Recall:

[Strong of law large numbers] let $X_1, X_2, \ldots X_n, \ldots$ be an i.i.d. sequence of random variables. $E[X_i] = \mu$ ←

Then

$$\frac{1}{n} \sum_{i=1}^{n} X_i := \bar{X}_n$$

$$\mathbb{P}\left(\lim_{n\to\infty} \bar{X}_n = \mu \right) = 1$$

Application: If we were interested in an event A in an experiment. Repeat the experiment n times

$$X_i = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not} \end{cases}$$

$$p = P(A) \quad (\in \text{model})$$

(Sample proportion) $$\hat{p}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{\#\text{times } A \text{ occurs}}{n}$$

$$\mathbb{P}\Big(\underbrace{\lim_{n\to\infty} \hat{p}_n}_{\text{limiting Relative frequency of } A} = p \Big) = 1 \qquad [\text{SLLN}]$$

Probability of A

Central limit Theorem

Recall :- - $X \sim \text{Normal}(\mu, \sigma^2)$, i.e it is a continuous random variable with p.d.f given by

$$f_X(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma}, \quad x \in \mathbb{R}$$

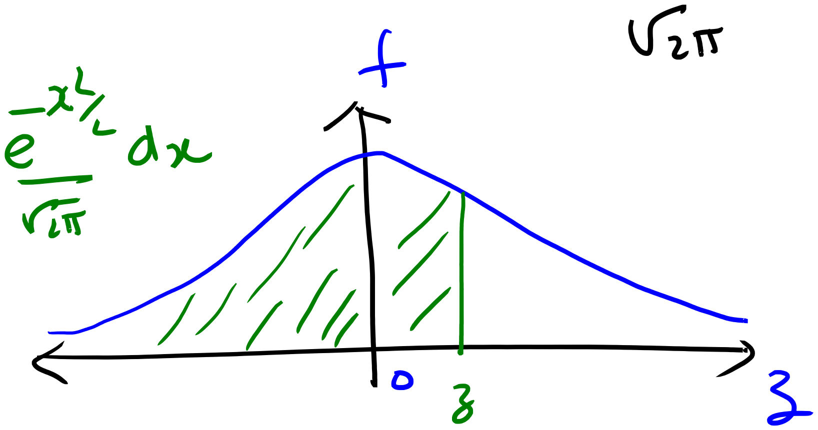
- $E[X] = \mu$, $\quad Var[X] = \sigma^2$

- $Z = \dfrac{X-\mu}{\sigma}$ then $Z \sim \text{Normal}(0,1)$

$$f_Z(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}} \quad z \in \mathbb{R}$$

- $\mathbb{P}(Z \leq z) = \displaystyle\int_{-\infty}^{z} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx$

(Numerically)



$\boxed{\text{Use symmetry \& Normal tables}}$

Applications :- The Normal distribution —

- occurs naturally in many places.
  E.g :- Heights of individuals
  # of leafs on a tree

- arises as a sums of independent random variables

  Some of the r.v. contribute low / middle / high

  yielding a frequency polygon that is symmetrical

  - this can be established theoretically.

Recall :- let $X_i \sim \text{Bernoulli}(p)$ $i \geq 1$ & independent
$0 < p < 1$

$$S_n = \sum_{i=1}^{n} X_i \qquad S_n \sim \text{Binomial}(n, p)$$

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad — \ (*)$$

- $p = \lambda/n$, $\lim_{n \to \infty} \mathbb{P}(S_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ $k \in \{0, 1, 2, ..$

[Poisson Approximation)
$np = \lambda$

- $(*)$ hard to compute for large n, $\mathbb{P}(a < S_n \leq b) = ?$
  $a, b \in \mathbb{R}_+$

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

$$\left.\begin{aligned} p &= E[X_1] \\ p(1-p) &:= \sigma = \text{Var}[X_1] \end{aligned}\right\} \Rightarrow \begin{aligned} E[S_n] &= np \\ \text{Var}[S_n] &= np(1-p) \end{aligned}$$

Standardized

- $$Z_n = \frac{S_n - np}{\sqrt{n}\,\sigma} \qquad E[Z_n] = 0, \quad \text{Var}[Z_n] = 1$$

$\alpha, \beta \in \mathbb{R}$

$$\mathbb{P}(\alpha < Z_n \le \beta)$$

$$= \mathbb{P}(\alpha\sqrt{n}\sigma + np < S_n \le \beta\sqrt{n}\sigma + np)$$

$$= \underbrace{\sum_{k=\lfloor \alpha\sqrt{n}\sigma + np \rfloor}^{\lfloor \beta\sqrt{n}\sigma + np \rfloor}}_{\text{Stirling's formula + Riemann Approximation}} \binom{n}{k} p^k (1-p)^{n-k}$$

$$\longrightarrow \int_\alpha^\beta \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx$$

$$\left| P(\alpha < Z_n \le \beta) - \mathbb{P}(\alpha < Z \le \beta) \right| \longrightarrow 0 \quad \text{as } n \to \infty$$

[DeMoivre – Laplace C.L.T.] Let $X_1, X_2, \dots, X_n$ be an i.i.d. sequence of Bernoulli(p) random variables, $0 < p < 1$.

Then $S_n = \sum_{i=1}^n X_i$

$$\circledast \quad \mathbb{P}\left(\alpha < \frac{S_n - np}{\sqrt{np(1-p)}} \le \beta\right) \xrightarrow[n\to\infty]{} \int_\alpha^\beta \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx$$

$\alpha, \beta \in \mathbb{R}$

$$\bar{X}_n = \frac{S_n}{n}; \quad \mathbb{P}\left(\alpha < \frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}} \le \beta\right) \xrightarrow[n\to\infty]{} \int_\alpha^\beta \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx$$

"Binomial converges in distribution to Normal"

Central limit theorem: let $X_1, X_2, \ldots X_n$ be i.i.d random variables. $E[X_1] = \mu$ and $\operatorname{var}[X_1] = \sigma^2$.

Then $$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\mathbb{P}\left(\alpha < \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq \beta \right) \xrightarrow[n \to \infty]{} \int_{\alpha}^{\beta} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx$$

i.e $\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ converges in distribution to Normal $(0,1)$.

Application of the result: $Z \sim \text{Normal}(0,1)$

$$- \mathbb{P}(-1.96 \leq Z \leq 1.96) \cong 0.95$$

From C.L.T for large n

$$- \mathbb{P}\left(-1.96 \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq 1.96\right) \cong 0.95$$

$$- \mathbb{P}\left(-\frac{\sigma}{\sqrt{n}} 1.96 \leq \bar{X} - \mu \leq 1.96 \frac{\sigma}{\sqrt{n}}\right) \simeq 0.95$$

$$\mathbb{P}\left(\mu \in \left(\overline{X} - \frac{1.96}{\sqrt{n}}\sigma,\ \overline{X} + \frac{1.96}{\sqrt{n}}\sigma\right)\right) \simeq 0.95$$

μ = true mean

$X_1, X_2, \ldots X_n$ i.i.d $E[X_1] = \mu$

$var[X_1] = \sigma^2$

$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ = sample mean

$$\left(\overline{X} - \frac{1.96}{\sqrt{n}}\sigma,\ \overline{X} + \frac{1.96}{\sqrt{n}}\sigma\right) \equiv \text{95\% Confidence interval for } \mu.$$

[SLLN] $\mathbb{P}\left(\lim_{n\to\infty} \overline{X} = \mu\right) = 1$ $\overline{X} \equiv$ Estimator for μ

[CLT] $\mathbb{P}\left(\mu \in \left(\overline{X} - \frac{1.96\sigma}{\sqrt{n}},\ \overline{X} + \frac{1.96\sigma}{\sqrt{n}}\right)\right) \simeq 0.95$

for large n

[Berry Essen bounds]

Questions :

- How large should n be to apply Central limit theorem?

$$\frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma} \approx \text{Normal}(0,1)$$

- Is 'n' universal no matter what the distribution of X_1 is?

- What does the confidence interval actually mean?

- What if σ is not known?

- Can we say something for small n?

- (X, Y) are two random variables, $\left(\frac{\text{discrete}}{\text{continuous}}\right)$

Distribution of $E[X|Y]$?

$$g(y) = E[X \mid Y = y] = \begin{cases} \sum_{t \in \text{Range}(X)} t\, \mathbb{P}(X = t \mid Y = y) \\ \int_{-\infty}^{\infty} x \frac{f(x,y)}{f_Y(y)}\, dx \end{cases}$$

$$E[X \mid Y] = g(Y) \equiv \text{random variable}$$

Finds its distribution.

- Median $\equiv m$

Definition — $\mathbb{P}(X < m) = \mathbb{P}(X \geq m) = \frac{1}{2}$