6.041SC Probabilistic Systems Analysis and Applied Probability, Fall 2013 Transcript – Lecture 8

JOHN TSITSIKLIS: OK. We can start. Good morning.

So we're going to start now a new unit. For the next couple of lectures, we will be talking about continuous random variables. So this is new material which is not going to be in the quiz. You are going to have a long break next week without any lecture, just a quiz and recitation and tutorial.

So what's going to happen in this new unit? Basically, we want to do everything that we did for discrete random variables, reintroduce the same sort of concepts but see how they apply and how they need to be modified in order to talk about random variables that take continuous values. At some level, it's all the same. At some level, it's quite a bit harder because when things are continuous, calculus comes in. So the calculations that you have to do on the side sometimes need a little bit more thinking.

In terms of new concepts, there's not going to be a whole lot today, some analogs of things we have done. We're going to introduce the concept of cumulative distribution functions, which allows us to deal with discrete and continuous random variables, all of them in one shot. And finally, introduce a famous kind of continuous random variable, the normal random variable.

OK, so what's the story? Continuous random variables are random variables that take values over the continuum. So the numerical value of the random variable can be any real number. They don't take values just in a discrete set.

So we have our sample space. The experiment happens. We get some omega, a sample point in the sample space. And once that point is determined, it determines the numerical value of the random variable.

Remember, random variables are functions on the sample space. You pick a sample point. This determines the numerical value of the random variable. So that numerical value is going to be some real number on that line.

Now we want to say something about the distribution of the random variable. We want to say which values are more likely than others to occur in a certain sense. For example, you may be interested in a particular event, the event that the random variable takes values in the interval from a to b. And we want to say something about the probability of that event.

In principle, how is this done? You go back to the sample space, and you find all those outcomes for which the value of the random variable happens to be in that interval. The probability that the

random variable falls here is the same as the probability of all outcomes that make the random variable to fall in there. So in principle, you can work on the original sample space, find the probability of this event, and you would be done.

But similar to what happened in chapter 2, we want to kind of push the sample space in the background and just work directly on the real axis and talk about probabilities up here. So we want now a way to specify probabilities, how they are bunched together, or arranged, along the real line. So what did we do for discrete random variables? We introduced PMFs, probability mass functions. And the way that we described the random variable was by saying this point has so much mass on top of it, that point has so much mass on top of it, and so on.

And so we assigned a total amount of 1 unit of probability. We assigned it to different masses, which we put at different points on the real axis. So that's what you do if somebody gives you a pound of discrete stuff, a pound of mass in little chunks. And you place those chunks at a few points.

Now, in the continuous case, this total unit of probability mass does not sit just on discrete points but is spread all over the real axis. So now we're going to have a unit of mass that spreads on top of the real axis. How do we describe masses that are continuously spread?

The way we describe them is by specifying densities. That is, how thick is the mass that's sitting here? How dense is the mass that's sitting there? So that's exactly what we're going to do. We're going to introduce the concept of a probability density function that tells us how probabilities accumulate at different parts of the real axis.

So here's an example or a picture of a possible probability density function. What does that density function kind of convey intuitively? Well, that these x's are relatively less likely to occur. Those x's are somewhat more likely to occur because the density is higher.

Now, for a more formal definition, we're going to say that a random variable X is said to be continuous if it can be described by a density function in the following sense. We have a density function. And we calculate probabilities of falling inside an interval by finding the area under the curve that sits on top of that interval.

So that's sort of the defining relation for continuous random variables. It's an implicit definition. And it tells us a random variable is continuous if we can calculate probabilities this way. So the probability of falling in this interval is the area under this curve. Mathematically, it's the integral of the density over this particular interval. If the density happens to be constant over that interval, the area under the curve would be the length of the interval times the height of the density, which sort of makes sense.

Now, because the density is not constant but it kind of moves around, what you need is to write down an integral. Now, this formula is very much analogous to what you would do for discrete random variables. For a discrete random variable, how do you calculate this probability?

You look at all x's in this interval. And you add the probability mass function over that range. So just for comparison, this would be the formula for the discrete case-- the sum over all x's in the interval from a to b over the probability mass function. And there is a syntactic analogy that's happening here and which will be a persistent theme when we deal with continuous random variables.

Sums get replaced by integrals. In the discrete case, you add. In the continuous case, you integrate. Mass functions get replaced by density functions. So you can take pretty much any formula from the discrete case and translate it to a continuous analog of that formula, as we're going to see.

OK. So let's take this now as our model. What is the probability that the random variable takes a specific value if we have a continuous random variable? Well, this would be the case. It's a case of a trivial interval, where the two end points coincide. So it would be the integral from a to itself. So you're integrating just over a single point.

Now, when you integrate over a single point, the integral is just 0. The area under the curve, if you're only looking at a single point, it's 0. So big property of continuous random variables is that any individual point has 0 probability.

In particular, when you look at the value of the density, the density does not tell you the probability of that point. The point itself has 0 probability. So the density tells you something a little different. We are going to see shortly what that is.

Before we get there, can the density be an arbitrary function? Almost, but not quite. There are two things that we want. First, since densities are used to calculate probabilities, and since probabilities must be non-negative, the density should also be non-negative. Otherwise you would be getting negative probabilities, which is not a good thing. So that's a basic property that any density function should obey.

The second property that we need is that the overall probability of the entire real line should be equal to 1. So if you ask me, what is the probability that x falls between minus infinity and plus infinity, well, we are sure that x is going to fall in that range. So the probability of that event should be 1. So the probability of being between minus infinity and plus infinity should be 1, which means that the integral from minus infinity to plus infinity should be 1. So that just tells us that there's 1 unit of total probability that's being spread over our space.

Now, what's the best way to think intuitively about what the density function does? The interpretation that I find most natural and easy to convey the meaning of a density is to look at probabilities of small intervals. So let us take an x somewhere here and then x plus delta just next to it. So delta is a small number.

And let's look at the probability of the event that we get a value in that range. For continuous random variables, the way we find the probability of falling in that range is by integrating the density over that range. So we're drawing this picture. And we want to take the area under this curve.

Now, what happens if delta is a fairly small number? If delta is pretty small, our density is not going to change much over that range. So you can pretend that the density is approximately constant. And so to find the area under the curve, you just take the base times the height.

And it doesn't matter where exactly you take the height in that interval, because the density doesn't change very much over that interval. And so the integral becomes just base times the height. So for small intervals, the probability of a small interval is approximately the density times delta. So densities essentially give us probabilities of small intervals.

And if you want to think about it a little differently, you can take that delta from here and send it to the denominator there. And what this tells you is that the density is probability per unit length for intervals of small length. So the units of density are probability per unit length.

Densities are not probabilities. They are rates at which probabilities accumulate, probabilities per unit length. And since densities are not probabilities, they don't have to be less than 1.

Ordinary probabilities always must be less than 1. But density is a different kind of thing. It can get pretty big in some places. It can even sort of blow up in some places. As long as the total area under the curve is 1, other than that, the curve can do anything that it wants.

Now, the density prescribes for us the probability of intervals. Sometimes we may want to find the probability of more general sets. How would we do that? Well, for nice sets, you will just integrate the density over that nice set.

I'm not quite defining what "nice" means. That's a pretty technical topic in the theory of probability. But for our purposes, usually we will take b to be something like a union of intervals.

So how do you find the probability of falling in the union of two intervals? Well, you find the probability of falling in that interval plus the probability of falling in that interval. So it's the integral over this interval plus the integral over that interval.

And you think of this as just integrating over the union of the two intervals. So once you can calculate probabilities of intervals, then usually you are in business, and you can calculate anything else you might want. So the probability density function is a complete description of any statistical information we might be interested in for a continuous random variable.

OK. So now we can start walking through the concepts and the definitions that we have for discrete random variables and translate them to the continuous case. The first big concept is the concept of the expectation. One can start with a mathematical definition.

And here we put down a definition by just translating notation. Wherever we have a sum in the discrete case, we now write an integral. And wherever we had the probability mass function, we now throw in the probability density function.

This formula-- you may have seen it in freshman physics-- basically, it again gives you the center of gravity of the picture that you have when you have the density. It's the center of gravity of the object sitting underneath the probability density function. So that the interpretation still applies.

It's also true that our conceptual interpretation of what an expectation means is also valid in this case. That is, if you repeat an experiment a zillion times, each time drawing an independent sample of your random variable x, in the long run, the average that you are going to get should be the expectation. One can reason in a hand-waving way, sort of intuitively, the way we did it for the case of discrete random variables. But this is also a theorem of some sort. It's a limit theorem that we're going to visit later on in this class.

Having defined the expectation and having claimed that the interpretation of the expectation is that same as before, then we can start taking just any formula you've seen before and just translate it. So for example, to find the expected value of a function of a continuous random variable, you do not have to find the PDF or PMF of g(X). You can just work directly with the original distribution of the random variable capital X.

And this formula is the same as for the discrete case. Sums get replaced by integrals. And PMFs get replaced by PDFs. And in particular, the variance of a random variable is defined again the same way.

The variance is the expected value, the average of the distance of X from the mean and then squared. So it's the expected value for a random variable that takes these numerical values. And same formula as before, integral and F instead of summation, and the P.

And the formulas that we have derived or formulas that you have seen for the discrete case, they all go through the continuous case. So for example, the useful relation for variances, which is this one, remains true. All right. So time for an example.

The most simple example of a continuous random variable that there is, is the so-called uniform random variable. So the uniform random variable is described by a density which is 0 except over an interval. And over that interval, it is constant. What is it meant to convey? It's trying to convey the idea that all x's in this range are equally likely.

Well, that doesn't say very much. Any individual x has 0 probability. So it's conveying a little more than that. What it is saying is that if I take an interval of a given length delta, and I take another interval of the same length, delta, under the uniform distribution, these two intervals are going to have the same probability.

So being uniform means that intervals of same length have the same probability. So no interval is more likely than any other to occur. And in that sense, it conveys the idea of sort of complete randomness. Any little interval in our range is equally likely as any other little interval.

All right. So what's the formula for this density? I only told you the range. What's the height? Well, the area under the density must be equal to 1. Total probability is equal to 1. And so the

height, inescapably, is going to be 1 over (b minus a). That's the height that makes the density integrate to 1. So that's the formula.

And if you don't want to lose one point in your exam, you have to say that it's also 0, otherwise. OK. All right? That's sort of the complete answer.

How about the expected value of this random variable? OK. You can find the expected value in two different ways. One is to start with the definition. And so you integrate over the range of interest times the density. And you figure out what that integral is going to be.

Or you can be a little more clever. Since the center-of-gravity interpretation is still true, it must be the center of gravity of this picture. And the center of gravity is, of course, the midpoint. Whenever you have symmetry, the mean is always the midpoint of the diagram that gives you the PDF. OK. So that's the expected value of X.

Finally, regarding the variance, well, there you will have to do a little bit of calculus. We can write down the definition. So it's an integral instead of a sum. A typical value of the random variable minus the expected value, squared, times the density. And we integrate. You do this integral, and you find it's (b minus a) squared over that number, which happens to be 12.

Maybe more interesting is the standard deviation itself. And you see that the standard deviation is proportional to the width of that interval. This agrees with our intuition, that the standard deviation is meant to capture a sense of how spread out our distribution is. And the standard deviation has the same units as the random variable itself. So it's sort of good to-- you can interpret it in a reasonable way based on that picture.

OK, yes. Now, let's go up one level and think about the following. So we have formulas for the discrete case, formulas for the continuous case. So you can write them side by side. One has sums, the other has integrals.

Suppose you want to make an argument and say that something is true for every random variable. You would essentially need to do two separate proofs, for discrete and for continuous. Is there some way of dealing with random variables just one at a time, in one shot, using a sort of uniform notation? Is there a unifying concept?

Luckily, there is one. It's the notion of the cumulative distribution function of a random variable. And it's a concept that applies equally well to discrete and continuous random variables. So it's an object that we can use to describe distributions in both cases, using just one piece of notation.

So what's the definition? It's the probability that the random variable takes values less than a certain number little x. So you go to the diagram, and you see what's the probability that I'm falling to the left of this. And you specify those probabilities for all x's.

In the continuous case, you calculate those probabilities using the integral formula. So you integrate from here up to x. In the discrete case, to find the probability to the left of some point, you go here, and you add probabilities again from the left.

So the way that the cumulative distribution function is calculated is a little different in the continuous and discrete case. In one case you integrate. In the other, you sum. But leaving aside how it's being calculated, what the concept is, it's the same concept in both cases. So let's see what the shape of the cumulative distribution function would be in the two cases.

So here what we want is to record for every little x the probability of falling to the left of x. So let's start here. Probability of falling to the left of here is 0-- 0, 0, 0. Once we get here and we start moving to the right, the probability of falling to the left of here is the area of this little rectangle. And the area of that little rectangle increases linearly as I keep moving. So accordingly, the CDF increases linearly until I get to that point.

At that point, what's the value of my CDF? 1. I have accumulated all the probability there is. I have integrated it.

This total area has to be equal to 1. So it reaches 1, and then there's no more probability to be accumulated. It just stays at 1. So the value here is equal to 1.

OK. How would you find the density if somebody gave you the CDF? The CDF is the integral of the density. Therefore, the density is the derivative of the CDF. So you look at this picture and take the derivative.

Derivative is 0 here, 0 here. And it's a constant up there, which corresponds to that constant. So more generally, and an important thing to know, is that the derivative of the CDF is equal to the density-- almost, with a little bit of an exception.

What's the exception? At those places where the CDF does not have a derivative-- here where it has a corner-- the derivative is undefined. And in some sense, the density is also ambiguous at that point. Is my density at the endpoint, is it 0 or is it 1?

It doesn't really matter. If you change the density at just a single point, it's not going to affect the value of any integral you ever calculate. So the value of the density at the endpoint, you can leave it as being ambiguous, or you can specify it.

It doesn't matter. So at all places where the CDF has a derivative, this will be true. At those places where you have corners, which do show up sometimes, well, you don't really care.

How about the discrete case? In the discrete case, the CDF has a more peculiar shape. So let's do the calculation. We want to find the probability of b to the left of here. That probability is 0, 0, 0. Once we cross that point, the probability of being to the left of here is 1/6. So as soon as we cross the point 1, we get the probability of 1/6, which means that the size of the jump that we have here is 1/6.

Now, question. At this point 1, which is the correct value of the CDF? Is it 0, or is it 1/6? It's 1/6 because-- you need to look carefully at the definitions, the probability of x being less than or equal to little x.

If I take little x to be 1, it's the probability that capital X is less than or equal to 1. So it includes the event that x is equal to 1. So it includes this probability here. So at jump points, the correct value of the CDF is going to be this one.

And now as I trace, x is going to the right. As soon as I cross this point, I have added another 3/6 probability. So that 3/6 causes a jump to the CDF. And that determines the new value. And finally, once I cross the last point, I get another jump of 2/6.

A general moral from these two examples and these pictures. CDFs are well defined in both cases. For the case of continuous random variables, the CDF will be a continuous function. It starts from 0. It eventually goes to 1 and goes smoothly-- well, continuously from smaller to higher values. It can only go up. It cannot go down since we're accumulating more and more probability as we are going to the right.

In the discrete case, again it starts from 0, and it goes to 1. But it does it in a staircase manner. And you get a jump at each place where the PMF assigns a positive mass. So jumps in the CDF are associated with point masses in our distribution. In the continuous case, we don't have any point masses, so we do not have any jumps either.

Now, besides saving us notation-- we don't have to deal with discrete and continuous twice-- CDFs give us actually a little more flexibility. Not all random variables are continuous or discrete. You can cook up random variables that are kind of neither or a mixture of the two.

An example would be, let's say you play a game. And with a certain probability, you get a certain number of dollars in your hands. So you flip a coin. And with probability 1/2, you get a reward of 1/2 dollars.

And with probability 1/2, you are led to a dark room where you spin a wheel of fortune. And that wheel of fortune gives you a random reward between 0 and 1. So any of these outcomes is possible. And the amount that you're going to get, let's say, is uniform.

So you flip a coin. And depending on the outcome of the coin, either you get a certain value or you get a value that ranges over a continuous interval. So what kind of random variable is it?

Is it continuous? Well, continuous random variables assign 0 probability to individual points. Is it the case here? No, because you have positive probability of obtaining 1/2 dollar. So our random variable is not continuous.

Is it discrete? It's not discrete, because our random variable can take values also over a continuous range. So we call such a random variable a mixed random variable.

If you were to draw its distribution very loosely, probably you would want to draw a picture like this one, which kind of conveys the idea of what's going on. So just think of this as a drawing of masses that are sitting over a table. We place an object that weighs half a pound, but it's an object that takes zero space. So half a pound is just sitting on top of that point. And we take another half-pound of probability and spread it uniformly over that interval.

So this is like a piece that comes from mass functions. And that's a piece that looks more like a density function. And we just throw them together in the picture. I'm not trying to associate any formal meaning with this picture. It's just a schematic of how probabilities are distributed, help us visualize what's going on.

Now, if you have taken classes on systems and all of that, you may have seen the concept of an impulse function. And you my start saying that, oh, I should treat this mathematically as a so-called impulse function. But we do not need this for our purposes in this class. Just think of this as a nice picture that conveys what's going on in this particular case.

So now, what would the CDF look like in this case? The CDF is always well defined, no matter what kind of random variable you have. So the fact that it's not continuous, it's not discrete shouldn't be a problem as long as we can calculate probabilities of this kind.

So the probability of falling to the left here is 0. Once I start crossing there, the probability of falling to the left of a point increases linearly with how far I have gone. So we get this linear increase. But as soon as I cross that point, I accumulate another 1/2 unit of probability instantly. And once I accumulate that 1/2 unit, it means that my CDF is going to have a jump of 1/2.

And then afterwards, I still keep accumulating probability at a fixed rate, the rate being the density. And I keep accumulating, again, at a linear rate until I settle to 1. So this is a CDF that has certain pieces where it increases continuously. And that corresponds to the continuous part of our randomize variable. And it also has some places where it has discrete jumps. And those district jumps correspond to places in which we have placed a positive mass.

And by the-- OK, yeah. So this little 0 shouldn't be there. So let's cross it out. All right.

So finally, we're going to take the remaining time and introduce our new friend. It's going to be the Gaussian or normal distribution. So it's the most important distribution there is in all of probability theory. It's plays a very central role. It shows up all over the place. We'll see later in the class in more detail why it shows up.

But the quick preview is the following. If you have a phenomenon in which you measure a certain quantity, but that quantity is made up of lots and lots of random contributions-- so your random variable is actually the sum of lots and lots of independent little random variables-- then invariability, no matter what kind of distribution the little random variables have, their sum will turn out to have approximately a normal distribution. So this makes the normal distribution to arise very naturally in lots and lots of contexts. Whenever you have noise that's comprised of lots of different independent pieces of noise, then the end result will be a random variable that's normal.

So we are going to come back to that topic later. But that's the preview comment, basically to argue that it's an important one. OK. And there's a special case. If you are dealing with a binomial distribution, which is the sum of lots of Bernoulli random variables, again you would expect that the binomial would start looking like a normal if you have many, many-- a large number of point fields.

All right. So what's the math involved here? Let's parse the formula for the density of the normal. What we start with is the function X squared over 2. And if you are to plot X squared over 2, it's a parabola, and it has this shape -- X squared over 2.

Then what do we do? We take the negative exponential of this. So when X squared over 2 is 0, then negative exponential is 1. When X squared over 2 increases, the negative exponential of that falls off, and it falls off pretty fast.

So as this goes up, the formula for the density goes down. And because exponentials are pretty strong in how quickly they fall off, this means that the tails of this distribution actually do go down pretty fast. OK. So that explains the shape of the normal PDF.

How about this factor 1 over square root 2 pi? Where does this come from? Well, the integral has to be equal to 1. So you have to go and do your calculus exercise and find the integral of this the minus X squared over 2 function and then figure out, what constant do I need to put in front so that the integral is equal to 1?

How do you evaluate that integral? Either you go to Mathematica or Wolfram's Alpha or whatever, and it tells you what it is. Or it's a very beautiful calculus exercise that you may have seen at some point. You throw in another exponential of this kind, you bring in polar coordinates, and somehow the answer comes beautifully out there.

But in any case, this is the constant that you need to make it integrate to 1 and to be a legitimate density. We call this the standard normal. And for the standard normal, what is the expected value? Well, the symmetry, so it's equal to 0.

What is the variance? Well, here there's no shortcut. You have to do another calculus exercise. And you find that the variance is equal to 1. OK. So this is a normal that's centered around 0.

How about other types of normals that are centered at different places? So we can do the same kind of thing. Instead of centering it at 0, we can take some place where we want to center it, write down a quadratic such as (X minus mu) squared, and then take the negative exponential of that. And that gives us a normal density that's centered at mu.

Now, I may wish to control the width of my density. To control the width of my density, equivalently I can control the width of my parabola. If my parabola is narrower, if my parabola looks like this, what's going to happen to the density? It's going to fall off much faster.

OK. How do I make my parabola narrower or wider? I do it by putting in a constant down here. So by putting a sigma here, this stretches or widens my parabola by a factor of sigma. Let's see. Which way does it go?

If sigma is very small, this is a big number. My parabola goes up quickly, which means my normal falls off very fast. So small sigma corresponds to a narrower density.

And so it, therefore, should be intuitive that the standard deviation is proportional to sigma. Because that's the amount by which you are scaling the picture. And indeed, the standard deviation is sigma. And so the variance is sigma squared.

So all that we have done here to create a general normal with a given mean and variance is to take this picture, shift it in space so that the mean sits at mu instead of 0, and then scale it by a factor of sigma. This gives us a normal with a given mean and a given variance. And the formula for it is this one.

All right. Now, normal random variables have some wonderful properties. And one of them is that they behave nicely when you take linear functions of them. So let's fix some constants a and b, suppose that X is normal, and look at this linear function Y.

What is the expected value of Y? Here we don't need anything special. We know that the expected value of a linear function is the linear function of the expectation. So the expected value is this.

How about the variance? We know that the variance of a linear function doesn't care about the constant term. But the variance gets multiplied by a squared. So we get these variance, where sigma squared is the variance of the original normal.

So have we used so far the property that X is normal? No, we haven't. This calculation here is true in general when you take a linear function of a random variable. But if X is normal, we get the other additional fact that Y is also going to be normal. So that's the nontrivial part of the fact that I'm claiming here. So linear functions of normal random variables are themselves normal.

How do we convince ourselves about it? OK. It's something that we will do formerly in about two or three lectures from today. So we're going to prove it. But if you think about it intuitively, normal means this particular bell-shaped curve. And that bell-shaped curve could be sitting anywhere and could be scaled in any way.

So you start with a bell-shaped curve. If you take X, which is bell shaped, and you multiply it by a constant, what does that do? Multiplying by a constant is just like scaling the axis or changing the units with which you're measuring it. So it will take a bell shape and spread it or narrow it. But it will still be a bell shape. And then when you add the constant, you just take that bell and move it elsewhere.

So under linear transformations, bell shapes will remain bell shapes, just sitting at a different place and with a different width. And that sort of the intuition of why normals remain normals under this kind of transformation. So why is this useful?

Well, OK. We have a formula for the density. But usually we want to calculate probabilities. How will you calculate probabilities? If I ask you, what's the probability that the normal is less than 3, how do you find it? You need to integrate the density from minus infinity up to 3.

Unfortunately, the integral of the expression that shows up that you would have to calculate, an integral of this kind from, let's say, minus infinity to some number, is something that's not known in closed form. So if you're looking for a closed-form formula for this-- X bar-- if you're looking for a closed-form formula that gives you the value of this integral as a function of X bar, you're not going to find it. So what can we do?

Well, since it's a useful integral, we can just tabulate it. Calculate it once and for all, for all values of X bar up to some precision, and have that table, and use it. That's what one does.

OK, but now there is a catch. Are we going to write down a table for every conceivable type of normal distribution-- that is, for every possible mean and every variance? I guess that would be a pretty long table. You don't want to do that.

Fortunately, it's enough to have a table with the numerical values only for the standard normal. And once you have those, you can use them in a clever way to calculate probabilities for the more general case. So let's see how this is done.

So our starting point is that someone has graciously calculated for us the values of the CDF, the cumulative distribution function, that is the probability of falling below a certain point for the standard normal and at various places. How do we read this table? The probability that X is less than, let's say, 0.63 is this number. This number, 0.7357, is the probability that the standard normal is below 0.63. So the table refers to the standard normal.

But someone, let's say, gives us some other numbers and tells us we're dealing with a normal with a certain mean and a certain variance. And we want to calculate the probability that the value of that random variable is less than or equal to 3. How are we going to do it? Well, there's a standard trick, which is so-called standardizing a random variable.

Standardizing a random variable stands for the following. You look at the random variable, and you subtract the mean. This makes it a random variable with 0 mean. And then if I divide by the standard deviation, what happens to the variance of this random variable? Dividing by a number divides the variance by sigma squared.

The original variance of X was sigma squared. So when I divide by sigma, I end up with unit variance. So after I do this transformation, I get a random variable that has 0 mean and unit variance.

It is also normal. Why is its normal? Because this expression is a linear function of the X that I started with. It's a linear function of a normal random variable. Therefore, it is normal. And it is a standard normal.

So by taking a general normal random variable and doing this standardization, you end up with a standard normal to which you can then apply the table. Sometimes one calls this the normalized score. If you're thinking about test results, how would you interpret this number? It tells you how many standard deviations are you away from the mean.

This is how much you are away from the mean. And you count it in terms of how many standard deviations it is. So this number being equal to 3 tells you that X happens to be 3 standard deviations above the mean. And I guess if you're looking at your quiz scores, very often that's the kind of number that you think about. So it's a useful quantity.

But it's also useful for doing the calculation we're now going to do. So suppose that X has a mean of 2 and a variance of 16, so a standard deviation of 4. And we're going to calculate the probability of this event. This event is described in terms of this X that has ugly means and variances. But we can take this event and rewrite it as an equivalent event.

X less than 3 is this same as X minus 2 being less than 3 minus 2, which is the same as this ratio being less than that ratio. So I'm subtracting from both sides of the inequality the mean and then dividing by the standard deviation. This event is the same as that event.

Why do we like this better than that? We like it because this is the standardized, or normalized, version of X. We know that this is standard normal. And so we're asking the question, what's the probability that the standard normal is less than this number, which is 1/4? So that's the key property, that this is normal (0, 1).

And so we can look up now with the table and ask for the probability that the standard normal random variable is less than 0.25. Where is that going to be? 0.2, 0.25, it's here. So the answer is 0.987.

So I guess this is just a drill that you could learn in high school. You didn't have to come here to learn about it. But it's a drill that's very useful when we will be calculating normal probabilities all the time. So make sure you know how to use the table and how to massage a general normal random variable into a standard normal random variable.

OK. So just one more minute to look at the big picture and take stock of what we have done so far and where we're going. Chapter 2 was this part of the picture, where we dealt with discrete random variables. And this time, today, we started talking about continuous random variables. And we introduced the density function, which is the analog of the probability mass function.

We have the concepts of expectation and variance and CDF. And this kind of notation applies to both discrete and continuous cases. They are calculated the same way in both cases except that in the continuous case, you use sums. In the discrete case, you use integrals.

So on that side, you have integrals. In this case, you have sums. In this case, you always have Fs in your formulas. In this case, you always have Ps in your formulas.

So what's there that's left for us to do is to look at these two concepts, joint probability mass functions and conditional mass functions, and figure out what would be the equivalent concepts on the continuous side. So we will need some notion of a joint density when we're dealing with multiple random variables. And we will also need the concept of conditional density, again for the case of continuous random variables.

The intuition and the meaning of these objects is going to be exactly the same as here, only a little subtler because densities are not probabilities. They're rates at which probabilities accumulate. So that adds a little bit of potential confusion here, which, hopefully, we will fully resolve in the next couple of sections.

All right. Thank you.

6.041SC Probabilistic Systems Analysis and Applied Probability
Fall 2013