Limit Theorems & Convergence Modes for random variables

- Law of Large Numbers (LLN) → Weak Law (WLLN), Strong Law (SLLN)
- Central Limit Theorem (CLT).

LLN ⇝ states that average of a large # of iid r.v.'s converges to the expected value.

CLT ⇝ states that (under some conditions), sum of a large # of r.v.'s has an approximately Normal distribution. (Gaussian)

Weak Law of Large Numbers (WLLN)

For iid r.v.'s $X_1, X_2, \dots, X_n$, the sample mean is denoted as
$$\bar{X} = \frac{X_1 + X_2 \cdots + X_n}{n}$$

* Sample mean is also a random variable.

$$E[\bar{X}] = \frac{E[X_1] + \cdots + E[X_n]}{n} = \frac{n\,E[X]}{n} = E[X]$$

(since X_i's are iid).

$$\text{* } Var[\bar{X}] = \frac{Var(X_1 + X_2 \cdots + X_n)}{n^2} \quad (\text{since } Var(aX) = a^2 Var(X))$$

$$= \frac{Var(X_1) + \cdots + Var(X_n)}{n^2} = \frac{n \cdot Var(X)}{n^2} \quad \left(\text{since } X_i\text{'s are iid}\right)$$

$$= \frac{Var(X)}{n}.$$

Weak Law of Large Numbers (WLLN)

Let $X_1, X_2, \ldots, X_n$ be iid r.v.'s with a finite expected value $E[X_i] = \mu < \infty$, then for any $\epsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\bar{X} - \mu\right| \geq \epsilon\right) = 0$$

Proof of WLLN: $\bar{X}$ (sample mean) is a r.v. with mean μ & Variance σ^2/n.

Applying Chebyshev's inequality

$$P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{Var(\bar{X})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0 \text{ as } n \to \infty.$$

$$\Rightarrow \boxed{\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0}$$

Central Limit Theorem (CLT).

Let $X_1, X_2, \ldots, X_n$ be iid r.v.'s with mean $\mu < \infty$, variance $0 < \sigma^2 < \infty$, Then, the random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 \cdots + X_n - n\mu}{\sqrt{n}\cdot\sigma}$$

Converges in distribution to a standard Gaussian r.v. as $n \to \infty$.

$$\lim_{n\to\infty} \underbrace{P(Z_n \leq z)}_{\text{CDF of } Z_n} = \underbrace{\Phi(z)}_{\text{CDF of } \mathcal{N}(0,1)} \quad \forall z$$

$\boxed{\text{CLT applies no matter what is the underlying distrib. of } X_i\text{'s (ie they need \underline{NOT} be Gaussian)}}$

Eg: Suppose X_i's are Bernoulli(p) ; $E[X_i] = p = \mu$

$\underbrace{Var[X_i]}_{\sigma^2} = p(1-p)$

$$Y_n = X_1 + X_2 \cdots + X_n$$

$$Y_n \sim \text{Binomial}(n, p)$$

$$Z_n = \frac{Y_n - n\mu}{\sqrt{n}\,\sigma} = \frac{Y_n - np}{\sqrt{np(1-p)}}$$

say $p = 1/3$

PMF of Z_1

$$Z_1 = \frac{X_1 - p}{\sqrt{p(1-p)}}$$

$$Z_2 = \frac{X_1 + X_2 - 2p}{\sqrt{2p(1-p)}}$$

increasing n

$$Z_3 = \frac{X_1 + X_2 + X_3 - 3p}{\sqrt{3p(1-p)}}$$

$\vdots$

$$Z_{30} = \frac{X_1 + X_2 \cdots + X_{30} - 30p}{\sqrt{30p(1-p)}}$$

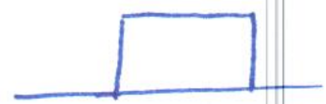
$\rightsquigarrow$ starts looking like $\mathcal{N}(0,1)$

<u>Eg</u>: X_i's $\sim$ unif $[0,1]$ $\quad \mu = 1/2$, $\sigma^2 = 1/12$

$$Z_n = \frac{X_1 + \cdots + X_n - \frac{n}{2}}{\sqrt{n/12}}$$

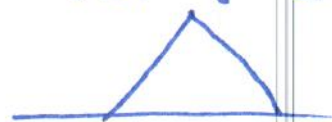
$$Z_1 = \frac{X_1 - 1/2}{\sqrt{1/12}}$$

PDF of Z_1



$$Z_2 = \frac{X_1 + X_2 - 1}{\sqrt{2/12}}$$

PDF of Z_2



$$Z_3 = \frac{X_1 + X_2 + X_3 - 2/3}{\sqrt{3/12}}$$

PDF of Z_3

$\vdots$

$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - \frac{30}{2}}{\sqrt{30/12}}$$

PDF of Z_{30}

Why is CLT useful?

- In many applications, a desired r.v. is often the sum of a large number of independent r.v.'s. & we can invoke CLT to justify the use of Normal approximation

- Examples:
  - Lab. measurement errors modeled as normal
  - Comms/Signal processing $\rightarrow$ Gaussian noise is the most frequently used model.
  - Finance, change/fluctuations in stocks/assets $\sim$ Normal.
  - Random sampling from a population to obtain statistical knowledge.

- CLT also significantly simplifies analysis & calculations.

How to apply CLT :

1. Write the r.v. Y (of interest) as a sum of n iid r.v.'s (X_i's)
$$Y = X_1 + \cdots + X_n$$

2. Find $E[Y] = n\mu$ $\qquad (\mu = E[X_i]\ ;\ \sigma^2 = Var[X_i])$
$$Var[Y] = n\sigma^2$$

3. From CLT, we know $\dfrac{Y - E[Y]}{\sqrt{Var[Y]}} = \dfrac{Y - n\mu}{\sqrt{n}\,\sigma} \sim \mathcal{N}(0,1)$

To find $P(y_1 \le Y \le y_2)$, we can write

$$P(y_1 \le Y \le y_2) = P\left(\frac{y_1 - n\mu}{\sqrt{n}\,\sigma} \le \frac{Y - n\mu}{\sqrt{n}\,\sigma} \le \frac{y_2 - n\mu}{\sqrt{n}\,\sigma}\right)$$
$$\underset{(\text{for large } n)}{\cong} \Phi\left(\frac{y_2 - n\mu}{\sqrt{n}\,\sigma}\right) - \Phi\left(\frac{y_1 - n\mu}{\sqrt{n}\,\sigma}\right)$$

Eg 1 : In a communication system, a data packet has 1000 bits. Due to noise, each bit is received in error w. prob. 0.1. Assuming that errors occur independently, what is the prob. that there are more than 120 errors in a data packet?

Solution : Let $X_i \to$ represent the error for i^{th} bit

$$X_i = \begin{cases} 1 & \text{if } i^{th} \text{ bit is received with error} \quad (w.p.)\ 0.1 \\ 0 & \text{error free} \quad (w.p.)\ 0.9. \end{cases}$$

$$\Rightarrow X_i \sim \text{Bernoulli}(P)$$
$$(P = 0.1)$$

Let $Y = X_1 + X_2 + \dots + X_n$

$\Rightarrow Y$ = Total number of bits received <u>with</u> error.

$E[X_i] = \mu = p = 0.1$

$Var(X_i) = \sigma^2 = p(1-p) = 0.09$ $(n = 1000)$

We are interested in

$$P(Y > 120) = P\Big(\underbrace{\frac{Y - n\mu}{\sqrt{n}\,\sigma}} > \frac{120 - n\mu}{\sqrt{n}\,\sigma}\Big)$$

$$= P\Big(Z_n > \frac{120-100}{\sqrt{90}}\Big)$$

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

$$\approx 1 - \Phi\Big(\frac{20}{\sqrt{90}}\Big) = 0.0175$$

Eg2: Packet transmission times on an internet link are iid with mean m & variance σ^2. Suppose that we transmit n packets & hence the total expected time is nm. Using CLT, approximate the probability that the transmission time for n packets is greater than twice the Expected time.

Solⁿ: $Y = X_1 + \dots + X_n$ $E[Y] = nm$

We are interested in

$$P(Y > 2E[Y]) = P(Y > 2nm)$$

$$= P\Big(\frac{Y - nm}{\sqrt{n}\,\sigma} > \frac{2nm - nm}{\sqrt{n}\,\sigma}\Big)$$

$$= P\Big(Z_n > \frac{nm}{\sqrt{n}\,\sigma}\Big) = 1 - \Phi\Big(\frac{m\sqrt{n}}{\sigma}\Big)$$

$$\Rightarrow \boxed{P(Y > 2E[Y]) \approx 1 - \Phi(m\sqrt{n}/\sigma)}$$