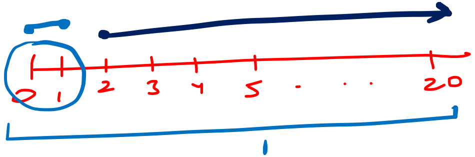
$\Rightarrow$ Sample , Sampling, Sample properties

uniform prob. dist

Probability Distributions 2



Quiz Question

cdf → cumulative prob. till that point

$P(x=0)+P(x=1)$

binom.cdf(k=1, n=20, p=0.1)

Sumit has a crush on Ankita but is too shy to ask her out. However, he hopes that Ankita may ask him out instead. Ankita is generally very friendly with Sumit, and each day there is a 10% chance that she will invite Sumit on a date. If there are 20 days in a particular month, what is the probability that Ankita asks Sumit out on at least 2 dates that month?

A) 0.52
B) 0.63
C) 0.36
D) 0.55

P(Sumit gets invited for a date) = 0.10

$n = 20$

① Is the outcome of this exp Binary?

Date < Yes / NO

x = # of dates

$x = \{0, 1, 2, 3 \dots\dots 20\}$

$P(x \geq 2) = 1 - [P(x=0) + P(x=1)]$

② prob. of success

P(Date) = 0.10

$= 1 - \left[{}^{20}C_0 (0.10)^0 (0.90)^{20} + {}^{20}C_1 (0.10)^1 (0.90)^{19} \right]$

③ fixed no. of trails

$n = 20$

= 1 − [binom.pmf(k=0, n=20, p=0.10) + binom.pmf(k=1, n=20, p=0.10)]

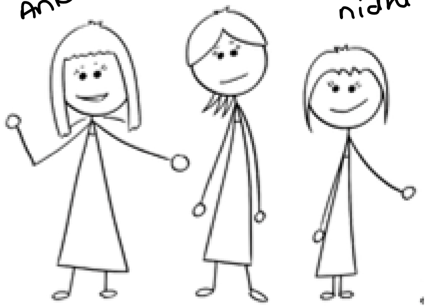
= 1 − [binom.cdf(k=1, n=20, p=0.10)]

$P(x \leq 1)$

Degree Of Freedom

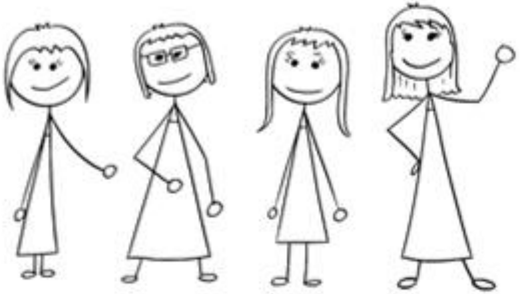
[1, 2, 3, 4, 5]   [3, 1]   [1]

Ankita   Rekha   nidhi



$\underbrace{3 \quad 1}$ $\boxed{1}$

free to choose any value



choose number such that the sum of the no. is 5

Is nidhi left with any choice = No

She is bounded to choose one

degree of freedom = n − 1

n = # of of observation

(n−1) = free to choose any value

free to deviate.

Sample

once the sample is fixed
the avg component will
not change

S

$n = 1000$

$[I_1, I_2, I_3, I_4 \ldots, I_{1000}]$

$\boxed{avg(S) = 3.79L}$

Population

India

$[I_1, I_2, I_3 \ldots I_{140cr}]$

$\downarrow$

$avg(P) = 5\ 25L$

$$\text{variance} = \sigma^2 = \frac{\sum_1 (x - \bar{x})^2}{n}$$

$\longrightarrow$ population variance

Sample statistic

$$\text{Sample variance} = \frac{\sum_1 (x - \bar{x})^2}{n-1}$$

$$\text{Sample std dev} = \sqrt{\frac{\sum_1 (x - \bar{x})^2}{n-1}}$$

$n \longrightarrow$ very high

then

Sample statistic $\longrightarrow$ population statistic

Problem.

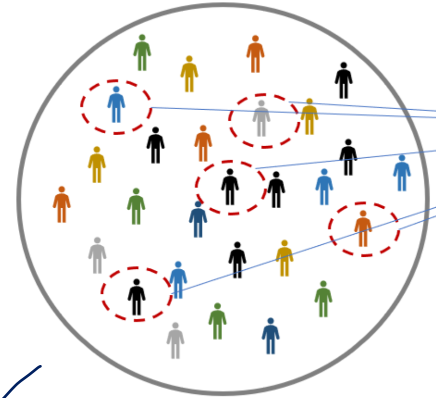
Yes the avg might change
because we are never
sure about the value of
n

&

It is impossible to
actually survey on
the entire population

50cr

Population



10,000 people

Sample

$$x = \begin{array}{cccccc} P1 & P2 & P3 & P4 & P5 & P_{10k} \\ [3, & 2, & 7, & 9, & 1 \ldots & 2] \end{array}$$

avg(x)

$\boxed{5.3}$

(the avg no. of times a person watched a movie in cinema hall the last year = 5.3)

10

can we use sample avg to infer regarding the entire population?

Q: You work for a small marketing firm and your manager wants you to find out on an average how many time a person in urban India watched a movie in a cinema hall the last year?

1. Reach out to each and every individual of the population and survey regarding how many times they have visited the movie hall the last year

It is impossible for any individual to conduct survey on such a wide range

can't expect from a small marketing firm
- → No money (Good investment req)
- → No time (Good time req.)
- → Resourcees

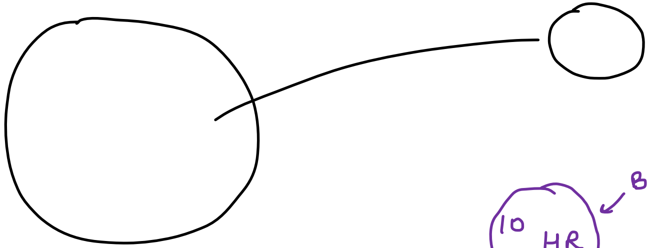
Condition → If I want the sample avg to represent the entire pop., we should get the sample to be rich → the sample should be representing the population

The procedure to select a sample out of the population is known as Sampling

Get a sample which is the representative of the population

Population

Sample



① unbiased sample

② good sample size

(get as large sample as possible so that your sample can become true representative of the population)

③ use sampling techniques

↓

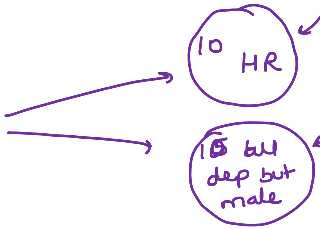
Simple Random Sampling

$$P(x) = \frac{1}{N}$$

Biased Sample

Biased sample

employees of TCS



Unbiased sample

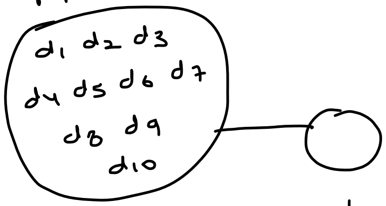
→ good prop. of people employees from each department including male and females

The prob. of selecting a data point x from the population to my sample is $\frac{1}{N}$

$$P(x) = \frac{1}{N}$$

N = total no. of ob. in the pop.

Population



$$P(d_1) = \frac{1}{10}$$

$$P(d_2) = \frac{1}{10}$$

$$P(d_3) = P(d_4) = P(d_5) = P(d_6) = P(d_7) = P(d_8) = P(d_9) = P(d_{10}) = \frac{1}{10}$$

The consequence of sample size

population std. deviation

$$\text{Standard Error (SE)} = \frac{\sigma}{\sqrt{n}}$$

# of observation in the sample (sample size)

Helps us to understand the amount of Error. we might do if we go with a specific sample size.

I will do in my estimation

Uniform prob. Dist.

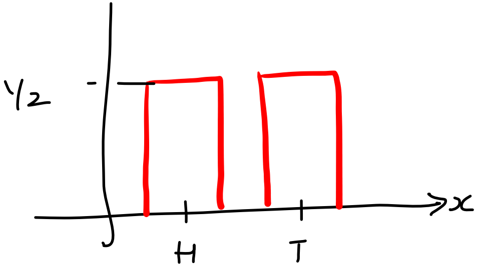
Discrete Random variable

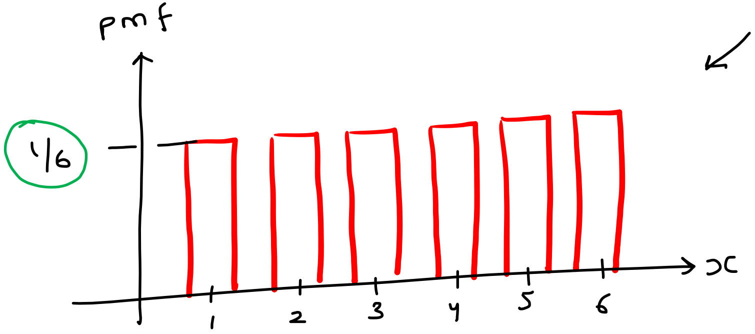
Exp: Roll of a dice

each outcome in my sample space is equally likely

$\Omega = \{1, 2, 3, 4, 5, 6\}$

$P(1) = P(2) = P(3) = P(4)$
$= P(5) = P(6) = 1/6$





Unifor prob. dist

Pmf of dist. uniform prob. dist

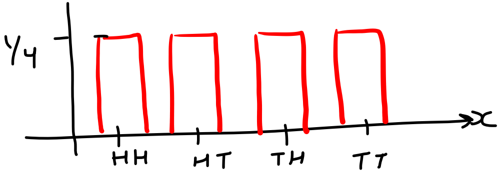
$$\text{pmf} = \frac{1}{(b-a+1)} = \frac{1}{6-1+1} = \frac{1}{6}$$

b = max outcome

a = min outcome

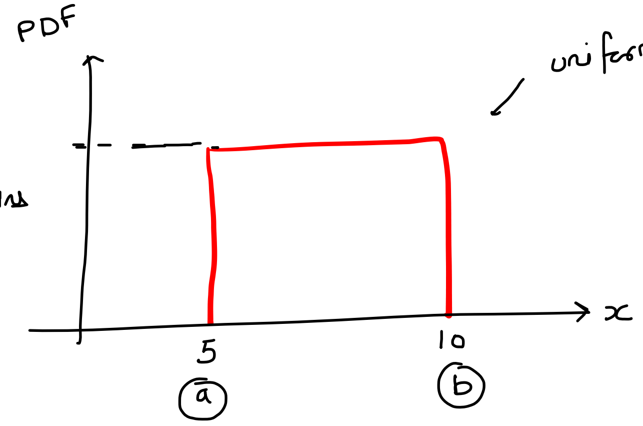
$b > a$

$\Omega = \{HH, HT, TH, TT\}$



continuous Random . var

The time taken by Sumit to commute from home to office is uniformly dist b/w 5-10mins



uniform prob. dist.

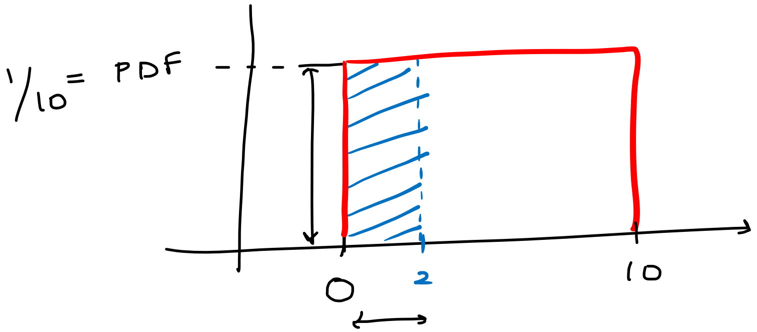
$$PDF = \frac{1}{b-a}$$

$$PDF = \frac{1}{10-5}$$

$$\boxed{PDF = \frac{1}{5}}$$

Uniform Distribution

Sumit regularly takes a break from work to go to the post office to send love letters to his girlfriend “Ankita”. The amount of time Sumit waits in the queue to be served at the post office has a continuous uniform distribution between 0 and 10 minutes. Find the probability that Sumit does not have to wait more than 2 minutes.



$$PDF = \frac{1}{b-a} = \frac{1}{10-0} = \frac{1}{10}$$

$$P(x \leq 2) = \frac{1}{10} \times 2 = \frac{2}{10} = 0.2$$

The uniform distribution has the following properties:

- Mean: $\dfrac{(a+b)}{2}$
- Variance: $\dfrac{(b-a)^2}{12}$, for Continuous uniform distributions
- Variance: $\dfrac{(b-a+1)^2-1}{12}$, for Discrete uniform distributions
- Standard Deviation: $\dfrac{(b-a)}{\sqrt{12}}$, for Continuous uniform distributions.
- Standard Deviation: $\sqrt{\frac{(b-a+1)^2-1}{12}}$, for Discrete uniform distributions.

Quiz-1: There are 45 students in a class. 5 students were randomly selected from this class and their heights (in cm) were recorded as follows: [131, 150, 140, 142, 152]
Calculate Sample mean and sample variance

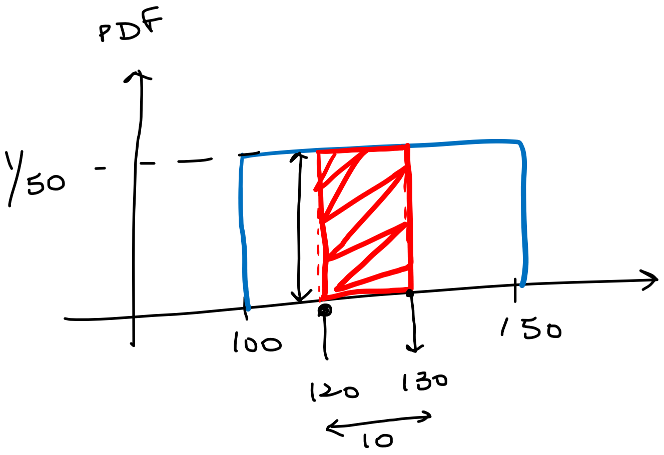
$$\frac{131+150+140+142+152}{5} = 143$$

$$\text{Sample variance} = \frac{(131-143)^2+(150-143)^2+(140-143)^2+(142-143)^2+(152-143)^2}{4}$$

Quiz-2: A sample of 30 latest returns on XYZ stock reveals a mean return of 4 with a sample standard deviation of 0.13.
Estimate the SE of the sample mean.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{0.13}{\sqrt{30}} = \underline{\underline{0.023}}$$

Quiz-3: Suppose the weight of dolphins is uniformly distributed between 100 pounds and 150 pounds.
If we randomly select a dolphin at random, then determine the probability that the chosen dolphin will weigh between 120 and 130 pounds.



$$\frac{1}{150-100}$$

$$P(120 \leq x \leq 130) = \frac{1}{50} \times 10 = \frac{1}{5} = 0.2$$