X, Y are 2 RVs

Rolling of dice

	1	2	3	4	5	6
X	0	1	0	1	0	1
Y	0	1	1	0	1	0

$X = 1$ if outcomes are 2, 4, 6
$\quad\;\; 0$ o.w.

$Y = 1$ if outcomes are prime 2, 3, 5
$\quad\;\; 0$ o.w.

$$p_X(x) = \sum_y p(x, y)$$

$$p_X(x=0) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

$$p_X(x=1) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

Correlation between X and Y.

jPMF: $P(x=0, y=0) \equiv P(x=0 \cap Y=0) = P(\{1\}) = \frac{1}{6}$

$P(x=0, Y=1) = \frac{1}{3}$

$P(x=1, Y=0) = \frac{1}{3}$

$P(X=1, Y=1) = \frac{1}{6}$

0 o.w.

$$Cov(X, Y) = E[XY] - E[X]\,E[Y] = \frac{1}{6} - \frac{1}{4} = \frac{2-3}{12} = -\frac{1}{12}$$

$$E[XY] = \sum_x \sum_y xy\, p(x, y) = \frac{1}{6}$$

$$E[X] = \sum_x \sum_y x\, p(x, y) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$$

$$E[Y] = \frac{1}{2}$$

Distribution of Sample Mean

$N = 5$

the age of 5 children

Children	Ages (x_i)
1	6
2	8
3	10
4	12
5	14

$$\mu = \frac{\sum_i x_i}{5} = 10$$

$$\sigma^2 = \frac{\sum_{i=1}^{5} (x_i - \mu)^2}{5} = \frac{40}{5} = 8$$

$$s^2 = \frac{\sum_{i=1}^{5} (x_i - \mu)^2}{N-1} = 10$$

We want to prepare dist. of sample means.

$n = 2$

$5^2 = 25$

1st draw	2nd draw: 6	8	10	12	14
6	(6, 6) 6	(6, 8) 7	(6, 10) 8	(6, 12) 9	(6, 14) 10
8	(8, 6) 7	(8, 8) 8	(8, 10) 7	(8, 12) 10	(8, 14) 12
10	(10, 6) 8	(10, 8) 9	(10, 10) 10	(10, 12) 11	(10, 14) 12
12				(12, 12) 12	(12, 14) 13
14					(14, 14) 14

Frequency table

Frequency Table

$\bar{x}$	Frequency	Relative Freq.
6	1	1/25
7	2	2/25
8	3	3/25
9	4	⋮
10	5	⋮
11	4	⋮
12	3	⋮
13	2	⋮
14	1	1/25

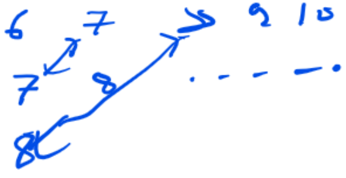
$$\bar{X} = \frac{\sum X_i}{n}$$

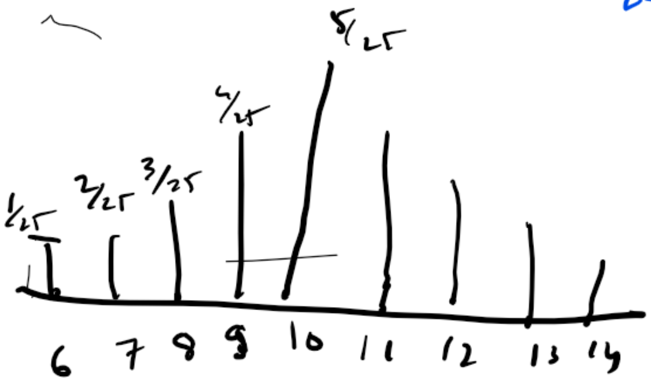
7

(8, 6)

(6, 8)

$$n \to \infty \quad N\left(\mu, \frac{\sigma^2}{\sqrt{n}}\right)$$





$$\mu_{\bar{X}} = 6 \times \frac{1}{25} + \cdots\cdots = 10$$

$$\frac{\sigma^2}{n} = \frac{8}{2}$$

$$\sigma_{\bar{X}}^2 = 4 \qquad \sigma^2 \neq \sigma_{\bar{X}}^2$$

$$\frac{\sigma}{\sqrt{n}} = \text{Standard error} \qquad = 4 = \sigma_{\bar{X}}^2$$

Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ be a sequence of identically and independently distributed RVs, each having mean μ and variance σ^2, ~~then~~ for n large the distribution of $X_1 + X_2 + \cdots + X_n$ is approximately normal with mean $n\mu$ and variance $n\sigma^2$.

$$X = \frac{X_1 + \cdots + X_n}{n}$$

$n \to \infty$, $\quad X \overset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$

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

$$X \sim Z(0, 1)$$

Sampling W/o replacement

5 $\qquad n = 2$

${}^5C_2 \rightarrow$ different

10 $\rightarrow$ upper triangular part of the sample mean

Matrix of.

(6, 8)

(8, 6)

6, 6	(6, 8)	(6, 10)	(6, 12)	(6, 14)
	(8, 8)	(8, 10)	(8, 12)	(8, 14)
			(10, 12)	(10, 14)
				(12, 14)

$\mu_{\bar{x}} = 10$

Check! $\sigma_{\bar{x}}^2 = 3$ $\qquad$ Finite size error

$\sigma^2 = 8$

$$\frac{N-n}{N-1} \cdot \frac{\sigma^2}{n} = \frac{5-2}{4} \cdot \frac{8}{2} = 3$$

Ex:- An insurance company has 25000 automobile policy holders.

yearly claim of a p.h $\Rightarrow$ R.V.
with mean = 320
S.d. = 540

approximate the probability that the total claim exceeds 8.3 million.

Soln.

X_i the claim of ith policy holder.

$N = 25000$ $\quad X = \sum_{i=1}^{25000} X_i$

$E[X_i] = \mu = 320$ $\quad E[X] = N \times \mu = 25000 \times 320$

$\sqrt{\sigma^2} = 540$ $\quad = 8 \times 10^6$

$N \geq 30$
CLT

$N(8 \times 10^6, n\sigma^2)$ $\quad P(X > 8.3 \times 10^6)$

$\sqrt{n}\,\sigma = 540 \times \sqrt{25000} = 8.538 \times 10^4$

$\Rightarrow P\left(\frac{X - 8 \times 10^6}{8.538 \times 10^4} > \frac{(8.3 - 8) 10^6}{8.538 \times 10^4}\right)$

Check. $\Rightarrow P(Z > 3.51) \approx 0.00023$