is the stick that I'm left with.

So I'm going to break it somewhere in between. So I pick a y between 0 and x. And of course, x is less than l. And I'm going to break it there. So y is uniform between 0 and x.

What does that mean, that the density of y, given that you have already told me x, ranges from 0 to little x? If I told you that the first break happened at a particular x, then y can only range over this interval. And I'm assuming a uniform distribution over that interval. So we have this kind of shape. And that fixes for us the height of the conditional density.

So what's the joint density of those two random variables? By the definition of conditional densities, the conditional was defined as the ratio of this divided by that. So we can find the joint density by taking the marginal and then multiplying by the conditional. This is the same formula as in the discrete case. This is our very familiar multiplication rule, but adjusted to the case of continuous random variables. So Ps become Fs.

OK. So we do have a formula for this. What is it? It's 1/l-- that's the density of X -- times 1/x, which is the conditional density of Y. This is the formula for the joint density.

But we must be careful. This is a formula that's not valid anywhere. It's only valid for the x's and y's that are possible.

And the x's and y's that are possible are given by these inequalities. So x can range from 0 to l, and y can only be smaller than x. So this is the formula for the density on this part of our space. The density is 0 anywhere else.

So what does it look like? It's basically a 1/x function. So it's sort of constant along that dimension. But as x goes to 0, your density goes up and can even blow up. It sort of looks like a sail that's raised and somewhat curved and has a point up there going to infinity. So this is the joint density.

Now once you have in your hands a joint density, then you can answer in principle any problem. It's just a matter of plugging in and doing computations. How about calculating something like a conditional expectation of Y given a value of x? OK. That's a concept we have not defined so far. But how should we define it?

Means the reasonable thing. We'll define it the same way as ordinary expectations except that since we're given some conditioning information, we should use the probability distribution that applies to that particular situation. So in a situation where we are told the value of x, the distribution that applies is the conditional distribution of Y. So it's going to be the conditional density of Y given the value of x.

Now, we know what this is. It's given by 1/x. So we need to integrate y times 1/x dy.

And what should we integrate over? Well, given the value of x, y can only range from 0 to x. So this is what we get. And you do your integral, and you get that this is x/2.

Is it a surprise? It shouldn't be. This is just the expected value of Y in a universe where X has been realized and Y is given by this distribution. Y is uniform between 0 and x. The expected value of Y should be the midpoint of this interval, which is x/2.

Now let's do fancier stuff. Since we have the joint distribution, we should be able to calculate the marginal. What is the distribution of Y? After breaking the stick twice, how big is the little piece that I'm left with? How do we find this?

To find the marginal, we just take the joint and integrate out the variable that we don't want. A particular y can happen in many ways. It can happen together with any x. So we consider all the possible x's that can go together with this y and average over all those x's.

So we plug in the formula for the joint density from the previous slide. We know that it's 1/lx. And what's the range of the x's? So to find the density of Y for a particular y up here, I'm going

to integrate over x's. The density is 0 here and there. The density is nonzero only in this part. So I need to integrate over x's going from here to there.

So what's the "here"? This line goes up at the slope of 1. So this is the line x equals y. So if I fix y, it means that my integral starts from a value of x that is also equal to y.

So where the integral starts from is at x equals y. And it goes all the way until the end of the length of our stick, which is l. So we need to integrate from little y up to l.

So that's something that almost always comes up. It's not enough to have just this formula for integrating the joint density. You need to keep track of different regions. And if the joint density is 0 in some regions, then you exclude those regions from the range of integration. So the range of integration is only over those values where the particular formula is valid, the places where the joint density is nonzero.

All right. The integral of 1/x dx, that gives you a logarithm. So we evaluate this integral, and we get an expression of this kind. So the density of Y has a somewhat unexpected shape. So it's a logarithmic function.

And it goes this way. It's for y going all the way to l. When y is equal to l, the logarithm of 1 is equal to 0. But when y approaches 0, logarithm of something big blows up, and we get a shape of this form.

OK. Finally, we can calculate the expected value of Y. And we can do this by using the definition of the expectation. So integral of y times the density of y. We already found what that density is, so we can plug it in here. And we're integrating over the range of possible y's, from 0 to l.

Now this involves the integral for y log y, which I'm sure you have encountered in your calculus classes but maybe do not remember how to do it. In any case, you look it up in some integral tables or do it by parts. And you get the final answer of l/4.

And at this point, you say, that's a really simple answer. Shouldn't I have expected it to be l/4? I guess, yes. I mean, when you break it once, the expected value of what you are left with is going to be 1/2 of what you started with. When you break it the next time, the expected length of what you're left with should be 1/2 of the piece that you are now breaking.

So each time that you break it at random, you expected it to become smaller by a factor of 1/2. So if you break it twice, you are left something that's expected to be 1/4. This is reasoning on the average, which happens to give you the right answer in this case. But again, there's the warning that reasoning on the average doesn't always give you the right answer. So be careful about doing arguments of this type.

Very good. See you on Wednesday.

6.041SC Probabilistic Systems Analysis and Applied Probability
Fall 2013