**PROFESSOR:** Now we're ready to prove the law of large numbers and while we're at it, to get a quantitative version, which will be the basis for theory of sampling an estimation.

So let's remember that the law of large numbers says that if you have n independent, identically distributed random variables with mean mu and we let An b their average, then for every positive number delta, the probability that the average differs from the mean by more than delta goes to 0 as the number of trials increases. Remember, that means that if you tell me what you think close means and what you think very likely means, then I can guarantee that by doing enough trials, the likelihood that the mean will be that close to the average will be outside the tolerance is as small as you thought small should be.

And we're ready for the proof. But in the proof, there's one extra fact that we're going to use that we didn't explicitly mention, which is that not only are all of these random variables identically distributed and independent, but we're actually going to assume that they have a variance. Now, not every random variable has a finite variance, even if it has a finite mean. In fact, there are random variables that don't even have finite means, and we'll look at them on the last day of class.

But so we're going to explicitly assume that all of these random variables have the same variance, namely the standard deviation sigma squared, and we'll be using that fact in the proof. Now, the first question to ask is, what is the expectation of the average? And the expectation of the average is simply the expectation. Let's prove that.

The expectation of the average is by definition the expectation of the sum of the R's over n, and by additivity of expectation, that's the sum of the expectation of each of the R's over n. But each of them has expectation mu, so the numerator is n mu and the n's cancel. And sure enough, the average has the same expectation as each of the individual variables, each of the trials.

Now, that [INAUDIBLE] let's us apply the Chebyshev bound to the random variable An because now we know what its mean is, and its mean is independent of n. We can apply Chebyshev to the probability that the average of n trials differs from its mean by more than delta. And according to Chebyshev, that's bounded by the variance of the average divided by n squared.

So I will have proved the law of large numbers if I can prove that the limit as n approaches infinity of the variance goes to 0 because that means that the right-hand side will be going to 0 over delta squared, namely going to 0, which is what the law of large numbers says. So we've reduced the proof of the law of large numbers to proving that the variance goes to 0 as n approaches infinity, where An is the average of n identically distributed variables with common mean mu and standard deviation sigma.

Well, let's calculate the variance of An where An, again, is the average, the sum of the R's over n. And since we're assuming independence of the R's, the variance sum rule, and it just tells us that this is the sum of the variances.

Now, if we factor out the R-- 1 over n-- now, this is a factor of 1 over times n this sum. When we factor a constant out of the variance, it squares, so the denominator here becomes n squared, and that's critical. And the numerator is the sum of the n variances.

Now, each variance is sigma squared and we've got n of them, so we wind up with n sigma squared over n squared, which is, of course, equal to sigma squared over n. And sigma squared is a constant, and n is going to infinity. So sure enough, the right-hand side goes to 0 as n increases, which is all we needed to do to conclude the weak law of large numbers.

Now, if we go back and look at this proof, the only thing that it used about the R's was that they had the same mean, mu, and they actually had the same variance, sigma squared, and the variances added. That was the key step in the proof, that the variance of the sum of the R's was equal to the sum of the variances.

Now, additivity of variances only requires pairwise independence and didn't even require the hypothesis that they were mutually independent, and it didn't require the previous proof that we went through-- did not ever use the fact that the R's had the same distribution, that they may not be identically distributed. It was sufficient that they have the same mean, and we can summarize what really proved.

When we thought we were proving the law of large numbers, we actually proved a precise quantitative theorem that says that if R1 through Rn are pairwise independent random variables with the same finite mean mu and variant sigma squared and we let An be the average of those n variables, then the probability that the average differs from the mean by more than delta is less than or equal to this definite number that we derived, 1 over n times

sigma over delta squared, and this was what we previously proven. We thought we were just proving the law of large numbers. We actually got this much tighter quantitative theorem.

Now, what this tells us here is that now if you tell me what delta is and you tell me how small you want this to be, well-- I'm given sigma and you give me the delta that you specified, and if you tell me how small you want this to be-- I will know how big an n to choose. So this tells me how big a sample I need, how many tries I have to make in order to get the probability that the mean is within a specified tolerance delta as small as you specified, and that is our independent sampling theorem.

That's why it's called independent sampling, because we now know how big a sample is needed to estimate the mean of any random variable with any desired tolerance and any desired probability, where of course the various has to be finite, the tolerance has to be positive-- tolerance is delta-- and the probability has to be less than 1.