The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high-quality, educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

PROFESSOR: Good morning. So today we're going to continue the subject from last time. So we're going to talk about derived distributions a little more, how to derive the distribution of a function of a random variable. So last time we discussed a couple of examples in which we had a function of a single variable. And we found the distribution of Y, if we're told the distribution of X.

So today we're going to do an example where we deal with the function of two random variables. And then we're going to consider the most interesting example of this kind, in which we have a random variable of the form W, which is the sum of two independent, random variables. That's a case that shows up quite often. And so we want to see what exactly happens in this particular case.

Just one comment that I should make. The material that we're covering now, chapter four, is sort of conceptually a little more difficult than one we have been doing before. So I would definitely encourage you to read the text before you jump and try to do the problems in your problem sets.

OK, so let's start with our example, in which we're given two random variables. They're jointly continuous. And their distribution is pretty simple. They're uniform on the unit square. In particular, each one of the random variables is uniform on the unit interval. And the two random variables are independent.

What we're going to find is the distribution of the ratio of the two random variables. How do we go about it? , Well, the same cookbook procedure that we used last time for the case of a single random variable. The cookbook procedure that we used for this case also applies to the case where you have a function of multiple random variables.

So what was the cookbook procedure? The first step is to find the cumulative distribution function of the random variable of interest and then take the derivative in order to find the density. So let's find the cumulative. So, by definition, the cumulative is the probability that the random variable is less than or equal to the argument of the cumulative. So if we write this

event in terms of the random variable of interest, this is the probability that our random variable is less than or equal to z.

So what is that? OK, so the ratio is going to be less than or equal to z, if and only if the pair, (x,y), happens to fall below the line that has a slope z. OK, so we draw a line that has a slope z. The ratio is less than this number, if and only if we get the pair of x and y that falls inside this triangle.

So we're talking about the probability of this particular event. Since this line has a slope of z, the height at this point is equal to z. And so we can find the probability of this event.

It's just the area of this triangle. And so the area is 1 times z times 1/2. And we get the answer, z/2.

Now, is this answer always correct? Now, this answer is going to be correct only if the slope happens to be such that we get a picture of this kind. So when do we get a picture of this kind? When the slope is less than 1.

If I consider a different slope, a number, little z -- that happens to be a slope of that kind -- then the picture changes. And in that case, we get a picture of this kind, let's say. So this is a line here of slope z, again. And this is the second case in which our number, little z, is bigger than 1.

So how do we proceed? Once more, the cumulative is the probability that the ratio is less than or equal to that number. So it's the probability that we fall below the red line.

So we're talking about the event, about this event. So to find the probability of this event, we need to find the area of this red shape. And one way of finding this area is to consider the whole area and subtract the area of this triangle.

So let's do it this way. It's going to be 1 minus the area of the triangle. Now, what's the area of the triangle? It's 1/2 times this side, which is 1 times this side.

How big is that side? Well, if y and the slope is z, now z is the ratio y over x. So if y over x-- at this point we have y/x = z and y =1. This means that z is 1/x.

So the coordinate of this point is 1/x. And this means that we're going to-- 1/z So here we get the factor of 1/z.

And we're basically done. I guess if you want to have a complete answer, you should also give the formula for z less than 0. What is the cumulative when z is less than 0, the probability that you get the ratio that's negative?

Well, since our random variables are positive, there's no way that you can get a negative ratio. So the cumulative down there is equal to 0. So we can plot the cumulative. And we can take its derivative in order to find the density.

So the cumulative that we got starts at 0, when z's are negative. Then it starts going up in proportion to z, at the slope of 1/2. So this takes us up to 1.

And then it starts increasing towards 1, according to this function. When you let z go to infinity, the cumulative is going to go to 1. And it has a shape of, more or less, this kind.

So now to get the density, we just take the derivative. And the density is, of course, 0 down here. Up here the derivative is just 1/2. And beyond that point we need to take the derivative of this expression.

And the derivative is going to be 1/2 times 1 over z-squared. So it's going to be a shape of this kind. And we're done.

So you see that problems involving functions of multiple random variables are no harder than problems that deal with the functional of a single random variable. The general procedure is, again, exactly the same. You first find the cumulative, and then you differentiate. The only extra difficulty will be that when you calculate the cumulative, you need to find the probability of an event that involves multiple random variables. And sometimes this could be a little harder to do.

By the way, since we dealt with this example, just a couple of questions. What do you think is going to be the expected value of the random variable Z? Let's see, the expected value of the random variable Z is going to be the integral of z times the density.

And the density is equal to 1/2 for z going from 0 to 1. And then there's another contribution from 1 to infinity. There the density is 1/(2z-squared). And we get the z, since we're dealing with expectations, dz.

So what is this integral? Well, if you look here, you're integrating 1/z, all the way to infinity. 1/z has an integral, which is the logarithm of z. And since the logarithm goes to infinity, this means

that this integral is also infinite.

So the expectation of the random variable Z is actually infinite in this example. There's nothing wrong with this. Lots of random variables have infinite expectations. If the tail of the density falls kind of slowly, as the argument goes to infinity, then it may well turn out that you get an infinite integral. So that's just how things often are. Nothing strange about it.

And now, since we are still in this example, let me ask another question. Would we reason, on the average, would it be true that the expected value of Z -- remember that Z is the ratio Y/X -- could it be that the expected value of Z is this number? Or could it be that it's equal to this number? Or could it be that it's none of the above?

OK, so how many people think this is correct? Small number. How many people think this is correct? Slightly bigger, but still a small number.

And how many people think this is correct? OK, that's-- this one wins the vote. OK, let's see.

This one is not correct, just because there's no reason it should be correct. So, in general, you cannot reason on the average. The expected value of a function is not the same as the same function of the expected values. This is only true if you're dealing with linear functions of random variables. So this is not-- this turns out to not be correct.

How about this one? Well, X and Y are independent, by assumption. So 1/X and Y are also independent. Why is this? Independence means that one random variable does not convey any information about the other.

So Y doesn't give you any information about X. So Y doesn't give you any information about 1/X. Or to put it differently, if two random variables are independent, functions of each one of those random variables are also independent.

If X is independent from Y, then g(X) is independent of h(Y). So this applies to this case. These two random variables are independent.

And since they are independent, this means that the expected value of their product is equal to the product of the expected values. So this relation actually is true. And therefore, this is not true. OK.

Now, let's move on. We have this general procedure of finding the derived distribution by

going through the cumulative. Are there some cases where we can have a shortcut? Turns out that there is a special case or a special structure in which we can get directly from densities to densities using directly just a formula. And in that case, we don't have to go through the cumulative.

And this case is also interesting, because it gives us some insight about how one density changes to a different density and what affects the shape of those densities. So the case where things easy is when the transformation from one random variable to the other is a strictly monotonic one. So there's a one-to-one relation between x's and y's.

Here we can reason directly in terms of densities by thinking in terms of probabilities of small intervals. So let's look at the small interval on the x-axis, like this one, when X ranges from-- where capital X ranges from a small x to a small x plus delta. So this is a small interval of length delta.

Whenever X happens to fall in this interval, the random variable Y is going to fall in a corresponding interval up there. So up there we have a corresponding interval. And these two intervals, the red and the blue interval-- this is the blue interval. And that's the red interval.

These two intervals should have the same probability. They're exactly the same event. When X falls here, g(X) happens to fall in there. So we can sort of say that the probability of this little interval is the same as the probability of that little interval. And we know that probabilities of little intervals have something to do with densities.

So what is the probability of this little interval? It's the density of the random variable X, at this point, times the length of the interval. How about the probability of that interval? It's going to be the density of the random variable Y times the length of that little interval.

Now, this interval has length delta. Does that mean that this interval also has length delta? Well, not necessarily.

The length of this interval has something to do with the slope of your function g. So slope is dy by dx. Is how much-- the slope tells you how big is the y interval when you take an interval x of a certain length.

So the slope is what multiplies the length of this interval to give you the length of that interval. So the length of this interval is delta times the slope of your function. So the length of the interval is delta times the slope of the function, approximately.

So the probability of this interval is going to be the density of Y times the length of the interval that we are considering. So this gives us a relation between the density of X, evaluated at this point, to the density of Y, evaluated at that point. The two densities are closely related.

If these x's are very likely to occur, then this is big, which means that that density will also be big. If these x's are very likely to occur, then those y's are also very likely to occur. But there's also another factor that comes in. And that's the slope of the function at this particular point.

So we have this relation between the two densities. Now, in interpreting this equation, you need to make sure what's the relation between the two variables. I have both little x's and little y's.

Well, this formula is true for an (x,y) pair, that they're related according to this particular function. So if I fix an x and consider the corresponding y, then the densities at those x's and corresponding y's will be related by that formula. Now, in the end, you want to come up with a formula that just gives you the density of Y as a function of y. And that means that you need to eliminate x from the picture.

So let's see how that would go in an example. So suppose that we're dealing with the function y equal to x cubed, in which case our function, g(x), is the function x cubed. And if x cubed is equal to a little y, If we have a pair of x's and y's that are related this way, then this means that x is going to be the cubic root of y.

So this is the formula that takes us back from y's to x's. This is the direct function from x, how to construct y. This is essentially the inverse function that tells us, from a given y what is the corresponding x. Now, if we write this formula, it tells us that the density at the particular x is going to be the density at the corresponding y times the slope of the function at the particular x that we are considering. The slope of the function is 3x squared.

Now, we want to end up with a formula for the density of Y. So I'm going to take this factor, send it to the other side. But since I want it to be a function of y, I want to eliminate the x's. And I'm going to eliminate the x's using this correspondence here.

So I'm going to get the density of X evaluated at y to the 1/3. And then this factor in the denominator, it's 1/(3y to the power 2/3). So we end up finally with the formula for the density of the random variable Y.

And this is the same answer that you would get if you go through this exercise using the cumulative distribution function method. You end up getting the same answer. But here we sort of get it directly.

Just to get a little more insight as to why the slope comes in-- suppose that we have a function like this one. So the function is sort of flat, then moves quickly, and then becomes flat again. What should be -- and suppose that X has some kind of reasonable density, some kind of flat density.

Suppose that X is a pretty uniform random variable. What's going to happen to the random variable Y? What kind of distribution should it have? What are the typical values of the random variable Y?

Either x falls here, and y is a very small number, or-- let's take that number here to be -- let's say 2 -- or x falls in this range, and y takes a value close to 2. And there's a small chance that x's will be somewhere in the middle, in which case y takes intermediate values. So what kind of shape do you expect for the distribution of Y?

There's going to be a fair amount of probability that Y takes values close to 0. There's a small probability that Y takes intermediate values. That corresponds to the case where x falls in here.

That's not a lot of probability. So the probability that Y takes values between 0 and 2, that's kind of small. But then there's a lot of x's that produces y's that are close to 2. So there's a significant probability that Y would take values that are close to 2.

So you-- the density of Y would have a shape of this kind. By looking at this picture, you can tell that it's most likely that either x will fall here or x will fall there. So the g(x) is most likely to be close to 0 or to be close to 2.

So since y is most likely to be close to 0 or close to most of the probability of y is here. And there's a small probability of being in between. Notice that the y's that get a lot of probability are those y's associated with flats regions off your g function. When the g function is flat, that gives you big densities for Y.

So the density of Y is inversely proportional to the slope of the function. And that's what you get from here. The density of Y is-- send that term to the other side-- is inversely proportional

to the slope of the function that you're dealing with.

OK, so this formula works nicely for the case where the function is one-to-one. So we can have a unique association between x's and y's and through an inverse function, from y's to x's. It works for the monotonically increasing case. It also works for the monotonically decreasing case. In the monotonically decreasing case, the only change that you need to do is to take the absolute value of the slope, instead of the slope itself.

OK, now, here's another example or a special case. Let's talk about the most interesting case that involves a function of two random variables. And this is the case where we have two independent, random variables, and we want to find the distribution of the sum of the two. We're really interested in the continuous case. But as a warm-up, it's useful to look at the discrete case first of discrete random variables.

Let's say we want to find the probability that the sum of X and Y is equal to a particular number. And to illustrate this, let's take that number to be equal to 3. What's the probability that the sum of the two random variables is equal to 3?

To find the probability that the sum is equal to 3, you consider all possible ways that you can get the sum of 3. And the different ways are the points in this picture. And they correspond to a line that goes this way. So the probability that the sum is equal to a certain number is the probability that -- is the sum of the probabilities of all of those points.

What is a typical point in this picture? In a typical point, the random variable X takes a certain value. And Y takes the value that's needed so that the sum is equal to W. Any combination of an x with a w minus x, any such combination gives you a sum of w.

So the probability that the sum is w is the sum over all possible x's. That's over all these points of the probability that we get a certain x. Let's say x equals 2 times the corresponding probability that random variable Y takes the value 1.

And why am I multiplying probabilities here? That's where we use the assumption that the two random variables are independent. So the probability that X takes a certain value and Y takes the complementary value, that probability is the product of two probabilities because of independence.

And when we write that into our usual PMF notation, it's a formula of this kind. So this formula is called the convolution formula. It's an operation that takes one PMF and another PMF-- p

we're given the PMF's of X and Y -- and produces a new PMF.

So think of this formula as giving you a transformation. You take two PMF's, you do something with them, and you obtain a new PMF. This procedure, what this formula does is -- nicely illustrated sort of by mechanically. So let me show you a picture here and illustrate how the mechanics go, in general.

So you don't have these slides, but let's just reason through it. So suppose that you are given the PMF of X, and it has this shape. You're given the PMF of Y. It has this shape. And somehow we are going to do this calculation.

Now, we need to do this calculation for every value of W, in order to get the PMF of W. Let's start by doing the calculation just for one case. Suppose the W is equal to 0, in which case we need to find the sum of Px(x) and Py(-x).

How do you do this calculation graphically? It involves the PMF of X. But it involves the PMF of Y, with the argument reversed. So how do we plot this?

Well, in order to reverse the argument, what you need is to take this PMF and flip it. So that's where it's handy to have a pair of scissors with you. So you cut this down. And so now you take the PMF of the random variable Y and just flip it.

So what you see here is this function where the argument is being reversed. And then what do we do? We cross-multiply the two plots. Any entry here gets multiplied with the corresponding entry there. And we consider all those products and add them up.

In this particular case, the flipped PMF doesn't have any overlap with the PMF of X. So we're going to get an answer that's equal to 0. So for w's equal to 0, the Pw is going to be equal to 0, in this particular plot.

Now if we have a different value of w -- oops. If we have a different value of the argument w, then we have here the PMF of Y that's flipped and shifted by an amount of w. So the correct picture of what you do is to take this and displace it by a certain amount of w.

So here, how much did I shift it? I shifted it until one falls just below 4. So I have shifted by a total amount of 5. So 0 falls under 5, whereas 0 initially was under 0. So I'm shifting it by 5 units.

And I'm now going to cross-multiply and add. Does this give us the correct-- does it do the correct thing? Yes, because a typical term will be the probability that this random variable is 3 times the probability that this random variable is 2. That's a particular way that you can get a sum of 5.

If you see here, the way that things are aligned, it gives you all the different ways that you can get the sum of 5. You can get the sum of 5 by having 1 + 4, or 2 + 3, or 3 + 2, or 4 + 1. You need to add the probabilities of all those combinations.

So you take this times that. That's one product term. Then this times 0, this times that. And so 1-- you cross-- you find all the products of the corresponding terms, and you add them together.

So it's a kind of handy mechanical procedure for doing this calculation, especially when the PMF's are given to you in terms of a picture. So the summary of these mechanics are just what we did, is that you put the PMF's on top of each other.

You take the PMF of Y. You flip it. And for any particular w that you're interested in, you take this flipped PMF and shift it by an amount of w. Given this particular shift for a particular value of w, you cross-multiply terms and then accumulate them or add them together.

What would you expect to happen in the continuous case? Well, the story is familiar. In the continuous case, pretty much, almost always things work out the same way, except that we replace PMF's by PDF's. And we replace sums by integrals.

So there shouldn't be any surprise here that you get a formula of this kind. The density of W can be obtained from the density of X and the density of Y by calculating this integral. Essentially, what this integral does is it fits a particular w of interest. We're interested in the probability that the random variable, capital W, takes a value equal to little w or values close to it.

So this corresponds to the event, which is this particular line on the two-dimensional space. So we need to find the sort of odd probabilities along that line. But since the setting is continuous, we will not add probabilities. We're going to integrate. And for any typical point in this picture, the probability of obtaining an outcome in this neighborhood is the-- has something to do with the density of that particular x and the density of the particular y that would compliment x, in order to form a sum of w.

So this integral that we have here is really an integral over this particular line. OK, so I'm going to skip the formal derivation of this result. There's a couple of derivations in the text. And the one which is outlined here is yet a third derivation.

But the easiest way to make sense of this formula is to consider what happens in the discrete case. So for the rest of the lecture we're going to consider a few extra, more miscellaneous topics, a few remarks, and a few more definitions. So let's change-- flip a page and consider the next mini topic.

There's not going to be anything deep here, but just something that's worth being familiar with. If you have two independent, normal random variables with certain parameters, the question is, what does the joined PDF look like? So if they're independent, by definition the joint PDF is the product of the individual PDF's.

And the PDF's each one of them involves an exponential of something. The product of two exponentials is the exponential of the sum. So you just add the exponents.

So this is the formula for the joint PDF. Now, you look at that formula and you ask, what does it look like? OK, you can understand it, a function of two variables by thinking about the contours of this function.

Look at the points at which the function takes a constant value. Where is it? When is it constant? What's the shape of the set of points where this is a constant? So consider all x's and y's for which this expression here is a constant, that this expression here is a constant.

What kind of shape is this? This is an ellipse. And it's an ellipse that's centered at-- it's centered at mu x, mu y. These are the means of the two random variables.

If those sigmas were equal, that ellipse would be actually a circle. And you would get contours of this kind. But if, on the other hand, the sigmas are different, you're going to get an ellipse that has contours of this kind.

So if my contours are of this kind, that corresponds to what? Sigma x being bigger than sigma y or vice versa. OK, contours of this kind basically tell you that X is more likely to be spread out than Y. So the range of possible x's is bigger.

And X out here is as likely as a Y up there. So big X's have roughly the same probability as certain smaller y's. So in a picture of this kind, the variance of X is going to be bigger than the

variance of Y.

So depending on how these variances compare with each other, that's going to determine the shape of the ellipse. If the variance of Y we're bigger, then your ellipse would be the other way. It would be elongated in the other dimension. Just visualize it a little more.

Let me throw at you a particular picture. This is one-- this is a picture of one special case. Here, I think, the variances are equal. That's the kind of shape that you get. It looks like a two-dimensional bell.

So remember, for a normal random variables, for a single random variable you get a PDF that's bell shaped. That's just a bell-shaped curve. In the two-dimensional case, we get the joint PDF, which is bell shaped again. And now it looks more like a real bell, the way it would be laid out in ordinary space.

And if you look at the contours of this function, the places where the function is equal, the typcial contour would have this shape here. And it would be an ellipse. And in this case, actually, it will be more like a circle.

So these would be the different contours for different-- so the contours are places where the joint PDF is a constant. When you change the value of that constant, you get the different contours. And the PDF is, of course, centered around the mean of the two random variables. So in this particular case, since the bell is centered around the (0, 0) vector, this is a plot of a bivariate normal with 0 means.

OK, there's-- bivariate normals are also interesting when your bell is oriented differently in space. We talked about ellipses that are this way, ellipses that are this way. You could imagine also bells that you take them, you squash them somehow, so that they become narrow in one dimension and then maybe rotate them.

So if you had-- we're not going to go into this subject, but if you had a joint pdf whose contours were like this, what would that correspond to? Would your x's and y's be independent? No.

This would indicate that there's a relation between the x's and the y's. That is, when you have bigger x's, you would expect to also get bigger y's. So it would be a case of dependent normals. And we're coming back to this point in a second.

Before we get to that point in a second that has to do with the dependencies between the

random variables, let's just do another digression. If we have our two normals that are independent, as we discussed here, we can go and apply the formula, the convolution formula that we were just discussing. Suppose you want to find the distribution of the sum of these two independent normals.

How do you do this? There is a closed-form formula for the density of the sum, which is this one. We do have formulas for the density of X and the density of Y, because both of them are normal, random variables.

So you need to calculate this particular integral here. It's an integral with respect to x. And you have to calculate this integral for any given value of w.

So this is an exercise in integration, which is not very difficult. And it turns out that after you do everything, you end up with an answer that has this form. And you look at that, and you suddenly recognize, oh, this is normal. And conclusion from this exercise, once it's done, is that the sum of two independent normal random variables is also normal.

Now, the mean of W is, of course, going to be equal to the sum of the means of X and Y. In this case, in this formula I took the means to be 0. So the mean of W is also going to be 0. In the more general case, the mean of W is going to be just the sum of the two means.

The variance of W is always the sum of the variances of X and Y, since we have independent random variables. So there's no surprise here. The main surprise in this calculation is this fact here, that the sum of independent normal random variables is normal. I had mentioned this fact in a previous lecture.

Here what we're doing is to basically outline the argument that justifies this particular fact. It's an exercise in integration, where you realize that when you convolve two normal curves, you also get back a normal one once more. So now, let's return to the comment I was making here, that if you have a contour plot that has, let's say, a shape of this kind, this indicates some kind of dependence between your two random variables.

So instead of a contour plot, let me throw in here a scattered diagram. What does this scattered diagram correspond to?

Suppose you have a discrete distribution, and each one of the points in this diagram has positive probability. When you look at this diagram, what would you say? I would say that when

y is big then x also tends to be larger.

So bigger x's are sort of associated with bigger y's in some average, statistical sense. Whereas, if you have a picture of this kind, it tells you in association that the positive y's tend to be associated with negative x's most of the time. Negative y's tend to be associated mostly with positive x's.

So here there's a relation that when one variable is large, the other one is also expected to be large. Here there's a relation of the opposite kind. How can we capture this relation between two random variables?

The way we capture it is by defining this concept called the covariance, that looks at the relation of was X bigger than usual? That's the question, whether this is positive. And how does this relate to the answer-- to the question, was Y bigger than usual?

We're asking-- by calculating this quantity, we're sort of asking the question, is there a systematic relation between having a big X with having a big Y? OK , to understand more precisely what this does, let's suppose that the random variable has 0 means, So that we get rid of this-- get rid of some clutter. So the covariance is defined just as this product.

What does this do? If positive x's tends to go together with positive y's, and negative x's tend to go together with negative y's, this product will always be positive. And the covariance will end up being positive. In particular, if you sit down with a scattered diagram and you do the calculations, you'll find that the covariance of X and Y in this diagram would be positive, because here, most of the time, X times Y is positive. There's going to be a few negative terms, but there are fewer than the positive ones.

So this is a case of a positive covariance. It indicates a positive relation between the two random variables. When one is big, the other also tends to be big.

This is the opposite situation. Here, when one variable-- here, most of the action happens in this quadrant and that quadrant, which means that X times Y, most of the time, is negative. You get a few positive contributions, but there are few. When you add things up, the negative terms dominate. And in this case we have covariance of X and Y being negative.

So a positive covariance indicates a sort of systematic relation, that there's a positive association between the two random variables. When one is large, the other also tends to be large. Negative covariance is sort of the opposite. When one tends to be large, the other

variable tends to be small.

OK, so what else is there to say about the covariance? One observation to make is the following. What's the covariance of X with X itself?

If you plug in X here, you see that what we have is expected value of X minus expected of X squared. And that's just the definition of the variance of a random variable. So that's one fact to keep in mind.

We had a shortcut formula for calculating variances. There's a similar shortcut formula for calculating covariances. In particular, we can calculate covariances in this particular way. That's just the convenient way of doing it whenever you need to calculate it.

And finally, covariances are very useful when you want to calculate the variance of a sum of random variables. We know that if two random variables are independent, the variance of the sum is the sum of the variances. When the random variables are dependent, this is no longer true, and we need to supplement the formula a little bit.

And there's a typo on the slides that you have in your hands. That term of 2 shouldn't be there. And let's see where that formula comes from.

Let's suppose that our random variables are independent of -- not independent -- our random variables have 0 means. And we want to calculate the variance. So the variance is going to be expected value of (X1 plus Xn) squared. What you do is you expand the square. And you get the expected value of the sum of the Xi squared.

And then you get all the cross terms. OK. And so now, here, let's assume for simplicity that we have 0 means. The expected value of this is the sum of the expected values of the X squared terms. And that gives us the variance.

And then we have all the possible cross terms. And each one of the possible cross terms is the expected value of Xi times Xj. This is just the covariance.

So if you can calculate all the variances and the covariances, then you're able to calculate also the variance of a sum of random variables. Now, if two random variables are independent, then you look at this expression. Because of independence, expected value of the product is going to be the product of the expected values. And the expected value of just this term is always equal to 0.

You're expected deviation from the mean is just 0. So the covariance will turn out to be 0. So independent random variables lead to 0 covariances, although the opposite fact is not necessarily true. So covariances give you some indication of the relation between two random variables.

Something that's not so convenient conceptually about covariances is that it has the wrong units. That's the same comment that we had made regarding variances. And with variances we got out of that issue by considering the standard deviation, which has the correct units.

So with the same reasoning, we want to have a concept that captures the relation between two random variables and, in some sense, that doesn't have to do with the units that we're dealing. We want to have a dimensionless quantity. That tells us how strongly two random variables are related to each other.

So instead of considering the covariance of just X with Y, we take our random variables and standardize them by dividing them by their individual standard deviations and take the expectation of this. So what we end up doing is the covariance of X and Y, which has units that are the units of X times the units of Y. But divide with a standard deviation, so that we get a quantity that doesn't have units.

This quantity, we call it the correlation coefficient. And it's a very useful quantity, a very useful measure of the strength of association between two random variables. It's very informative, because it falls always between -1 and +1. This is an algebraic exercise that you're going to see in recitation.

And the way that you interpret it is as follows. If the two random variables are independent, the covariance is going to be 0. The correlation coefficient is going to be 0. So 0 correlation coefficient basically indicates a lack of a systematic relation between the two random variables.

On the other hand, when rho is large, either close to 1 or close to -1, this is an indication of a strong association between the two random variables. And the extreme case is when rho takes an extreme value.

When rho has a magnitude equal to 1, it's as big as it can be. In that case, the two random variables are very strongly related. How strongly? Well, if you know one random variable, if

you know the value of y, you can recover the value of x and conversely.

So the case of a complete correlation is the case where one random variable is a linear function of the other random variable. In terms of a scatter plot, this would mean that there's a certain line and that the only possible (x,y) pairs that can happen would lie on that line. So if all the possible (x,y) pairs lie on this line, then you have this relation, and the correlation coefficient is equal to 1. A case where the correlation coefficient is close to 1 would be a scatter plot like this, where the x's and y's are quite strongly aligned with each other, maybe not exactly, but fairly strongly. All right, so you're going to hear a little more about correlation coefficients and covariances in recitation tomorrow.