[SQUEAKING]

[RUSTLING]

[CLICKING]

**ERIK DEMAINE:** All right, let's get started. Today we continue probability, the last theme of the class. So we have three more lectures about probability. Today's lecture is about expectation, which is a nice summary. It's a one-number summary of a random variable.

Remember from last time, a random variable is neither random nor variable. It's really a function from your sample space to something, usually real numbers. So we use this to represent some computation about an experiment.

We had a similar setup with derived variables, if you remember back to state machines, which were also not variables, they were really functions from the state of the machine to numbers. So random variable is the same idea. If I give you-- if I tell you what sample-- like, what random events happened, you can compute a single value for this variable. That's the meaning.

And we had a particular type of random variable that's going to be really useful today, which is an indicator random variable where the output is always 0 or 1, indicating something happened-- that's the 1-- or it didn't happen.

So functions are complicated. If I want to write down a function, I got to write down a lot of numbers-- one for every possible sample. So expectation is a way to take that complicated function and reduce it down to one number.

Obviously, it's not going to tell you everything about the function, but it tells you what you expect to happen on average. What is the-- it's not the most likely thing. But if you take all the things that could happen and you average them, that's expectation. So let's define that precisely. Match my notation here.

So let's call random variable X over a sample space S. Then the expectation of X, which we write Ex of X. Although, often it's written just E of x. So if you look at other probability sources, you might see that as well. Both of these mean the same thing. It's going to be a weighted average-- weighted by probabilities.

So we're going to sum over all possible samples, which we call omega and S of the probability of that outcome times X-- the function X evaluated at that income-- outcome-- omega for outcome.

So if I-- so this is a particular type of average because the probabilities sum to 1. And in particular if the probabilities are all equal, this is literally the average value of X. But, of course, we would like-- what we care about is the weighted by the probability. If there's something that's more likely to occur, that should get a bigger multiplier in front. So this is going to be some average of X.

Sometimes the expectation is also called the average. Let's do a very simple example, which is we roll a fair six-sided die. I happen to have such a die right here. So we roll it. It has values 1 to 6, each equally likely. Here, I got a 1. What is the expected value?

So the sample space here is 1 up to 6. All the probabilities are 1, 6. Probability of I occurring is 1 over 6. And so the expected value-- I should probably give this a name. Let's call it D. Probability that D equals I for any particular I between 1 and 6 is 1 over 6.

And so the expected value of D, the die roll is from that formula. Well, it's just the probability that we get a 1 times 1 plus the probability that we get a 2 times 2 plus and so on. I'm going to rewrite that a little bit.

So these probabilities are all 1/6. So we've got 1/6 times 1 plus 1/6 times 2 plus up to 1/6 times 6. These numbers is our favorite summation-- the triangular number. This is Gauss, right?

So if I factor out the 1/6 in front, then we have that times the sum 1 through 6, which you should remember is 6 times 7 over 2. It's getting a little crowded. This is 1 over 6 times 6 times 7 over 2. Let me get a color chalk.

It's time for the exciting cancellation moment. The 6's cancel. And so we're left with 7 over 2, also known as 3.5, which is, as you would expect, the average of numbers 1 through 6.

But notably, a point of this example is that the expectation is not actually a valid outcome. You never get a 3.5 on a die. But you get 1, 2, 3, 4, 5, and 6, and the average of those is 3 and 1/2. So don't expect the expectation to be an actual point in your-- an actual value that X could take on.

All right. Let's do a more interesting example, which we're going to use a lot. What if I have an indicator random variable, also X-- no, call it I-- for event A. So remember this means I is 1 if event A occurred. If the sample that happened is in A, and it's zero if A did not occur. So let's try to compute the expected value of I.

So this is a little tricky with this formula because S-- even though I only takes on two values, S can be size a million. Who knows? There's a lot of different random outcomes that could happen. And then I kind of projects them all down to this very simple 0,1 space. So I'm just going to write the definition first.

Omega in S-- probability of omega times I of omega. But this thing is either 0 or 1. So this is a summation. And remember, one of the techniques for dealing with summations is to rearrange the terms in some useful way. And given that every term here is either multiplied by 0, in which case it disappears, or multiplied by 1, let's maybe collect like terms.

So we can split this summation into two parts. There's the sum where W is in a complement probability of omega times 0, because I of something in a complement here is 0. Or it could be omega inside A without the complement. And then we get probability of omega times 1.

So these disappear, and we're just left with-- and the 1 cancels, or it doesn't do anything. So we get the sum of omega and A times probability of omega. And this is our definition of probability of A. So that's nice and simple.

Another way to write this that's more directly about I is this is the probability that I equals 1. So the expected value of an indicator random variable is always the probability that that indicator random variable equals 1.

This is just a fun fact because I only takes on 0 and 1. The zeros don't show up-- they disappear from the expectation formula, and the 1's basically add up just the probabilities of the event occurring. We're going to use this fact, several times today. So it's an example and a useful lemma.

Probability, I'm told, was originally invented to analyze gambling. So let's put that to good use and do some gambling. So let me define a game. And then I'm going to need a volunteer-- a couple of volunteers-- to play this game with me. And it doesn't have a fancy name. We're just going to call it the gambling game.

So we're going to have three players. And we're going to follow these sequence of steps. Everybody has some pile of money, which we're going to represent by candy, but I'll write it as dollars.

So every round, every player puts $2 into a shared pot. Then each player picks heads or tails. You may have played this game in the single-player version, where you just try to guess the outcome of a coin. Then we're going to flip a fair coin.

With multiplayer, it's just more interesting, let's say. Now we could have multiple winners. Maybe two people guessed heads, and the outcome of the coin was heads. Then what we're going to do is take the pot of money that everyone added to and evenly distribute it among the winners, among the people that guessed the coin outcome correctly. So we split pot equally among winners.

There's one special case, which is maybe no one guesses the correct value. In that case, we'll also equally split the pot, which is equivalent to just giving your money back and trying again. So we're going to play this game many rounds. And your goal is to, of course, maximize the amount of money you get. Who would like to play this game. Oh, so many people.

All right. I'm afraid it's only a three-player game. So I'm going to pick Jenny, my accomplice. And how about you?

**AUDIENCE:** Don't tell them where I got this.

**ERIK DEMAINE:** [LAUGHS] Yeah. No. Totally fair game. Don't worry. What's your name?

**AUDIENCE:** Cleo.

**ERIK DEMAINE:** Cleo. All right. So everyone welcome Cleo and Jenny to the stage. Yeah.

[APPLAUSE]

I want you to be in like equally hard and good position. So we're going to put you in the middle. C-L-E-O?

**AUDIENCE:** Yeah.

**ERIK DEMAINE:** All right. Try to keep track here. We've got a coin. Except I don't have a coin. I have a giant die. So we're going to say was it heads, H? H is huge. I've got to write this down, or I'm going to forget. H is huge, and T is tiny. OK. So 1, 2, 3 is a tails. 4, 5, 6 is an H. All right.

So I'm going to guess first. Then why don't you stand in the middle so we're like in order here. So I'll guess tails, my usual.

**AUDIENCE:** Heads.

**AUDIENCE:** Tails.

**ERIK DEMAINE:** Tails. OK. Let's roll this dies. That F doesn't feel very fair, but OK. Well, I'll do different types of rolls. So we got a 4, which is H. Ah-ha. Cleo wins. Oh, I forgot to distribute the money. OK, these are all equal. So pick your pile of cash, which are poker chips. And so it's, I think, more fun if everyone can see the pile. So let me--

So we were supposed to start out by putting two in to the pot. And then you are the only winner for the pot. So you get all this money. That's already looking good. All right.

So I'm going to flip the-- I should flip after guessing. A little too forced, otherwise. All right. I'll try heads this time.

**AUDIENCE:** Tails.

**ERIK DEMAINE:** Tails.

**AUDIENCE:** Heads.

**ERIK DEMAINE:** Heads. OK. Now we flip. OK. 6 is high. Oh, I forgot to do our thing. Two in, and then we evenly split the pot. All right. Great.

**AUDIENCE:** Again.

**ERIK DEMAINE:** Very consistent coin. Play this game properly. Supposed to put the money-- you're supposed to put your money where your mouth is before you play the game. I'll try heads.

**AUDIENCE:** Heads.

**ERIK DEMAINE:** Heads.

**AUDIENCE:** There's no way it's three heads in a row. Tails.

**ERIK DEMAINE:** [LAUGHS] Yeah, three heads in a row is never useful. Because whatever happens, we will just take our money back. All right. So we flip. This is quite fun. Another heads. Wow, it's a very consistent coin here.

All right. So here is the distribution. We put our money back in-- two each. I'll do tails.

**AUDIENCE:** Heads.

**ERIK DEMAINE:** Heads.

**AUDIENCE:** Tails.

**ERIK DEMAINE:** Tails. I'm very efficient. This is great. I'm the slow part here, rolling Oh, yeah. Another heads. Great.

[LAUGHS]

I love this. There's another problem we'll be working on, which is how many times do you have to flip a coin until you get tails? The answer is two. [LAUGHS]

All right. So money back in, and I'll do tails.

**AUDIENCE:** Sorry, heads.

**ERIK DEMAINE:** Heads.

**AUDIENCE:** Tails.

**ERIK DEMAINE:** Tails. All right. Based on how this has been going, good guess. All right. Have we played enough? Do you get the idea?

**AUDIENCE:** Not yet.

**ERIK DEMAINE:** Oh, not yet. You're almost depleted there, Jenny. All right. Let's see. One more. We should keep going until we get a tails, right. That can't take forever. Famous last words. I've done this before. All right. Heads.

**AUDIENCE:** Tails.

**ERIK DEMAINE:** Tails.

**AUDIENCE:** Heads.

**ERIK DEMAINE:** Heads. Now, the gambler's ruin would say we're due for a tails now. It's more likely than 50%. Oh, gambler is wrong. It's 50/50 every time. All right. Heads.

**AUDIENCE:** Tails.

**AUDIENCE:** Heads.

**ERIK DEMAINE:** Yes. Finally, a tails. All right. Cleo here has won basically all the candy. You can take as many as you like because I'm actually lactose intolerant, so I can't eat these. So both of you are welcome to take as many as you like. So that's the gambling game.

Now, is it coincidence that Cleo won and I lost? No.

**AUDIENCE:** Jenny was picking the other one. The other from you, indeed. Very observant. This is why Jenny was so fast at guessing. Jenny is one of our TAs, by the way. That's how I convinced her to cheat. But she cheated in order to destroy me.

So there's this fun strategy, which is if you guess the opposite of one of the previous players, which, if you guess in sequence like this, is easy to do. Cleo could have done it. Cleo actually did it most of the time. I think there were only a couple times-- one time when we matched. And that's a good strategy. In particular-- well, let's analyze. shall we?

How do we analyze a game like this. So small-- we just draw the decision tree or the probability tree. All right. Now, it's a little bigger than some of the ones we've done. Luckily, the whole thing is symmetric. It doesn't really matter which one you do first. The whole thing is symmetric.

So I'm going to do it in the order of let's branch on what player one does, player two, player three. And then what the coin does. And then we're going to get some outcome, which I'll define in a second.

So it turns out whether the first player guesses heads or tails because the coin is random, it's going to be symmetric. So let's just ignore the whole right half of the space and pretend we're in this left half. So suppose I always guessed heads, or somebody always guesses heads. I think it's probably going to be me that I care about. Yeah.

Then player two guesses heads or tails. Now, we haven't said what the probabilities are here. I was trying to be roughly 50/50. But I wasn't flipping a coin each time to decide how I should guess the coin. I was just making some arbitrary choice. Cleo was making some arbitrary choice. Jenny was making some less arbitrary choice.

And then the coin-- this part is random. So this one is one half either way. So that's helpful. And then we get some outcome, which I'm going to measure the net for player one. That's this number. And I wrote it down here because it's very easy to get these wrong.

So the first case is that we all guessed heads. And then it doesn't matter what we do. Just our money is going to go back-- or the pot is going to go back to even distribution. Whether we win or lose, the money goes back to where it started. So the net for player one, in particular, is zero. So these are boring.

Then we have heads, heads, tails for the three players. And so then if it's heads, these two players won. And so they evenly distribute. So we're always putting 6 in each time. So they each get 3.

But they started by putting in 2. And so the net gain is only 1. This is 3 minus 2. We gain 3, but we had to pay 2 to get there. Or we lose. Whenever we lose-- and all of these tails branches player 1 is going to lose because they guessed heads. And, in that case, you get minus 2 because you had to put in the 2.

So in this case is the same. There's 2 heads. And so you each get 3. And so it's a net of 1. And then there's this case. This is the case you're really aiming for. This is the best for player one. Because in this case, player one is the unique winner.

So here, there was heads. And the other two guessed tails, and the coin came up heads. And so that's what-- you're really trying to be the unique winner in this game, the less common winner. That's when you get lots of points.

So let's compute the expectation. To compute the expectation, we need to know what the probabilities are in these branches or something about them. And I'm going to start by saying, let's suppose that all players are uniform and independent. So 50-50 on every branch.

Then the probabilities down here-- let me use color-- are going to be 1/16 in the whole tree or 1/8 in this subtree. So I'm going to just condition on the H in the beginning. And so I'll write 1/8 down here. But if you were drawing the whole tree, you'd have to write 1/16. Each of these things occurs with 1/8 probability. You get the idea. Each of these 1/8.

And so we get the expected value is 1/8 times the sum of these values-- 0 plus 0, plus 1, minus 2, plus 1, minus 2, plus 4, minus 2, which I hope is equal to 0. Because the negative is 6, and the positive is also 6.

So this is what you might call a fair game. Your expected winnings are 0. So you're just as likely-- well, it doesn't mean you're just as likely to go above or go below. But, on average, you will break even.

But that's not what we did. We considered the case where, let's say, player two does not equal player three, which is something that player three could enforce. And that's what Jenny was doing, always choosing opposite of Cleo here.

In that case, it's really bad for me. Because whatever I guess out here-- if there's both a heads and tails out here, if they're always guessing opposite, one of them's heads. One of them's tails. So that means I always match one of them guaranteed.

And the other of them is in the unique pile. And so they have a 50% chance. Like in this situation, Cleo has a 50% chance of getting that plus 4. There's still a 50% chance that it came up tails. Of course, not with this die. But, in principle, there's a 50% chance that it comes up tails. But then we each only get 1 point.

So if you're focusing on my winnings, I will never get $4 in a round with Jenny's strategy. I will always get either 1 or minus 2. That doesn't sound like a very good game for me. If I randomly get-- a 50% chance I get a 1, 50% chance I get a minus 2.

So this was, I guess, expected net for player 1. That's what we've been computing. So in this situation the expected net for player 1, If we want to be precise, you can say it's this subtree here. This is when the last two guesses are heads tails or tails heads is the middle part of this subtree. And you can see my possible outcomes are limited to these guys.

And it turns out they're equally likely. In fact, independent of what player two does. I won't try to prove that here. But because all that matters is this last coin flip, if it's heads, then I get a plus 1. If it's tails, I get a minus 2 every time in this half of the tree. And so it ends up being half-- in other words, the average of 1 minus 2, also known as minus a half.

So, given how Jenny played, in fact, I expect to lose. And you saw that happening. Jenny also happened to lose. The way this is practical, if you're a gambler, is if you have two players that collude, they decide whatever happens, we're going to take our total winnings and split it evenly between us at the end of the game after all the rounds are done.

Then this is not quite a guarantee, but it's a pretty good bet that together, Cleo and Jenny would get most of the coins, most of the money, because their expected outcome is plus a half, I believe, should be the complement. Plus-- yeah. Anyway, it's positive. And my expectation is negative. So yeah, don't play this game if you don't trust players two and three not to collude. All right. Cool.

You might think this is a silly game. Who would ever play this game? Well, the Massachusetts lottery played this game. It's a little more complicated. I think you had up to four digits, and you tried to guess a sequence of four digits that would come up randomly. I mean, it's pretty similar to lotteries today. They just have more numbers.

And there's this great MIT professor, Herman Chernoff, who we'll talk about some of his probability results later in the class. And he wrote this paper, "How to Beat the Massachusetts Numbers Game," which is what the lottery was called originally. And he just looked at some data. And it's a similar situation.

If there are many lottery winners, they evenly split the pot. Also, the state takes like 40% or 60% cut. I forget which way it goes. But even then, he made money off of the lottery. And there have been more recent versions of this run by MIT students, I hear, to make a lot of money off the Massachusetts lottery. And they've since changed the rules to make those strategies hard to run.

But, in particular, he executed this, made a lot of money. The whole-- the idea is if you can guess numbers that are less likely to be guessed by other people, and all numbers are equally likely, then you expect to win. And it turns out people are so predictable.

Let's just say zeros and 9's were unpopular. Smaller digits were popular, and so he ended up just going with random numbers from 5 to 8, I think and 0 to 9 maybe. And he won not a lot of money, but he won some money.

You can read the paper for more details. It's a pretty fun read. It goes through all the probability and stuff. This is like a simplified version. So cool, practical. Now you know how to make money. Let's go on to some other ways to compute expectation and some more examples and so on. Cool.

All right. We talked about-- so here, we computed if we roll a single six-sided die, the expected value of that single die is 3.5. What if I rolled two dice, A and B, and I want to know the expectation of dice A, die A. Should also be 3.5, right? If I roll two dice, it shouldn't change anything.

But if I computed it explicitly in this way, it would be really tedious because the sample space would go from 1 through 6 to 1 through 6 squared. So now there's 36 different possible outcomes. And for each of them, I need to write down a term and add it all up. It will come out to 3.5, but that's super annoying to do.

So here's a nice theorem that makes it easy to do this with six terms instead of 36 terms. If I have a random variable x, and I want its expectation, this is equal to the sum where x is in the range of big X of probability X equals X times little x.

Let me erase this-- my terrible losing game. How many coin flips till tails? 1, 2, 3, 4, 5, 6, 7. All right, Keep that in mind-- seven coin flips to get a tails.

All right. So here's a formula. It's really exactly what we did here. So we wrote down, here are all the possible outcomes, and we multiply them by what the random variable could be. And when there was an indicator, random variable is either 0 or 1. So there were two groups.

In general, there's whatever the range of x groups. So if I-- yeah. So this is exactly this kind of clustering where we split it into two sums. Each of them was the probability of being in that group times the outcome of that group. This sum is exactly probability of not naught A. And this sum is probability of A.

And in general, if we split into not two groups, but all the different possible values of big X-- it can be little x for every little x in the range of big X, then this is-- we get a sum of probability of outcomes, which is exactly the probability that big X equals small x. And then we multiply that by small x because that's what we have in the definition of expectation. So this is an easy theorem. I won't write down a proof because it's this thing generalized a little bit.

And let's do an example maybe on a different board.

Suppose I want to know the number of hearts in a 3-card hand. So what that means is I have 52 cards in a deck. I randomly choose 3 cards without replacement. So just like you're drawing from the top of a randomly shuffled deck. I want to know how many hearts do I get in expectation?

Let's call this-- yeah-- expected number of hearts is, if we just use the definition, probability of w times the number of hearts in omega. This would be the definition. And S, here, is size 52 cubed. Not quite. 52 times 51 times 50. It's really big.

So I don't want to have to compute this explicitly. Instead, I'm going to use this formula and say, well look, the number of hearts in a 3-card hand is 0, 1, 2, or 3. So I can write this as the probability that number of hearts equals 0 times 0. OK, that's not very interesting.

We can take the probability number of hearts equals 1 and multiply that by 1, plus the probability the number of hearts equals 2 times 2, plus probability number of hearts equals 3 times 3. And so now I have three terms-- relatively easy. And these are, hopefully, things you know how to compute.

For example, probability I get three hearts-- I'm going to cheat. This is like 13 choose 3 divided by 52 choose 3. Because there are 13 different heart cards. So the number of ways to choose three hearts out of those 13 heart cards is 13 choose 3. The total number of ways to choose three cards is 52 choose 3. And so that's a probability. And you need to multiply it by 3 to get this formula.

And there's some similar things for these terms. I think I maybe won't write them down. Maybe I'll write down the middle one just for kicks. I went through the effort of figuring things out. This point should be something that you could do just like on your problem set. These are combinatorics problems you know how to solve.

So you get some sum of binomial terms. But hey, it has some reasonably small size. Of course, these have huge factorials in them. But it's only choose 1, 2, or 3. So they're not too big. You plug this into Wolfram Alpha, say, and you get 3/4-- very nice number. We might come back to this.

A lot of the time, because this is 6 1,200, we're working with integers, natural numbers. All of the random variables I've defined so far, I believe, are integers. We happen to get an expectation that was not an integer. But all of the values that D could take on or hearts could take on was an integer.

So let's consider the case where I'll say the codomain of X is the natural number-- 0, 1, 2, 3. Remember, codomain is just what you declare this function could take on for values. It doesn't have to take on those values.

Range here-- the intent is these are just the values that are taken on by this function. And so I didn't have to write-- even though the number of hearts could, I don't know, be 100 in some other card, in this sample space, we know it's just 0, 1, 2, or 3. And so we just need four terms.

But let's suppose potentially all the natural numbers are there, but not necessarily all of them. Then expected value of X has a nice form. And this is a formula I encourage you not to memorize. You should write it down on your cheat sheet and always look it up from there because it's easy to get wrong, especially the next one.

1 to infinity, probability x equals I times I. This is just exactly this theorem. Nothing exciting. Here, I wrote a sum over a range. Now I'm writing a sum over an explicit range.

The one clever idea I had is skip 0 because 0 is going to get-- the 0 case is going to get multiplied by 0, so that never matters. So we get to start at 1. Wow, I saved this one step for evermore. And then sum up to infinity if we might have to go that high So that is true.

But there's an even cooler form, which is sum I equals 1 to infinity of probability of X greater than or equal to I. Or there's no I here-- I equals 0 to infinity probability X greater than I.

So this part is just this theorem. This step is I just changed greater than or equal to greater than and changed where I started I. So here, it started at I,1. And here, it started at I, 0.

And so here, I'm effectively adding 1 to I. And so I get to do-- or subtracting 1 from I. So I do greater than instead of greater than or equal to. So these two arrow equalities should be obvious. And this one is the interesting one. So let's prove that.

OK, proof.

OK. I want to compute this sum and claim that it equals this one. And I'm going to do it by writing it in an obvious way. First, I'm just going to write out the infinite series with dot, dot, dot. So we've got X equal to 1 plus X gradient equal to 2 plus X gradient equal to 3. Just give me a moment, and I'll write all the remaining terms. Might take a while.

And now I want to expand these out into separate pieces in terms of X equal I. So this is, what is the probability that X is greater or equal to 1. Given that, again, that they're all natural numbers, we know this is probability X equals 1 plus probability X equals 2 plus probability X equals 3 plus and so on up to infinity.

What's this one? Well, same thing, but starting here. Probability X equals 2 plus probability X equals 3 plus and so on. This one starts over here. Probability X equals 3.

So if I now sum this way down the columns, I get 1 times probability X equals 1 plus 2 times probability X equals 2. Because there are two values here-- there are two ranges here for X that include X equals 2-- these two. 2 is not greater than or equal to 3, 4, 5, and so on. So I get exactly two of them. I get exactly three probability X equals 3, and so on.

And if you work out the dot, dot, dots via proof by induction, you get some I probability X equals I. And that's what we wanted. 1 to infinity. So that's why these two sums are the same.

So the point is to do this work once, so that in the future if you have a problem, you can use any of these formulas you want. You could use the definition of expectation, which is a sum over outcomes in the sample space. Or you could use this formula-- sum over all the things in the range with equality here. And then you multiply it by x.

Or if in the special case where your function only takes on integer values-- so this doesn't work if you have other real numbers-- then you can use one of these three formulas. Probably the last one or the first one are the useful ones.

If it's easy to compute the probability X equals something, then maybe use this formula. If it's easy to compute the probability that X is bigger than something-- this was CDF and this was PMF, remember, from last time.

It's easy, if you have a nice form for PMF, then use that. If you have a nice form for CDF, use that, provided you work with integers. Do not try to mix these formulas together. Don't compute the sum of I equals 1 to infinity, probability X greater than or equal to I times I. That would give you something that is not expectation. But surprisingly, students try. So don't do that. Use this one, or this one, or this one, or the definition.

Cool. Let's do some more interesting examples.

Lonely brace.

Is there questions? All right.

Next problem is called mean time to failure. It's a little negative. You can use it to compute successful things as well. It's related to this problem of how many times-- it is essentially this problem of how many times do I have to flip a coin until I get tails, or until I get heads? Either way. I'll write it as heads because that's what my notes say.

It's less interesting-- it's more interesting for a biased coin. So, let's say, the probability of getting heads is p. I want to know what is the expected number of flips until I get heads.

So you could use this if you're analyzing coin flipping. But you could also use it for things like my laptop, every day there's a 1 in 1,000,000 chance that the hard drive dies or something. So how many days do I expect until the hard drive dies?

Hard drives are usually designed-- engineered to last, I don't know, five years or so. And then all your hard drives-- that's for spinning disks. SSDs are probably different. They depend on usage. So that's something you might care about. And that's why this is called mean time to failure.

It could be something more positive, like maybe you have an algorithm. You're doing randomized algorithms in the future, and you have an algorithm that either solves your problem with some probability 1 minus P. With probability P, it solves your problem. And with probability 1 minus P it says, I don't know. There are lots of algorithms like this out there.

How many times do I have to run this algorithm until it actually gives me an answer? So I can run it, and it gives-- it might say, I don't know. But if I run it again, it might then solve the problem, depending on the random coin flips that it does.

So then it's mean time to success. How many times do I have to flip this coin, do I have to run this algorithm, given some probability of it succeeding-- that's P. How many times do I have to keep going until I get a successful answer? Lots of other examples. We will get to them in a moment.

Now, our traditional approach to solving tough problems like this is to draw a tree. In this case, the tree is infinite. So that's going to be a little harder to draw. We have P versus 1 minus P. Whenever we get a heads, we stop. But this tree keeps going off to the right. So that's-- maybe let's not look at the infinite tree. Let's use some formulas.

So I'm going to call this number-- the number of flips I'm going to call F. And then we want to compute-- remember, F is a function. So there are various outcomes here which correspond to the executions-- the various numbers of flips. It's going to be tails, tails, tails, tails, tails, tails I times and then heads. Those are the various outcomes that could happen because we stop whenever we get a heads. So it could be 0 tails and then heads, 1 tails then heads, and so on. So that's our sample space.

And then F is a function that maps that sample space to count how many total flips were there-- heads and tails? So it's always going to be at least one. And then expect-- so F is a function-- can take on some integer value between 0 and 1, I guess, and infinity.

And now we're going to reduce this to a single number via expectation. What is the average number of flips that we expect to take? So that's expectation of F.

Now, we will use our latest and greatest formula-- this one. Let me memorize it for a second. It's just sum X greater than I. Already forgotten it. It's too far from over there to over here. Was it probability F is greater than I, I hope. Yes.

So this is that CDF. And I claim this is not too hard to figure out using what we know. Because we're assuming each flip is independent from subsequent flips, like the flips of this coin are mutually independent. It's not some magical coin that always comes up heads or always comes up tails or something. Well, that would be represented by the P. But there's no correlation between the different flips.

So for there to be at least greater than I flips before I get a heads, that means that the first I flips are tails. These are equivalent events. This F greater than I is the same as first I flips are tails. And then after that, I don't know, I might get a heads right away. Then F is I plus 1. Or it might take longer. But F being greater than I is exactly this situation.

So what is the probability that the first I flips are tails? 1 minus P to the I. I saw you whispering it. I feel like-- yeah, cool. Now, the hard part is working out this sum, or remembering what this sum is. This is a geometric sum. So 1 minus P to the I because this is the chance of getting a tails. And then they're each independent, so we just take the product-- via product rule.

This is a geometric thing. Sum of X to the I from I equals 0 to infinity is 1 over 1 minus X. So this is 1 over 1 minus X here is 1 minus p. So 1 minus 1 minus p, that's also known as p. So this whole thing is 1 over p. Amazingly simple formula for the mean time to failure. This is why mean time to failure is such a famous example because it has such a beautiful solution.

So if I'm flipping a fair coin, 50/50. P is a half. I expect to have to flip it twice to get a heads or a tails. Now, reality is not equal to the average, sadly, so I had to flip seven times to get a heads or tails. But if I was gaming for heads, it would be very fast. It happened in the very first step. So that's the expectation.

It doesn't tell you everything about the distribution. This distribution has a name Alluded to by this geometric series. This was the CDF. We call this the geometric distribution. It comes up somewhat frequently in real life distributions.

If I were to write down the PMF, it's 1 minus p to the I times p. This is the chance that I get I tails and then a heads. That's-- well, OK. It's either that or probability-- the way I defined it before, it was F equals I plus 1.

So one of these is true. Pick one. These are both called the geometric distribution. If you look at the Wikipedia definition of geometric distribution, it says there are two distributions called geometric distribution-- this one and this one. So pick your poison. Both are called that.

So here is an application, another application of mean time to failure. I don't know about you, but I like collecting things. Everyone likes collecting something. To me, the epitome of collecting things is gacha. Oh, that's the video game version of *Gacha,* where you're collecting characters in a video game.

Please do not play these games. They're annoyingly addictive-- I spend most of my nights these days. But gacha is named after this notion of-- you can see it in the top right there-- is a gashapon, which is like these little-- if you grew up in the US, you have vending machines that give you candy.

Well, these are like the really cool vending machines in Japan. This was a Gachapon festival. And you pick your favorite weird thing that you can get in a little ball out of this machine. You put in your 200 yen, and you get a randomly selected ball from those sets of weird toys.

So this is like a toy delivery-- or not delivery service. It's a toy random selection service. And if you want to collect all the different things on this list, how many times do you have to do it? This is a problem called coupon collectors, or the Gotcha problem, I would say. But we'll solve it in general in recitation tomorrow. But I'll solve a very simple version here, which is a nice application of this-- of mean time to failure.

Let's solve the special case where there are only two toys. gacha-- gacha is the sound that when you crank the thing, it's like, gacha, gacha, gacha, puh, is when the ball drops down-- gacha with two toys or characters or whatever you're collecting. And so you get a random choice. Let's say it's 50/50 between the two.

So in your first pull is what we usually call these, at least in the video game context, our first crank in the physical context, you get one of the toys. So now it's really just, how many steps do I need until I get the second toy? Two.

This is the coin-flipping problem. So how many more pulls until you get a distinct toy? And in this case, it's going to be 50/50 each pull. And so you expect two more, which is a total of three.

Now, when you have k toys you want to collect, it's annoying. Because in the beginning, it's really easy to get lots of distinct toys because you're very likely. But when you're trying to collect the last toy, it's very unlikely that you pull the last toy because there's only like a 1 in k chance that you get it. So you end up with some harmonic series, and you get a log factor. You'll see that tomorrow.

OK. Let's do a little more theory and then some more examples. Where should I go? Let's erase gambling. I feel like I should have a disclaimer. Please don't gamble. It's not good for you. Please just simulate gambling. That's better.

I was writing a simulation of this game this morning to see what were my chances of actually winning despite the odds being against my favor? And with 10 rounds, I had a 20% chance of winning. So luckily, the demo went in my favor. But there was a decent chance that I could have won even though the expectation was against me. So I guess.

If there is one thing in this class that you remember about probability, I suggest linearity of expectation. It is simultaneously the coolest thing, in my opinion, about probability and the most useful thing for solving problems. So what is it? The expectation function is linear. Let me formalize.

Suppose you have two random variables, X and Y. And you take their sum, and then you take their expectation. It turns out that equals the sum of the expectations. Or if you take a random variable and you multiply it by a constant, you can pull that constant out.

And so, more generally, this is the full linearity form. If I have a-- where am I here-- sum I equals 1 to n of Ci times Xi, this is the sum I equals 1 to n of Ci times the expected value of Xi.

It takes almost longer to write it down than to prove it. I mean, it's actually not hard to prove if you stare at this definition enough. It's because this is a linear function. If you look at how it depends on the X's, it's a linear function in X.

You've got some constant weights. I mean, they depend on your probability setup. But if you mess around with X and don't mess around with your sample space or its probabilities, then this is just a sum of constants times X applied to things.

So, in particular, if I replace X here with X plus Y, and I just jam that into the sum, I can split this into two sums. And lo and behold, I get expectation of X plus expectation of Y. And these other things follow in the same way. So I won't write down a proof. There's one in the notes if you want to look.

So it's almost an obvious thing once I tell it to you. But it's extremely powerful, mainly because it has no conditions. You don't need that the Xi's are independent. You don't need that they're disjoint. You don't need anything except that they are functions over the same sample space. I guess that's the one assumption you need.

Compare this with-- remember, there's the product rule. This only works if A and B are independent. Or there's the sum rule. We use these all the time. But you have to be very careful when you use them. You have to check. Are those really independent? For the sum rule it only works if A and B are disjoint.

So when we do law of total probability, we're using this all the time. But you have to be very careful that you use it correctly. Whereas, linearity of expectation, you don't need to be careful other than making sure you define your random variables well.

Let's do an example-- common one in games, gambling games and other games, dice games. Let's say we roll 2 fair two-side-- or sorry-- six-sided dice. What is the expected value of the sum?

Well, the expected value of the sum is the expected value of the individual parts. Let me rewrite this as a-- instead of words, let's use some notation. Let's say D1 is the random variable that just gives me what is the value of the first die. And D2 is the random variable that gives me what's the second value.

The sample space here is a pair, an ordered pair, of the first die, the second die. So this extracts the first part, and this extracts the second part. The expected value of this sum by linearity of expectation is the sum of the expectations. And we already did this example. This was 3.5, 3.5. So two total is 7. Easy.

And what's cool is this didn't even need that the dice were independent. You could glue the dice together so that they always roll together. As long as that doesn't change the probabilities of the outcomes, which may be a little questionable with gluing.

But imagine, I don't know, they're there quantumly entangled. And however I roll this die, the other one comes out the same way, this will still be true. It changes the distribution, but it does not change the expectation, which is cool. So we just needed the probabilities of each outcome were equal to get the 3.5. Question?

**AUDIENCE:** [INAUDIBLE].

**ERIK DEMAINE:** Gluing dice together. I'm imagining-- like if I take two dice. I glue them together. Then when I roll that unit, however the first die comes up, it forces the other one to come up in a particular orientation. So they're literally tied together so that the-- D1 determines D2 if they're glued together. But this formula applies in that situation, provided it's still equally likely that you get 1, 2, 3, 4, 5, or 6 for each die individually.

It's maybe a little clearer with coins. If I glue two coins together, then whichever one comes up heads, the other one also comes up heads or comes up tails, depending if I glue the head sides together or heads to tails.

So I want to do more cool examples. And they illustrate how to use linearity of expectation with a-- the key to using linearity of expectation is this technique. If you want to compute the expectation of some random variable, decompose that random variable into a sum of other random variables, possibly indicator random variables.

You see it here already. I mean, we said, what's the expectation of the sum? Well, let's decompose that into the first die and the second die. And then the sum is the first one plus the second one. And then we could split it up using linearity of expectation. And then compute each of these individually instead of having to look at the giant sample space, which is the pairs of die rolls.

So this is the technique we're going to use now several times, whatever remains of our time. So first example.

So previously, we analyzed how many times do I flip a biased coin until I get a heads? Now, let's suppose I want two heads-- natural generalization. This seems quite a bit harder because you keep flipping. And then at some point, you get the first heads. You keep flipping. At some point, you get the second heads.

Well, that sounds kind of like two steps. There's how many times do I flip until I get the first head? And then there's how many more times I have to flip until I get the second heads? Both of those problems sound like mean time to failure. So let's write that out carefully with random variables.

So let's call this number of flips F. Then F equals F1 plus F2, where F1 is the number of flips from the beginning until we get the first head. And F2 is the number of flips starting after the first head until the second head.

So to be precise, the sample space here is strings of H's and T's, where there's exactly two H's, and the last character is an H. Because I keep flipping heads, tails, heads, tails, heads, tails until I get the second H, then I stop. So that's my sample space.

And F1 is just counting how many flips are there up to including the first head? And F2 is counting how many flips are there starting right after the first head and counting up to and including the second head, which is the end of the string? So together, F1 and F2 count the total number of flips.

But crucially, F1 and F2, if you think about it, are exactly the mean time to failure problems because they were counting-- I mean, F1 is most obvious. It's how many flips do I do until the first head?

So we know here the expected value of F Is the sum of the expectations. F1 is almost literally the definition of mean time to failure. So it's 1 over P in expectation. F2, if you think about it a little bit, this is the same experiment, but starting after the first heads. And because all the flips are independent, it doesn't matter what happened before now. It's still going to be the same situation for F2. And so the expectation is 2 over P. Almost as simple.

And, in general, if I want K heads, then the expected number of flips I need to do is K over P by linearity of expectation. Beautiful. So simple.

Let's do another coin-flipping example.

So here, I stopped the experiment when I got a desired number of heads. What if I just flip a coin n times? No matter what, I always flip n times. How many heads do I expect to get? That's kind of the dual problem.

What do se want to do? Let's do biased again. n biased coins-- again, independent.

Let H be the number of heads in those n coin flips. We can rewrite this as a sum of n indicator random variables, where Hi is indicator random variable for the ith coin coming up heads.

So expected value of H by linearity is the sum of the expected values of the Hi's. But the coins are all identical because they're independent, and they have the same bias. I didn't mention it, but we're supposing, again, probability of heads is p.

So what's the-- so this is just going to be n times the expectation of Hi, because they all behave the same. And what's the expectation of Hi? Well, we worked that out over here. If you have an indicator random variable, the expectation of Hi is the probability that it equals 1. So the probability of the event, which here was the probability of heads. So that's just p. So this is n times probability Hi equals 1, which is n times p. OK, maybe obvious, but--

Yeah. We didn't even need that these were independent. That's where it's less obvious. It just-- all we need is that for each coin, the probability of it coming up heads is p. But it could be all the coins always flip the same. They're always all heads or all tails, but heads with probability p.

Or it could be either independent. These are different setups. But the expected number of heads in all cases will be n times p by linearity of expectation. So that's pretty cool.

You can do a similar trick for-- did I erase it-- the hearts problem. I think I erased it. Yeah. I'll just tell you briefly.

So with the hearts, you can say the number of hearts is a sum of 13 different indicator random variables. Part i is the ith heart card chosen. Again, this is a setup where we have 52 cards. We choose three without replacement, and I want to know how many hearts there are.

So this is number of hearts. And this is just did I choose the ith heart card in this setup? So, of course, the number of hearts is exactly the sum of 1 for every heart card that I chose. So one-- this is like Ace, and this is King. And I've got all the 1's in between.

So this is-- whereas, here was the ith coin flip or whatever. Here's the ith heart that we potentially choose. And the probability of choosing a given heart is easy. It's-- it's not so easy. It's some small expression.

It could be you choose it in the first step. Or you don't choose it in the first step, but then you choose it in the second step. Or you don't choose it in the second step, and then you'd have to choose it in the third step. Those are the three ways you can get it.

And conveniently, all the hard numbers here cancel. That's exciting. And you get 2 over 52? 3 over 52. I can count. So you add them up, 13 of them. So you get 13 times 3 over 52, and that's 3/4. So kind of nice. No binomial coefficients required.

One more pair of examples.

Yeah.

I have in my notes that this is a situation that happens in*All of Us are Dead,* which is a great zombie high school TV show. I imagine it just happens in lots of schools. But I never went to school, so I don't know. You can tell me whether this is the case. Probably less so in the US.

Maybe less so when you're older. Maybe when you're young and you have a phone, you have to submit your phone when you arrive at school. And here's the weird part. And this is maybe only in a zombie scenario. At the end of the day, phones are returned uniformly at random. [LAUGHS] I hope this didn't happen to you. How many people get the right phone? Natural question.

**AUDIENCE:** The natural question is, do I get my phone?

**ERIK DEMAINE:** You care about whether you get your phone? That's also true. That's a closely-- we're going to have to solve that problem, in fact. So we're assuming here each-- there's a one-to-one correspondence between phones and people. No co-ownership.

So the number of kids that get their own phone we can write it as a sum. Over all kids, we had n kids. So k for kids. 1 up to n of the probability that kid k gets their phone.

OK. I'm just writing this because I've preworked out these examples. Of course, when you're solving them, you'll probably know in this part of the class that you should use this technique. Always-- if you're asked for an expectation, probably you might need to decompose the random variable into sums of other random variables.

You might not have to. This hard problem, we solved in two different ways. But if you can, it's going to be easier. So I encourage you to try to do this. But how do you decompose into sum of random variables? I mean, that's divine inspiration.

You just got to try a bunch of things until you find something that works, that gives you-- think about what you know how to compute, and then try to build it up. Or think about how to decompose, top down, whatever.

But as Brinmore said, what I care about is whether I get my phone, and I'm kid seven. I don't know. What's the probability that kid seven gets their phone, or kid one or whatever.

It turns out they're all the same because-- oh. Yeah. I did two steps here. This is how many kids get their phone. We can call this X. And this is going to be the sum-- sorry. I skipped way too many steps here.

The expectation of x is going to be this. But the number of kids that get their phone is going to be equal to the sum of Xk. That equals 1 to n, where Xk is an indicator random variable for whether kid k got their phone.

OK. This is important because this was a random variable. So if I'm going to write equals, I should have some random variables over there. When I wrote a probability, that means I'm working with a number. And so that means the best I'm hoping for is the expectation.

So this first equation I wrote is wrong. Please do not do that in your homework. It's easy to do, as you can see. But you should first set up the random variables, which are functions. Decompose them into sums of other functions. And then say, OK, let's compute the expectation. Use linearity of expectation.

This is equal to by linearity, the sum of the expectations of the Xi's. And then by the properties of indicator random variables, the expected value of an indicator random variable is the probability of that event.

So then, what's the probability that kid k gets their phone? Well, how many different distributions are there of the n phones? Help me out.

**AUDIENCE:** n.

**ERIK DEMAINE:** n. Great answer. [LAUGHS] n different ways-- there's n different possibilities for the first kid to get their phone back. And then--

**AUDIENCE:** [INAUDIBLE].

**ERIK DEMAINE:** Hmm?

**STUDENT:** A factorial.

**ERIK DEMAINE:** A factorial. Yeah. There's n factorial different distributions of the phones.

**AUDIENCE:** That's what I said.

**ERIK DEMAINE:** Oh. Oh, you said n really loud, which is n factorial. Thank you. Bad pun. So that's the denominator. Those are all the possible rearrangements of the phones to the kids. n, very emphatically.

And then how many are there where kid k gets their phone? Well this is easy, right? This is just saying, well, kid k gets their phone. And then among all the other n minus 1 phones, they're whatever they are. How many different ways are there to distribute those phones? n minus 1.

[LAUGHTER]

Hopefully, I didn't max out the mic. OK. N minus 1 factorial over n factorial. Oh, my gosh. Such big factorials. But basically, everything cancels. And this is 1 over n.

1 over n summed n times-- that's 1. Ha, ha, ha. One kid expects to get their phone back. Hopefully, it's Brinmore. [LAUGHS] Who is the-- I mean, it doesn't mean that one person will get their phone back. Maybe none of them do.

Here's a related game. Suppose n diners put their phones-- put their phone singular on a turntable, sometimes called a Lazy Susan, for distributing food, if you're eating family style. And then you spin the table, spin uniformly at random.

And then you take the phone in front of you. Let's assume that this is a discrete table. So when you spin it, there's only n different rotations that could come out. And so there's always a phone immediately in front of you. This is a different way to be terrible with phones.

In this one, anything could happen. It could be no one gets their phone back. That's maybe-- it's not quite what you expect. You expect one person to get their phone back. Or it could be everyone gets exactly their phone back. If you do it randomly, anything's possible. It's just a 1 over n factorial chance that any particular outcome happens.

Here, it's actually guaranteed. Either everyone gets their phone back or no one does. Because if I rotate by 0, then everyone gets their phone back. If I rotate by any other amount, then everyone gets a different phone.

So it's a different distribution. It's a different probability setup. But the expectation is the same. Expected number of matching people phones, whatever-- matching phones. It's still 1.

Here, it's really easy to compute because there are only n different outcomes. And in one of those scenarios, you get-- so 1 over n chance everyone gets their phone back. And in all other scenarios, no one gets their phone back. So that disappears. You get 1 over n times n. You get 1.

That's by the definition of expectation, how to compute this. But you can also set it up in exactly the same way. You can say, well, you could sum over all the diners and check what's the probability that a particular diner gets their phone back. But now you don't have to do this n minus 1 factorial over n factorial. It's just 1 over n by the setup. That was this term here. So this is also the probability of Xi. And so you can also do it with linearity of expectation.

So this is a funny scenario where I have two rather different setups for redistributing phones. They have very different behaviors. But in expectation, they're the same. So expectation doesn't tell you everything about the distribution. And in the next couple of lectures, we'll talk more about understanding the rest of the distribution beyond just the mean. All right.