Convergence.

Convergence of Deterministic Sequences:

Let $a_1, a_2, \ldots$ be a seq. of real #s.

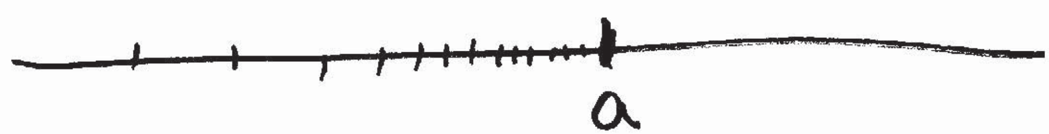
Let a be another real #.

Say a_n converges to a or

$$\lim_{n \to \infty} a_n = a \quad \text{if}$$

$$\forall \epsilon > 0, \quad \exists n_0 \text{ s.t. } |a_n - a| \le \epsilon \quad \forall n \geq n_0$$

Ex: $a_i = \frac{1}{i}$ $\qquad \frac{1}{i^2}$



Convergence in Probability:

let $Y_1, Y_2, \ldots$ be a seq of R.V.

a = Real #.

Say Seq Y_n converges to a in probability if

$\forall \epsilon > 0$ we have.

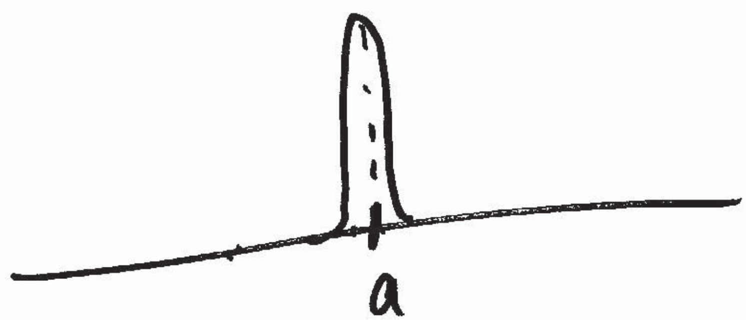
$$\lim_{n \to \infty} P(|Y_n - a| > \epsilon) = 0$$

Aliter $\forall \epsilon > 0$, $\delta > 0$ $\exists n_0$ s.t.

$$P(|Y_n - a| > \epsilon) \leq \delta \quad \forall n \geq n_0$$

accuracy level. Confidence level.

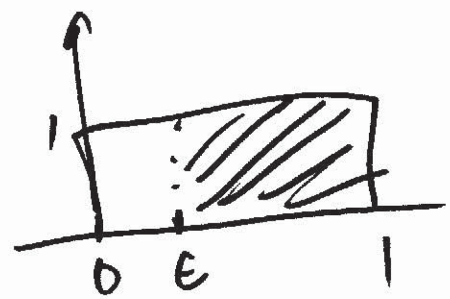
Intuitively: if $Y_1 \ldots Y_2 \ldots$ converge in probability to a, then almost all the PMF or PDF of Y_n is concentrated in an ϵ interval around a, for large values of n



Ex Consider $X_1, X_2 \ldots, X_n$ independent uniformly distributed R.V in $[0, 1]$.

$$Y_n = \min\{X_1, \ldots, X_n\}$$

$$\lim_{n \to \infty} Y_n = 0$$



for $\epsilon > 0$, use independence X_n:

$$P(|Y_n - 0| \geq \epsilon) = P(X_1 \geq \epsilon, X_2 \geq \epsilon, X_3 \dots, X_n \geq \epsilon)$$

indp

$$= P(X_1 \geq \epsilon) P(X_2 \geq \epsilon) \dots P(X_n \geq \epsilon)$$

$$= (1-\epsilon)^n$$

Take limit as $n \to \infty$

$$P \lim_{n \to \infty} (|Y_n - 0| \geq \epsilon) = \lim_{n \to \infty} (1-\epsilon)^n = 0$$

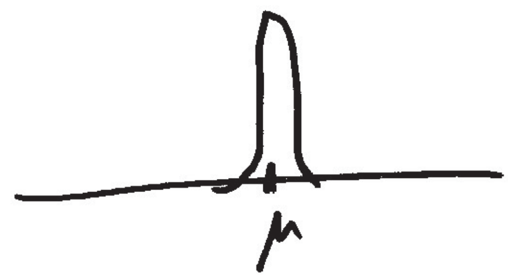
Y_n converge to zero in Probability.

Central Limit Theorem (CLT)

X_i I.I.D. R.V. $\longrightarrow$ mean μ, sd. σ

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

Shows us that pdf for M_n is concentrated in the near vicinity of μ.



Weak law of large #s:

$P(|M_n - \mu| \geq \epsilon) \to 0$ as $n \to \infty$.

How about: $S_{\underline{n}} = X_1 + X_2 \cdots + X_n = n M_n$.

$\mathrm{Var}(S_n) \to \infty$ as $n \to \infty$

How $\dfrac{S_n - n\mu}{\sqrt{n}}$

$$\mathrm{Var}(\ ) = \sigma^2$$

How about $\boxed{Z_n = \dfrac{S_n - \mu n}{\sigma \sqrt{n}}}$ mean 0 Var 1

$$\mathrm{Var}(Z_n) = 1.$$

mean$[Z_n]=0$

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

$$\mathrm{Var}[Z_n] = \frac{1}{\sigma^2 n}\left[\mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \cdots \mathrm{Var}[X_n]\right] = 1$$

CLT

Let $X_1, X_2, \ldots X_n$ be a seq of I.I.D. R.V with mean μ var σ^2.

$$Z_n = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

~~the~~ a CDF of Z_n converges in probability to CDF of a normal (gaussian) with mean 0, var 1.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx$$

$$\lim_{n \to \infty} P(Z_n < z) = \Phi(z) \quad \forall z$$

Ex loading plane 100 package weights are indp. R.V. uniform 5 → 50 lbs.

$\mu = 27.5$

$\sigma^2 = 168.75$

Compute P(Total weight > 3000 lbs)

S_{100} = Sum of 100 packages.

$P(S_{100} < 3000) =$

$$\boxed{Z_n = \frac{X_1 + X \dots + X_n - 100 \cdot \times 27.5}{\sqrt{168} \times 100}}$$

$$P(X_1 + X_2 + \dots + X_{100} < 3000) =$$

$$P(X_1 + X_2 + \dots + X_{100} - (100 \cdot \times 27.5) < 3000 - (100 \times 27.5)$$

$=$

$$= P\Bigg(\underbrace{\frac{X_1 + X_2 + \cdots + X_{100} - (100 \times 27.5)}{\sqrt{168 \times 100}}}_{Z_n \text{ normal}(0,1)} < \frac{3000 - (100 \times 27)}{\sqrt{168 \times 100}} \Bigg)$$

$$= P(Z_n < 1.92)$$

Φ Table

$$= 0.9726$$

$$\Rightarrow P(S_{100} < 3000) = 0.9726$$

$$P(S_{100} > 3000) = 1 - 0.9726 = \underline{0.0274}$$

2% ←

Polling

p = % of population voting for Kerry.

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

X_i = indp. Bernoulli R.V. with parameter p.

$E[M_n] = p$. $\quad Var[M_n] = \frac{p(1-p)}{n}$

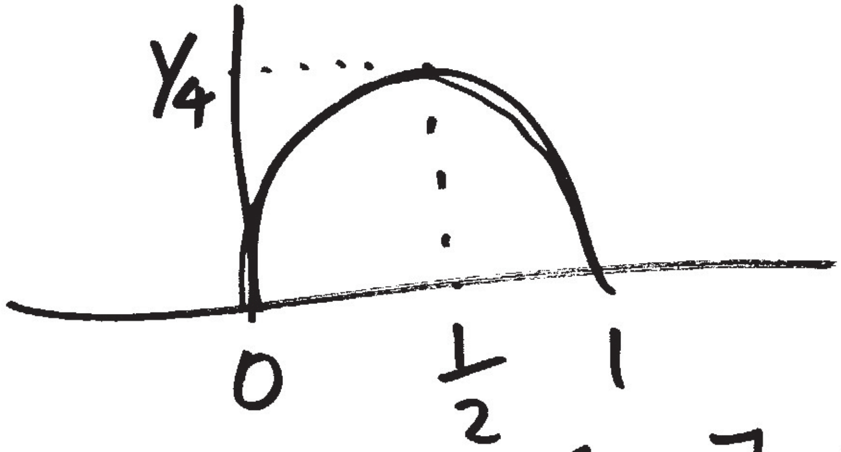
CLT: $X_1 + X_2 + \cdots + X_n \sim$ Normal.

$M_n \sim$ normal.

Last Time $\boxed{\begin{aligned} &P(|M_n - p| > \epsilon) \\ &= 2P(M_n - p > \epsilon) \end{aligned}}$: Chebyshev inequality

Apply CLT: But don't know variance

$p(1-p)$



$p(1-p) \le 1/4 \qquad \mathrm{Var}[M_n] \le \frac{1}{4n}$

$$\boxed{Z_n = \frac{\overbrace{X_1+X_2+\cdots}^{M_n}\ \overbrace{+n\mu}^{p}}{\sigma\sqrt{n}}}$$

$$P(M_n - p > \epsilon) = P\left(\underbrace{\frac{M_n - p}{\text{s.d of } M_n}}_{\text{normal}(0,1)} \ge \frac{\epsilon}{\text{s.d. of } M_n}\right)$$

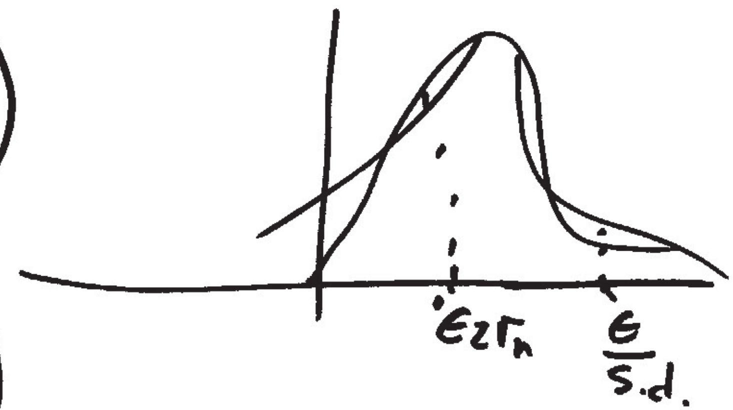
CLT

$$P(M_n - p > \epsilon) \approx 1 - \Phi\left(\frac{\epsilon}{\text{s.d}}\right)$$

$$p(1-p) \le \frac{1}{4} \Longrightarrow Var[M_n] \le \frac{1}{4n} \Longrightarrow$$

$$s.d.M_n = s.d\ of\ M_n < \frac{1}{2\sqrt{n}} \Longrightarrow$$

$$\frac{1}{s.d.M_n} > 2\sqrt{n} \Longrightarrow \frac{\epsilon}{s.d M_n} > \epsilon 2\sqrt{n} \Longrightarrow$$



$$\Longrightarrow \Phi\left(\frac{\epsilon}{s_d}\right) > \Phi\left(\underline{\epsilon 2\sqrt{n}}\right)$$

$$1 - \Phi\left(\frac{\epsilon}{s_d}\right) < 1 - \Phi(\epsilon 2\sqrt{n})$$

$$\boxed{P(M_n - p > \epsilon) < 1 - \Phi(\epsilon 2\sqrt{n})}$$

$\underline{Ex}$ $n = 100$ $\epsilon = 0.1$

$$P(|M_n - p| > 0.1) < 2\left(1 - \Phi(0.1 \times 2 \times 10)\right)$$

$$P(|M_n - P| > 0.1) < 0.046$$

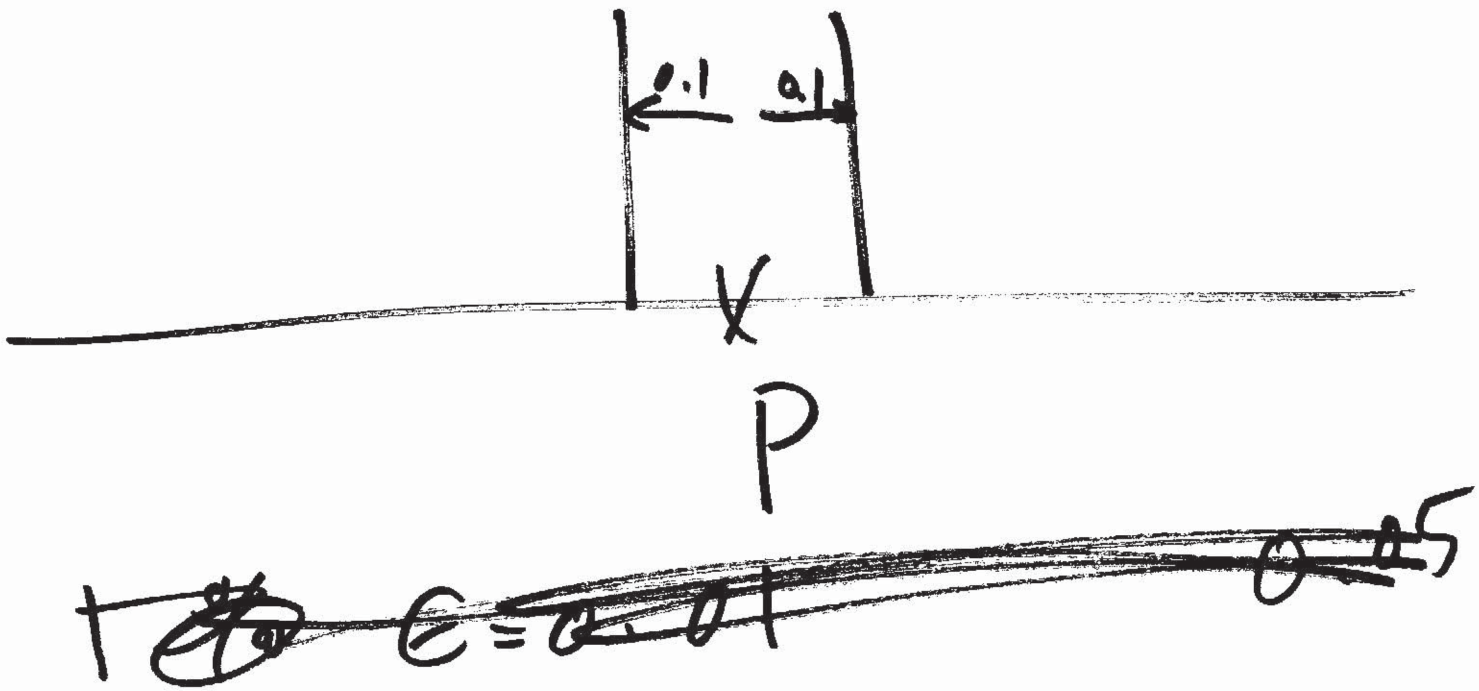
30%

$\uparrow$

$$M_n = \frac{X_1 + \dots X_n}{n}$$ mean $= 40\%$

$\downarrow$

$p = 50\%$



Q howlage of acaple do I need To be with 1 % $\epsilon = 0.01$ of p
with prob of at least 95 %

$$2 - 2\,\Phi(0.01 \cdot 2 \cdot \sqrt{n}) \leq 0.05$$

lookup $\Phi(1.96) = 0.975$.

$$2 \times 0.01 \times \sqrt{n} \geq 1.96$$

$$\Longrightarrow \boxed{n \geq 9604}$$

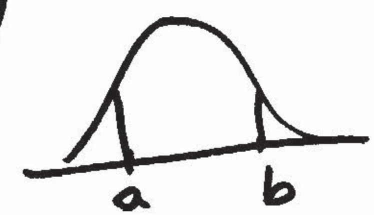
CLT

Consider Bernoulli Process with parameter p.

n Trials $K = \#$ of successes.

$$P_K(k_0) = \binom{n}{k_0} p^{k_0} (1-p)^{n-k_0}$$

$$P_{or}(a \le K \le b) = \sum_{k_0=a}^{b} \binom{n}{k_0} p^{k_0} (1-p)^{n-k_0}$$



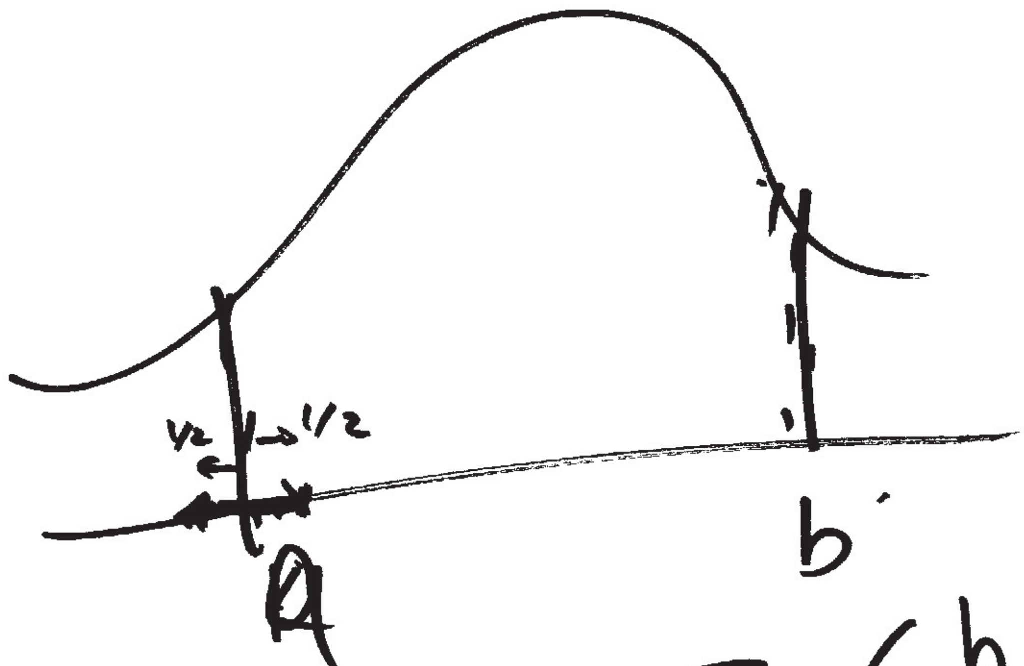
$$K = X_1 + X_2 + \cdots + X_n.$$

iid. Bernoulli

$\xrightarrow{CLT}$ $$\text{Prob}(a \le K \le b) = \Phi\left(\frac{b - E(k)}{\sigma_k}\right) - \Phi\left(\frac{a - E(k)}{\sigma_k}\right)$$

$E(k) = np$ $\sigma_k = \sqrt{np(1-p)}$

$a = b$



$$\Pr(a \le K \le b) = \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

De Moivre-Laplace th.

limit.