THE LAW OF LARGE NUMBERS

THEOREM (The Law of Large Numbers)
Suppose $X_1, X_2, \ldots$ are i.i.d. random variables, each with expected value μ. Then for every $\epsilon > 0$,

$$\lim_{n \to \infty} \text{Prob}\left[\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right] = 0 .$$

****

Before we can appreciate this theorem, we must understand its statement. Informally, we think of X_i as being a number obtained from the ith repetition of an experiment.

1. "i.i.d." This means "independent identically distributed".

2. "Independent" means that for every n, and for all intervals $[a_i, b_i]$,

$$\text{Prob}[a_1 \leq X_1 \leq b_1, \ a_2 \leq X_2 \leq b_2, \ \ldots, \ a_n \leq X_n \leq b_n]$$
$$=\text{Prob}[a_1 \leq X_1 \leq b_1] \times \text{Prob}[a_2 \leq X_2 \leq b_2] \times \cdots \times \text{Prob}[a_n \leq X_n \leq b_n] .$$

Informally, this means that knowing the results of some of the experiments gives you no additional information about the results of the others.

3. "identically distributed" means that for any numbers a and b, $\text{Prob}[a \leq X_i \leq b]$ is the same for all X_i. (Let D name this probability distribution which every one of the X_i have.)
Informally, think of repeating the experiment under the same conditions.

4. The Primordial Example. Imagine that a coin is heads with probability p and tails with probability $1 - p$. Imagine flipping the coin over and over, with the different flips having nothing to do with each other. Define the r.v. X_i to be 1 if the ith flip is heads, and 0 otherwise. Then the sequence $X_1, X_2, \ldots$ is i.i.d.

5. "Prob". The probabilities with which $(X_1 + \ldots X_n)/n$ assumes certain values is determined by the common distribution D and the i.i.d. condition.
   (E.g., if $p = 1/2$ in the primordial example, then the probability that $(X_1 + X_2 + X_3)/3$ equals 0 is 1/8.)

6. "lim" This means: given any $\epsilon > 0$, it holds for every $\delta > 0$, that there exists a positive integer N, such that whenever $n \geq N$,

$$\text{Prob}\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right] < \delta .$$

Informally, this just means that for n large, that average is very close to the expected value with probability very close to 1.

7. Remember, the model of a probability space and expected value is something which we set up with the HOPE that it will be some kind of model of experiments (at least). After we set it up, we WANT there to be some kind of theorem which we can interpret as saying that our setup really does describe the long-term likelihood of events in some reasonable sense.

The Law of Large Numbers (LLN) is exactly such a theorem. It is a justification of our use of the theory of probability. Of course, perfectly independent experiments are an idealization, but we can imagine a model of independent experiments as a reasonable approximation of some actual activities (e.g. flipping coins — and more serious activities). The LLN says that if the model of i.i.d. experiments is a reasonable approximation, then we really can think of the expected value μ as the long term expected average.

8. We can also think of the rv's X_i as representing random samples of some population distribution. For example, X_i might be the annual salary of a person randomly chosen from the class. Then we can think of the average $\overline{X} = (X_1 + \cdots + X_n)/n$ as an estimate of the true population average μ (salary in this case). Now we can interpret the LLN as saying that for large n, the estimate $\overline{X}$ is very close to the true population average, with probability very close to 1.

So again we have a theorem that gives a mathematical justification for applying probability to a real world situation: estimates by sampling.

9. The law of large numbers is not true if the X_i don't have finite expected value, even if they are i.i.d.

10. Actually the LLN above is the "weak" law of large numbers. There is also a "strong" law of large numbers, which implies the weak law. Understanding the statement of the SLLN involves higher math, so we will skip it.