PROFESSOR: The law of large numbers gives a precise formal statement of the basic intuitive idea that underlies probability theory, and in particular, our interest in random variables and their expectations-- their means. So let's begin by asking what the mean means. Why are we so interested in it, for example.

If you roll a fair die, with faces one through six, the mean value, its expected value is 3 and 1/2. And you'll never roll 3 and 1/2 because there is no 3 and 1/2 face. So why do we care about what this mean is if we're never going to roll it? And the answer is that we believe that after many rolls, if we take the average of the numbers that show on the dice, that average is going to be near the mean. The mean is going to be near 3 and 1/2.

Let's look at an even more basic example. If it's a fair die, the probability of rolling a six, as with any other number, is one sixth. And the very meaning of the fact that the probability of rolling a six is one sixth is that we expect that if you roll a lot of times, if you roll about n times, the fraction of sixes is going to be around n/6.

The fraction of six is going to be about one sixth. Of n rolls, you'll get about n/6 6s. That's almost the definition-- or the intuitive idea behind what we mean when we assign probability to some outcome. It's that if we did it repeatedly, the fraction of times that it came up would be equal to its probability-- or at least closely equal to it in the long run.

So let's look at what Jacob Bernoulli, who is the discoverer of the law of large numbers, had to say on the subject. He was born in 1659 and died in 1705. And his famous book, *The Art of Guessing-- Ars Conjectandi--* was actually published posthumously by his cousin. And Bernoulli says, "Even the stupidest man-- by some instinct of nature per se and by no previous instruction-- this is truly amazing-- knows for sure that the more observations that are taken, the less the danger will be of straying from the mark."

All right. What does he mean? Well, it's what we said a moment ago. If you roll the fair die n times and the probability of a roll is a sixth, then the average number of sixes, which is the number of sixes rolled divided by n, we believe intuitively that that number is going to approach one sixth as n approaches infinity.

That's what Bernoulli is saying, that everybody understands that they intuitively are sure of it. And who knows how they figured that out. But that's what everyone thinks. And he might be

right.

Now of course, when you're doing this experiment of rolling n times and counting the number of sixes and seeing if the fraction is close to a sixth, you might be unlucky. And it's possible that you'd get an average that actually was way off one sixth. But that would be unlucky. And the question is, how unlikely is it to be that you'd get a fraction of sixes that wasn't really close to a sixth? And with the law of large numbers is getting a grip on that, and in fact, subsequently, we'll get a more even quantitative grip on it, which will be crucial for applications in sampling and hypothesis testing. But let's go on.

So let's look at some actual numbers which I calculated. And if you roll a die n times, where n is 6, 60, 600, 1,200, 3,000 or 6,000, the probability that you're going to be within 10% of the expected number of sixes is given here. So it turns out, of course, that in order to be within 10-- if you're going to roll six times, the only way to be within 10% of the one expected six that you should roll, is to roll exactly one six in six tries. And the probability of that is about 40%, 0.4 as you can check yourself easily.

Then it turns out that if you roll 60 times, the probability of being-- the expected number in 60 rolls is going to be 10. So the probability of there being within 10% of 10, or nine to 11 sixes is 0.26. And likewise, the probability of there being within 10% of 100, which is the expected number of sixes when you roll 600 times, is 0.72.

And so on until finally the probability of being within 10% of 1,000, which is the expected number when you roll 6,000 times, that is between 900 and 1,100 sixes in 6,000 rolls, is 0.999-- triple nines. In fact, it's a little bit bigger. So it's really only about one chance in 1,000 that your number of sixes won't fall in that interval, within 10% of the expected number.

Well, suppose I ask for a tighter tolerance and I'd like to know what's the probability of being within 5%. Well first of all, notice of course, that as the number of rolls get larger, the probability of being in this given interval is getting higher and higher, which is what Bernoulli said and what we intuitively believe. The more rolls, the more likely you are to be close to what you expect.

If you tighten the tolerance, of course, then the probabilities wind up getting smaller that you'll do so well. So if you want to be within 5% of the average in six rolls, it means you still have to roll exactly one sixth, which means the probability is still 0.4. But if you're trying to be within 5%

of the expected number 10 and 60 rolls, meaning between five and 15, that probability is only 0.14 compared to the probability of 0.26 of being within 10%.

And if we jump down here, say, to 3,000 rolls, the probability of being within 10% of 500, which is the expected number in 3,000 rolls, within 10% is 0.98. But within being within 5% of 500, it's 0.78, or about a little over 3/4. So what does that tell us? Well, it means that if you rolled 3,000 times and you did not get within 10% of the expected number 500, that is you did not get in the interval between 450 and 556 sixes, you can be 98% confident that your die is loaded. It's not weighted one sixth to show a six.

And similarly, if you did not get within 425 and 525 sixes in 3,000 rolls, you can be 78% sure that your die is loaded. And this is exactly why the law of large numbers is so important to us because it allows us to do an experiment and then assess whether what we think is true is verified by the outcome that we got in this experiment.

All right. Let's go on to see what else Bernoulli was concerned with in his time. "It certainly remains to be inquired whether after the number of observations has been increased, the probability of obtaining the true ratio finally exceeds any given degree of certainty, or whether the problem has, so to speak, its own asymptote-- that is, whether some degree of certainty is given, which one can never exceed."

Now, that's 17th century English that may be a little bit hard to parse. So let's translate it into math language. What is it that Bernoulli is asking?

So what Bernoulli means is that he wants to think about taking a random variable R with an expectation or mean of mu. And he wants to make n trial observations of R and take the average of those observations and see how close they are to mu.

All right. What is making n trial observations mean? Well, formally, the way we're going to capture it is we're going to think of having a bunch of mutually independent, identically distributed random variables R1 through Rn. This phrase "independent, identically distributed" comes up so often that there's a standard abbreviation i.i.d random variables. So we're going to have n of them.

And think of those as being the n observations that we make of a given random variable R. So R1 through Rn each have exactly the same distribution as R. And they're mutually independent. And again, since they have identical distributions, they all have the same mean,

mu, as the random variable R that we were trying to investigate. So we model n independent trials, repeated trials, by saying that we have n random variables that are i.i.d. OK.

Now, what Bernoulli's proposing is that you take the average of those n random variables. So you take the sum of R1, R2, up through Rn, and divide by n. That's the average value. Call that A sub n-- the average of the n observations or the n rolls.

And Bernoulli's question is, is this average probably close to the mean, mu, if n is big? What exactly does that mean? Probably close to mu means that the probability that the distance between the average and mu is less than or equal to delta. Is what? So delta is talking about how close you are. Delta is a parameter. We expect it's got to be positive.

Think of whatever "close" means to you. Does it mean 0.1? Does it mean 0.01? What amount would persuade you that the average was close to what it ought to be?

And we ask then, whether the distance between the average and the mean is close-- less than or equal to delta. And Bernoulli wants to know, what is the probability of that? And what it goes on to say is, "Therefore this is the problem which I now set forth and make known after I have pondered over it for 20 years. Both its novelty and its very great usefulness, coupled with its great difficulty, can exceed in weight and value all the remaining chapters of this thesis."

Now, Bernoulli was right on about the usefulness of this result, at least in its quantitative form. And at the time, it was really pretty difficult for him. It took him like 200 pages to complete his proof in *Ars Conjectandi.* Nowadays, we are going to do it in about a lecture worth of material. And you'll be seeing that in some subsequent video segment.

So that's what happens with 350 years to tune up a result. What took 200 pages then, now takes 10 or less pages. In fact, if it was really concise, it could be done in three pages.

All right. So again, coming back to Bernoulli's question. Bernoulli's question is, what is the probability that the distance between the average and the mean is less than or equal to delta as you take more and more tries, as n goes to infinity?

And Bernoulli's answer to the question is, that the probability is 1. That is, if you want to have a certain degree of certainty of being close to the mean, if you take enough trials you can be as certain as you want, that you'll be as close as you want. And that is called the weak law of large numbers. And it's one of the basic, transcendent rules and theorems of probability theory.

It's usually stated in the other way, as that the limit of the probability that the average is a distance away from the mean delta is zero. It's the probability that it's extremely unlikely. It can be as unlikely as you want to make it that it's more than any given tolerance from the mean, if you take a large enough number of trials.

Now, in this form, it's not yet really useful. This is a romantic qualitative limiting result. And to really use it, you need to know something or other about the rate at which it approaches the limit, which is what we're going to be seeing in a subsequent video. And in fact, the proof of this is going to follow easily from the Chebyshev inequality bound and variance properties when we go about trying to get the quantitative version that explains the rate at which the limit is approached.