PROFESSOR: Hi. Today I'd like to introduce a new module. In previous modules we've talked about how to model particular systems, how to make predictions about certain kinds and classifications of systems, and also how to design both theoretical and physical systems.

If our systems are operating in a deterministic universe, then we're all set. We've accounted for all the things we could possibly account for. But if we're making systems that are going to operate in the real world, then we need to be able to deal with some level of uncertainty.

Today I'm going to talk about probability, which is the method by which we're going to model uncertainty in our world. And later we're going to talk about different strategies we can take to deal with that uncertainty to hopefully increase the level of autonomy of our systems as they operate in the real world.

The first thing I need to talk about is how to properly model probability, or how to probably talk about probability such that we can use it to talk about uncertainty. When you're talking about probability, typically you'll end up talking about the sample space or U. This refers to your entire universe, where your universe is the values that you care about, or the possible assigned values of the variables that you care about.

I'm already talking about variables, but what I really mean to say is events. There are different states that your universe can take, and those states can be sort of exhaustively enumerated or be atomic, or they can be factored into different variables.

Let me elaborate on this a little. If I were talking about four coin flips, if those events were specified atomically then I would have to exhaustively specify all possible sequences of four coin flips. 4 heads, 3 heads and 2 tail, 2 heads and 2 tails, that sort of exercise. Or flipping 4 heads would be one event, flipping 3 heads and then 1 tail would be a separate event, flipping 2 heads 1 tail and another head would be

a third, distinct event.

If we're talking about a factored state space, then we'll have a different variable representing each one of these coin flips. And each one of those variables will say, here is the value associated with that particular sub-event. And the accumulation of all those values, or the specification for all those values, will end up referring to the same states that were addressed by the atomic events.

Let me talk about variables, because they're the thing that allow us to not exhaustively enumerate every possible event in the universe, and talk about our universe in meaningful ways, or in ways that they can be effectively communicated. And why am I talking about random variables? Why aren't they just variables? Well, random variables is the way that you specify the fact that you're talking about probabilistic variables.

When you're just dealing with regular algebra, variables have some sort of assigned value, and you're not forced to be in the space of 0 to 1. When you're talking about from 0 to 1, when you're talking about probabilities, you're forced to be in the spaces of 0 to 1 and forced to remain within the reals. If you want to talk about all the possible assigned values of that random variable, then you're talking about the distribution over that random variable. This means that A could be anything that A could be. You're talking about the function that says give me a value of A, and I will tell you a probability associated with that value of A.

If you're talking about the probability of A being assigned to a particular value, then you'll return out the probability. If A represents the color of the shirt I'm wearing, and I want to know the probability that A is some value, then I look at the color of the shirt I'm wearing and try to determine whether or not it's 1 or 0. Or possibly look at the colors of shirts that I've worn over the past year and make some sort of ratio of the number of pink shirts I've worn over the past year.

If I'm dealing with a factored state space, I'm going to end up talking about more than one random variable. There are two main ways to talk about more than one random variable at the same time. One is joint probabilities, where all the random

variables are collectively specified at the same time. And the other is conditional probabilities, where you've already decided that you specified some value for one or more random variables. And then within that scope you're going to talk about the probabilities associated with other random variables.

I want to demonstrate this graphically, but there are two more things that I need to mention right now. One is the difference between frequentist and Bayesian interpretations of probability. Right now they don't seem particularly relevant, but people will use these words, and it's good to know, approximately, what they're talking about. The frequentist interpretation of probability is more relevant when you're talking about actions that happen a lot of different times. For instance, how frequently it rains.

If I say that today is a Wednesday, and I want to know the probability that it rains on Wednesdays, then that probability is open to frequentist interpretation because there are a lot of Wednesdays, and it's rained a lot in the universe of Wednesdays or the space of possible Wednesdays. So thinking about the fact that there's a 70% chance of rain, or a 30% chance of rain, or whatever probability of rain there is on Wednesdays, makes sense, or is open to frequentist interpretation.

The other interpretation that you'll hear talked about with respect to probabilities is the Bayesian interpretation. Bayesian interpretation tends to be more relevant when you're talking about spaces that are more atomic as opposed to factored, or represent events that are specified to the point that it does not make sense to talk about them in the frequentist sense. When we talk about probabilities in the Bayesian interpretation, we frequently use the term likelihood.

For instance, if I'm talking about whether or not it's likely to rain on August 24, 2011 in the afternoon, the specificity of that event is so high that at that point it doesn't make sense to talk about the frequency of Wednesday, August 24, 2011 in the afternoons, at least for now. At that point we're talking more about likelihood and less about frequency. That event is more conducive to Bayesian interpretation.

No whirlwind tour of probability would be complete without a mention of the three

axioms of probability. The first axiom is that the likelihood of the universe happening is one, or all random variables are going to be specified at some point. Relatedly, the likelihood of nothing happening is 0. What these two really do is establish the boundaries of a graphical representation up here.

The other axiom of probability is that if you're going to be talking about the union of two events, or the probability associated with one or the other event having an assignment, then you're talking about the probability of one of those variables and the probability of the other variable, and then removing the section that you've double-counted.

So if I were to attempt to demonstrate this on this graph, I would be talking about the probability of A, added the probability of B. And at this point I've double-counted this section the middle where they're both one. So I would subtract this exactly once. And then I'm talking about the size of this space.

If I'm talking about the probability of A being equal to 1, then I'm talking about the space in which A is specified to be equal to 1, divided by the area associated with my universe. When I'm talking about joint distributions, I'm going to find the space in which both of these things are true, scoped to the size of the universe, or the entire sample space. So in this space both A and B are true, and I'm looking at it relative to the size of the universe.

In contrast, when I'm talking about conditional probability, or the probability of A given B, I'm going to look at where my specifications are true, scoped to where my givens are true. So if I'm already dealing in the space restricted to B, I'm just looking at this size, or this space. But because I'm scoped to B, I'm going to end up dividing by the area of B, instead of the area of U.

This is the main difference between the joint and conditional distributions. And a lot of people get hung up on it, which is why I'm exhaustively walking through it.

A couple of other things before we talk about the first way in which we can use our models for uncertainty to do some amount of addressing of the fact that we're going

to have to deal with uncertainty in the future. If we start off with a joint distribution and we want to reduce the number of variables that we're actually talking about, we can do so by exhaustively walking through all the possible assigned values for the variable that we want to disregard, and then summing up the values appropriately.

An easy example for this is if we had the joint probabilities of all the colors of shirts that I wear and all the colors of pants that I wear, and we only wanted to talk about all the colors of shirts that I wear, then we could exhaustively cover all the different colors of pants that I wear and accumulate all the different values of shirts that I wear simultaneously, and then collect that distribution.

Related is the concept of total probability, which is the same kind of accumulation. It just operates in the conditional space, as opposed to in the joint space. So in this case, if I'm already operating in A given B land, I have to scope myself back out to the space of the universe by then accounting for the fact that I've only been operating in the scope of B. Then exhaustively enumerate all the possible values of B, sum the probabilities associated with those values, and then I've reduced the number of dimensions that I'm talking about.

The final thing we have to talk about before we move on to state estimation is Bayes' evidence. Or you've probably seen this demonstrated as Bayes' rule. If I want to talk about B scoped to A, and all I have is A scoped to B, B and A. When we walked through total probability, we saw that the conditional probability multiplied by the probability associated with the variable that we're conditioning on, is equal to the joint probability.

When we multiply P of A given B by P of B, we're going to end up with the joint probability associated with the two variables. When we then divide back out by A or scope our joint probability to A, we end up talking about conditional probabilities again, which is where B given A comes from. Bayes' evidence, or Bayes' rule, is the basis for inference, which is going to be really important for state estimation, which we'll cover next time.