5.9 Law of Large Numbers

Unit 5: Averages

Adapted from Blitzstein-Hwang Chapters 4, 5, and 10.

We turn next to two theorems, the law of large numbers and the central limit theorem, which describe the behavior of the sample mean of i.i.d. r.v.s as the sample size grows. Let $X_1, X_2, X_3, \ldots$ be i.i.d. with finite mean μ and finite variance σ^2. For all positive integers n, let

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

be the sample mean of X_1 through X_n. The sample mean is itself an r.v., with mean μ and variance σ^2/n:

$$E(\bar{X}_n) = \frac{1}{n}E(X_1 + \cdots + X_n) = \frac{1}{n}(E(X_1) + \cdots + E(X_n)) = \mu,$$
$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n) = \frac{1}{n^2}(\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)) = \frac{\sigma^2}{n}.$$

The law of large numbers (LLN) says that as n grows, the sample mean $\bar{X}_n$ converges to the true mean μ (in a sense that is explained below). LLN comes in two versions, which use slightly different definitions of what it means for a sequence of random variables to converge to a number. We will state both versions.

> THEOREM 5.9.1 (STRONG LAW OF LARGE NUMBERS).
>
> The sample mean $\bar{X}_n$ converges to the true mean μ pointwise as $n \to \infty$, with probability 1. In other words, the event $\bar{X}_n \to \mu$ has probability 1.

> THEOREM 5.9.2 (WEAK LAW OF LARGE NUMBERS).
>
> For all $\epsilon > 0$, $P(|\bar{X}_n - \mu| > \epsilon) \to 0$ as $n \to \infty$. (This form of convergence is called *convergence in probability*.)

The law of large numbers is essential for simulations, statistics, and science. Consider generating "data" from a large number of independent replications of an experiment, performed either by computer simulation or in the real world. Every time we use the proportion of times that something happened as an approximation to its probability, we are implicitly appealing to LLN. Every time we use the average value in the replications of some quantity to approximate its theoretical average, we are implicitly appealing to LLN.

Example 5.9.3 (Running proportion of Heads).

Let $X_1, X_2, \ldots$ be i.i.d. $\mathrm{Bern}(1/2)$. Interpreting the X_j as indicators of Heads in a string of fair coin tosses, $\bar{X}_n$ is the proportion of Heads after n tosses. SLLN says that with probability 1, when the sequence of r.v.s $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots$ crystallizes into a sequence of numbers, the sequence of numbers will converge to $1/2$. Mathematically, there are bizarre outcomes such as *HHHHHH*$\ldots$ and *HHTHHTHHTHHT*$\ldots$, but collectively they have zero probability of occurring. WLLN says that for any $\epsilon > 0$, the probability of $\bar{X}_n$ being more than ϵ away from $1/2$ can be made as small as we like by letting n grow.

As an illustration, we simulated six sequences of fair coin tosses and, for each sequence, computed $\bar{X}_n$ as a function of n. Of course, in real life we cannot simulate infinitely many coin tosses, so we stopped after 300 tosses. The figure below plots $\bar{X}_n$ as a function of n for each sequence.

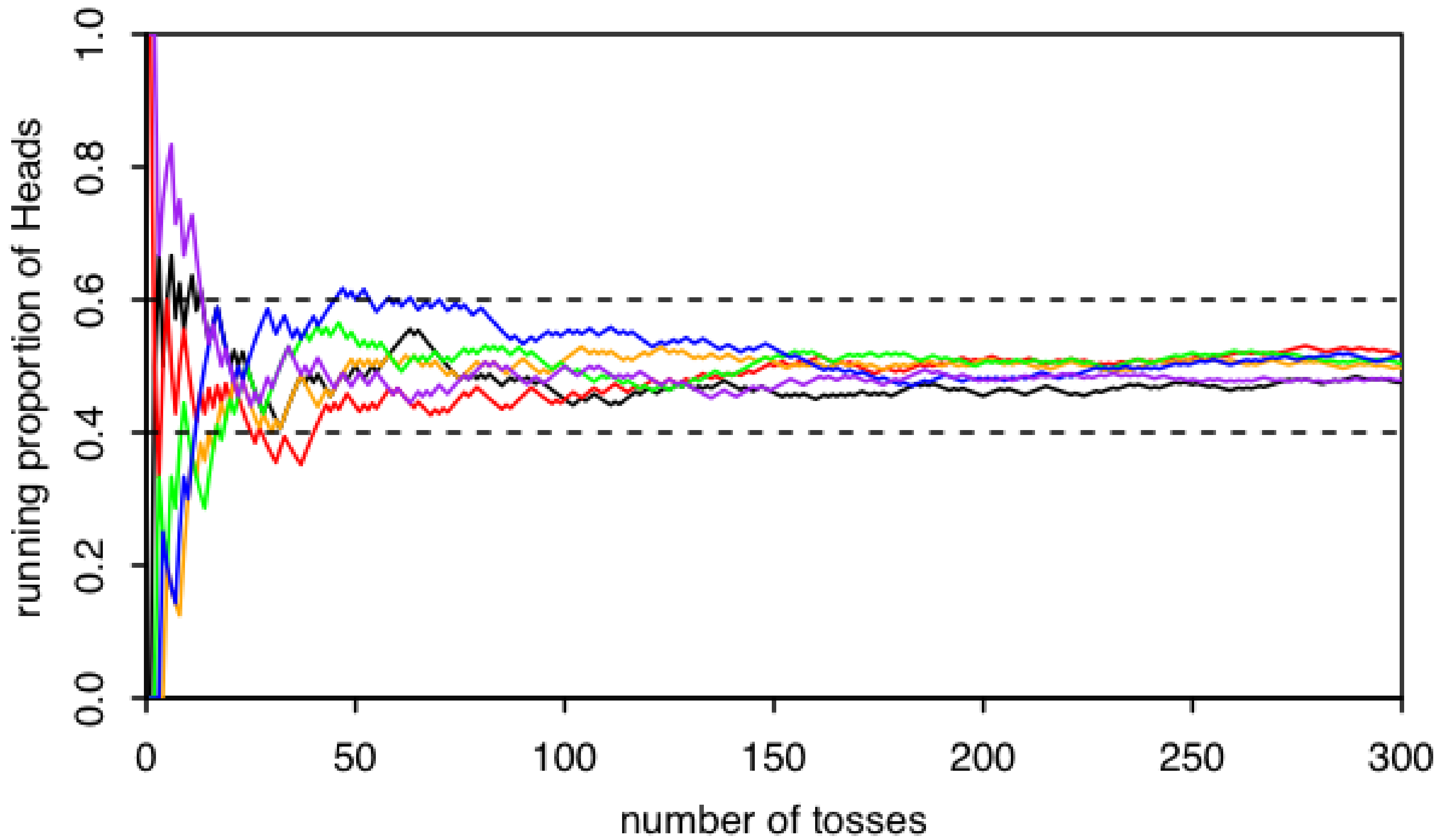

Figure 5.9.4: Running proportion of Heads in 6 sequences of fair coin tosses. Dashed lines at 0.6 and 0.4 are plotted for reference. As the number of tosses increases, the proportion of Heads approaches $1/2$.

View Larger Image

Image Description

At the beginning, we can see that there is quite a bit of fluctuation in the running proportion of Heads. As the number of coin tosses increases, however, $\mathbf{Var}(\bar{X}_n)$ gets smaller and smaller, and $\bar{X}_n$ approaches $1/2$.