

PROFESSOR: Random walks provide probabilistic models for a bunch of settings. In fact, we've seen a couple already, so let's examine what they are in general. So the set up for a random walk is that you have a digraph, and we can also often think and talk about the digraph as though it was a state diagram for a machine with state, so here's a three-state digraph-- blue, orange, and green-- and the part that becomes probabilistic is that we think of the process of which edge to follow when you're at a given state is made probabilistically.

And the only rules are that we're going to assign probabilities to the edges in a way like this where, for example, what I'm telling you is there's a $1/3$ probability that I'll follow the edge from O to O, and a $2/3$ probability that I'll follow the edge from O to green, and the rule is simply that the sum of the probabilities on the outgoing edges has to sum to 1. So let's fill in the rest of the graph in a legal way. So here we have B with a $1/2$ probability of going from B to B, a $1/4$ from B to O, and a quarter from B to G. And the green state is certain in one step if you're at green to go to blue next. There's only one edge out. It has probability 1.

Now, gambler's ruin can be seen as an example of this kind of a random walk. The states where the amount of money that you had, ranging from 0 when you're bankrupt to T when you've reached your target and N is the start state, which is your initial stake, the green edges are weighted with a probability P of winning a bet, so we have transitions from K to K plus 1 for K less than T, with a weight probability P and, likewise, the red edges are weighted with the probability of losing a bet, Q, or 1 minus P.

So there is a digraph, or a state machine, that describes the gambler's ruin problem as a probabilistic walk on a graph. And the typical kind of question that we would ask about a random walk on a graph would be, what's the probability of reaching T, the target, before reaching 0, bankrupt, given that you're starting at some state And.

So in walks come up in a bunch of quite different settings. For example, in physics Brownian motion is the random motion of a particle that's being buffeted by atomic forces, and its modeled by saying that this particle can move in any direction in 3D space and chosen uniformly at random. And the theory of Brownian motion-- it had been observed first without being understood. Einstein was the first one to come up with a random walk model and corresponding theorems about the behavior of particles under Brownian motion. In fact, that was one of the main components of his Nobel Prize. He wasn't given a Nobel Prize for

relativity at the time, because it had not yet been firmly proven, although it was widely celebrated.

Another case is in finance. We've already seen how gambler's ruin reflects, or seems to reflect, the biased random oscillation of stock prices over time, and we will see at the end of this set of videos an application of random walks on a graph to model web search and clustering of a focus on vertices in a digraph.

So the general kinds of questions that come up when you're talking about random walks on graphs are illustrated by this simple three-state example with blue, orange, and green. We might ask, for example, starting at state B, what's the probability of reaching state O in seven steps? And that would be easy enough to calculate in this small example, but it would be a typical question.

A more interesting general question would be, what's the average number of steps that it takes to get from B to O? I mean, you could with probability of $1/4$, you go there in one step, but with probability $1/8$, you go there in three steps, and so on. You can calculate again explicitly and easily enough what the average number of steps from B to O is in this simple example, and we'll shortly remark about general ways to solve that problem.

And finally, you can ask a gambler's ruin type question, what's the probability of starting at B of getting to G before O? Well, in this trivial example, you can just read off the answer. You are going to get to G before O with 50-50 probability, because from B you have to go one place or the other with equal probability. But in general, this becomes a more interesting and complicated question, which you can solve by methods that we're about to lay out.

Let me just remind you that we've already seen an interesting and illustrative example of random walk on a graph when we were looking at coin tosses. The problem, for example, of if I toss a fair coin and I wait for three consecutive tosses that form the pattern HTH or the pattern TTH, and I want to know what's the probability of winning because HTH comes before TTH, I can model that with a-- we said with an infinite tree diagram, using our tree method for forming probability spaces, but the tree was very recursively defined. Sub-trees were isomorphic to the original tree, which allowed us, in fact, to come up with a finite description of the infinite tree that amounted to a finite state machine or finite graph. So let's look at that example in more detail.

If I'm trying to model the coin flipping thing, we start off in a state where the previous two flips

don't exist. I haven't flipped anything yet. The state's going to record the values of the previous two flips, and with no prior flips, there's a 50-50 chance that the first flip will be H, in which case I'm in the state with just an H and nothing preceding, or there's a 50-50 chance I flip a T, in which case I'm in the state in which there's been a T and nothing previous.

But I can already say something then about the probability of tossing HTH before TTH, namely the probability of winning. The probability of winning is, of course, the probability of winning, given that I start at the start state with no prior flips, but the probability that I win starting here is simply the probability that I win starting at the state nothing H and where the probability that I win at the state started nothing T, with the two probabilities weighted equally since this is a fair coin, and there's a 50-50 chance of going each way.

That is the probability of winning given no prior tosses is $\frac{1}{2}$ the probability of winning if the first toss is an H plus $\frac{1}{2}$ the probability of winning if the first toss is a T. This is just an application of the Law of Total Probability. So continuing in this way, let's expand more of the digraph. So suppose that I have tossed a head and then after that I toss a head, and I go to state HH or I toss a T, and I go to state HT. So here I'm just recording the previous two flips, with the most recent one on the right.

This structure of the state diagram tells us that if I want to know what the probability of winning, given that I flipped exactly one head at the start, the probability is simply by, again, total probability. The probability of winning from HH weighted by $\frac{1}{2}$ and the probability of winning from HT weighted by $\frac{1}{2}$, and I wind up again with a simple linear equation that connects the probability of winning in one state with the probability of winning in the states that it goes to.

Let's continue and do another example. So, likewise, if I expand what happens after I flip a T or an H after having flipped the first head, I get a corresponding equation that the probability of winning after a single tail is the same as $\frac{1}{2}$ the probability of winning with a tail followed by an H or a tail followed by a tail. This is a more interesting part of the diagram, where suppose that my past two flips have been two H's.

Well, if I flip an H again, then I'm back in state where the previous two flips were H's, or if I flip a T, then I'm in this state HT, where the previous flips were an H and a T, in that order. And that tells me, if I want to know about the probability of winning given HH, now it's the probability of winning given HH plus the probability of winning given HT. And, again, it's a linear equation

connecting up the probability of winning in one state with the probability of winning in other states, possibly itself, but there's no circularity here. It's just a system of linear equations.

Well, there's what the whole diagram looks like. In particular, once you flipped HT, if you then flip an H you've won because you got to HTH first, and you stay in the win state forever or, alternatively, once you flip TT, if you flip an H you've lost because TT has come up first. If you flip the T again, you stay in state TT.

And what we can say is the probability of winning if we're in the win state is 1, and the probability of winning if you're in a lose state is 0. And overall, I simply have this system of linear equations for the probability of winning in one state given other states, and I can solve these linear equations to find the probability of winning in the start state, which is simply the probability of winning.

So looking back at our questions for random walks, where we ask whether the probability of reaching O in seven steps starting at B, what's the probability of that? What's the average number of steps to go from B to O? What's the probability of reaching G before O, starting at B? In every case, these questions can be formulated simply as solving systems of linear equations whose structure directly reflects the structure of the digraph.