1 Law of Large Numbers

Recall the *frequentist* interpretation of probability. Suppose you run an experiment, and let E be some event of interest (e.g., flip a fair coin, where E is the event that you flip heads). If you run this experiment many times, and let $X_i = 1$ if E occurs on the ith trial (and $X_i = 0$ otherwise), then intuitively we would expect that the proportion

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

of times that it has occurred after n trials should converge to $\mathbb{P}(E)$ as $n \mapsto \infty$. Indeed, this is the Law of Large Numbers in a nutshell.

Note that, while we only considered the case that the X_i are IID Bernoulli/Indicator random variables of a given event E, which have means $\mu = \mathbb{P}(E)$, this fact is true in general for any sequence of IID RVs (provided that their means μ and variances σ^2 exists).

Theorem 1.1: Law of Large Numbers

Suppose that $X_1, X_2, \ldots$ are IID random variables with finite means $\mu = \mathbb{E}(X) < \infty$ and variances $\sigma^2 = \text{Var}(X) < \infty$. Let $S_n = \sum i = 1^n X_i$. Then, as $n \mapsto \infty$, then $\frac{S_n}{n} \mapsto \mu$ in the sense that, for any real number $\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) \mapsto 1$$

as $n \mapsto \infty$.

Remarks:

- Note that $\frac{S_n}{n}$ is a sequence of random variable. In fact, $\mathbb{P}\left(\lim_{n\mapsto\infty} \frac{S_n}{n} = \mu\right) = 1$.
- Sometimes, the Law of Large Numbers is known as the *Law of Averages*.

Recall the following:

- **Chebyshev's Inequality**: Let X be a random variable with mean μ and standard deviation σ. Then, for any real number $a > 0$, we have that

$$\mathbb{P}(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}.$$

  Setting $a = \epsilon/\sigma$, we have $\mathbb{P}(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$.
- **Markob's Inequality**: Let X be a non-negative RV (this means that $\mathbb{P}(X \geq 0) = 1$) with mean μ, and $b > 0$ a positive number. Then, $\mathbb{P}(X \geq b) \leq \frac{\mu}{b}$.

Proof. Since the X_i are IID with means μ and variances σ^2, it follows that

$$E\left(\frac{S_n}{n}\right) = \frac{1}{n} n\mu = \mu$$

and $\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$. Hence, by Chebyshev's Inequality, we have

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right).$$

Note now that

$$1 - \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \geq 1 - \frac{\sigma^2}{n\epsilon^2} \mapsto 1$$

as $n \mapsto \infty$. □

Theorem 1.2: Strong Law of Large Numbers

$\frac{S_n}{n}$ converges to μ, in the sense that

$$\mathbb{P}\left(\lim_{n\mapsto\infty} \frac{S_n}{n} = \mu\right) = 1.$$

In other words, the random variable $\lim_{n\mapsto\infty} \frac{S_n}{n}$ has a very simple distribution. It only takes one value (namely μ) with probability 1.

(Example.) Let $X_1, X_2, \ldots$ be an IID sequence of fair die rolls. We know that their means are $\mu = \frac{7}{2}$ (Their variances σ^2 also exists). Hence, by the LLN, our long run average number observed will be $\mu = \frac{7}{2}$.

(Example.) Recall the method of Monte Carlo Integration, discussed in Lecture 3, used to estimate an integral

$$\int_0^1 g(x)dx.$$

We will assume that g is continuous and that $0 \le g(x) \le 1$ for all $0 \le x \le 1$. If you select a large number of uniformly random points in the square $[0,1] \times [0,1]$ and then count the proportion of these that are under the curve $y = g(x)$.

However, there is an even better way of estimating the integral. Select a large number $X_1, \ldots, X_n$ of IID Uniform[0, 1] random variables, and consider the IID random variables $g(X_1), \ldots, g(X_n)$. Then, note that

$$\mu = \mathbb{E}(g(X_i)) = \int_0^1 g(x)dx$$

by the LotUS, which is exactly what we want. Now, similarly,

$$\sigma^2 = \mathbb{E}[(g(X_i) - \mu)^2] = \int_0^1 (g(x) - \mu)^2 dx.$$

Recall that $g(x) \in [0,1]$ for all $x \in [0,1]$. Therefore, $\mu \in [0,1]$ and so also $|g(x) - \mu| \in [0,1]$ for all such x, hence $\sigma^2 < 1$. So, by LLN, for a large n, the average

$$I_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

will be a good approximation to

$$I = \int_0^1 g(x)dx.$$

Moreover, by Chebyshev's Inequality,

$$\mathbb{P}(|I_n - I| < \epsilon) \ge 1 - \frac{1}{n\epsilon^2}.$$

So, if we want the error to be less than 0.02, with probability at least 90% (we want to be at least 90% sure that our approximation is within 0.02 away from the true value), then we should take at least $n = 25000$ points.