5.9 Law of Large Numbers

- Let $X_1, X_2, X_3, \ldots$ be i.i.d r.v.s with finite mean μ and finite variance σ^2

For all positive integer n, let:

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}$$

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

$$E(\overline{X}_n) = \frac{1}{n} E(X_1 + \ldots + X_n) = \frac{1}{n}(EX_1 + \ldots EX_n) = \mu$$

$$Var(\overline{X}_n) = \frac{1}{n^2} Var(X_1 + \ldots + X_n) = \frac{1}{n^2}\left(Var(X_1) + \ldots + Var(X_n)\right) = \frac{\sigma^2}{n}$$

The Law of Large Numbers (LLN) says that as n grow larger, the sample mean $\overline{X}_n$ converges to true mean μ.

- LLN comes in 2 versions, stated below:

Theorem 5.9.1 (Strong Law of Large Numbers)

- The sample mean, $\overline{X}_n$ converges to the true mean μ pointwise as $n \to \infty$, with probability 1.
- In other words, the event $\overline{X}_n \to \mu$ has probability 1.

Theorem 5.9.2 (Weak Law of Large Numbers)

For all $\varepsilon > 0$, $P(|\overline{X}_n - \mu| > \varepsilon) \to 0$ as $n \to \infty$. (This form of convergence is called convergence in probability)

Example 5.9.3 (Running proportion of Heads)

Let $\begin{cases} X_1, X_2, \ldots \text{ be i.i.d Bern}(1/2). \\ \overline{X}_n \text{ be the proportion of Heads after } n \text{ tosses.} \end{cases}$

- The Strong Law of Large Numbers (SLNN) says that with probability 1, when the sequence of r.v.s $\overline{X}_1, \overline{X}_2, \overline{X}_3, \ldots$ crystallizes into a sequence of numbers, the sequence of numbers will converge to $\frac{1}{2}$.

- The Weak Law of Large Numbers (WLNN) says that for any $\varepsilon > 0$, the probability of the difference between $\overline{X}_n$ and $\mu\left(\frac{1}{2}\right)$ being large than ε can be made smaller as we increase n.

5.10 Central Limit Theorem

- Recall that for $X_1, X_2, X_3, \ldots$ be i.i.d with mean μ and variance σ^2, the LLN says that as $n \to \infty$, $\overline{X}_n$ converges to constant μ.
- The Central Limit Theorem (CLT) says that in process mentioned above, the distribution of $\overline{X}_n$ after standardization approaches Standard Normal distribution.

Theorem 5.10.1 (Central Limit Theorem)

As $n \to \infty$,

$$\underbrace{\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right)}_{\text{Standardize } \overline{X}_n} \longrightarrow N(0, 1) \quad \text{in distribution}$$

In other words, the CDF of the left-hand side approaches Φ, the CDF of the Standard Normal distribution.

- The CLT also tells us about the approximate distribution of $\overline{X}_n$ when n is a finite large number.

Central Limit Theorem, approximation form

- For large n, the distribution of $\overline{X}_n$ is approximately $N(\mu, \sigma^2/n)$.

Note:
- The distribution of individual X_j can be anything, with large n, the act of averaging will cause Normality to emerge.
- If the distribution of individual X_j is highly skewed or multimodal distribution, we will need larger n before the Normal approximation becomes accurate than if X_j is already i.i.d Normals

Central Limit Theorem, sum of individual i.i.d X_j

- Since $W_n = X_1 + X_2 + \ldots + X_n = n\overline{X}_n$ is just a scaled version of $\overline{X}_n$, the CLT also implies W_n is approximately Normal.
- If X_j have mean μ and variance σ^2, then:

$$W_n \sim N(n\mu, n\sigma^2)$$

- This is also completely equivalent to the approximation for $\overline{X}_n$

Example 5.10.3 (Poisson convergence to Normal)

- Let $Y \sim \text{Pois}(n)$. By theorem 5.7.6 (sum of independent Poissons), we can consider Y to be a sum of n i.i.d $\text{Pois}(1)$ r.v.s.

Therefore, for large n:

$$Y \sim N(n, n)$$

Example 5.10.4 (Binomial covergence to Normal)

- Let $Y \sim \text{Bin}(n, p)$, we consider Y to be sum of n i.i.d $\text{Bern}(p)$ r.v.s

Therefore, for large n:

$$Y \sim N(np, np(1-p))$$

Note for example 5.10.4:

- The Normal approximation for Binomial distribution is complementary to the Poisson approximation discussed in previous Section.
  - The Poisson approximation works best when n is large, p is small
  - the Normal approximation works best when n is large, p is around $\frac{1}{2}$

Example 5.10.5 (Volatile stock)

Each day, a volatile stock rises 70% or drops 50% in price with the same probability, each day is independent.

Let Y_n be the stock price after n days, starting from initial value of $Y_0 = 100$.

a) Explain why $\log Y_n$ is approximately Normal for large n, and state its parameters.

- We can write $Y_n = Y_0 \underbrace{(0.5)}_{\text{drop } 50\%}^{\overbrace{n-U_n}^{\text{\# times it drops}}} \underbrace{(1.7)}_{\text{rise } 70\%}^{\overbrace{U_n}^{\text{\# times it rises}}}$, where $U_n \sim \text{Bin}(n, \frac{1}{2})$ is the number of times the stock rises in the first n days. This gives:

$$\log Y_n = \log Y_0 - n\log 2 + U_n \log 3.4$$

which is a location-scale transformation of U_n.

- By CLT, U_n is approximately $N(\frac{n}{2}, \frac{n}{4})$ for large n, so $\log Y_n$ is also approximately Normal with:
  - Mean: $E(\log Y_n) = \log 100 - n\log 2 + (\log 3.4) E(U_n)$
  $\approx \log 100 - 0.081 n$
  - Variance: $\text{Var}(\log Y_n) = (\log 3.4)^2 \cdot \text{Var}(U_n)$
  $\approx 0.374 n$

b) What happens to $E(Y_n)$ as $n \to \infty$?

- We can write: $Y_n = Y_0 0.5^{n-U_n} 1.7^{U_n} = Y_0 0.5^n 3.4^{U_n}$
- Using the fact that if X and Y are independent then $E(XY) = E(X)E(Y)$, so:

$$E(Y_n) = Y_0 0.5^n \cdot E(3.4^{U_n}) = Y_0 0.5^n (E(3.4)^B)^n,$$

where $B \sim \text{Bern}(\frac{1}{2})$.

This simplifies to:

$$E(Y_n) = Y_0 (0.5)^n \left(3.4 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}\right)^n = 1.1^n Y_0$$

So $\boxed{E(Y_n) \to \infty \text{ as } n \to \infty}$

c) Use the Law of Large Numbers to find out what happens to Y_n as $n \to \infty$

- Let $U_n \sim \text{Bin}(n, \frac{1}{2})$ be the number of times the stock rises in first n days

- Let say if the stock rise 70% and then falls 50%, overall the stock price still dropped 15% since $1.7 \cdot 0.5 = 0.85$. So after large n days, the stock price Y_n will be very small according to LLN
- Writing Y_n in terms of $\frac{U_n}{n}$ in order to apply LLN, we have:

$$Y_n = Y_0 (0.5)^{n-U_n} (1.7)^{U_n} = Y_0 \left(\frac{(3.4)^{U_n/n}}{2} \right)^n$$

Since $\frac{U_n}{n} \to 0.5$ with probability 1.

$$\Rightarrow (3.4)^{U_n/n} \longrightarrow \sqrt{3.4} < 2 \text{ with probability } 1.$$

So $Y_n \to 0$ with probability 1.

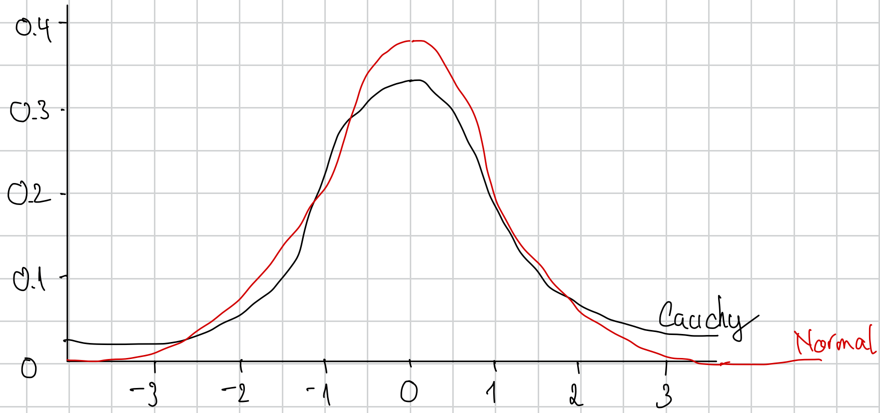
- Paradoxically, $E(Y_n) \to \infty$ but $Y_n \to 0$ with probability 1.

Warning 5.10.6 (The Evil Cauchy)

The Cauchy distribution is the distribution of X/Y, where X and Y are i.i.d. $N(0,1)$. The Cauchy has PDF:

$$f(x) = \frac{1}{\pi(1+x^2)}, \text{ for all real } x$$

The Cauchy has much heavier tails than Standard Normal PDF:



- Properties of Cauchy:
  - Doesn't have finite mean or variance, so the LLN and CLT don't apply
  - The sample mean of n Cauchys is still Cauchy, no matter how large n gets