7. The Law of Large Numbers and the Central Limit Theorem

The law of large numbers and the central limit theorem play very important roles in statistics.

7.1. **The Weak Law of Large Numbers (WLLN).**

Two probability inequalities are needed in the proof of WLLN.

Theorem 7.1. *(Markov's Inequality) Let X be a non-negative random variable and $\mathbb{E}(X) < \infty$. For any $t > 0$,*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Proof. We only prove for the continuous case.

$$\mathbb{E}(X) = \int_0^\infty xf(x)dx = \int_0^t xf(x)dx + \int_t^\infty xf(x)dx \geq \int_t^\infty xf(x)dx \geq t\mathbb{P}(X > t).$$

□

We can easily prove the result using the concept of indicator function. Note that $I(X \geq t) \leq (X/t)$ is always true, and the result is obtained by taking expectation on both sides.

Theorem 7.2. *(Chebyshev's Inequality) Let X be a random variable with finite mean $\mu = \mathbb{E}(X)$ and variance $\sigma^2 = \mathbb{V}(X)$. Then for $t > 0$ we have*

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

Proof. We apply Markov's inequality to the random variable $|X - \mu|^2$:

$$\mathbb{P}(|X - \mu| \geq t) = \mathbb{P}(|X - \mu|^2 \geq t^2) \leq \frac{\mathbb{E}(|X - \mu|^2)}{t^2} = \frac{\sigma^2}{t^2}.$$

□

Similarly we can use indicator function to prove Chebyshev's inequality by taking expectation over both sides of $I(|X - \mu| \geq t) \leq \frac{|X-\mu|^2}{t^2}$.

Theorem 7.3. *(The Weak Law of Large Numbers (WLLN)) If $X_1, X_2, \ldots$ is a sequence of iid random variables with finite expectation μ, then for any $\epsilon > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right) = 0.$$

Proof. We only prove a simple case where $X_1, X_2, \ldots$ have finite variance $\sigma^2 < \infty$. Let $\bar{X}_n = \sum_{i=1}^n X_i/n$ and we have $\mathbb{E}(\bar{X}_n) = \mu$ and $\mathbb{V}(\bar{X}_n) = \sigma^2/n$. Using Chebyshev's inequality,

$$0 \leq \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0.$$

□

The WLLN says that the distribution of the average of iid random variables $\bar{X}_n$ will be more and more concentrated on its mean μ as n gets large. In other words, the observed sample mean $\sum_{i=1}^n x_i/n$ will be a very good approximation to μ as n gets large.

7.2. **The Central Limit Theorem (CLT).**

Theorem 7.4. *(The Central Limit Theorem) Let $X_1, X_2, \ldots$ be a sequence of iid random variables with finite mean μ and variance σ^2. Define*

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

where $\bar{X}_n = \sum_{i=1}^n X_i/n$, then the distribution of Z_n converges to a standard normal distribution as $n \to \infty$. What it means is that, for any z we have

$$\lim_{n\to\infty} \mathbb{P}(Z_n \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

The CLT says that the mean of a random sample $X_1, \ldots, X_n$ from **any** distribution can be approximated by a normal distribution as long as μ and σ^2 are finite and n gets large. It is a little bit surprising and has tremendous applications.

Example 7.5. Let $X_i \sim \mathsf{Bernoulli}(p)$ and $\bar{X}_n = \sum_{i=1}^n X_i/n$. Notice that the exact distribution of $n\bar{X}_n$ is $\mathsf{Binomial}(n,p)$. In Figure 7.1 we illustrate that the distribution of $n\bar{X}_n$ becomes increasingly similar to a normal pdf as $n \to \infty$, which verifies the CLT.

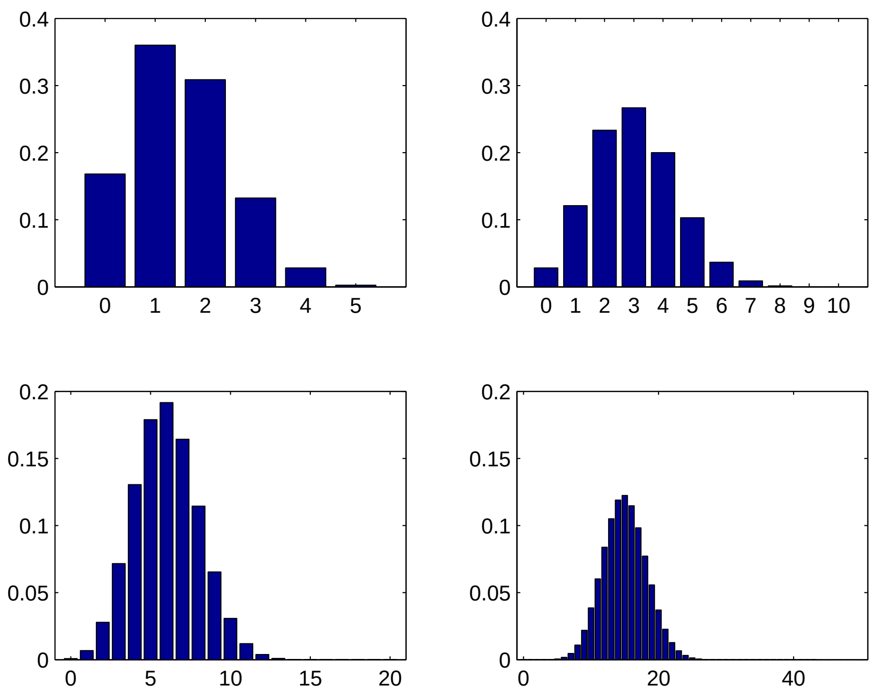

FIGURE 7.1. The distribution of $\mathsf{Binomial}(n, p = 0.3)$ for $n = 5, 10, 20$ and 50.

Example 7.6. (Example 7.8 in [1]) Let Y_i be the service time for the ith customer, and we have $\mu = 1.5$, $\sigma^2 = 1$. So $\mathbb{P}(\sum_{i=1}^{100} Y_i \le 120) = \mathbb{P}(\bar{Y} \le 1.2)$. Since $n = 100$ is large, CLT tells that $\bar{Y}$ is approximately $N(\mu, \sigma^2/n) = N(1.5, 0.01)$. So $\mathbb{P}(\bar{Y} \le 1.2) = \mathbb{P}(\frac{\bar{Y}-1.5}{0.1} \le \frac{1.2-1.5}{0.1}) \approx \mathbb{P}(Z \le -3) = \Phi(-3) = 0.00135$. Here Z is a standard normal and $\Phi(z)$ can be found from Table 4, Appendix III in [1].