The purpose of this segment is to give you a little bit of the bigger picture.

We did discuss some inequalities, we did discuss convergence of the sample mean-- that's the weak law of large numbers-- and we did discuss a particular notion of convergence of random variables, convergence in probability.

How far can we take those topics?

Let's start with the issue of inequalities.

Here, one would like to obtain bounds and approximations on tail probabilities that are better than the Markov and Cherbyshev inequalities that we have seen.

This is indeed possible.

For example, there is a so-called Chernoff bound that takes the following form.

The Chernoff bound tells us that the probability that the sample mean is away from the true mean by at least a, where a is a positive number, this probability is bounded above by a function that falls exponentially with n and where the exponent depends on the particular number, a, that we are considering.

But in any case, this term in the exponent is a positive quantity.

Notice that this is much better, much stronger than what we obtained from the Cherbyshev inequality because in the Cherbyshev inequality, we only obtain an inequality for this probability that falls off as the rate of 1 over n.

So this falls much faster, and so it tells us that this probability is indeed much smaller than what the Cherbyshev inequality might predict.

However, this inequality requires some additional assumptions on the random variables involved.

Another type of approximation on this tail probability can be obtained through the central limit theorem, which will actually be the next topic that we will be studying.

Very loosely speaking, the central limit theorem tells us that the random variable M sub n, which is the sample mean, behaves as if it were a normal random variable with the mean and the variance that it should have.

We know that this is the mean and the variance of the sample mean, but the central limit theorem tells us that in addition to that, we can also pretend that the sample mean is normal and carry out approximations as if this were a normal random variable.

Now, this statement that I'm making here is only a loose statement.

It is not mathematically completely accurate.

We will see later a more accurate statement of the central limit theorem.

In a different direction, we can talk about different types of convergence.

We did define convergence in probability, but that's not the only notion of convergence that's relevant to random variables.

There's an alternative notion, which is convergence with probability one.

Here is what it means.

We have a single probabilistic experiment.

And within that the experiment, we have a sequence of random variables and another random variable, and we want to talk about this random variable converging to that random variable.

What do we mean by that?

We consider a typical outcome of the experiment, that is, some omega.

Look at the values of the random variable Y_n under that particular omega, and look at that sequence of values, the values of the different random variables under that particular outcome.

Under that particular outcome, Y also has a certain numerical value, and we're interested in whether this convergence takes place as n goes to infinity.

Now for some outcomes, omega, this will happen.

For some, it will not happen.

We will say that we have convergence with probability one if this event has probability equal to 1.

That is, there is probability one, that is, essential certainty, that when an outcome of the experiment is obtained, the resulting sequence of values of the random variables Y_n will converge to the value of the random variable Y.

Now, this definition is easy to write down, but to actually understand what it really means and the ways it is different from convergence in probability is not so easy.

It does take some conceptual effort, and we will not discuss it any further at this point.

Let me just say that this is a stronger notion of convergence.

If you have convergence with probability one, you also gets convergence in probability.

And it turns out that the law of large numbers also holds under this stronger notion of convergence.

That is, we have that the sample mean converges to the true mean with probability one.

This is the so-called strong law of large numbers, and because this is a stronger notion of convergence, a more demanding one, that's why this is called the strong law.

Incidentally, at this point, you might be quite uncertain and confused as to what is really the difference between these two notions of convergence.

The definitions do look different, but what is the real difference?

This is quite subtle, and it does take quite a bit of thinking.

It's not supposed to be something that is obvious.

So the purpose of this discussion is only to point out these further directions but without, at this point, going into it in any depth.

Finally, there is another notion of convergence in which we're looking at the distributions of the random variables involved.

So we may have a sequence of random variables.

Each one of them has a certain distribution described by a CDF, and we can ask the question, does this sequence of CDFs converge to a limiting CDF?

If that happens, then we say that we have convergence in distribution, and this is more or less the type of convergence that shows up when we deal with the central limit theorem because this is really a statement about distributions, that the distribution of the sample mean in some sense starts to approach the distribution of a normal random variable.