

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: Now today, we're going to talk about random walks. And in particular, we're going to look at a classic phenomenon known as Gamblers Ruin. It's a great way to end the term, because the solution requires several of the techniques that we've developed since the midterm.

So it's actually a good review. We'll review recurrences. We'll review a lot of probability laws. And it's actually a nice problem to look at. It's another example where you get a non-intuitive solution using probability. And if you like to gamble, it's really good that you look at this problem before you go to Vegas or down to Foxwoods.

Now the Gambler's Ruin problem, you start with n dollars. And we're going to do a simplified version, where in each bet, you win \$1 or you lose \$1. Now, these days, there are not many bets in a casino for \$1. It's more like \$10.

But just to make it simple for counting, we're going to assume that each bet you win \$1 with probability p , and you lose \$1 with probability $1 - p$. And in this version, we're going to assume you keep playing until one of two things happens-- you get ahead by m dollars, or you lose all the money you came with-- all n dollars.

So you play until you win m more-- net m plus-- or you lose n . And that's where you go broke. You run out of money. And we're going to assume you don't borrow anything from the house.

All right, and we're going to look at the probability that you come out a winner versus going home broke-- that you made m dollars. Now, the game we're going to analyze is roulette, but the technique works for any of them. How many people have played roulette before in some form or another?

OK, so this is a game where there's the ball that goes around the dish, and you spin the wheel. And there's 36 numbers from 1 to 36. Half of them are red, half are black. And then there's the zero and the double zero that are green.

And we're going to look at the version where you just bet on red or black. And you win if the

ball lands on a slot that's red. And there's 18 of those. And you lose otherwise.

So in this case, the probability of winning, p , is there's 18 chances to win. And it's not 36 total. It's 38 total because of the zero and the double zero. All right so this is $9/19$ chance of winning and a $10/19$ chance of losing.

And so this is a game that has a chance of winning of about 47%, so it's almost a fair game. It's not 50-50. And that's because the casino's got to make some money.

I mean, they have the big facility. They're giving you free drinks, and all the rest. So they got to make money somehow.

And they make money on this bet because they're going to make \$0.03 on the dollar here. You're going to wager. And then you're going to come back with 47%. And people generally are fine with that. They don't expect to have the odds in their favor when you're gambling in a casino.

Now, in an effort to sort of come home a winner, the way people do that-- knowing that the odds are a little against them-- is they might put more money in their pocket coming in than they expect to win. So often, you'll see people come into the casino with the goal of winning 100, but they start with 1,000 in their pocket. So they're willing to risk \$1,000, but they're going to quit happy if they get up 100.

OK so you either go home with \$1,100, or you're going home with \$0, in this case. And you came with \$1,000. And this means that you're-- at least the thinking goes-- this means you're more likely to go home happy. If you quit when you get up by 100, you're more likely to land there, because it's almost a fair game, than you are to lose all 1,000. That's the thinking anyway.

In fact, my mother-in-law plays roulette, red and black, and she follows the strategy. And she claims that she does this for that reason-- that she almost always wins. She goes home happy almost always.

And that's the important thing here. And it does reasonable, because after all, roulette is almost a fair game. So what do you think?

How many people think she's right that she almost always wins? Anybody? I have sort of set it up. It's my mother-in-law, after all, so probably she's going to be wrong.

Well, how many people think it's better than a 50% chance you win \$100 before you lose \$1,000? That's probably more-- how many people think you're more likely to lose \$1,000 before you win \$100? Wow, OK, so you've been to 6.04 too long now.

OK, what about this-- how many people think you're more likely to lose \$10,000 than to win \$100? All right, how many people think you're more likely to lose \$1 million? A bunch of you still think that.

OK, well, you're right. In fact, it is almost certain you will go broke, no matter how much money you bring, before you win \$100. In fact, we're going to prove today that the probability that you win \$100 before losing \$100 million if you stayed long enough-- that takes a while-- the chance you go home a winner is less than 1 in 37,648. You have no chance to go home happy.

So my mother-in-law's telling me the story about how she always goes home happy. And I'm saying, no, no, wait a minute, you can't. You never went home happy. Let's be honest. It can't be.

She goes, no, no, no, it's true. I go, no, look, there's a mathematical proof. I have a proof. I can show you my proof-- very unlikely you go home a winner.

So somehow, she's not very impressed with the mathematical proof. And she keeps insisting. And I keep trying to show her the proof. And anyway, I hope I'll have more luck with you guys today in showing you the proof that the chance you go home happy here is very, very small. Now, in the end, I didn't convince her, but we'll see how we do here today.

Now, in order to see why this probability is so stunningly small-- you would just never guess it's that low-- we've got to learn about random walks. And they come up in all sorts of applications. In fact, page rank-- that got Google started-- it's all based on a random walk through the Web or through the links on web pages on that graph.

Now, for the gambling problem, we're going to look at a very special case-- probably the simplest case of a random walk-- and that's a one-dimensional random walk. In a one-dimensional random walk, there's some value-- say the number of dollars you've got in your pocket. And this value can go up, or go down, or stay the same each time you do something like make a bet. And each of this happens with a certain probability.

Now in this case, you either go up by one, or you go down by one, and you can't stay the

same. Every bet you win \$1 or you lose \$1. So it's really a special case.

And we can diagram it as follows. We can put time, or the number of bets, on this axis. And we can put the number of dollars on this axis.

Now in this case, we start with n dollars. And we might win the first bet, so we go to $n + 1$. We might lose a bet, might lose again, could win the next one, lose, win, lose, lose. So this corresponds to a string-- win, lose, lose, lose, lose, win, lose, win, lose, lose, lose.

And when we win, we go up \$1. When we lose, we go down \$1. And it's called one-dimensional, because there's just one thing that's changing. You're going up and down there.

Now, the probability of going up is p . And that's no matter what happened before. It's a memoryless independent system.

The probability you win your i -th bet has nothing to do-- is totally independent, mutually independent-- of all the other bets that took place before. So let's write that down. So the probability of an up move is p . The probability of a down move is $1 - p$. And these are mutually independent of past moves.

Now, when you have a random walk where the moves are mutually independent, it has a special name. It's called a martingale. All random walks don't have to have mutually independent steps.

Say you're looking about winning and losing a baseball game in a series. We looked at a scenario where, if you lost yesterday, you're feeling lousy, more likely to lose today. Not true in the gambling case here. It's mutually independent. And that's the only case we're going to study for random walks.

Now, if p is not $1/2$, the random walk is said to be biased. And that's what happens in the casino. It's biased in favor of the house. If p equals $1/2$, then the random walk is unbiased.

Now, in this particular case that we're looking at, we have boundaries on the random walk. There's a boundary at 0, because you go home broke if you lost everything. If the random walk ever hit \$0, you're done.

And we're also going to put a boundary at $n + m$. So I'm going to have a boundary here. So that if I win m dollars here, I stop and I go home happy. If the random walk ever goes here,

then I stop.

Those are called boundary conditions for the walk. And what we want to do is analyze the probability that we hit that top boundary before we hit the bottom boundary. So we're going to define that event to be W^* .

W^* is the event that the random walk hits T , which is $n + m$, before it hits 0 . In other words, you go home happy without going broke. Let's also define D to be the number of dollars at the start. And this is just going to be n in our case.

We're interested in, call it X_n is the probability that we go home happy given we started with n dollars. And that's a function of n . So we'll make a variable called X_n .

And we want to know what that probability is. And of course, the more you come with, you'd think it's a higher chance of winning the more you have in your pocket, because you can play for more. So the goal is to figure this out.

Now to do this, we could use the tree method. But it gets pretty complicated, because the sample space is the sample space of all one-loss sequences. And how big is that sample space?

AUDIENCE: Infinite.

PROFESSOR: Infinite. I could play forever. All right, now it turns out the probability of playing forever is 0 . And we won't prove that, but there are an infinite number of sample points. So doing the tree method is a little complicated when it's infinite.

So what we're going to do is use some of the theorems we've proved over the last few weeks and set up a recurrence to find this probability. Now, I'm going to tell you what the recurrence is, and then prove that that's right. So I claim that X_n is 0 probability if we start with $\$0$. It's 1 if we start with T dollars. And it's p times X_{n-1} plus $(1-p)$ times X_{n+1} if we start with between $\$0$ and T dollars.

All right, so that's what I claim X_n is. And it's, of course, a recursion that I've set up here. So let's see why that's the case.

OK, so let's check the 0 case. X_0 is the probability we go home a winner given we started with $\$0$. Why is that 0 ?

AUDIENCE: [INAUDIBLE].

PROFESSOR: What's that?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah, you started broke. You never get off the ground, because you quit as soon as you have \$0. So you have no chance to win, because you're broke to start.

Let's check the next case, X_T -- case n equals T -- is the probability you go home a winner given you started with T dollars. Why is that 1? Why is that certain, sort of from the definition?

AUDIENCE: [INAUDIBLE].

PROFESSOR: You already have your money. You already hit the top boundary, because you started there. Remember, you quit and you're happy. Go home happy if you hit T dollars. All right, so you're guaranteed to go home happy, because you never make any bets. You started with all the money you needed to go home happy.

Then we have the interesting case, where you start with between 0 and T dollars. And now you're going to make some bets. And then X_n is the probability -- just the definition -- of going home happy -- i.e. winning and having T dollars, if you start with n .

Now, there's two cases to analyze this, based on what happens in the first bet. You could win it, or you could lose it. And then we're going to recurse.

So we're going to define E to be the event that you win the first bet. And \bar{E} is the event that you lose the first bet. Now, by the theory of total probability, which we did in recitation maybe a couple weeks ago, we can rewrite this depending on whether E happened or the complement of E happened.

And you get that the probability is simply the probability of going home happy and winning the first bet times -- and I've got to put the conditioning in. That doesn't go away. So I'm breaking into two cases. The first one is you win the first bet given D equals n , and the case where you lose the first bet, given D equals n . Any questions here?

The probability of going home happy given you start with n dollars is the probability of going home happy and winning the first bet given D equals n plus the probability of going home happy and losing the first bet given D equals n -- just those are the two cases. Now I can use

the definition of conditional probability to rewrite these. This is the probability-- you've got two events-- that the first one happens given D equals n times the probability the second one happens given that the first one happened and D equals n .

This is just the definition of conditional probability, when I've got an intersection of events here. The probability of both happening is the probability of the first happening times the probability of the second happening given that the first happened. And of course, everything is in this universe of D equals n . So I've used it in a little different twist than we had it before.

The same thing over here-- this now is the probability of E prime given D equals n times the probability of W star-- winning, going home happy-- given that you lost the first bet and D equals n . That's D equals n there. So it looks like it's got more complicated, but now we can start simplifying.

What's the probability of winning the first bet given that you started with n dollars?

AUDIENCE: p .

PROFESSOR: p -- in fact, does this have anything to do with the probability of winning the first bet? No, this is just p . Now, what about this thing?

I am conditioning on winning the first bet given and I start with n dollars. What's another way of expressing I won the first bet and I started with n dollars? Yeah?

AUDIENCE: You have n plus \$1.

PROFESSOR: I now have n plus \$1 going forward. And because I have a martingale, and everything is mutually independent, it's like the world starts all over again. I'm now in a state with n plus \$1, and I want to know the probability that I go home happy.

It doesn't matter how I got the n plus \$1. It's just going forward-- I got n plus \$1 in my pocket, I want to know the probability of going home happy. So I reset to D equals n plus 1. So I replace this with that, because however long it took me to get there and all that stuff doesn't matter for this analysis. It's all mutually dependent.

Probability of losing the first bet given that I started with n dollars-- 1 minus p . Doesn't matter how much I started with. And here, I want to know the probability of going home happy given-- well, if I lost the first bet and I started with n , what have I got? n minus 1.

It doesn't matter how I got to $n - 1$. Now this is going to get really simple. What's another name for that expression? X_{n+1} .

And another name for this expression? X_{n-1} . So we proved that X_n equals $p X_{n+1} + 1 - p X_{n-1}$. And that's what I claimed is true. So we finished the proof. Any questions?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Did I screw it up?

AUDIENCE: [INAUDIBLE].

PROFESSOR: I claim probability of winning-- so let's see if I have a wrong in here. I might have screwed it up. I think I proved it's $n + 1$, right?

Yep, sure enough, I think this is a $+1$. That's a -1 . Now, it's always good to check to you proved what you said you were going to prove.

So I needed to change this. That's what I proved. Any other questions? That was a pretty important question.

All right, so we have a recurrence for X_n . Now, it's a little funny looking at first, because normally with a recurrence, X_n would depend on X_i that are smaller-- the i 's are smaller than n . So it looks a little wacky.

But is that a problem? I can just solve for X_{n+1} -- just subtract this and put it over there. So let's do that.

OK, so if I solve for X_{n+1} up there, I'll put $p X_{n+1}$ on its own side, I get $p X_{n+1} - X_{n+1} + p X_{n-1} - X_{n-1} = 0$. And I know that X_0 is 0. And I know that X_1 equals 1.

Now, what type of recurrence is this?

AUDIENCE: Linear.

PROFESSOR: Linear, good, so it's a linear recurrence. And what type of linear recurrence is it?

AUDIENCE: Homogeneous.

PROFESSOR: Homogeneous-- that's the best case, simple case, that's good. The boundary conditions are a little weird, because the recurrences we all saw before, if we had two boundary conditions it would be X_0 and X_1 . Here it's X_0 and X_T .

But all's you need are two. Doesn't matter where they are. So how do I solve that thing?
What's the next thing I do? What is it?

AUDIENCE: Characterize the equation.

PROFESSOR: Characterize the equation. And what do you do with that equation?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Solve it, get the roots. This'll be good practice for the final, because you'll probably have to do something like this. So that's the characteristic equation.

And what's the order of this equation-- the degree? That's going to be 2, right? I'm going to have $pr^2 - r + 1 - p = 0$. That's my characteristic equation.

Remember that? So I make this be the constant term. Then I have the first-order term, then the second-order term.

All right, now I solve it. And that's easy for a second-order equation. $1 \pm \sqrt{1 - 4p}$ over $2p$. Let's do that.

OK, so this is $1 \pm \sqrt{1 - 4p}$ over $2p$. Just using the quadratic formula and simplifying. And it works out really nicely, because that is the square root of-- this is just $1 - 2p$ squared.

So that's $1 \pm \sqrt{1 - 2p}$ over $2p$. And that is $2 - 2p$ over $2p$ or $1 - p$ over p . The 1 cancels, then $-2p$ is -2 over $2p$.

So the answers, the roots are divide by 2 on this one. I get $1 - p$ over p and 1 . Those are the roots. Are these roots different? Do I have the case of a double root? Are the roots always different?

They're usually different. What's the case where these roots are the same?

AUDIENCE: 0.5.

PROFESSOR: 0.5, which is sort of an interesting case in this game. Because if p equals $1/2$, we have an unbiased random walk. You got a fair game.

And so it says right away, well, maybe the result is going to be different for a fair game than the game we're playing in the casino, where it's biased. So let's look at the casino game where p is not $1/2$. Then the roots are different. Later, we'll go back and analyze the case when the roots of the same for the fair game.

So if p is not $1/2$, then we can solve for X_n . X_n is some constant times the first root to the n th power plus a constant times the second root to the n th power. Remember, that's how it works for any linear homogeneous recurrence.

And that's easy, because the second root was 1. This is just plus $B \cdot 1$ to the n is 1.

How do I figure out what A and B are?

AUDIENCE: Boundary conditions.

PROFESSOR: Boundary conditions, very good. So let's look at the boundary conditions.

OK, so the first boundary condition is at 0. So we have 0 equals X_0 . Plugging in there-- oops I forgot the n up here. Plugging in n equals 0-- well, this to the 0 is just 1. That is $A + B$. That means that B equals minus A .

Then the second boundary condition is 1 equals X_T . And that is $A \cdot 1 - p$ over p to the T plus B , but B was minus A . And now I can solve for A .

So that means that A equals 1 over $1 - p$ over p to the T minus 1. And B is negative A -- minus 1 over $1 - p$, over p to the T minus 1. And then I plug those back in to the formula for X_n .

So here's my constant A . I multiply that times $1 - p$ over p to the n , plus I add this in. So this means that the probability of going home a winner is $1 - p$ over p to the n over that thing-- $1 - p$ over p to the T minus 1, plus the B term, which really is a minus term here, is just minus 1. Put that on top here.

That sort of looks messy, but there's a simplification to get an upper bound that's very close. In particular, if you have a biased game against you-- so if p is less than $1/2$, as it is in roulette, then this is a number bigger than 1. That means that $1 - p$ over p is bigger than 1.

So this is bigger than 1. This is bigger than 1. T is the upper limit. It's n plus m .

So I've got a bigger number down here than I do here. So overall, it's a fraction less than 1. And when you have a fraction less than 1, if you add 1 to the numerator and denominator, it gets closer to 1. It gets bigger.

So this is upper-bounded by just adding 1 to each of these. Its upper-bounded by this over that, which is $1 - p$ over p to the n minus T . And T is just n plus m . So this equals-- why don't I turn it upside down? Make it p over $1 - p$ to get a fraction that's less than 1.

T minus n , and that equals p over $1 - p$ to the m . And this is how much you're trying to get ahead-- \$100 in the case of my mother-in-law. So what we've proved-- let me state what we proved as a theorem.

So we proved that if p is less than $1/2$ -- if you're more likely to lose a bet than win it-- then the probability that you win m dollars before you lose n dollars is at most p over $1 - p$ to the m . That's what we just proved. And so now you can plug in values-- for example, for roulette.

p equals $9/19$, which means that p over $1 - p$ -- that's going to be $9/19$ over $10/19$, which is just $9/10$. And if m -- the amount you want to win-- is \$100, and n is \$1,000-- that's what you start with and you're willing to lose-- well, the probability you win-- you go home happy-- W star you win \$100-- is less than or equal to $9/10$ raised to the m , which is 100. So it's $9/10$ of 100, and that turns out to be less than 1 in 37,648, which is where that answer came from.

Now you can see why my mother-in-law may have got lost somewhere here now in the calculations. But this is a proof that the chance you win \$100 before you lose \$1,000 is very, very small. Now, do you see why the answer is no better than if you came with \$1 million in your pocket?

Say you came with n equals \$1 million. Why is the answer not changing? Yeah.

AUDIENCE: Once you lose, say, \$1,000, you're already in a really deep hole.

PROFESSOR: That's the intuition. That's right. We're going to get to that in a minute. I want to know from the

formula, why is it no difference if I come with \$1,000 versus \$1 million? Yeah.

AUDIENCE: The formula doesn't have n .

PROFESSOR: Yeah, the formula has nothing to do with n . You could come with \$100 trillion in your wallet, and it doesn't improve this bound. This bound only depends on what you're trying to win, not on how much you came with. So no matter how much you come with, the chance you win \$100 before you lose everything is at most 1 in 37,000.

Now, we can plug in some other values just for fun-- different values of m . If you thought 1 in 37,000 was unlikely, the chance of winning \$1,000, or 1,000 bets worth before you're broke-- that's less than $9/10$ to the 1,000. That's less than 2 times 10 to the minus 46-- really, really, really unlikely. Even winning \$10 is not likely.

Just plug in the numbers. The probability you win \$10 betting \$1 at a time is less than $9/10$ to the 10th power. That's less than 0.35.

You can come to the casino with \$10 million, bet \$1 at a time, and you quit if you just get up 10 bets-- get up \$10. The chance you get up \$10 before you lose \$10 million is about 1 in 3 you're twice as likely to lose \$10 million as you are to win 10. That just seems weird, right?

Because it's almost a fair game. It's almost 50-50. Any questions about the analysis?

Yes, I find that shocking. Just the intuition would seem say otherwise. So I guess there's a moral here.

If you're going to gamble, learn how to count cards in blackjack, or some game where you can make it even. Because even in a game where it's pretty close, you're doomed. You're just never going to go home happy.

Now, if you could have a fair game, the world changes-- much better circumstance. So actually, let's do the same analysis for a fair game, because that's where our intuition really comes from. Because we're thinking of this game as almost fair.

And in a fair game, the answer's going to be very different. And it all goes back to the recurrence and the roots of the characteristic equation. Because in a fair game, p is $1/2$.

And then you have a double root. 1 minus $1/2$ over $1/2$ equals 1 , and that means a double root at 1 . And that changes everything.

So let's go through now and do all this analysis in the case of a fair game. And this will give us practice with double roots and recurrences. Because as you see now, it does happen.

Let's figure out the chance that we go home a winner. OK, so let's see. In this case, we know the roots. Can anybody tell me what formula we're going to use for the solution? Got a double root at 1. So there's going to be a 1 to the n here. I don't just put a constant A in front. What do I do with a double root?

AUDIENCE: [INAUDIBLE].

AUDIENCE: A^n .

PROFESSOR: What is it?

AUDIENCE: A^n .

PROFESSOR: A^n -- not quite A^n . You got an A^n here.

AUDIENCE: Plus B .

PROFESSOR: Plus B -- that's what you do for a double root, because you make a first degree polynomial in n here. So we plug that in. The root's at 1 , so it's real easy.

The solution's really easy now. No messy powers or anything. It's just A^n plus B . And I can figure out A and B from the boundary conditions.

All right, X_0 is 0 . X_0 is just B , because it's A times 0 goes away. And that means that B equals 0 . This is getting really simple.

1 is X_T . And that's A plus B , but B was 0 . So that's A times 1 plus B . That's just A .

It means A equals A^n . Good, n 's not 1 . n 's T . So it's A^T plus B .

This is A^T here. So A^T equals 1 . That means A is 1 over T .

All right, that means that X_n is n over T . And T is the total. The top limit is n plus m , because you quit if you get ahead m dollars.

This is just now n over n plus m . All right, so let's write that down. It's a theorem.

If p is $1/2$, i.e., you have a fair game, then the probability you win m dollars before you lose n dollars is just n over n plus m . And this might fit the intuition better. So for the mother-in-law strategy, if m is 100, and n is 1,000, what's the probability you win-- you go home a winner?

Yeah, 1,000 over 1,000 plus 100. 1,000 over 1,000 is 10 over 11. So she does go home happy most of the time-- 10 out of 11 nights-- if she's playing a fair game.

Any questions about that? So the trouble we get into here is that the fair game results match our intuition. You know if you have 10 times as much money in a fair game, you'd expect to go home happy 10 out of 11 nights.

That makes a lot of sense. You go home happy 10, and then you lose the 11th. That's a 10 to 1 ratio, which is the money you brought into the game.

The trouble we get into is, the fair game is very close to the real game. Instead of 50-50, it's 47-53. And so our intuition says the results-- the probability of going home happy in a fair game-- should be close to the probability of going home happy in the real game.

And that's not true. There's a discontinuity here because of the double root. And the character completely changes.

So instead of being close to 10 out of 11, you're down there at 1 in 37,000-- completely different behavior. OK, any questions? All right, so let me give you an-- yeah.

AUDIENCE: So what happens if you make n 1, and then you do that repeatedly?

PROFESSOR: Now, if I did n equals 1, I could use that as an upper bound, and it's not so interesting as, say, 90%. But I would actually go plug it back in here. So this would be n plus 1, and it would depend how much money I brought.

But there is a pretty good chance I go home a winner for m equals 1. Because I've got a pretty good chance that I either-- 47% chance I win the first time. Then I go home happy.

If I lost the first time, now I've just got to win twice. And I might win twice in a row. That'll happen about 20% of the time.

If I lose that, now I've got to win three in a row. That'll happen around 10% of the time. So I've got 10 plus 20 plus almost 50. Most of the time, I'm going to go home happy if I just have to get ahead by \$1.

But it doesn't take much more than one before you're not likely to go home happy. Getting ahead 10 is not going to happen, very likely. Now, you want to recurse on that? I'm pretty likely to get ahead by one.

Well, OK, get ahead by one. I'm pretty likely to do it again. And I did it again. Now I'm pretty likely to do it again.

And there's this thing called induction that we worried a lot about. So by induction, are we likely to go home happy with 10? No, because every time you don't get there, you're dead. You had a little chance of dying and not reaching one, and a little chance of dying and not going from one to two. And you add up all those chances of dying, and you're toast, because that'll be adding up to everything, pretty much.

So that's a good question. If you're likely to get up by one, why aren't you likely to get up by 10? It doesn't work that way. That's a great question.

Let me show you the phenomenon that's going on here, as to why it works out this way. We had the math. So we looked at it that way.

We notice that one case is a double root and the other case isn't. And that exponential, in the case where you didn't have that second root at 1 makes an enormous difference. Qualitatively, we can draw the two cases. So in the case of an unbiased or fair game, if we track what's going on over time, and we start with n dollars, sort of this is our baseline.

And here's our target-- T is n plus m . And so we quit if we ever get here. And we quit if we ever hit the bottom.

And we've got a random walk. It's going around, just doing this kind of stuff. And eventually, it's going to hit one of these boundaries. And if m is small compared to n , we're more likely to hit this boundary.

And in fact, the chance we hit this boundary first is the ratio of these sizes. It's n over the total. It's the chance that we hit that one first. Now in the biased case, the picture looks different.

So in the biased case-- so this is now biased. And we're going to assume it's downward biased. You're more likely to lose.

So you start at n , you've got your boundary up here at T equals n plus m . Time is going this way. The problem is, you've got a downward sort of baseline, because you expect to lose a little bit each time.

And so you're taking this random walk. And you collide here. And these things are known as the swings. This is known as the drift.

And the drift downward is $1 - 2p$. That's what you expect to lose if you get the expected loss on each bet-- $1 - 2p$. Because you're going to not be a fair game.

This one has zero drift up there. It stays steady. And in random walks, drift outweighs the swings.

These are the swings here. And they're random. The drift is deterministic. It's steadily going down.

And so almost always in a random walk, the drift totally takes over the swings. The swings are small compared to what you're losing on a steady basis. And that's why you're so much more likely to lose when you have the drift downward.

Just as an example, maybe putting some numbers around that. The swings are the same in both cases, so that gives you some qualification for how big the swings tend to be. We can sort of do that with standard deviation notation.

After X bets or X steps, the amount you've drifted, or the expected losses, $1 - 2pX$. Maybe we should just understand why this is the case. The expected return on a bet is 1 with probability p , and -1 with probability $1 - p$. And so that is-- did I get that right? I think that's right.

Oh, expected loss-- [INAUDIBLE] drifts down. Instead of expected return, let's do the loss, because that's the drift. It's a downward thing.

So the expected loss-- now you lose $\$1$ with $1 - p$. And you gain $\$1$, which is negative loss, with probability p . And so you get $1 - p - p$ is $1 - 2p$.

So that's your expected loss. Your expected winnings are the negative of that. So after x steps, you expect to lose-- well, I just add up the linearity of expectation.

You expect to lose this much x times. So that's your expected drift. You're expected to lose

that much.

Now, the swing-- and we won't prove this-- the swing is expected to be square root of x times a constant. So I've used the theta notation here. And the constant is small.

If I take x consecutive bets for \$1, I'm very likely to be about square root of x off of the expected drift. And you can see that this is square root. That is linear. So this totally dominates that.

So the swings are generally not enough to save you. And so you're just going to cruise downward and crash, almost surely. OK, any questions about that?

All right, so we figured out the probability of winning m dollars before going broke. That's done with. Now, this means it's logical to conclude you're likely go home broke in an unfair game.

Actually, before we do that, there's one other case we've got to rule out. We've proved you're likely not to go home a winner. Does that necessarily mean you're likely to go broke? I've been saying that, but there's some other thing we should check. What's one way you might not go home broke?

AUDIENCE: [INAUDIBLE].

PROFESSOR: What is it?

AUDIENCE: You don't go home.

PROFESSOR: You don't go home. And why would you not go home? Yeah?

AUDIENCE: You're playing forever.

PROFESSOR: You're playing forever-- we didn't rule out that case-- you're playing forever. But it turns out, if you did the same analysis, you can analyze the probability of going home broke. And when you add it to the probability of going home a winner, it adds to 1, which means the probability playing forever is 0.

Now, there are sample points where you play forever. But when you add up all those sample points, if their probability is 0, we ignore them. And we say it can't happen.

Now, we're bordering on philosophy here, because there is a sample point here. You could win, lose, win, lose, win, lose forever. But because you add them all up at 0, measure theory

and some math we're not going to get into tells you it doesn't happen. It's probability 1 you're a winner or a loser. All right, so I'm not going to prove that the probability you play forever is 0.

But let's look at how long you play. How long does it take you to go home one way or another-- go broke? And to do this, we're going to set up another recurrence. So we know eventually we hit a boundary.

I want to know how many bets does it take to hit the boundary? How long do we get to play before we go home unhappy? So S will be the number of steps until we hit a boundary.

And I want to know the expected number-- I'll call it E_n here-- is the expected value of S given that I start with n dollars. I mean, the reason you could think about this is, we know we're going to go home broke-- pretty likely. Do we at least have some fun in the meantime? Do we get a lot gambling in and free drinks, or whatever, before we're killed here?

Now, this also has a recurrence. And I'm going to show you what it is, then prove that that's correct. So I claim that the expected number of steps given we start with n dollars is 0 if we start with no money, because we are already broke.

It's 0 if we start with T dollars, because then we just go home happy. There's no bets, because we've already hit the upper boundary. And the interesting case will be it's $1 + p E_n - 1 + (1-p) E_{n-1}$, if we start with between 0 and T dollars.

OK, so let's prove that. Actually, the proof is exactly the same as the last one. So I don't think I need to do it.

The proof is pretty simple, because we look at two cases. You win the first bet-- happens with probability p . And then you're starting with $n + \$1$ over again. Or you lose the first bet-- happens with probability $1 - p$. And you're starting over with $n - \$1$ now-- same as last time.

In fact, this whole recurrence is identical to last time except for one thing. What's the one thing that's different now?

AUDIENCE: [INAUDIBLE].

PROFESSOR: What is it?

AUDIENCE: You have [INAUDIBLE].

PROFESSOR: You have--

AUDIENCE: So it's not [INAUDIBLE] any more.

PROFESSOR: That's different. There's another difference. That's one difference that's going to make it inhomogeneous. That's sort of a pain.

What's the other difference from last time? This part's the same otherwise.

AUDIENCE: Boundaries.

PROFESSOR: What is it?

AUDIENCE: Boundary conditions.

PROFESSOR: Boundary conditions-- that was a 1 before. Now it's a 0. OK, so a little change here, and I added a 1 here. But that's going to make it a pretty different answer.

So let's see what the recurrence is. I'll rearrange terms here to put it into recurrence. I get $p E_{n+1} - E_n + 1 - p E_n = -1$, not 0. And the boundary conditions are $E_0 = 0$ and $E_T = 0$.

OK, what's the first thing you do when you have an inhomogeneous linear recurrence? Solve the homogeneous one. And the answer there-- well, it's the same as before. This is the part we analyzed.

And we'll do it for the case when p is not $1/2$ -- so the unfair game. So the homogeneous solution is E_n just from before-- same thing-- $\frac{1-p}{p} \left(\frac{1-p}{p}\right)^n + B$. And this is the case with two roots. p does not equal $1/2$.

What's the next thing you do for inhomogeneous recurrence? Are we plugging in boundary conditions yet? No. So what do I do next? Particular solution.

And what's my first guess? We have the recurrence like this here. What do I guess for E_n ? I'm trying to guess something that looks like that. So what do I guess?

Constant, yeah. That's a scalar. I just guess a constant. And if I plug a constant a into here, it's

going to fail. Because I'll just pull the a out. I'll get p minus 1 plus 1 minus p is 0, and 0 doesn't equal minus 1. So it fails.

So I guess again. What do I guess next time? a n plus b . All right, and I don't think I'll drag you through all the algebra for that, but it works.

And when you do it, you find that a is minus 1 over $2p$ minus 1. And b could be anything. So let me just rewrite this as 1 over 1 minus $2p$. And b can be anything, so we'll set b equal to 0.

So we've got our particular solution. It's not hard to go compute that. You just plug it back in and solve. Now we add them together to get the general solution.

This is A n plus B . B was 0, and here's A as 1 over 1 minus $2p$. And now what do we do to finish? I've got my general solution here by adding up the homogeneous and the particular solution. Plug in the boundary conditions.

All right, I'm not going to drag you through solving this case, but I'm going to show you the answer. E_n equals n over 1 minus $2p$ minus T , the upper boundary, over 1 minus $2p$ times 1 minus p over p to the n minus 1 over 1 minus p over p to the T minus 1. So actually, this looks a little familiar from the last time when we did this recurrence, figuring out the probability we go home a winner.

Here this is the expected number of steps to hit a boundary, to go home. If we plug in the values, it's a little hairy, but you can compute it. So for example, if m is 100, n is 1,000, T would be 1,100 in that case. p is $9/19$ playing roulette.

Then the expected number of bets before you have to go home is 1,900 from this part, minus 0.56 from that part. So actually 19,000, sorry. So it's very close to 19,000 bets you've got to make.

So it takes a long time to lose \$1,000. And it sort of comes very close to the answer you would have guessed without thinking and solving the recurrence. If you expect to lose 1 minus $2p$ every bet, and you want to know how long the expected time to lose n dollars, you might well have said, I think it's going to be n over the amount I lose every time.

That would be wrong, technically, because you'd have left off this nasty thing. But this nasty thing doesn't make much of a real difference, because it goes to 0 really fast for any numbers

like 100 and 1,000-- makes no difference at all. So the intuition in that case comes out to be pretty close, even though technically, it's not exactly right.

Now, to see why this goes to 0, if T equals n plus m here-- this is n plus m -- and your upper limits, say m goes to infinity-- it's 100 in this case-- then that just zooms to 0, and you're only left with that. Which means that we can use asymptotic notation here to sort of characterize the expected number of bets. And it's totally dominated by the drift. So as m goes to infinity, the expected time to live here is \tilde{n} over $1 - 2p$. If you've got n dollars, losing $1 - 2p$ every time, then you last for n over $1 - 2p$ steps.

OK, now, actually, what situation in words does m going to infinity mean? Say I set m to be infinity? What is that kind of game if m is infinity? How long am I playing now? Yeah.

AUDIENCE: Now you're playing for as long as it takes you to lose all of your money.

PROFESSOR: Yes, because there is no stopping condition up here-- going home happy. I'm going to play forever or until I lose everything. And this says how long you expect to play. It's a little less than n over $1 - 2p$.

So if you play until you go broke, that's how long you expect to play. So that sort of makes sense in that scenario. That's not one where it surprises you by intuition.

It is interesting to consider the case of a fair game. Because there's something that's non-intuitive that happens there. So in a fair game, p is $1/2$.

Now, if I plug in $1/2$ here, well, I divide by 0. I expect to play forever. That's not a good way to do the analysis, that you get to a divide by 0.

Let's actually go back and look at this for the case when p is $1/2$. And see what happens in a fair game-- how long you expect to play in a fair game. Then the homogeneous solution is the simple case.

E is A^n plus B . You have a double root at 1, which we don't have to worry about 1 to the n . When you do your particular solution, you'll try a single scalar, and it fails. I'll use lowercase a -- fails.

You will then try a degree one polynomial, and that will fail. What are you going to try next? Second-degree polynomial, and that will work.

OK, and the answer you get when you do that is that-- I'll put the answer here. It turns out that a is minus 1 and b and c can be 0. So it's just going to be minus n squared for the particular solution. That means your general solution is $A n$ plus B minus n squared.

Now you do your boundary condition. You have $E 0$ is 0. Plug in 0 for n . That's equal to B . So B is 0. That's nice.

$E T$ is 0. And I plug in T here, I get AT , B is 0 minus T squared. So I solve for A here.

That means that A equals T . AT squared minus T squared is 0. A has to be T .

So that means that $E n$ is Tn minus n squared. Now, T is the upper bound. It's just n plus m . n plus m times n minus n squared-- this gets really simple.

The m squared cancels. I just get n out. That says if you're playing a fair game, until you win m or lose n , you expect to play for nm steps, which is really nice. This is p is $1/2$ -- very clean.

Now, if you let m equal to infinity, you're going to expect to play forever. So with a fair game, if you play until you're broke, the expected number of bets is infinite. That's nice. You can play forever is the expectation.

Now, here's the weird thing. If you expect to play forever, does that mean you're not likely to go home broke? You expect to play forever. And as long as you're playing, you're not going home broke.

Now, there's some chance of going home broke, because you might just lose every bet-- not likely. Here's the weird thing-- the probability you go home broke if you play until you go broke is 1. You will go home broke.

It's just that it takes you expected infinite amount of time to do it-- sort of one of these weird things in a fair game. So here we proved the expected number bets is nm . If m is infinite, that becomes an infinite number of bets.

One more theorem here-- this one's a little surprising. This theorem is called Quit While You're Ahead. If you start with n dollars, and it's a fair game, and you play until you go broke, then the probability that you do go broke, as opposed to playing forever, is 1.

It's a certainty. You'll go broke, even though you expect it to take an infinite amount of time. All right, so let's prove that.

OK, the proof is by contradiction. Assume it's not true. And that means that you're assuming that there exists some number of dollars that you can start with, and some epsilon bigger than 0, such that the probability that you lose the n dollars-- in which case you're going home broke-- let me write the probability you go broke-- is at most $1 - \epsilon$.

In other words, if the theorem is not true, there's some amount of money you can start with such that the chance you go broke is less than $1 - \epsilon$. OK, now that means that for all m , where you might possibly stop but you're not going to, the probability you lose n before you win m is at most $1 - \epsilon$. Because we're saying the probability you lose n no matter what is at most that. So it's certainly less than $1 - \epsilon$ that you lose n before you win m dollars.

And we know what that probability is. This probability is just m over $n + m$. We proved that earlier. So that has to be less than $1 - \epsilon$ for all m .

And now I just multiply through for all m . That means that m is less than or equal to $(1 - \epsilon)(n + m)$. And then we'll solve that.

OK, so just multiply this out. So for all m less than or equal to $n + m - \epsilon n - \epsilon m$, and now pull the m terms out here, I get for all m , ϵm is less than or equal to $(1 - \epsilon)n$. That means for all m , m is smaller than $(1 - \epsilon)/\epsilon$ times n .

And that can't be true. It's not true for all m , this is less than that, because these are fixed values. That's a contradiction.

All right, so we proved that if you keep playing until you're broke, you will go broke with probability 1. So even if you're playing a fair game, quit while you're ahead. Because if you don't, you're going to go broke. The swings will eventually catch up with you. So if we draw the graph here, we'll see why that's true.

All right, if I have time going this way, and I start with n dollars, my baseline is here. The drift is 0. I'm going to have swings.

I might have some really big, high swings, but it doesn't matter, because eventually I'm going to get a really bad swing, and I'm going to go broke. Now, if you ever play a game where

you're likely to be winning each time, and the drift goes up, that's a good game to play, obviously. It just keeps getting better. And that's a whole math change there.

So that's it. Remember, we have the ice cream study session Monday. So come to that if you'd like. And definitely come to the final on Tuesday. And thanks for your hard work, and being such a great class this year.

[APPLAUSE]