Law of Large Numbers

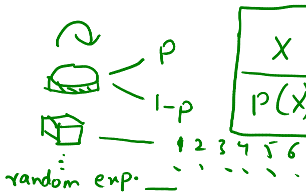
$$E(X) \approx \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N}$$

N ← large

Central Limit Theorem

① i.i.d random variables

independent, identically distributed



Bernoulli R.V

X	0	1
P(X)	1-p	p

$X_1 \quad X_2 \quad X_3 \quad X_4 \quad \cdots \quad X_N$

$\underbrace{\qquad\qquad\qquad}_{\text{i.i.d random variables.}}$

② Very often, random phenomena can be represented as a sum of other random phenomena.

$Y \sim Bin(n, p)$

$$\underset{\substack{\uparrow \\ \text{Bin}}}{Y} = \underbrace{X_1 + X_2 + \cdots + X_n}_{\text{sum of i.i.d Bernoulli r.v's.}} = \# \text{ of heads}$$

μ and σ^2 of a Bernoulli R.V

X	0	1
P(X)	1-p	p

$E(X) = 0(1-p) + 1(p)$
$= p$

$Var(X) = (0-p)^2(1-p) + (1-p)^2 p$
$= p^2(1-p) + p(1-p)^2$
$= p(1-p)(p + (1-p))$
$= p(1-p)$

$(X-\mu)^2$	$(0-p)^2$	$(1-p)^2$
P	1-p	p

③ (a) $X_1, X_2, \cdots, X_N$ i.i.d r.v's with mean μ and variance σ^2.

(b) $Y = X_1 + X_2 + \cdots + X_N$

(c) According to the Central Limit Theorem (CLT)

$Y \sim \mathcal{N}(N\mu, N\sigma^2)$ as $N \to \infty$.

④ How large should N be?

If $N \geq 50$, it's usually safe to apply CLT.

Often it applies even for smaller N.

$Y \sim Bin(n, p)$

$$Y = \underset{\substack{\downarrow \\ \mu = p \\ \sigma^2 = p(1-p)}}{X_1} + \underset{\substack{\downarrow \\ \mu = p \\ \sigma^2 = p(1-p)}}{X_2} + \cdots + X_n$$

X	- - - - -
p(x)	- - - - -

According to the CLT,

$Y \sim \mathcal{N}(n\mu, n\sigma^2)$

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

$\underset{(n,p)}{Bin} \xrightarrow{\text{approx.}} \underset{(np,\, np(1-p))}{\mathcal{N}}$

①

X	0	1
$p(X)$	0.9	0.1

9 pieces

$\boxed{0}\boxed{0}\boxed{0}\cdots\boxed{0}\ \boxed{1}$

Box

X	0	1
$P(X)$	0.9	0.1

same random experiment

② Roll a fair die N times.
Take the sum.

$\longleftrightarrow$ same random experiment

$\boxed{1}\boxed{2}\boxed{3}\boxed{4}\boxed{5}\boxed{6}$

- Randomly draw from this box N times with replacement
- Take the sum

X_1 X_2 X_3

1 2 3 4 5 6

$$Y = X_1 + \underbrace{X_2 + \cdots + X_N}_{\text{i.i.d}}$$

$\mu = 3.5$
$\sigma^2 = 2.97$

1	2	3	4	5	6
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$\mu = 3.5$

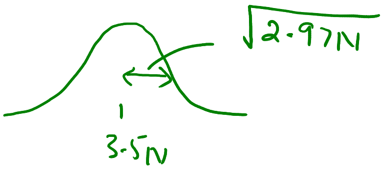
Accordy to CLT

$$Y \sim N(3.5N,\ \underbrace{2.97N}_{\text{Variance}})$$

$(1-3.5)^2$	$(2-3.5)^2$	$\cdots$	$(6-3.5)^2$
$\frac{1}{6}$	$\frac{1}{6}$	$\cdots$	$\frac{1}{6}$

$\sigma^2 = 2.97$

Y



Example

Sum of 25 draws from the box $\boxed{0}\boxed{2}\boxed{3}\boxed{4}\boxed{6}$

X_i = value of i-th draw

X_i	0	2	3	4	6
$P(X_i)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

$\boxed{0\ 2\ 2\ 4\ 6}$

Y	0	2	4	6
$P(Y)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

$E(X_i) = ?$

$Var(X_i) = ?$

$StdDev(X_i) = \sqrt{Var(X_i)}$

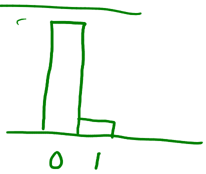
$$P(65 \le \text{sum} \le 85) = ?$$

---

9 pieces

$\boxed{0}\boxed{0}\boxed{0}\cdots\boxed{0}\ \boxed{1}$ $\longrightarrow$

Box

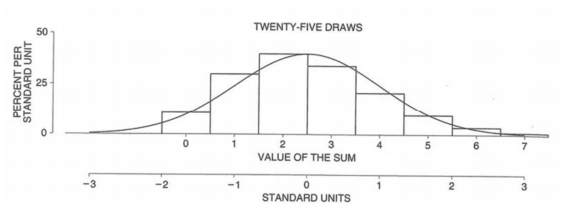


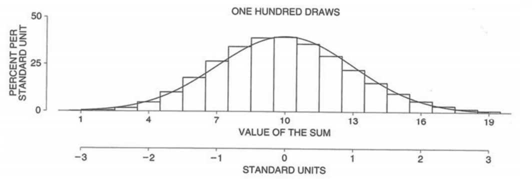
$S = 0, 1, 2, 3, \cdots, 25$

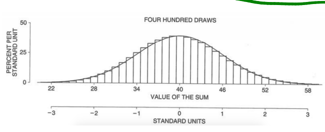
$X_1 + X_2 + \cdots + X_{25}$ — 25 draws — sum $\leftarrow$ repeatedly

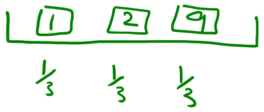
$X_1 + X_2 + \cdots + X_{100}$ — 100 draws — sum $\leftarrow$ repeatedly

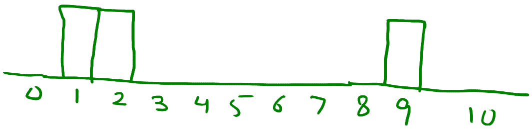
$X_1 + X_2 + \cdots + X_{400}$ — 400 draws — sum $\leftarrow$ repeatedly

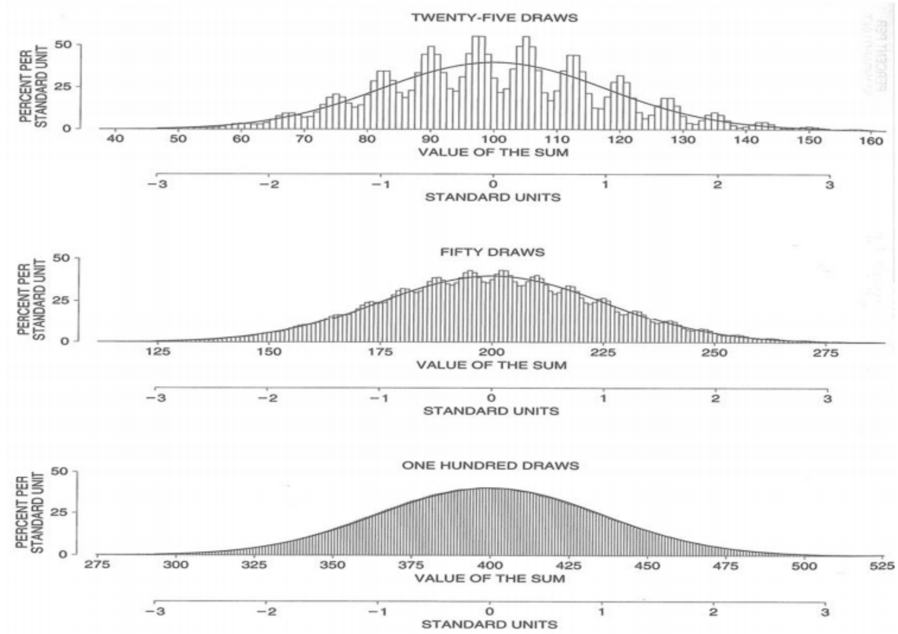










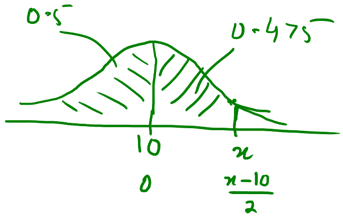


σ^2

$X \sim N(10, 4)$ $\quad \sigma = 2$

$P(X < x) = 0.975$

Find x?



$10 \;—\; x \;—\; 0.475$

$0 \longleftarrow \frac{x-10}{2}$ area is 0.475

$0 \longleftarrow 1.96$ area is 0.475

$\Rightarrow \frac{x-10}{2} = 1.96 \Rightarrow x = (1.96)2+10$
$= 13.92$