

In this segment, we derive and discuss the weak law of large numbers.

It is a rather simple result, but plays a central role within probability theory.

The setting is as follows.

We start with some probability distribution that has a certain mean and variance, which we assume to be finite.

We then draw independent random variables out of this distribution so that these X_i 's are independent and identically distributed, i.i.d.

for short.

What's going on here is that we're carrying out a long experiment during which all of these random variables are drawn.

Once we have drawn all of these random variables, we can calculate the average of the values that have been obtained, and this gives us the so-called sample mean.

Notice that the sample mean is a random variable because it is a function of random variables.

It should be distinguished from the true mean, μ , which is the expected value of the X_i 's, which is a number.

It is not random.

And μ is some kind of average over all the possible outcomes of the random variable X_i .

The sample mean is the simplest and most natural way for trying to estimate the true mean, and the weak law of large numbers will provide some support to this notion.

Let us now look at the properties of the sample mean.

Let us calculate its expectation.

By the way, this object here involves two different kinds of averaging.

The sample mean averages over the values observed during one long experiment, whereas the expectation averages over all possible outcomes of this experiment.

The expectation is some kind of theoretical average because we do not get to observe all the possible outcomes of this experiment, but the sample mean is something that we actually calculate on the basis of our observations.

In any case, the expected value of the sample mean, by linearity, it is the expected value of the numerator divided by the denominator.

Using linearity once more, the expected value of the sum is the sum of the expected values, and since each one of those expected values is equal to μ , we obtain n times μ divided by n , which leaves us with μ .

So the theoretical average, the expected value of the sample mean, is equal to the true mean.

Let us now calculate the variance of the sample mean.

The variance of a random variable divided by a number is the variance of that random variable divided by the square of that number.

Now, since the X_i 's are independent, the variance is the sum of the variances.

And therefore, we obtain n times the variance of each one of them.

And after we simplify, this leaves us with σ^2 over n .

We're now in a position to apply the Chebyshev inequality.

The Chebyshev inequality tells us that the distance of a random variable from its mean, being larger than a certain number, has a probability that's bounded above by the variance of the random variable of interest divided by the square of the number that we have here.

We have already calculated the variance, and so this quantity is σ^2 over n times ϵ^2 .

And now, if we consider ϵ as a fixed number and let n go to infinity, then what we obtain is a limiting value of 0.

So the probability of falling far from the mean diminishes to zero as we draw more and more samples.

That's exactly what the weak law of large numbers tells us.

If we fix any particular ϵ , which is a positive constant, the probability that the sample mean falls away from the true mean by more than ϵ , that probability becomes smaller and smaller and converges to 0 as n goes to infinity.

Let us now interpret the weak law of large numbers.

As I already hinted, we have to think in terms of one long experiment, and during that experiment, we draw

several independent random variables drawn from the same distribution.

One way of thinking about those random variables is that each one of them is equal to the mean, the true mean, plus some measurement noise, which is a term that has zero expected value.

And all of these noises are independent.

So we have a collection of noisy measurements, and then we take those measurements and form the average of them.

What the weak law of large numbers tells us is that the sample mean is unlikely to be far off from the true mean.

And by far off, we mean at least epsilon distance away.

So the sample mean is, in some ways, a good way of estimating the true mean.

If n is large enough, then we have high confidence that the sample mean gives us a value that's close to the true mean.

As a special case let us consider a probabilistic model in which we repeat independently many times the same experiment.

There's a certain event A associated with that experiment that has a certain probability, and each time that we carry out the experiment, we use an indicator variable to indicate whether the outcome was inside the event or outside the event.

So X_i is 1 if A occurs, and it is 0 otherwise.

The expected value of the X_i 's, the true mean in this case, is equal to the number p .

In this particular example, the sample mean just counts how many times the event A occurred out of the n experiments that we carried out, so it's the frequency with which the event A has occurred.

And we call it the empirical frequency of event A . What the weak law of large numbers tells us is that the empirical frequency will be close to the probability of that event.

In this sense, it reinforces or justifies the interpretation of probabilities as frequencies.