PROFESSOR: We ask about averages all the time. And in the context of random variables, averages get abstracted into a lovely concept called the expectation of the random variable. Let's begin with a motivating example which, as is often the case, will come from gambling.

So there's a game that's actually played in casinos called Carnival Dice where you have three dice, and the way you play is you pick your favorite number from 1 to 6, whatever it happens to be. And then you roll the three dice. The dice are assumed to be fair, so each one of them has a one in six chance of coming up with any given number. And then the payoff goes as follows.

For every match of your favorite number, you get $1.00. And if none of your favorite-- if none of the die show your favorite number, then you lose $1.00. OK. Let's do an example.

Suppose your favorite number was five. You announce that to the house, or the dealer, and then the dice start rolling. Now if your roll happened to come up with the numbers two, three, and four, well, there's no fives there, so you've lost $1.00. On the other hand, if your rolls came out five, four, six, there's one five, you've one $1.00. If it came out five, four, five, there's two fives, you've won $1.00. And if it was all fives, you've actually won $3.00.

Now real carnival dice is often played where you either win or lose $1.00 depending on whether there's any match at all, but we're playing a more generous game where, if you double match, you get $2.00. If you triple match, you get $3.00. So the basic question about this is, is this a fair game. Is this worth playing, and how can we think about that? Well, we're going to think about it probabilistically.

So let's think about the probability of rolling no fives. If five is my favorite number, what's the probability that I roll none of them? Well, there's a five out of six chance that I don't roll a five on the first die, and on the second die and on the third die. And since the die rolls are assumed to be independent, the dies are independent, the probability of no fives is 5/6 to the third, which comes out to be 125/216. I'm writing this out because we're going to put all the numbers over 216 to make them easier to compare.

OK. What's the probability of one five? Well, the probability of any single sequence of die rolls with a single five is 5/6 of no five times 5/6 of no five times 1/6 of one five. And there are 3 choose 1 possible sequences of dice rolls with one five, and the others non-fives. Likewise, for

two fives, there's 3 choose 2 times 5/6 to the 1, which is one way of choosing the place that does not have a five. And 1/6 times 1/6, which is the probability of getting fives in the other places. I didn't say that well, but you can get it straight. OK.

The probability of three fives is the probability of 1/6 of getting a five on the first die, 1/6 of getting a five on the second die, 1/6 of getting a five on the third die. It's simply 1/6 cubed. OK, so we can easily calculate these probabilities. This is a familiar exercise. Let's put them in a chart.

So what we've figured out is that 0 matches has a probability of 125 over 216. And in that case, I lose $1.00. One match turns out to have a probability of 75 out of 216, and I win $1.00. Two matches is 15 out of 216, I win $2.00. And three matches, there's one chance in 216 that I win the $3.00.

So now I can ask about what do I expect to win. Suppose I play 216 games, and the games split exactly according to these probabilities. Then what I would expect is that I would wind up with 0 matches about 125 times. That was the probability of there being no matches. It was 125/216. So if I played 216 games, I expect about 125 are going to-- I'm going to win nothing. Or, I'm going to get no matches, which actually means I'll lose $1.00 on each.

One match I expect about 75 times. 2 matches, 15 times. 3 matches, once. So my average win is going to be 125 times minus 1, 75 times 1, 15 times 2 plus 1 times 3 divided by 216. So these numbers on the top were how the 216 rolls split among my choices of losing $1.00, winning $1.00, winning $2.00, and winning $3.00. And it comes out to be slightly negative. It's actually minus $0.08-- minus 17/216 of $1.00, which is about minus $0.08. So I'm losing, on the average, $0.08 per roll. This is not a fair game. It's really biased against me. And if I keep playing long enough, I'm going to find that I average out a kind of steady loss of about $0.08 a play.

So we would summarize this by saying that you expect to lose $0.08, meaning that your average loss is $0.08 and you expect that that's going to be the phenomenon that comes up if you keep playing the game repeatedly and repeatedly. It's important to notice, of course, you never actually lose $0.08 on any single play. So what you-- this notion of your expecting to lose $0.08, it never happens. It's just your average loss. Notice every single play you're either going to lose $1, win $1, win $2, win $3. There's no $0.08 at all showing up. OK.

So now let's abstract the expected value of a random variable R. So a random variable is this thing that probabilistically takes on different values with different probabilities. And its expected value is defined to be its average value where the different values are weighted by their probabilities. We can write this out as a precise formula. The expectation of a random variable R is defined to be the sum over all its possible values-- it doesn't indicate what the summation is, but that's over all possible values v-- of v times the probability that v comes up, the probability that R equals v. So this is the basic definition of the expected value of a random variable.

Now let me mention here that this sum works because since we're assuming accountable sample space, R is defined on only countably many outcomes, which means it can only take countably many values. So this is a countable sum over all the possible values that R takes, because there are only countably many of them. And what we've just concluded, then, is the expected win in the carnival dice game is minus 17/216. Check this formal definition of the expectation of a random variable versus the random variable defined to be how much you win on a given play of carnival dice, and it comes out to be that average. Minus 17/216.

Now there's a technical result that's useful in some proofs that says that there's another way to get the expectation. The expectation can also be expressed by saying it's the sum over all the possible outcomes in the sample space-- S is the sample space-- of the value of the random variable at that outcome times the probability of that outcome. So this is another alternative definition of compared to saying it's the sum over all the values times the probability of that value. Here, it's the sum over all the outcomes of the value of the random variable, the outcome times the probability of the outcome. It's not entirely obvious that those two definitions are equivalent.

This form of the definition turns out to be technically helpful in some proofs, although outside of proofs you don't use it so much in applications. But it's not a bad exercise to prove this equivalence. So I'm going to walk you through it. But if it's boring-- it's kind of a boring series of equations on slides, and you're welcome to skip past it. It is a derivation that I expect you to be able to carry out. So let's just carry out this derivation. I'm going to prove that the expectation is equal to the sum over all the outcomes of the value of the random variable at the outcome times the probability of the outcome. And let's prove it.

In order to prove it, let's begin with one little remark that's useful. Remember that this notation R equals v describes the event that the random variable takes the value v, which by definition

is an event is the set of outcomes where this property holds. So it's the set of outcomes omega where R of omega is equal to v. So let's just remember that, that brackets R equals v is the event that R is equal to v, meaning the set of outcomes where that's true. So what that tells us in particular is that the probability of R equals v is, by definition, the sum of the probabilities of the outcomes in the event. So it's the sum over all those outcomes.

Now let's go back to the original definition of the expectation of R. The original definition is-- and the standard one is-- it's the sum over all the values of the value times the probability that the random variable is equal to the value. Now on the previous slide, we just had a formula for the probability that R is equal to v. It's simply the sum over all the outcomes of where R is equal to v, of the probability of that outcome. So I can replace that term by the sum over all the outcomes of the probability of the outcome.

OK. So I'm trying to head towards an expressions that's only outcomes, which is kind of the top-level strategy here. So the first thing I did was I got rid of that probability of v and replaced it by the sum of all these probabilities-- of the probabilities of all the outcomes where R is v. Well, next step is I'm going to just distribute the v over the inner sum. And I get that this thing is equal to the sum, again, over all those outcomes in R equals v of v times the probability of the outcome. But look, these outcomes are the outcomes where R is equal to v. So I could replace that v by R of omega. That one slipped sideways a little bit, so let's watch that.

This v is simply going to become an R of omega. I'm still [INAUDIBLE] over the same set of omegas, but now I've gotten rid of pretty much everything but the omegas. So I've got this inner sum of over all possible omegas in R of v of R of omega times the probability of omega. And I'm summing over all possible v. But if I'm summing over all possible v and then all possible outcomes where R is equal to v, I wind up summing over all possible outcomes. And so I've finished the proof that the expectation of R is equal to the sum over all the outcomes of R of omega times the probability of omega.

Now I'd never do a proof like this in a lecture, because I think watching a lecturer write stuff on the board, a whole bunch of symbols and manipulating equations, is really insipid and boring. Most people can't follow it anyway. I'm hoping that in the video, where you can go back if you wish and replay it and watch it more slowly, or at your own speed, the derivation will be of some value to you. But let's step back a little bit and notice some top-level technical things that we never really paid attention to in the process of doing this manipulative proof.

So the top-level observation, first of all, is that this proof, like many proofs in basic foundations of probability theory and random variables, in particular, involves taking sums and rearranging the terms in the sums a lot. So the first question is, why sums? Remember here we were summing over all the possible variables, all the possible values of some random variable. Why is that a sum? Well it's a sum because we were assuming that the sample space was countable. There were only a countable number of values R of omega 0, R of omega 1, R of omega n, and so on. And so we can be sure that the sum over all the possible values of the random variable is a countable sum. It's a sum, and we don't have to worry about integrals, which is the main technical reason why we're doing discrete probability and assuming that there are only a countable number of outcomes.

There's a second very important technicality that's worth mentioning. All the proofs involved rearranging terms in sums freely and without care. But that means that we're implicitly assuming that it's safe to do that, and that, in particular, that the defining sum for expectations needs to be absolutely convergent. And all of these sums need to be absolutely convergent in order for that kind of rearrangement to make sense. So remember that absolute convergence means that the sum of the absolute values of all the terms in the sum converge.

So if we look at this definition of expectation, it said it was the sum over all the values in the range. We know that's a countable sum of the value times the probability that R was equal to that value. But the very definition never specified the order in which these terms, v times probability R equals v, got added up. It better not make a difference. So we're implicitly assuming absolute convergence of this sum in order for the expectation to even be well-defined.

As a matter of fact, the terrible pathology that happens-- and you may have learned about this in first-time calculus, and we actually have a problem in the text about it-- is that you can have sums like this, that are not absolutely convergent, and then you pick your favorite value and I can rearrange the terms in the sum so that it converges to that value. When you're dealing with non-absolute value sums, rearranging is a no-no. The sum depends crucially on the ordering in which the terms appear, and all of the reasoning and probability theory would be inapplicable. So we are implicitly assuming that all of these sums are absolutely convergent.

Just to get some vocabulary in place, the expected value is also known as the mean value, or the mean, or the expectation of the random variable. Now let's connect up expectations with averages in a more precise way. We said that the expectation was kind of an abstraction of

averages, but it's more intimately connected to averages than that, even. Let's take an example where suppose you have a pile of graded exams, and you pick one at random. Let's let S be the score of the randomly picked exam. So I'm turning this process, this random process of picking an exam from the pile, is defining a random variable, S, where by definition of picking one at random, I mean uniformly. So S is actually not a uniform random variable, but I'm picking the exams with equal probability. And then they have different scores, so the outcomes are of uniform probability. But S is not, because there might be a lot of outcomes, a lot of exams with the same score. All right. S is a random variable defined by this process of picking a random exam.

And then you can just check that the expectation of S now exactly equals the average exam score, which is the typical thing that students want to know when the exam is done, what's the average score. Actually, the average score is often less informative than the median score, the middle score, but people somehow rather always want to know about the averages. The reason why the average may not be so informative is because-- well, it has some weird properties that I'll illustrate in a second. But the point here of what we did where we took the-- we got at the average score on the exam by defining a random variable based on picking a random exam. So that's a general process. We can estimate averages in some population of things by estimating the expectations of random variables based on picking random elements from the thing that we're averaging over. That's called sampling, and it's a basic idea of probability theory that we're going to be able to get a hold of averages by abstracting the calculation of an average into taking-- defining a random variable and calculating its expectation.

Let's look at an example. It's obviously impossible for all the exams to be above average, because then the average would be above average. That's absurd. So if you translate that into a formal statement about expectations, it translates directly-- by the way, I don't know how many of you listen to the Prairie Home Companion, but one of the sign-offs there is at the town of Lake Woebegone in Wisconsin, where all the children are above average. Well, t'ain't possible. That translates into this technical statement that the probability that a random variable is greater than its expected value is less than 1. It can't always be greater than its expected value. That's absurd.

On the other hand, it's actually possible for the probability that the random variable is bigger than its expected value to be as close to 1 as you want. And one way to think about that is

that, for example, almost everyone has an above average number of fingers. Think about that for a second. Almost everyone has an above average number of fingers. Well, the explanation is really simple. It's simply because amputation is much more common than polydactylism. And if you can't understand what I just said, look it up and think about it.