* Population vs Sample
* Sample Statistics
    - Mean
    - Variance
    - SD

* Point Estimates
* Sampling Distribution
* Standard Error
* Uniform Distribution
    - PMF
    - PDF
    - CDF

$\boxed{\text{Population vs Sample}}$

BLR ⟹ 1,00,000 Data Scientist

Avg salary and σ of Salary

Survey ⟹

Population ⇒ 1,00,000

1) Reaching Everyone for survey is impractical and Costly

→ avg Salary μ

→ Avg $\underline{\text{Salary}}$ $\bar{x}$

* Sampling is the Solution for large populations

Sample Statistics

Population Stats	Sample Stats
① $\mu = \sum_{i=1}^{n} \frac{X_i}{n}$	① $\bar{X} = \sum_{i=1}^{n} \frac{x_i}{n}$
② σ^2 (pop Variance) $\downarrow$ $\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{n}$	② s^2 (Sample Var) $\downarrow$ $\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}$
③ $\sigma = \sqrt{\sigma^2}$	③ $s = \sqrt{s^2}$

* $n-1$ is called **Bessel's correction** and it is use for correcting the Bias due to sample

Sample Data

	x_i	$x_i-\bar{x}$
1	1000	-1000
2	2000	0
3	3000	+1000

$\bar{X} = 6/3 = 2K$

$n = 3$

[131, 150, 140, 142, 152]

$$\bar{X} \Rightarrow \frac{(131+150+140+142+152)}{5}$$

$\downarrow$

143

$$S^2 \Rightarrow \frac{x_i - \bar{X}}{4}$$

A	$\bar{X} = 143, \sigma^2 = 71$
B	$\bar{X} = 142.4, \sigma^2 = 66.3$
C	$\bar{X} = 147, \sigma^2 = 73.2$
D	$\bar{X} = 152, \sigma^2 = 64.5$

* Estimate the Avg Salary of all DS in Bengaluru

$\downarrow$

$\boxed{\text{Point Estimate}}$

$\Rightarrow$ Estimating population statistics based on Sample Statistic

BiasNess in Sampling

1) All people we sampled can have lower Salary then AVG

2) All people we sampled can have Higher Salary then AVG

Sampling Technique

① Probabilistic Sampling

② Non-Probabilistic Sampling

$\Rightarrow \frac{1}{1,00,000}$

$\leftarrow \frac{1}{1,00,000}$

Sachin

Pop ⇒ 1,00,000

Prob of 1 DS ⇒ $\frac{1}{1,00,000}$

Simple Random Sampling
(Every Entity in population has
Equal chance of Getting picked)

There are four main types of probability samples.

1. **Simple random sampling** ✓
2. **Systematic sampling**
3. **Stratified sampling**
4. **Cluster sampling**

ML

Steps for point Estimates

Step 1 : Define population

Step 2 : Determine Sample size

① certainity ⇒ 95%

② Error of Margin ⇒ 5%

Let's say $\xrightarrow[\text{for } 95\%]{}$ 5,000

Step 3 : Randomly Sample the population Based on Size

Step 4 : Collect Data of all the Samples

↓ Outcome of 4 steps

$\boxed{\text{Sample Data of Size K}}$ → Sample Statistic

Q What if we do all 4 steps multiple times

* Later

CLT ⇒ Central Limit Theorem

Sample 1
$\bar{X}_1$
S^2

Sample 2
$\bar{X}_2$
S^2

Sample 3
$\bar{X}_3$
S^2

$$\bar{\mu} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{3}$$

How far Avg of Avg is from POP mean

mean of (mean of Sample)

Sampling $\underbrace{\text{Distribution}}$ Statistic

* Standard Error
  - Quantifies Variability among multiple Sample
  - How much Sample Mean is expected to Deviate from population Mean

① If population SD (σ) is given

$$\boxed{SE = \frac{\sigma}{\sqrt{n}}}$$

n : Sample Size

② If Sample SD (s) is given

$$\boxed{SE = \frac{s}{\sqrt{n}}}$$

$\approx \sigma$

n : Sample Size

Key Take-away

$$SE \propto \frac{1}{n}$$

* Law of Large Numbers

① As sample size increases, the sample mean gets closer to population Mean

Diff b/w

SE and S

Standard Error (Sampling Distribution)

Sample 1 | Sample 2 | Sample 3

S_1 | S_2 | S_3

Quiz

① $n \Rightarrow 30$

② $\overline{x} \Rightarrow 4$

③ $SD \Rightarrow 0.13\ (\sigma)$

Q $SE \Rightarrow \frac{\sigma}{\sqrt{n}}$ or $\frac{s}{\sqrt{n}}$

$\frac{0.13}{\sqrt{30}}$

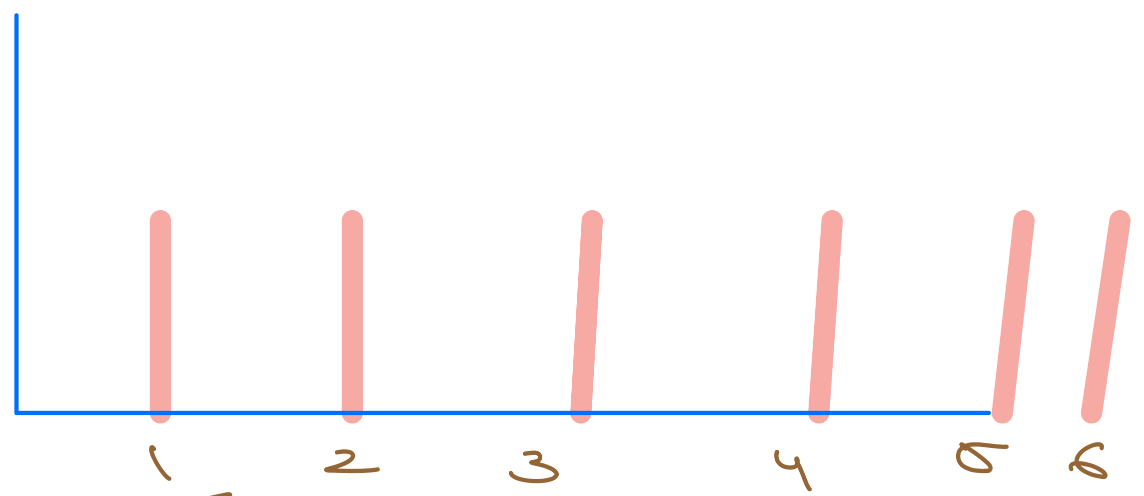
Uniform Distribution

- Probability of all Outcomes is Equal

Ex: Die roll $\rightarrow \{1, 2, 3, 4, 5, 6\}$

$\downarrow$

Uniform Discrete Distribution



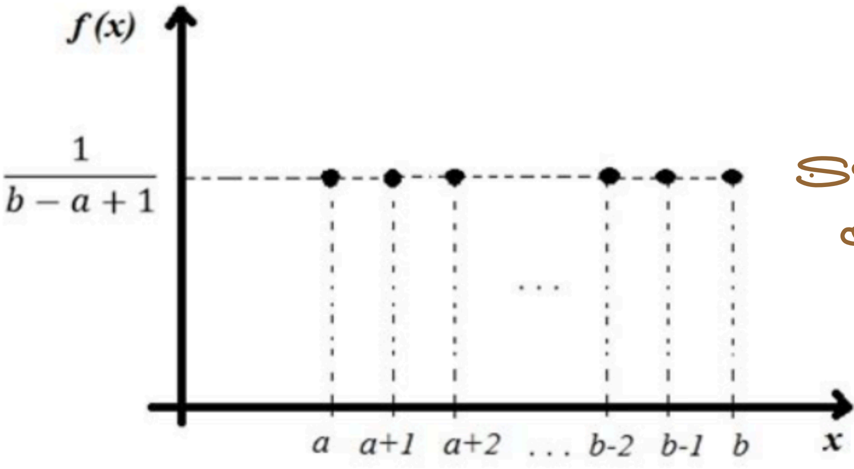
$$\boxed{\text{PMF} : \frac{1}{b-a+1}}$$

for $a \le x \le b$

$a \downarrow$ min val, $b \downarrow$ max. val

Dice Roll: $\frac{1}{6-1+1} \Rightarrow \frac{1}{6}$

Coin Toss: $\frac{1}{1-0+1} \Rightarrow \frac{1}{2} = 0.5$

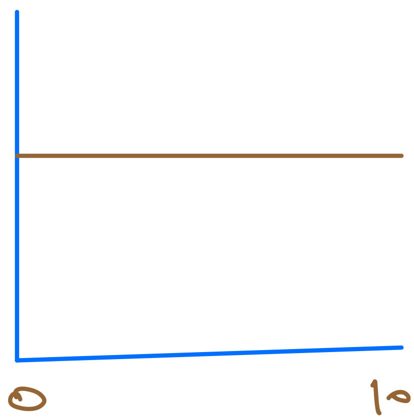


Same prob across each outcome

a = 1
b ⇒ 6 } Same as Dice

a ⇒ 0
b ⇒ 1 } Same as Coin Toss

② 0 – 10 ← Random Num Generator
↓
floating ⇒



PDF ⇒ $\frac{1}{b-a}$ $(a \leq x \leq b)$ 9.9

9.8