We will now take a step towards abstraction, and discuss the issue of convergence of random variables. Let us look at the weak law of large numbers. It tells us that with high probability, the sample mean falls close to the true mean as n goes to infinity.

We would like to interpret this statement by saying that the sample mean converges to the true mean. However, before we can make such a statement, we should first define carefully the word "converges." And we need a notion of convergence that refers to convergence of random variables.

Here's a definition. Suppose that we have a sequence of random variables that are not necessarily independent. We say that this sequence of random variables converges in probability-- that's a particular notion of convergence we're defining.

It converges to a certain number if the following is true-- no matter what epsilon is, as long as it is a positive number, the probability that the random variable falls far from this number-- that is, epsilon or further away from that number-- that probability converges to 0 as n increases. That is, as n increases, there is higher and higher probability that Yn will be close to this particular number a.

This is the notion of convergence that we have defined. And notice that this notion of convergence corresponds exactly to what is happening in the weak law of large numbers. And so in particular, we can describe the weak law of large numbers as saying that Mn, the sample mean, converges to mu as n goes to infinity, but in a particular sense-- in the sense of convergence in probability.

Let us now try to understand a little better what convergence in probability really amounts to. And we will do that by making a comparison with the ordinary notion of convergence of real numbers. When we're dealing with convergence of numbers, we start with a sequence of numbers, and we are interested in the statement that this sequence converges to a certain limit.

What does that mean? What we mean is that the elements of the sequence eventually-- that is, when n is large enough-- will get and stay arbitrarily close to this particular number a, which is the limit. In terms of a picture, here is a, the limit. Here is n.

We take a small band around this number a. And what we require is that the elements of the sequence eventually get within this band around the number a. They might get outside the band, get inside again.

But eventually-- that is, after some time-- the elements of the sequence will only stay inside this band.

Now to translate this into a more formal mathematical statement, which is the mathematical definition of the notion of convergence, we have the following-- if I give you some epsilon, epsilon could be a very small number. I form a band around a that goes from a minus epsilon to a plus epsilon. What I want is that there exists a certain time, n0-- in this picture, n0 would be here-- such that for all times after n0, our sequence stays within epsilon from a. That is, our sequence stays inside this band.

Now let us move to the case of random variables, and see what convergence in probability is talking about. Here, instead of a sequence of numbers, we have a sequence of random variables. And we're interested in the meaning of the convergence of the sequence of random variables to a particular number.

In words, what this means is that if I fix a certain epsilon, as in this picture, then the probability that the random variable falls outside this band converges to 0. So the picture would be as follows.

We have, again, our limit. We fix some epsilon, which could be an arbitrarily small number. For any fixed choice of epsilon, we take this band, and for any given n, we look into the probability that our random variable falls inside that band.

So if I am to draw the distribution of our random variable, a distribution might be something like this-- so there is a certain probability that it falls outside this band. But when n becomes large, this probability of falling outside this band becomes very small.

So the probability of falling outside the band becomes tiny. So the bulk of the distribution-- that is, most of the probability-- is concentrated inside this band. And this is true, no matter how small epsilon is.

If I take a much narrower band around a, I still want all of the probability to be eventually concentrated inside that band. Of course, it might take longer. It might take a larger value of n, but I want that when n is very large, the bulk of the distribution is, again, concentrated inside this narrow band. So in words, convergence in probability means that almost all of the probability mass of the random variable Yn, when n is large, that probability mass get concentrated within a narrow band around the limit of the random variable.

We finally point out a few useful properties of convergence in probability that parallel well-known

properties of convergence of sequences. Suppose that we have a sequence of random variables that converges in probability to a certain number a, and another sequence that converges in probability to some other number b. We do not make any assumptions about independence. We do not assume the Xn's are independent of each other. We do not assume that the sequence of Xn's is independent of Yn.

We then have the following properties-- if g is a continuous function, then the function of the random variables converges to the function of the number. So for example, the sequence of random variables Xn squared is going to converge to a squared. Another fact is that the sum of these two random variables converges to the sum of their limits.

Both of these properties are analogous to what happens with ordinary convergence of numbers. And they tell us that, in some sense, convergence in probability is not a very different notion. We will not prove those properties at this point, but they're useful to know.

However, there's an important caveat. Xn might converge to a certain number in probability. However, the expected value of Xn does not necessarily converge to that same limit. So convergence of random variables does not imply convergence of expectations. And we will be seeing an example where this convergence does not take place.