MA 203

1. Weak Law of Large Numbers
2. Strong Law of Large Numbers
3. Central Limit Theorem

Weak Law of Large Numbers (WLLN): Let $X_1, X_2, \cdots, X_i, \cdots$ be a sequence of IID RVs with mean μ and variance σ^2. Then, for any $\in > 0$, we have

$$\lim_{n\to\infty} P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \in\right\} = 0$$

$$\boxed{\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \mu\right| > \in\right\} = 0}$$

$$\left\{\frac{S_n}{n}\right\} \xrightarrow{P.} \mu$$

Theorem 1: Let $X_1, X_2, \cdots, X_i, \cdots$ be a sequence of IID RVs having a Bernoulli distribution with parameter p. Then, for any $\in > 0$, show that $\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| \geq \in\right) = 0$.

Proof: $P\{X_i = 1\} = p$; $P\{X_i = \cancel{1}\,0\} = \cancel{p}\ 1-p$; $n = 1,2,\cdots$

$X_i \to 1 \to p$

$X_i \to 0 \to 1-p$

$P\{X_i = 1\} = \text{Success}$

$P\{X_i = 0\} = \text{fail}$

Sample: $\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$

$E[X_i] = p$

$Var(X_i) = p(1-p)$

$$E\left[\frac{S_n}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = p = \frac{1}{n} \times np$$

$$var\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n} = \frac{p(1-p)}{n}$$

Apply Chebyshev Inequality

$X \quad \mu \quad \sigma^2 \quad \varepsilon > 0$

$P\{|X-\mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$

$$P\left\{\left|\frac{S_n}{n} - p\right| \geq \epsilon\right\} \leq \frac{var\left(\frac{S_n}{n}\right)}{\epsilon^2}$$

$$P\left\{\left|\frac{S_n}{n} - p\right| \geq \epsilon\right\} \leq \frac{p(1-p)}{n\,\epsilon^2} \qquad (1)$$

$$\boxed{P\left\{\left|\frac{S_n}{n} - p\right| < \epsilon\right\} > 1 - \frac{p(1-p)}{n\,\epsilon^2}}$$

Apply limit

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - p\right| \geq \epsilon\right\} \leq \lim_{n\to\infty} \frac{p(1-p)}{n\,\epsilon^2}$$

$$\boxed{\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - p\right| \geq \epsilon\right\} = 0}$$

$\frac{S_n}{n} \xrightarrow{P} p$ as $n \to \infty$

Example 1: E: Rolling a Dice

Event A: Getting a number 5

For $\in = 0.01$, what is the minimum number of Bernoulli trials such that

$$\boxed{P\left\{\left|\frac{S_n}{n} - p\right| < \in\right\} > 0.95.}$$

Error

Confidence

$X_1 \quad X_2 \quad \cdots \quad X_n$

5   o.w.

$1/6 \quad 5/6$

$A \longrightarrow$ getting number 5

getting a number $\{1, 2, 3, 4, 6\}$

$P(A) = 1/6$

$$\boxed{P\{X_i = 6\} = 1/6 = p}$$

$$P\left\{\left|\frac{S_n}{n}-p\right| \geq \varepsilon\right\} \leq \frac{p(1-p)}{n\varepsilon^2}$$

$$P\left\{\left|\frac{S_n}{n}-p\right| < \varepsilon\right\} > 1-\frac{p(1-p)}{n\varepsilon^2} \quad \text{--- (1)}$$

$$P\left\{\left|\frac{S_n}{n}-p\right| < \varepsilon\right\} > 0.95 \quad \text{--- (2)}$$

By Comparing (1) & (2)

$p = 1/6$

$$1-\frac{p(1-p)}{n\varepsilon^2} = 0.95 \Rightarrow$$

$$\boxed{n = 27,778}$$

Example 2: Suppose someone gives you a coin and claims that this coin is biased; that it lands on heads only 48% of the time. You decide to test the coin for yourself. If you want to be 95% confident that this coin is indeed biased, how many times must you toss the coin?

$\varepsilon = 0.02$

n

$$P\left\{\left|\frac{S_n}{n} - \mu\right| < 0.02\right\} > 0.95$$

$X_1 \quad X_2 \quad \cdots \quad X_n$

$$P\left\{\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right\} > 1 - \frac{p(1-p)}{n\varepsilon^2}$$

$H \to 1$, $T \to 0$

$p = 0.48$

$(1-p) = 0.52$

$$1 - \frac{p(1-p)}{n\varepsilon^2} = 0.95$$

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1 + 0 + \cdots + 1}{n} = \checkmark$$

or $$n = \frac{p(1-p)}{(0.05) \times (0.02)^2}$$

$$\boxed{n = 12{,}480}$$

Example 3: A survey of 1500 people is conducted to determine whether they prefer Pepsi or Coke. The results show that 27% of people prefer Coke while the remaining 73% favour Pepsi. Estimate the margin of error in the poll with a confidence of 90%.

Sol:-

$X_1 \quad X_2 \quad \cdots \quad X_{1500}$

Pepsi → 1, 0.73

Coke → 0, 0.27

$p = 0.73$

$(1-p) = 0.27$

$P\{X_1 = 1\} = p = 0.73$

$$P\left\{\left|\frac{S_n}{n} - p\right| < \varepsilon\right\} > 1 - \frac{p(1-p)}{n\varepsilon^2}$$

$\to 0.90$

$$1 - \frac{p(1-p)}{n\varepsilon^2} = 0.90$$

$n = 1500$

$p = 0.73$

$$\boxed{\varepsilon = 0.0362}$$

Example 4: Let p be the fraction of population that will poll in an election. The objective is to estimate p using sample mean $\left(\frac{S_n}{n}\right)$. Find the minimum sample size (n) such that the probability of less than .01 error is more than 0.95.

Sol:-

$$P\left\{\left|\frac{S_n}{n} - p\right| < \varepsilon\right\} > 1 - \frac{p(1-p)}{n\varepsilon^2}$$

$p = 0.60$

M

60%

$X_1 \quad X_2 \; \cdots \; X_n$

$\varepsilon = 0.01$

$p = p$

n 0.95

V MV

p $1-p$

$\boxed{n = ?}$

$$P\left\{\left|\frac{S_n}{n}-p\right|<\varepsilon\right\} > 1-\frac{p(1-p)}{n\varepsilon^2} \to 0.95$$

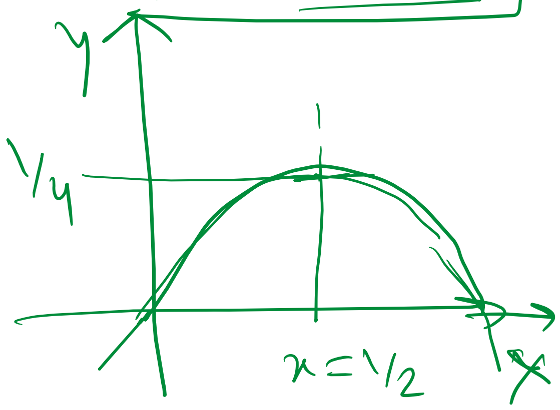
$$P\left\{\left|\frac{S_n}{n}-p\right|\ge\varepsilon\right\} \le \frac{p(1-p)}{n\varepsilon^2} \le \frac{1}{4n\varepsilon^2}$$

$$y = x(1-x)$$
$$= x - x^2$$
$$\frac{dy}{dx} = 1-2x = 0$$
$$x = 1/2$$

$$p(1-p)$$
$$\boxed{y = x(1-x)}$$

$\varepsilon = 0.01,\quad 95\%$

$$P\left\{\left|\frac{S_n}{n}-p\right|\ge 0.01\right\} \le \frac{1}{4n\varepsilon^2}$$



$$\frac{1}{4\times n\times\varepsilon^2} = 0.05$$

$$n = \frac{1}{4\times 0.05\times(0.01)^2} = \frac{10^6}{20}$$

$$n = 50{,}000$$

Strong Law of Large Numbers (SLLN): Let $X_1, X_2, \cdots, X_i, \cdots$ be a sequence of iid RVs with finite mean μ and finite variance σ^2. Then

$$\frac{S_n}{n} \xrightarrow{a.s} \mu$$

Where

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

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

Central Limit Theorem (CLT)

Theorem: Let $X_1, X_2, \cdots, X_n, \cdots$ be a sequence of IID RVs defined on the probability space (S, F, P). Assume that $E[X_i] = \mu$ and $var(X_i) = \sigma^2$.

Define,

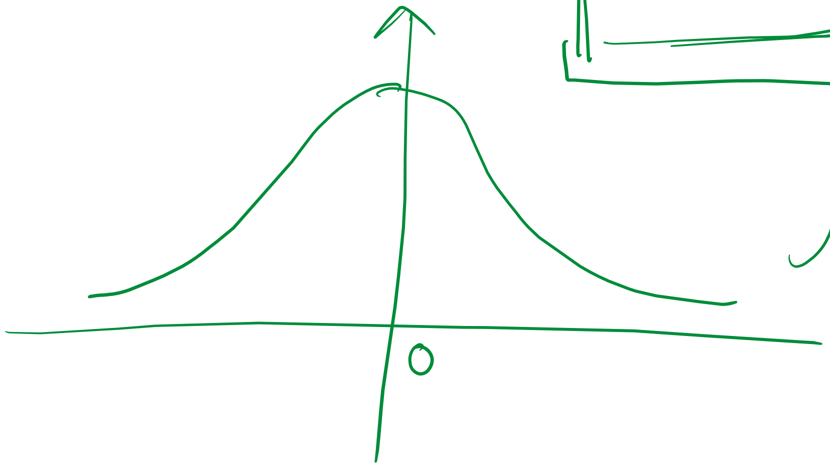
$$Z_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E[X_i]}{\sqrt{var(\sum_{i=1}^{n} X_i)}}; \; n = 1,2,\cdots$$

$$Z_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E[X_i]}{\sqrt{Var(\sum_{i=1}^{n} X_i)}}$$

Then, for larger n, Z_n (approximately) follow the standard normal distribution,

$n \to \infty$

$$P(Z_n \leq z) \approx \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$



$\mu = 0$

$Var = 1$

$n = 1$

$$Z_1 = \frac{X_1 - E[X_1]}{\sqrt{Var(X_1)}}$$

$$Z_2 = \frac{(X_1 + X_2) - (E(X_1 + X_2))}{\sqrt{Var(X_1 + X_2)}}$$

$\vdots$

Z_n

Proof: Assume that MGF of $X_i's$ exist

$$M_{Z_n}(t) = E\left[e^{\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}\right)t}\right]$$

$$= e^{-\frac{\sqrt{n}\mu t}{\sqrt{n}\sigma}} E\left[e^{\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}\sigma}\right)t}\right]$$

$$= e^{-\frac{\sqrt{n}\mu t}{\cancel{\sqrt{n}}\sigma}} \left\{E\left[e^{\left(\frac{X_1}{\sqrt{n}\sigma}\right)t}\right]\right\}^n$$

We know that

$$M_X(t) = 1 + E[X]t + \frac{E[X^2]}{2} + \cdots$$

$$\ln M_X(t) = \ln\left\{1 + \left(E[X]t + \frac{E[X^2]}{2} + \cdots\right)\right\}$$

$$lnM_{Z_n}(t) = \frac{-\sqrt{n}\,\mu t}{\sigma} + \mathrm{n}\ln\left(1 + \frac{\mu t}{\sqrt{n}\,\sigma} + \frac{(\sigma^2+\mu^2)t^2}{2\,n\sigma^2} + \cdots\right)$$

$$= \frac{-\sqrt{n}\,\mu t}{\sigma} + \mathrm{n}\,\left\{\left(\frac{\mu t}{\sqrt{n}\,\sigma} + \frac{(\sigma^2+\mu^2)t^2}{2\,n\sigma^2} + \cdots\right) - \frac{1}{2}\left(\frac{\mu^2 t^2}{n\sigma^2} + \cdots\right) + \frac{1}{3}(\cdots) - \cdots\right\}$$

$$\lim_{n\to\infty} lnM_{Z_n}(t) = \frac{t^2}{2}$$

$$M_{Z_n}(t) = e^{\frac{t^2}{2}}$$

$$Z_n \sim N(0,1)$$

Example 1: Suppose someone gives you a coin and claims that this coin is biased; that it lands on heads only 48% of the time. You decide to test the coin for yourself. If you want to be 95% confident that this coin is indeed biased, how many times must you toss the coin?