[SQUEAKING]

[RUSTLING]

[CLICKING]

SARA ELLISON: OK, a couple of weeks ago, we ended with an example involving computing probabilities from a joint PDF. And it generated a lot of questions you guys had. So I thought it would be useful if I went through another example.

And this one is taken from a DeGroot and Schervish's book. But I think it's a really nice example. It lets me draw some nice 3D pictures of what's going on, which I hope will help your intuition a little bit.

OK, so suppose we have a joint PDF, little f, of x and y. And it's just equal to some constant times x squared y for these values of x and y. And it's equal to 0, otherwise.

Well, let's first start-- before we do any kind of calculations, let's draw a picture of the support of this distribution. What is the support of a distribution? Well, it's the region of the x/y-plane over which there is positive probability.

So just as a first step, let's make sure we understand what the support is. And how do we get from the specification of the joint distribution that I've given you here-- how do we get the support? Where do we look for information on the support?

Right. So it's this statement right here. So this PDF is nonzero for values of y and x, such that x squared is less than or equal to y, which is less than or equal to 1. So the first step is, all we do is draw that support. And this is what it looks like.

So is that clear to everyone where that came from? So basically, I just drew the function y equals x squared. Here's y equals x squared. And then I drew the line y equals 1. And the support is basically everything that's contained between those two functions.

OK, so this is our joint PDF. It's equal to x squared y over that support. So what I've done here is given you a 3D drawing of the actual joint PDF. I've taken the x/y-plane, the y-axis, and rotated it down, so now it's on this horizontal plane along with the x-axis. So this is the support that I drew in the previous slide, right here.

And then the joint PDF is the surface above that describes the probability with which x and y take on various values. So how did I get this? How did I get this picture? Where did this come from?

Right. So it came from the function c times x squared times y. So that's all I've done here is just taken c, some constant, times x squared times y and drawn it over the support.

OK, so, so far, I haven't told you what c is. And we might want to know because this isn't necessarily a completely specified joint PDF. How do we find c?

Well, the way that's the foolproof way that we have for finding c is to remember that a joint PDF always has to integrate to 1. That's just a property a joint PDF has. And so if we integrate this function over the support, set that equal to 1, set that integral equal to 1, then we can solve for c. So that's just what that says.

Now, note that we only need to integrate that function. We don't need to integrate this function over the entire x/y-plane. We only have to integrate it over the support of the distribution because, outside of that support, the PDF is equal to 0.

So here's what we do. I set up the integral. So here is the joint PDF. I've set up the double integral, chosen these limits of integration so that I'm only integrating over the support of the PDF, and perform that integral. We get that it's equal to 4 over 21 times c. And we set that equal to 1, and so that we get c is equal to 21 over 4.

So you might be a little puzzled as to how I got these limits of integration if, for instance, you haven't taken calculus in the last year or so. Maybe you are a little rusty about that. So let's take a look again at the picture of the support and see where these limits of integration come from.

OK, so here is the support. And what I'm doing is I'm integrating first out over y. And y takes on values from x squared up to 1. And then the outer integral is over x, and x takes on values from negative 1 to 1. And I could have done the integration in the other order, as well, and that would have worked just fine.

OK, so now we've got this completely specified joint PDF. We plugged in the value for the constant that we computed. And so we've got this completely specified joint PDF of x and y. And so now we can actually start doing things with it.

And one of the things that we can do is calculate probabilities of certain regions. So in particular, let's suppose that we're interested in the probability that x is greater than y. How would we do that? We've got the joint PDF, so we have all the information we need.

All we do is we just set up an integral so that we're integrating the joint PDF over the region that's described by this probability statement. So how do we do that? Well, we have to figure out the region of the x/y-plane over which to integrate, which is what I just said.

And how we do that is we draw the line for x equals y. So this is the line here-- x equals y. This is part of the support of the joint PDF. I didn't draw the entire support. I could have continued and drawn the whole thing, but the rest of it was irrelevant for this probability question. So I only drew part of the support.

And basically, what I want to do is integrate over this slice because this slice is the part-- it's the intersection between the values where x is greater than y. So all the values sort of southeast of this line are values such that x is greater than y. And it's the intersection between those and the support.

So is that clear? OK, so once I've identified what region we need to integrate over, then all we have to do is just remember our multivariable calculus and figure out how to set the limits of integration, do the integral, and we're all done.

And if we-- so this is-- we set the limits of integration to ensure that we're integrating over this little sliver here. And if we perform that integration, we get 3 over 20. So here, just to reiterate, y is going from x squared up to x, and then x is going from 0 to 1. And that's where the limits of integration come from.

OK, so now we're starting to get comfortable with the notion of a joint distribution being a surface or a set of point masses over the x/y-plane that describe the probability with which a random vector takes on certain-- is in certain regions of the x/y-plane. And we've also seen examples of how to calculate probabilities by integrating the PDF over the relevant regions.

Now we're going to see some other things that we can do with joint PDFs or joint distributions. And one of the things we can do is we can take the information that's embodied in a joint PDF or joint probability function, and we can extract the information that tells us about the individual distributions of those constituent random variables.

So let's suppose we've got a joint PDF of X and Y. But we really only care about X, for some reason. So how do we get the individual or the marginal distribution of X from that joint distribution? So let me, first, tell you how to do it, give you the definition, and then we'll do some examples.

OK, so for discrete random variables, all we do is, if we want the marginal distribution of x, and we have the joint distribution of x and y, we sum out over y. And I'll give you the intuition for that in just a second. And for continuous random variables, we do, not surprisingly, the continuous analog of that, which is that we integrate out over y.

So what's the intuition behind this? Well, for discrete random variables, for a particular value of x, what we want to do is we just want to sum up the joint distribution over all the values of y to obtain the marginal distribution of X at that point.

So if we want to know the probability that a random variable X is equal to little x, for instance, what we want to do is we want to sum up the probability from the joint distribution that at the pair-- well, for any possible value of y, all of the pairs for any possible value of y, we want to sum up those probabilities. And the sum of those probabilities gives us the marginal or individual distribution or probability for x.

And then for continuous random variables, we just do the continuous analog of that. So this may seem a little vague at this point. So now, I think, would be a good time to do an example or two.

OK, so let me turn to tennis. So Esther and I play tennis. And like a lot of sports, tennis players tend to rise or fall to the level of their opponent. And so it wouldn't be surprising to learn that we're more likely to both be playing well in a particular day or both be playing poorly.

Oh, by the way, that's the MIT Tennis Bubble. Have you guys played there? It's a very nice place to play tennis. If you ever want to play tennis during the winter, look into playing there.

OK, so let's see. Where was I? So anyhow, it's not surprising that, on any given day, Esther and I, if we're playing each other, we might both be playing well or both be playing poorly. What would that observation suggest about the shape of the joint probability function of our unforced errors by game?

OK, so for those of you who don't play tennis, just a tiny bit of background-- a game is completed when a player wins four points by at least two points. And the score-keeping has this funny, historical character where we don't count by-- we don't just count 0, 1, 2, 3. But an unforced error is basically when Esther hits me the ball, and it's a shot that I can get back, but I hit it into the net, instead.

So that's an unforced error. So obviously, if we're both playing well, we would have relatively few unforced errors. If we're both playing poorly, we'd have more unforced errors.

So what does this observation that we're typically both going to be playing well or poorly suggest about the shape of this joint PDF or joint PF of our unforced errors by game? Any guesses? Well, if you think about it, the probability is going to be concentrated around values of x and y. So let's say x is the number of unforced errors that Esther hits or that she's responsible for, and y is my unforced errors, the number of my unforced errors.

So if we're both playing well or poorly, we would expect the probability to be concentrated around points in the joint PF where x and y have similar values. So we would expect there not to be a lot of probability for points where I have a lot of unforced errors, and Esther has few, and vice versa. But we would expect a lot of probability to be concentrated around points where we have similar numbers.

So exactly what might something like this look like? Well, here's one example. So I just made this up, obviously. This is not based on any long-term data that we've been collecting.

Let's say X is my number of unforced errors, and y Is Esther's number of unforced errors. And what we see is there's a concentration of probability around this axis where X and Y have similar values and very little probability here and very little probability up there-- actually, 0, the way I've drawn it. Does that make sense? OK. And apparently, because we don't hit that many unforced errors, we're pretty good.

OK, so here's the graphical representation of exactly the joint PF I just showed you. So I just drew what it would look like graphically. And now, let's calculate the marginal distributions for both X and Y.

So remember what we said a couple of slides ago. All we want to do is add up over values of the other random variable to get the value of the marginal distribution for a particular x or a particular y. So specifically, the probability that I make two unforced errors in a game is the probability that I make two unforced errors, and Esther makes zero, plus the probability that I make two, and Esther makes one, plus the probability that I make two, and Esther makes two, et cetera.

So all I'm doing is just adding up along a particular value for x. So then we get-- if I do that, and I do that for all the values of x, I get the following marginal distribution. So the probability that big X is equal to 3/8 is-- sorry-- the probability that big X is equal to 0 is equal to 3/8. The probability that it's equal to 1 is equal to 5/16 and so forth.

And it turns out that the marginal distributions of our errors are exactly the same. So I've just written this-- instead of writing both of the marginal distributions out, I just wrote it that f sub x of little x is equal to f sub y of little y.

And I want to also mention that you will actually-- I will never reveal the true distributions of our unforced errors. You can speculate, if you want. Oh, and here, it is graphically. So both of the marginal distributions look the same, obviously.

OK, so now we've done an example using a discrete PF, probability function. And let's do the analogous thing using a continuous distribution or continuous joint PDF. So returning to our example from the beginning of lecture, we have this PDF that I've reminded you of there. And recall, this is the shape of the support of the PDF. And this is my three-dimensional rendering of what the joint PDF looks like.

So just applying the formula that I gave you a few slides ago, if we want the marginal PDF of x, say, all we have to do is integrate out y. And so what we do is we want to integrate-- in general, we're going to integrate the joint PDF out over the entire support. For us, y only takes on values as a function of x from x squared up to 1. And so that's the region we're going to integrate over.

And you can see it here. I've included the support down here to remind you. So we're integrating for a particular value of x. We just integrate from x squared up to 1.

And if I wanted to be really pedantic, and I wanted to be very precise about my notation, what I would do, actually, is I'd put an indicator function here because this PDF is only defined-- or it's only nonzero over this region here, where x is between negative 1 and 1. So just to be super careful, I might include an indicator function here, just to remind us of that.

And if we perform the integration, then this is what we get-- 21/8 times x squared times 1 minus x to the fourth. And again, just to be very careful, I put the indicator function there. And this is what it looks like graphically.

OK, so we can do the same thing. We just calculated the marginal PDF of x. We can do the same thing with y and calculate the marginal PDF of y. So here, for a particular value of-- sorry-- yes, for a particular value of x-- for a particular value of y-- I want to make sure I get this right-- the limits of integration are going from a negative square root of y to a square root of y.

And again, just to be careful. I want to put the indicator function there, although, a lot of times, you might see it not used. And we perform that integration, and we get that it's equal to 7/2y to the 5 over 2. And what does this look like graphically? There we go.

OK, so we've seen a couple of examples, now, where we take a joint distribution, and we want to extract the information on the marginal or individual distributions of the random variables that are part of that joint distribution. So we have seen examples of how we can do that, and we know, in general, that we can do that.

So I want to pose the following question if the marginals, can you construct the joint distribution? So if I gave you, for instance, this marginal PDF and this marginal PDF, do you think that you could reconstruct the joint distribution of x and y? Guesses?

OK, well, the answer is no-- in general, no. We need another crucial piece of information, and that's the relationship between the random variables. So in other words, another way to say this is, basically, that a joint distribution contains three different types of information.

It contains the information about how each one of the two-- if it's a bivariate distribution-- about how each one of the two random variables is distributed on its own. And then it also contains information about how they're related to each other. And without all three of these pieces of information, we can't reconstruct the joint PDF.

So that provides a nice segue into our discussion of independence because independence is, in fact, a special kind of relationship between random variables. And random variables can be related in a variety of ways, but independence is-- well, it's an important and special type of relationship. So it's a good place to start.

OK, so let me, first, define what independence of random variables is. So X and Y are independent if the probability that X is in some region A and that Y is in some region B is equal to the product of the probabilities that each are in those respective regions for all regions of A and B.

So if you think about having two random variables and having their joint distribution perhaps and trying to check something like this, that's virtually impossible. I mean, how do you check that something is true for all possible regions, subregions A and B? But in fact, we can come up with some more useful results that we can use to test whether random variables are independent.

So this definition here implies that, in particular, the joint CDF of two random variables is equal to the product of the marginal CDFs. So that can be useful. And furthermore, one can prove-- and we won't go through the details here-- that X and Y are independent if and only if that's also true of the PDFs. So in other words, the joint PDF factors as the two marginal PDFs.

So if we have that, then we know that these random variables are independent. And this condition is easy to check, and it's very useful. So we'll use it a lot, and we'll come back to it.

And in fact, we can even go further. We can say something even stronger about the joint PDF of independent random variables. If they're both continuous, and they have a joint PDF, they're independent if and only if the joint PDF factors into two functions where g is a non-negative function of x alone, and j is a non-negative function of y alone.

So that can be a useful way to check independence of random variables. We just look at the joint PDF. And if we can factor it-- if it's continuous, and we can factor it in this way, we know that the random variables are independent.

OK, so do you guys remember the headache example from a few lectures ago? OK. So let's return to that example and ask ourselves about the relationship between the two random variables in this example, and in particular, whether they're independent. OK. So remember, I said that this was what the joint PDF of the effective period of the two tablets was.

So the effective period of naproxen is random variable X-- yeah, random variable X-- and the effective period of-- what's the other one-- acetaminophen is Y. And this is their joint distribution. So are X and Y are independent here? What do you think?

So we look at this joint PDF, and we see if it fits any of these conditions here, like, for instance, whether we can factor it into non-negative functions of X alone and Y alone. And in fact, we can. So this PDF can be factored. And in fact, X and Y-- if we went to the trouble of computing the marginal PDF of X and the marginal PDF of Y, we would see that they were identical.

The X and Y have the same distribution. But we don't need to actually compute the marginal PDFs to test or to check whether these are independent. We can just check the factor and result.

OK, second example-- let's go back to the example that I started the lecture with today. And here is the joint PDF of X and Y. Are X and Y independent in this case? And before you answer, let me just tell you, it's a bit of a trick question.

OK, well, it's a bit of a trick question because if you just look at the joint PDF, this part of the joint PDF, it certainly looks like that function factors into a function of X alone and a function of Y alone. However, the support implies that the values that the random variable Y can take depend on the value of X and vice versa.

So in other words, this function-- or this joint PDF cannot, in fact, be factored into a function of X alone and a function of Y alone. The function of X is going to depend on Y through its support and vice versa. So these random variables are not independent.

And here is the third example. This is also just from a few minutes ago. So here is the joint PF of unforced errors in mine and Esther's tennis game. Are X and Y independent here? What do you think?

No, they're not. And why is that? Well, remember when I told you what the definition of independence of random variables was? I said-- actually, let's just flip back there quickly. X and Y are independent if the probability that X is in some region, and Y is in some region is equal to the product of the probabilities for all regions A and B.

And here, in this example, it's easy to think of a region where that definition is violated. So in particular, if you look at this region, there's no probability in the joint probability function-- zero probability for all for that entire region.

And yet, the product of the two marginal distributions is nonzero. If you think back to the marginal distributions that we calculated for this example, take their product, the marginal distributions are nonzero. So this violates the definition.

OK, so let me do one, additional-- tell you one, additional fact and do one, additional example that I hope will solidify your intuition a little bit about independent random variables and joint distributions. So here's the fact. For discrete random variables, if you have a table representing their joint probability function, the two variables are independent if and only if the rows of the table are proportional to one another, if and only if the columns of the table are proportional to one another.

So basically, if you look at a table of probabilities, every row has to be a multiple of the other rows. Every column-- sorry. I'm doing it the wrong way. Every column has to be a multiple of the other columns. Every row has to be a multiple of the other rows. And if that's not the case, then the random variables are not independent.

So why is that? Well, independence means that the product of the marginals is equal to the joint. So each column of the table is just a multiple of every other column, the multiple being the ratio of the marginal probabilities associated with the two columns.

That's kind of a mouthful, so let me do an example. OK, so here is a table representing probabilities of two random variables X and Y. So this is the joint PDF of the two random variables. And you'll notice something about these columns and the rows of this table.

This column is just a multiple of this column. Every single value is the same ratio with this value. This column is, in fact, exactly the same as this one and this one. And then this column is also multiple of all the other columns. You can look down the rows and see exactly the same patterns existing.

So what might this be a distribution of? Well, again, it's just made up. But we could think of this as a possible joint probability function of unforced errors if mine and Esther's errors were independent.

So if, in fact, we didn't have this situation where, when I played well, Esther played well, and when I played poorly, Esther played poorly, or vice versa. So if our unforced errors were completely independent, maybe this is what the joint distribution would look like.

OK, so everyone feel comfortable with independence? Well, similar to the idea of conditional probability-- so now we'll move on to talk about conditional distributions. Similar to the idea of conditional probability, we want to introduce the conditional distribution now.

And basically, the conditional distribution allows us to update the distribution of a random variable, if necessary, given relevant information. So if I tell you, for instance, that Esther made two unforced errors in a particular game of tennis that we played, then that lets you-- given the joint distribution of our unforced errors, that lets you figure out what the distribution of my unforced errors was, conditional on the fact that Esther made two.

So how do we calculate these conditional distributions? Well, for continuous conditional PDF, we just calculate it as-- so this is how we denote it in terms of the notation. And it's just the joint PDF divided by the marginal PDF, the marginal PDF of the conditioning variable. And for a discrete random variables, it's just equal to that.

So one thing-- it's a little easy to get confused with the notation of conditional distributions because they're often written as a function of both x and y. And so if you didn't think carefully about it, you might get confused and think it was a joint distribution.

But for a particular-- but it doesn't behave the same way as a joint distribution. So for a particular value of the conditioning distribution or a conditioning variable, it behaves just like a marginal PDF. So it may be written as a function of X and Y. But then if X is your conditioning variable, you plug in a particular value of X, and then you just have a marginal distribution.

So another way to think-- and we'll see this graphically in a second. Another way that I find useful to think about conditional distributions is that what we're doing is we're taking a slice of the joint distribution. We're taking the relevant slice which corresponds to a value of the conditioning variable.

And then once we take that slice, we have something that looks like a PDF, perhaps, but it might not or it's not, in general, going to integrate to 1. And so then we have to blow it up. We've got to multiply it by some factor so that it integrates to 1.

So this is exactly how I think of this formula for a conditional PDF. We're just taking a particular slice, a particular value of x of the joint distribution and then blowing it up by this factor so that it integrates to 1. So let's see some examples.

OK, so here's the now-familiar original joint probability function of unforced errors. And let's take a slice of it. What do I mean by take a slice? I mean, choose a particular value of Y. Y is going to be our conditioning variable this time.

So choose a particular value of Y, say, Y equals 2. And then that slice. So basically, that tells us something about-- so if Esther has two unforced errors in a particular game, that's what this slice represents-- Esther having two unforced errors. And then these numbers tell us something about the relative probabilities of me having zero, one, two, three, or four unforced errors in that game.

But as it stands-- I mean, that gives us some information. But as it stands, this is not a PDF on its own. This slice is not a PDF because it doesn't integrate or doesn't sum to 1. So we need to blow it up.

So the factor we blow it up by is just the probability that Y is equal to 2. And we can go back and look at the marginal distribution that we calculated several slides ago from this joint distribution, the marginal distribution of Y.

And we look up that the probability that Y is equal to 2 is 5 over 32. And that's the factor that we use to multiply-- or divide, actually-- all of these probabilities so that they sum to 1. OK, so here we have it.

Once we've divided by that factor, then we get to conditional distribution of x, conditional on Y equaling 2, is 2/5 for x equals 1 or 2, and 1/5 for x equals 3, and 0 otherwise. So is that clear?

So basically, what we've done here-- we're conditioning on Y equals 2. We care about the distribution of x given that Y is equal to 2. So we take the slice, the relevant slice of the joint PDF. But then that thing, on its own, is not a marginal PDF because it doesn't sum to 1. So we've got to blow it up so it sums to 1. And that's exactly what we did.

Let's also compute a conditional PDF. So that was a discrete example. Let's do a continuous example now. And we'll return to the earlier example we had this lecture-- this one.

So now what we want to do is we want to take a slice of this joint PDF at some particular value. And then that slice is not going to be a true PDF. And so then we've got to figure out what we multiply it by so that it integrates to 1.

So let's take the slice x equals 1/2. OK, so I've drawn it here. Maybe you can kind of picture in your mind what that slice is going to look like. And so I think I already said this. We're just going to take the slice and then suitably normalize it so it is a valid PDF on its own.

So recall that the joint-- or sorry, the marginal-- we calculated the marginal PDF of x. And we'll need this in just a second to calculate this conditional PDF because, remember, the formula for the conditional PDF is the joint PDF over the marginal PDF of the conditioning variable. And so I'm just reminding you here, this is the marginal PDF we computed several minutes ago.

And note that this PDF is nonzero for all x in this interval here, except for x equals 0. At x equals 0, this PDF is 0. So that means, in particular, if we're going to condition on x, we can only condition-- our conditional distribution conditioned on x is only going to be defined for nonzero values of the PDF of x.

And so a PDF conditional on x will only be defined for values in this interval except for 0. So just note that, and we'll come back to that at the end. We just have to make sure we don't forget that. OK, so we've got the marginal PDF of x.

Recall that this is the formula for calculating the conditional PDF. And this is the joint PDF up here. So we have everything we need to calculate the conditional PDF. And this is what we have. This is what we come up with. And again, keep in mind, it's only defined for nonzero values of x.

So now what we can do-- so that's fine. Recall I said a few minutes ago that sometimes the notation is a little confusing because conditional PDFs are expressed as a function of both x and y, so they look kind of joint PDFs. Well, that's the case here.

So this conditional PDF is a function of both x and y-- little x and little y. But what we can do is we can plug in a particular value of x, as long as it's a value where the probability of x is nonzero. We plug in a particular value of x, and then that gives us a conditional distribution of y that's no longer a function of x.

So plug in any legitimate value of x. And let's just say x equals 1/2. That's the little red arrow that I had on the previous slide. And we just plug that into the general form for the conditional distribution on the top, and we get 32 over 15y for y between 1/4 and 1 and 0 otherwise.

So what does this look like graphically? So here's the picture from before. Remember, I said what we were going to do is we would take a slice of this joint PDF and then just blow it up so that it was a legitimate PDF. And this is precisely what it looks like when we do the calculation.

OK, so not surprisingly, there is a relationship between conditional distributions and independence. So we saw this when we were just-- before we were even talking about random variables, when we were just talking about probability, we talked about the relationship between independent events and conditional probabilities. And some very similar relationship exists here.

So in particular, the conditional distribution of y is equal to the unconditional distribution of y, so just the marginal distribution of y, if and only if x and y are independent. So the intuition behind this is that, if you have two random variables, and they're independent, then knowing something about the realizations of one of the random variables doesn't tell you anything about the distribution of the other.

And remember how I said that a conditional distribution was like a way of updating the distribution of a random variable given some relevant information? Well, if you're told x is equal to 2, and x and y are independent, well, then that's actually not relevant. The fact that x is equal to 2 isn't relevant information for the distribution of y. So that's basically what this relationship says.

OK, so now, in the next few minutes, we're going to switch gears or start the next topic. And I want to just give you a little bit of a preview, I guess, a little bit of a segue. So as I've emphasized before, we need to start with a foundation in probability, even though we may be ultimately interested in analyzing data, and we might be ultimately interested in statistics and how we can test a various hypotheses using data sets.

But we need to start with a foundation in probability because we can't talk about how functions of random variables behave until we know about how random variables behave. And we can't talk about statistics, such as the sample mean, or a T-test, or whatever until we know about functions of random variables because that's precisely what a statistic is.

So there are good reasons to have a foundation in probability otherwise, but at least for us, this may be the most important reason because what we're going to be doing-- when we talk about statistics, we're talking about functions of random variables, and we can't talk about functions of random variables unless we know how random variables behave. And so that's what we've been doing for the better part of this semester.

So now we move on to our discussion of functions of random variables. And this is going to lead us into our discussion of statistics. OK, so the basic idea is that we have a random variable X, and we know its distribution. And we want to know how a new random variable, Y, which is some known function of X, is distributed.

So a more complicated version of this problem is that we have a whole bunch of random variables-- X1, X2, X3-- and we want to know how X-- sorry-- how h of X, a function of the entire random vector X, is distributed. So here, this is maybe not the most careful notation in the world, but I'll sometimes use capital X to refer not just to a random variable but an entire random vector. And that's what I did here.

OK, so let's start. So I'll give you, probably next time, specific information about how you figure out the distribution of functions of random variables in different situations. But I think it's helpful to start with a graphical example, where we just kind of do it intuitively.

So let's suppose we want the distribution of some new random variable Y, which is equal to the absolute value of 2 times some random variable X plus 3. So we've got this random variable X. We know its distribution. I'm putting it right there.

So we know its distribution. And what we want to do is figure out what the distribution of this function of the random variable is. So first, what I've done is drawn the PDF of X. So I gave it to you in formula here, and then I'm just drawing what it looks like graphically.

And let's transform it or consider this function step by step. So let's first multiply the random variable by 2. And what happens to the PDF when we multiply the random variable by 2? Well, what happens is it stretches out.

So basically, the random variable used to have support negative 1 to positive 1. It now has support negative 2 to positive 2. And because it still has to have a distribution, the PDF still has to integrate to 1, this got squished down. So basically, the PDF just got stretched out.

So that's the first step. We've just multiplied by 2. Now let's take the absolute value of this random variable and see what happens to the PDF. All of the density over the negative values that we saw in the last slide, so basically, all of this density here, gets flipped into positive values when you take the absolute value.

So then our PDF looks something like this. So now the support only goes from 0 to positive 2. But it's twice as tall as the previous PDF because-- I guess there are a couple of ways to think about it. First of all, this is a PDF, so it has to integrate to 1. So that specifies what height it has to have.

But, also, you're basically taking all of this density and adding it to the density over the positive values that was already there. And now, the final step is-- so we've multiplied it by 2. We've taken the absolute value. And now we want to add 3 to it.

And what does adding 3 to a random variable do? Well, it just shifts the entire distribution over. So instead of going-- instead of the support being from 0 to 2, the support is now from 3 to 5. OK, so keep in mind that, throughout this process, the distribution always retain the properties of a PDF-- in particular, it integrated to 1.

And let me just do one more example that should help firm up your intuition of what a function of a random variable does. So suppose we have a random variable X, which has a uniform 0, 1 distribution. What function g can transform X to be binomial with parameters 2 and 0.5?

So this is kind of a puzzling question. So the last example, I gave you the function, and then we just went step by step, transforming the random variable by the function. Here, I didn't give you the function. I asked you to come up with the function.

So we've got two random variables, one has a uniform 0, 1 distribution, and one has a binomial 2, 0.5 distribution. What function might transform one into the other? Well, maybe I can give you a hint.

The first thing-- so first thing to do that might be helpful is to write down what these PDFs look like. So the PDF for uniform 0, 1 random variable is just equal to 1 over the unit interval and 0 otherwise. And what is the PDF for a binomial 2.5 random variable look like?

Well, we can calculate it. You might be able to figure this out off the top of your head. But if not, you can go back and remind yourself what a binomial distribution looks like. And you get that it's equal to 1/4 where x equals 0 or 2 and equal to 1/2 where x equals 1.

So remember, this is just a coin flip. You flip the coin two times. It's a fair coin, so there's probability 1/2 of a head and 1/2 of a tail. What's the number of heads you get? Well, it's 0 or 2 with probability 1/4, and it's equal to 1 with probability 1/2.

So now we've got these two PDFs, and we still want the function that's going to map one into the other, that's going to transform the first into the second. Well, how about just chopping up the unit interval and mapping the appropriate-sized subintervals into each of these point masses?

So the other thing I didn't emphasize already but I should is that this is a continuous variable here, and this is a discrete variable. So we have the PDF and the PF of the two random variables. And what we're going to do is chop up the unit interval into sections of various sizes and map those, the appropriate-sized subintervals, to each point mass.

So here, graphically, is the PDF of x. There's the PF of y. And let's suppose we divide up the subinterval that way. Keep in mind this isn't the only way we could do it. I mean, we could divide it up into any way such that a quarter of the probability got mapped into 0, a quarter got mapped into 2, and a half got mapped into 1.

But this is as good a way as any. And then we map the subintervals like that. Does that make sense?

So our function, then-- how do we specify this function? Well, we say that Y, this new random variable Y, is equal to 0 if x is less than a quarter. It's equal to 1 if x is between a quarter and 3/4. And it's equal to 2 if x is greater than 3/4.

OK, so I mentioned this before. This is one possible way to do it-- certainly not unique. And we'll see this over and over again when we talk about functions of random variables. Sometimes there will be multiple functions of a random variable that will map into the distribution of a new random variable. And so these are not always unique.

So one final word before we stop. There are various methods, some of which I'm going to talk about, we can use to figure out the distribution of a function of random variables. Which methods one can use on a particular problem depend on whether the original random variable is discrete or continuous, whether there's just one random variable or random vector, and whether the function is invertible or not.

And we're not going to learn all the methods. If you were taking a semester-long probability class, you would presumably learn all of these methods, but here we don't have that luxury. Instead, we're going to learn one important method and also see a lot of examples that we're going to be able to apply pretty generally and to questions of functions of random variables that come up. OK, so I think that's it. Yep.