

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:

Today we're going to study stochastic processes and, among them, one type of it, so discrete time. We'll focus on discrete time. And I'll talk about what it is right now.

So a stochastic process is a collection of random variables indexed by time, a very simple definition. So we have either-- let's start from 0-- random variables like this, or we have random variables given like this. So a time variable can be discrete, or it can be continuous.

With these ones, we'll call discrete time stochastic processes, and these ones continuous time. So for example, a discrete time random variable can be something like-- and so on.

So these are the values, x_0 , x_1 , x_2 , x_3 , and so on. And they are random variables. This is just one-- so one realization of the stochastic process. But all these variables are supposed to be random.

And then a continuous time random variable-- a continuous time stochastic process can be something like that. And it doesn't have to be continuous, so it can jump and it can jump and so on. And all these values are random values.

So that's just a very informal description. And a slightly different point of view, which is slightly preferred, when you want to do some math with it, is that-- alternative definition-- it's a probability distribution over paths, over a space of paths.

So you have all a bunch of possible paths that you can take. And you're given some probability distribution over it. And then that will be one realization. Another realization will look something different and so on.

So this one-- it's more a intuitive definition, the first one, that it's a collection of

random variables indexed by time. But that one, if you want to do some math with it, from the formal point of view, that will be more helpful. And you'll see why that's the case later.

So let me show you some more examples. For example, to describe one stochastic process, this is one way to describe a stochastic process. $f(t)$ with-- let me show you three stochastic processes, so number one, $f(t) = t$. And this was probability 1.

Number 2, $f(t)$ is equal to t , for all t , with probability $1/2$, or $f(t)$ is equal to $-t$, for all t , with probability $1/2$. And the third one is, for each t , $f(t)$ is equal to t or $-t$, with probability $1/2$.

The first one is quite easy to picture. It's really just-- there's nothing random in here. This happens with probability 1. Your path just says $f(t) = t$. And we're only looking at t greater than or equal to 0 here. So that's number 1. Number 2, it's either this one or this one.

So it is a stochastic process. If you think about it this way, it doesn't really look like a stochastic process. But under the alternative definition, you have two possible paths that you can take. You either take this path, with $1/2$, or this path, with $1/2$.

Now, at each point, t , your value $x(t)$ is a random variable. It's either t or $-t$. And it's the same for all t . But they are dependent on each other. So if you know one value, you automatically know all the other values.

And the third one is even more interesting. Now, for each t , we get rid of this dependency. So what you'll have is these two lines going on. I mean at every single point, you'll be either a top one or a bottom one. But if you really want draw the picture, it will bounce back and forth, up and down, infinitely often, and it'll just look like two lines.

So I hope this gives you some feeling about stochastic processes, I mean, why we want to describe it in terms of this language, just a tiny bit. Any questions?

So, when you look at a process, when you use a stochastic process to model a real

life something going on, like a stock price, usually what happens is you stand at time t . And you know all the values in the past-- know. And in the future, you don't know.

But you want to know something about it. You want to have some intelligent conclusion, intelligent information about the future, based on the past. For this stochastic processes, it's easy. No matter where you stand at, you exactly know what's going to happen in the future.

For this one, it's also the same. Even though it's random, once you know what happened at some point, you know it has to be this distribution or this line, if it's here, and this line if it's there. But that one is slightly different.

No matter what you know about the past, even if know all the values in the past, what happened, it doesn't give any information at all about the future. Though it's not true if I say any information at all. We know that each value has to be t or minus t . You just don't know what it is.

So when you're given a stochastic process and you're standing at some time, your future, you don't know what the future is, but most of the time you have at least some level of control given by the probability distribution.

Here, it was, you can really determine the line. Here, because of probability distribution, at each point, only gives t or minus t , you know that each of them will be at least one of the points, but you don't know more than that.

So the study of stochastic processes is, basically, you look at the given probability distribution, and you want to say something intelligent about the future as t goes on.

So there are three types of questions that we mainly study here. So a, first type, is what are the dependencies in the sequence of values. For example, if you know the price of a stock on all past dates, up to today, can you say anything intelligent about the future stock prices-- those type of questions.

And b is what is the long term behavior of the sequence? So think about the law of large numbers that we talked about last time or central limit theorem.

And the third type, this one is left relevant for our course, but, still, I'll just write it down. What are the boundary events? How often will something extreme happen, like how often will a stock price drop by more than 10% for a consecutive 5 days-- like these kind of events. How often will that happen?

And for a different example, like if you model a call center and you want to know, over a period of time, the probability that at least 90% of the phones are idle or those kind of things.

So that's was an introduction. Any questions? Then there are really lots of stochastic processes. One of the most important ones is the simple random walk.

So today, I will focus on discrete time stochastic processes. Later in the course, we'll go on to continuous time stochastic processes. And then you'll see like Brownian motions and-- what else-- Ito's lemma and all those things will appear later.

Right now, we'll study discrete time. And later, you'll see that it's really just-- what is it-- they're really parallel. So this simple random walk, you'll see the corresponding thing in continuous time stochastic processes later.

So I think it's easier to understand discrete time processes, that's why we start with it. But later, it will really help if you understand it well. Because for continuous time, it will just carry over all the knowledge.

What is a simple random walk? Let Y_i be IID, independent identically distributed, random variables, taking values 1 or minus 1, each with probability 1/2. Then define, for each time t , X_t as the sum of Y_i , from i equals 1 to t .

Then the sequence of random variables, and X_0 is equal to 0. X_0, X_1, X_2 , and so on is called a one-dimensional, simple random walk. But I'll just refer to it as simple random walk or random walk. And this is a definition. It's called simple random walk.

Let's try to plot it. At time 0, we start at 0. And then, depending on the value of Y_1 , you will either go up or go down. Let's say we went up. So that's at time 1. Then at time 2, depending on your value of Y_2 , you will either go up one step from here or

go down one step from there. Let's say we went up again, down, 4, up, up, something like that. And it continues.

Another way to look at it-- the reason we call it a random walk is, if you just plot your values of X_t , over time, on a line, then you start at 0, you go to the right, right, left, right, right, left, left, left.

So the trajectory is like a walk you take on this line, but it's random. And each time you go to the right or left, right or left, right or left. So that was two representations. This picture looks a little bit more clear. Here, I just lost everything I draw. Something like that is the trajectory.

So from what we learned last time, we can already say something intelligent about the simple random walk. For example, if you apply central limit theorem to the sequence, what is the information you get?

So over a long time, let's say t is way, far away, like a huge number, a very large number, what can you say about the distribution of this at time t ?

AUDIENCE: Is it close to 0?

PROFESSOR: Close to 0. But by close to 0, what do you mean? There should be a scale. I mean some would say that 1 is close to 0. Some people would say that 100 is close to 0, so do you have some degree of how close it will be to 0? Anybody?

AUDIENCE: So variance will be small.

PROFESSOR: Sorry?

AUDIENCE: The variance will be small.

PROFESSOR: Variance will be small. About how much will the variance be?

AUDIENCE: 1 over n .

PROFESSOR: 1 over n . 1 over n ?

AUDIENCE: Over t .

PROFESSOR: 1 over t ? Anybody else want to have a different?

AUDIENCE: [INAUDIBLE].

PROFESSOR: 1 over square root t probably would.

AUDIENCE: [INAUDIBLE].

AUDIENCE: The variance would be [INAUDIBLE].

PROFESSOR: Oh, you're right, sorry. Variance will be 1 over t . And the standard deviation will be 1 over square root of t . What I'm saying is, by central limit theorem.

AUDIENCE: [INAUDIBLE]. Are you looking at the sums or are you looking at the?

PROFESSOR: I'm looking at the X_t . Ah. That's a very good point-- t and square root of t . Thank you.

AUDIENCE: That's very different.

PROFESSOR: Yeah, very, very different. I was confused. Sorry about that. The reason is because X_t , 1 over the square root of t times X_t -- we saw last time that this, if t is really, really large, this is close to the normal distribution, $0, 1$. So if you just look at it, X_t over the square root of t will look like normal distribution.

That means the value, at t , will be distributed like a normal distribution, with mean 0 and variance square root of t . So what you said was right. It's close to 0 . And the scale you're looking at is about the square root of t . So it won't go too far away from 0 .

That means, if you draw these two curves, square root of t and minus square root of t , your simple random walk, on a very large scale, won't like go too far away from these two curves. Even though the extreme values it can take-- I didn't draw it correctly-- is t and minus t , because all values can be 1 or all values can be minus 1 .

Even though, theoretically, you can be that far away from your x-axis, in reality, what's going to happen is you're going to be really close to this curve. You're going to play within this area, mostly.

AUDIENCE: I think that [INAUDIBLE].

PROFESSOR: So, yeah, that was a very vague statement. You won't deviate too much. So at the 100 square root of t , you will be inside this interval like 90% of the time. If you take this to be $10,000$ times square root of t , almost 99.9% or something like that.

And there's even a theorem saying you will hit these two lines infinitely often. So if you go over time, a very long period, for a very, very long, you live long enough, then, even if you go down here. Even, in this picture, you might think, OK, in some cases, it might be the case that you always play in the negative region.

But there's a theorem saying that that's not the case. With probability 1, if you go to infinity, you will cross this line infinitely often. And in fact, you will meet these two lines infinitely often.

So those are some interesting things about simple random walk. Really, there are lot more interesting things, but I'm just giving an overview, in this course, now. Unfortunately, I can't talk about all of these fun stuffs. But let me still try to show you some properties and one nice computation on it.

So some properties of a random walk, first, expectation of X_k is equal to 0. That's really easy to prove. Second important property is called independent increment. So if look at these times, t_0, t_1, \dots, t_k , then random variables $X_{t_i+1} - X_{t_i}$ are mutually independent.

So what this says is, if you look at what happens from time 1 to 10, that is irrelevant to what happens from 20 to 30. And that can easily be shown by the definition. I won't do that, but we'll try to do it as an exercise.

Third one is called stationary, so it has the property. That means, for all h greater or equal to 0, and t greater than or equal to $0-h$ is actually equal to $1-h$ the distribution

of X_{t+h} minus X_t is the same as the distribution of X_{h} .

And again, this easily follows from the definition. What it says is, if you look at the same amount of time, then what happens inside this interval is irrelevant of your starting point. The distribution is the same.

And moreover, from the first part, if these intervals do not overlap, they're independent. So those are the two properties that we're talking here.

And you'll see these properties appearing again and again. Because stochastic processes having these properties are really good, in some sense. They are fundamental stochastic processes. And simple random walk is like the fundamental stochastic process.

So let's try to see one interesting problem about simple random walk. So example, you play a game. It's like a coin toss game. I play with, let's say, Peter. So I bet \$1 at each turn. And then Peter tosses a coin, a fair coin. It's either heads or tails.

If it's heads, he wins. He wins the \$1. If it's tails, I win. I win \$1. So from my point of view, in this coin toss game, at each turn my balance goes up by \$1 or down by \$1.

And now, let's say I started from \$0.00 balance, even though that's not possible. Then my balance will exactly follow the simple random walk, assuming that the coin it's a fair coin, 50-50 chance. Then my balance is a simple random walk.

And then I say the following. You know what? I'm going to play. I want to make money. So I'm going to play until I win \$100 or I lose \$100. So let's say I play until I win \$100 or I lose \$100.

What is the probability that I will stop after winning \$100?

AUDIENCE: 1/2.

PROFESSOR: 1/2 because?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yes. So happens with $1/2$, $1/2$. And this is by symmetry. Because every chain of coin toss, which gives a winning sequence, when you flip it, it will give a losing sequence. We have one to one correspondence between those two things.

That was good. Now if I change it. What if I say I will win \$100 or I lose \$50? What if I play until win \$100 or lose \$50? In other words, I look at the random walk, I look at the first time that it hits either this line or it hits this line, and then I stop.

What is the probability that I will stop after winning \$100?

AUDIENCE: [INAUDIBLE].

PROFESSOR: $1/3$? Let me see. Why $1/3$?

AUDIENCE: [INAUDIBLE].

PROFESSOR: So you're saying, hitting this probability is p . And the probability that you hit this first is p , right? It's $1/2$, $1/2$. But you're saying from here, it's the same. So it should be $1/4$ here, $1/2$ times $1/2$. You've got a good intuition. It is $1/3$, actually.

AUDIENCE: [INAUDIBLE].

PROFESSOR: And then once you hit it, it's like the same afterwards? I'm not sure if there is a way to make an argument out of it. I really don't know. There might be or there might not be. I'm not sure. I was thinking of a different way. But yeah, there might be a way to make an argument out of it. I just don't see it right now.

So in general, if you put a line B and a line A, then probability of hitting B first is A over A plus B . And the probability of hitting this line, minus A, is B over A plus B . And so, in this case, if it's 100 and 50, it's 100 over 150, that's $2/3$ and that's $1/3$.

This can be proved. It's actually not that difficult to prove it. I mean it's hard to find the right way to look at it.

So fix your B and A. And for each k between minus A and B define f of k as the probability that you'll hit-- what is it-- this line first, and the probability that you hit the

line B first when you start at k .

So it kind of points out what you're saying. Now, instead of looking at one fixed starting point, we're going to change our starting point and look at all possible ways. So when you start at k , I'll define f of k as the probability that you hit this line first before hitting that line.

What we are interested in is computing f of 0 . What we know is f of B is equal to 1 , f of A is equal to 0 . And then actually, there's one recursive formula that matters to us. If you start at k , you either go up or go down.

You go up with probability $1/2$. You go down with probability $1/2$. And now it starts again. Because of this-- which one is it-- stationary property. So starting from here, the probability that you hit B first is exactly f of k plus $1/2$.

So if you go up, the probability that you hit B first is f of k plus $1/2$. If you go down, it's f of k minus $1/2$. And then that gives you a recursive formula with two boundary values. If you look at it, you can solve it. When you solve it, you'll get that answer.

So I won't go into details, but what I wanted to show is that simple random walk is really this property, these two properties. It has these properties and even more powerful properties. So it's really easy to control. And at the same time it's quite universal.

It can model-- like it's not a very weak model. It's rather restricted, but it's a really good model for like a mathematician. From the practical point of view, you'll have to twist some things slightly and so on. But in many cases, you can approximate it by simple random walk.

And as you can see, you can do computations, with simple random walk, by hand. So that was it. I talked about the most important example of stochastic process. Now, let's talk about more stochastic processes.

The second one is called the Markov chain. Let me right that part, actually. So Markov chain, unlike the simple random walk, is not a single stochastic process. A

stochastic process is called a Markov chain if it has some property.

And what we want to capture in Markov chain is the following statement. These are a collection of stochastic processes having the property that-- whose effect of the past on the future is summarized only by the current state.

That's quite a vague statement. But what we're trying to capture here is-- now, look at some generic stochastic process at time t . You know all the history up to time t . You want to say something about the future.

Then, if it's a Markov chain, what it's saying is, you don't even have to know all about this. Like this part is really irrelevant. What matters is the value at this last point, last time.

So if it's a Markov chain, you don't have to know all this history. All you have to know is this single value. And all of the effect of the past on the future is contained in this value. Nothing else matters.

Of course, this is a very special type of stochastic process. Most other stochastic processes, the future will depend on the whole history. And in that case, it's more difficult to analyze. But these ones are more manageable.

And still, lots of interesting things turn out to be Markov chains. So if you look at simple random walk, it is a Markov chain, right? So simple random walk, let's say you went like that. Then what happens after time t really just depends on how high this point is at.

What happened before doesn't matter at all. Because we're just having new coin tosses every time. That this value can affect the future, because that's where you're going to start your process from. Like that's where you're starting your process. So that is a Markov chain. This part is irrelevant. Only the value matters.

So let me define it a little bit more formally. A discrete time stochastic process is a Markov chain if the probability that X at some time, $t + 1$, is equal to something, some value, given the whole history up to time n is equal to the probability that X_t

plus 1 is equal to that value, given the value X_n for all n greater than or equal to t greater than or equal to 0 and all s .

This is a mathematical way of writing down this. The value at X_{t+1} , given all the values up to time t , is the same as the value at time $t+1$, the probability of it, given only the last value.

And the reason simple random walk is a Markov chain is because both of them are just 1/2. I mean, if it's for-- let me write it down. So example, random walk probability that X_{t+1} equal to s , given t is equal to 1/2, if s is equal $X_t + 1$ or $X_t - 1$, and 0 otherwise.

So it really depends only on the last value of X_t . Any questions? All right. If there is case when you're looking at a stochastic process, a Markov chain, and all X_i have values in some set s , which is finite, a finite set, in that case, it's really easy to describe Markov chains.

So now denote the probability ij as the probability that, if at that time t you are at i , the probability that you jump to j at time $t+1$ for all pair of points i, j . I mean, it's a finite set, so I might just as well call it the integer set from 1 to m , just to make the notation easier.

Then, first of all, if the sum over all j and s , P_{ij} , that is equal to 1. Because if you start at i , you'll have to jump to somewhere in your next step. So if you sum over all possible states you can have, you have to sum up to 1.

And really, a very interesting thing is this matrix called the transition probability matrix, defined as. So we put P_{ij} at [INAUDIBLE] and [INAUDIBLE].

And really, this tells you everything about the Markov chain. Everything about the stochastic process is contained in this matrix. That's because a future state only depends on the current state.

So if you know what happens at time t , where it's at time t , look at the matrix, you can decode all the information you want. What is the probability that it will be at--

let's say, it's at 0 right now. What's the probability that it will jump to 1 at the next time? Just look at 0 comma 1, here. There is no 0, 1, here, so it's 1 and 2. Just look at 1 and 2, 1 and 2, i and j.

Actually, I made a mistake. That should be the right one. Not only that, that's a one-step. So what happened is it describes what happens in a single step, the probability that you jump from i to j. But using that, you can also model what's the probability that you jump from i to j in two steps.

So let's define q_{ij} as the probability that X at time t plus 2 is equal to j, given that X at time t is equal to i. Then the matrix, defined this way, can you describe it in terms of the matrix A? Anybody?

Multiplication? Very good. So it's A square. Why is it? So let me write this down in a different way. q_{ij} is-- you sum over all intermediate values-- the probability that you jump from i to k, first, and then the probability that you jump from k to j.

And if you look at what this means, each entry here is described by a linear-- what is it-- the dot product of a column and a row. And that's exactly what occurs. And if you want to look at the three-step, four-step, all you have to do is just multiply it again and again and again.

Really, this matrix contains all the information you want if you have a Markov chain and its finite. That's very important. For random walk, simple random walk, I told you that it is a Markov chain. But it does not have a transition probability matrix, because the state space is not finite. So be careful.

However, finite Markov chains, really, there's one matrix that describes everything. I mean, I said it like it's something very interesting. But if you think about it, you just wrote down all the probabilities. So it should describe everything.

So an example. You have a machine, and it's broken or working at a given day. That's a silly example. So if it's working today, working tomorrow, broken with probability 0.01, working with probability 0.90.

If it's broken, the probability that it's repaired on the next day is 0.8. And it's broken at 0.2. Suppose you have something like this. This is an example of a Markov chain used in like engineering applications.

In this case, s is also called a sample state space, actually. And the reason is because, in many cases, what you're modeling is these kind of states of some system, like broken or working, rainy, sunny, cloudy as weather. And all these things that you model represent states a lot of time. So you call it state set as well. So that's an example.

And let's see what happens for this matrix. We have two states, working and broken. Working to working is 0.99. Working to broken is 0.01. Broken to working is 0.8. Broken to broken is 0.2.

So that's what we've learned so far. And the question, what happens if you start from some state, let's say it was working today, and you go a very, very long time, like a year or 10 years, then the distribution, after 10 years, on that day, is A to the 3,650. So that will be that times $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ will be the probability p, q . p will be the probability that it's working at that time. q will be the probability that it's broken at that time.

What will p and q be? What will p and q be? That's the question that we're trying to ask. We didn't learn, so far, how to do this, but let's think about it.

I'm going to cheat a little bit and just say, you know what, I think, over a long period of time, the probability distribution on day 3,650 and that on day 3,651 shouldn't be that different. They should be about the same. Let's make that assumption. I don't know if it's true or not. Well, I know it's true, but that's what I'm telling you.

Under that assumption, now you can solve what p and q are. So approximately, I hope, p, q -- so $A^{3,650} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is approximately the same as $A^{3,651} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. That means that this is p, q . p, q is about the same as A times p, q .

Anybody remember what this is? Yes. So p, q will be the eigenvector of this matrix. Over a long period of time, the probability distribution that you will observe will be

the eigenvector. And what's the eigenvalue? 1, at least in this case, it looks like it's 1.

Now I'll make one more connection. Do you remember Perron-Frobenius theorem? So this is a matrix. All entries are positive. So there is a largest eigenvalue, which is positive and real. And there is one all positive eigenvector corresponding to it. What I'm trying to say is that's going to be your p, q .

But let me not jump to the conclusion yet. And one more thing we know is, by Perron-Frobenius, there exists an eigenvalue, the largest one, λ greater than 0, and eigenvector v_1, v_2 , where v_1, v_2 are positive. Moreover, λ has a multiplicity of 1. I'll get back to it later.

So let's write this down. $A \times v_1, v_2$ is equal to λ times v_1, v_2 . $A \times v_1, v_2$, we can write it down. It's $0.99 v_1$ plus $0.01 v_2$. And that $0.8 v_1$ plus $0.2 v_2$, which is equal to v_1, v_2 . You can solve v_1 and v_2 , but before doing that-- sorry about that. This is flipped.

Yeah, so everybody, it should have been flipped in the beginning. So that's 8. So sum these two values, and you get λ of this, v_1, v_2 .

On the left, what you get is v_1 plus v_2 , so sum two coordinates. On the left, you get v_1 plus v_2 . On the right, you get λ times v_1 plus v_2 . That means your λ is equal to 1.

So that eigenvalue, guaranteed by Perron-Frobenius theorem, is 1, eigenvalue of 1. So what you'll find here will be the eigenvector corresponding to the largest eigenvalue-- eigenvector will be the one corresponding to the largest eigenvalue, which is equal to 1.

And that's something very general. It's not just about this matrix and this special example. In general, if you have a transition matrix, if you're given a Markov chain and given a transition matrix, Perron-Frobenius theorem guarantees that there exists a vector as long as all the entries are positive.

So in general, if transition matrix of a Markov chain has positive entries, then there

exists a vector π such that $A\pi = \pi$. I'll just call it v . Av is equal to v . And that will be the long term behavior as explained.

Over a long, if it converges to some state, it has to satisfy that. And by Perron-Frobenius theorem, we know that there is a vector satisfying it. So if it converges, it will converge to that. And what it's saying is, if all the entries are positive, then it converges. And there is such a state. We know the long-term behavior of the system.

So this is called the stationary distribution. Such vector v is called. It's not really right to say that a vector has stationary distribution. But if I give this distribution to the state space, what I mean is consider probability distribution over s such that probability is-- so it's a random variable X -- X is equal to i is equal to π_i .

If you start from this distribution, in the next step, you'll have the exact same distribution. That's what I'm trying to say here. That's called a stationary distribution. Any questions?

AUDIENCE: So [INAUDIBLE]?

PROFESSOR: Yes. Very good question. Yeah, but Perron-Frobenius theorem say there is exactly one eigenvector corresponding to the largest eigenvalue. And that turns out to be 1. The largest eigenvalue turns out to be 1. So there will a unique stationary distribution if all the entries are positive.

AUDIENCE: [INAUDIBLE]?

PROFESSOR: This one?

AUDIENCE: [INAUDIBLE]?

PROFESSOR: Maybe. It's a good point. Huh? Something is wrong. Can anybody help me? This part looks questionable.

AUDIENCE: Just kind of [INAUDIBLE] question, is that topic covered in portions of [INAUDIBLE]? The other eigenvalues in the matrix are smaller than 1. And so when you take

products of the transition probability matrix, those eigenvalues that are smaller than 1 scale after repeated multiplication to 0. So in the limit, they're 0, but until you get to the limit, you still have them. Essentially, that kind of behavior is transitional behavior that dissipates. But the behavior corresponding to the stationary distribution persists.

PROFESSOR: But, as you mentioned, this argument seems to be giving that all λ has to be 1, right? Is that your point? You're right. I don't see what the problem is right now. I'll think about it later. I don't want to waste my time on trying to find what's wrong.

But the conclusion is right. There will be a unique one and so on. Now let me make a note here.

So let me move on to the final topic. It's called martingale. And this is another collection of stochastic processes. And what we're trying to model here is a fair game, stochastic processes which are a fair game.

And formally, what I mean is a stochastic process is a martingale if that happens.

Let me iterate it. So what we have here is, at time t , if you look at what's going to happen at time $t + 1$, take the expectation, then it has to be exactly equal to the value of X_t . So we have this stochastic process, and, at time t , you are at X_t .

At time $t + 1$, lots of things can happen. It might go to this point, that point, that point, or so on. But the probability distribution is designed so that the expected value over all these are exactly equal to the value at X_t . So it's kind of centered at X_t , centered meaning in the probabilistic sense. The expectation is equal to that.

So if your value at time t was something else, your values at time $t + 1$ will be centered at this value instead of that value. And the reason I'm saying it models a fair game is because, if this is like your balance over some game, in expectation, you're not supposed to win any money at all

And I will later tell you more about that. So example, a random walk is a martingale. What else? Second one, now let's say you're in a casino and you're playing roulette.

Balance of a roulette player is not a martingale.

Because it's designed so that the expected value is less than 0. You're supposed to lose money. Of course, at one instance, you might win money. But in expected value, you're designed to go down. So it's not a martingale. It's not a fair game. The game is designed for the casino not for you.

Third one is some funny example. I just made it up to show that there are many possible ways that a stochastic process can be a martingale. So if Y_i are IID random variables such that Y_i is equal to 2, with probability $1/3$, and $1/2$ is probability $2/3$, then let X_0 equal 1 and X_k equal. Then that is a martingale.

So at each step, you'll either multiply by 2 or $1/2$ by 2-- just divide by 2. And the probability distribution is given as $1/3$ and $2/3$. Then X_k is a martingale. The reason is-- so you can compute the expected value. The expected value of the X_k plus 1, given X_k up to [INAUDIBLE], is equal to-- what you have is expected value of Y_k plus 1 times Y_k up to Y_1 . That part is X_k .

But this is designed so that the expected value is equal to 1. So it's a martingale. I mean it will fluctuate a lot, your balance, double, double, double, half, half, half, and so on. But still, in expectation, you will always maintain.

I mean the expectation that all time is equal to 1, if you look at it from the beginning. You look at time 1, then the expected value of x_1 and so on. Any questions on definition or example?

So the random walk is an example which is both Markov chain and martingale. But these two concepts are really two different concepts. Try not to be confused between the two. They're just two different things.

There are Markov chains which are not martingales. There are martingales which are not Markov chains. And there are somethings which are both, like a simple random walk. There are some stuff which are not either of them. They really are just two separate things.

Let me conclude with one interesting theorem about martingales. And it really enforces your intuition, at least intuition of the definition, that martingale is a fair game.

It's called optional stopping theorem. And I will write it down more formally later, but the message is this. If you play a martingale game, if it's a game you play and it's your balance, no matter what strategy you use, your expected value cannot be positive or negative.

Even if you try to lose money so hard, you won't be able to do that. Even if you try to win money so hard, like try to invent something really, really cool and ingenious, you should not be able to win money. Your expected value is just fixed.

That's the concept of the theorem. Of course, there are technical conditions that have to be there. So if you're playing a martingale game, then you're not supposed to win or lose, at least in expectation.

So before stating the theorem, I have to define what a stopping point means. So given a stochastic process, a non-negative integer, a valued random variable, τ , is called a stopping time, if, for all integer k greater than or equal to 0, τ , lesser or equal to k , depends only on X_1 to X_k . So that is something very, very strange.

I want to define something called a stopping time. It will be a non-negative integer valued random variable. So it will be 0, 1, 2, or so on. That means it will be some time index. And if you look at the event that τ is less than or equal to k -- so if you want to look at the events when you stop at time less than or equal to k , your decision only depends on the events up to k , on the value of the stochastic process up to time k .

In other words, if this is some strategy you want to use-- by strategy I mean some strategy that you stop playing at some point. You have a strategy that is defined as you play some k rounds, and then you look at the outcome. You say, OK, now I think it's in favor of me. I'm going to stop.

You have a pre-defined set of strategies. And if that strategy only depends on the

values of the stochastic process up to right now, then it's a stopping time. If it's some strategy that depends on future values, it's not a stopping time.

Let me show you by example. Remember that coin toss game which had random walk value, so either win \$1 or lose \$1. So in coin toss game, let τ be the first time at which balance becomes \$100, then τ is a stopping time.

Or you stop at either \$100 or negative \$50, that's still a stopping time. Remember that we discussed about it? We look at our balance. We stop at either at the time when we win \$100 or lose \$50. That is a stopping time.

But I think it's better to tell you what is not a stopping time, an example. That will help, really. So let τ -- in the same game-- the time of first peak. By peak, I mean the time when you go down, so that would be your τ .

So the first time when you start to go down, you're going to stop. That's not a stopping time. Not a stopping time. To see formally why it's the case, first of all, if you want to decide if it's a peak or not at time t , you have to refer to the value at time t plus 1. For just looking at values up to time t , you don't know if it's going to be a peak or if it's going to continue.

So the event that you stop at time t depends on t plus 1 as well, which doesn't fall into this definition. So that's what we're trying to distinguish by defining a stopping time.

In these cases it was clear, at the time, you know if you have to stop or not. But if you define your stopping time in this way and not a stopping time, if you define τ in this way, your decision depends on future values of the outcome. So it's not a stopping time under this definition. Any questions? Does it make sense? Yes?

AUDIENCE: Could you still have τ as the stopping time, if you were referring to t , and then t minus 1 was greater than [INAUDIBLE]?

PROFESSOR: So.

AUDIENCE: Let's say, yeah, it was [INAUDIBLE].

PROFESSOR: So that time after peak, the first time after peak?

AUDIENCE: Yes.

PROFESSOR: Yes, that will be a stopping time. So three, τ is $\tau_0 + 1$, where τ_0 is the first peak, then it is a stopping time. It's a stopping time.

So the optional stopping theorem that I promised says the following. Suppose we have a martingale, and τ is a stopping time. And further suppose that there exists a constant t such that τ is less than or equal to t always.

So you have some strategy which is a finite strategy. You can't go on forever. You have some bound on the time. And your stopping time always ends before that time. In that case, then expectation of your value at the stopping time, when you've stopped, your balance, if that's what it's modeling, is always equal to the balance at the beginning.

So no matter what strategy you use, if you're a mortal being, then you cannot win. That's the content of this theorem. So I wanted to prove it, but I'll not, because I think I'm running out of time.

But let me show you one, very interesting corollary of this applied to that number one. So number one is a stopping time. It's not clear that there is a bounded time where you always stop before that time. But this theorem does apply to that case.

So I'll just forget about that technical issue. So corollary, it applies not immediately, but it does apply to the first case, case 1 given above. And then what it says is expectation of X_τ is equal to 0.

But expectation of X_τ is-- X at τ is either 100 or negative 50, because they're always going to stop at the first time where you either hit \$100 or minus \$50. So this is 100 times some probability plus $1 - p$ times minus 50.

There's some probability that you stop at 100. With all the rest, you're going to stop

at minus 50. You know it's set. It's equal to 0. What it gives is-- I hope it gives me the right thing I'm thinking about. p , 100, yes. It's $150p$ minus 50 equals 0. p is $1/3$. And if you remember, that was exactly the computation we got.

So that's just a neat application. But the content of this, it's really interesting. So try to contemplate about it, something very philosophically. If something can be modeled using martingales, perfectly, if it really fits into the mathematical formulation of a martingale, then you're not supposed to win.

So that's it for today. And next week, Peter will give wonderful lectures. See you next week.