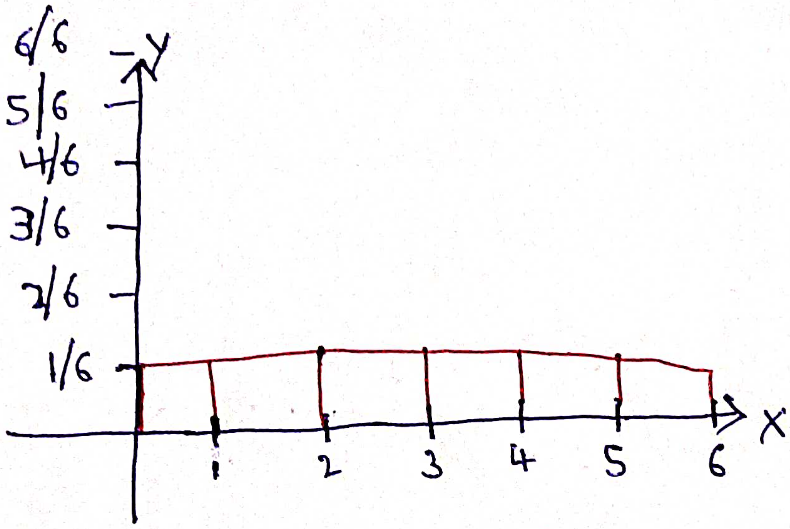
Probability Mass Function:

* used to Represent Discrete Random variables

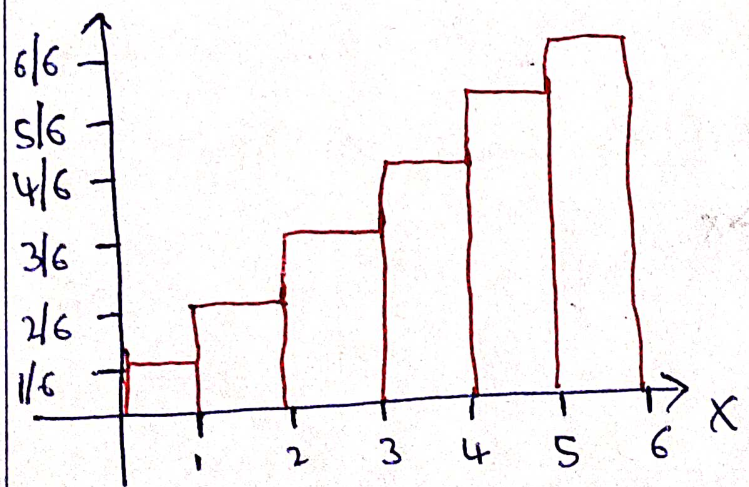
For eg: Rolling a Dice

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

$Pr(1) = 1/6$, $Pr(2) = 1/6$, ...



cumulative probability:

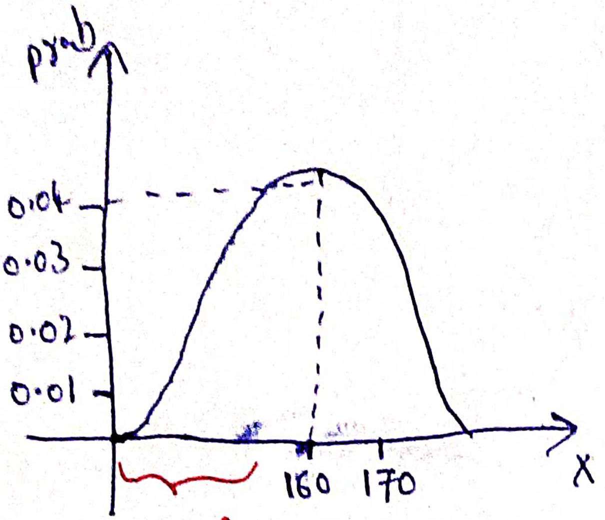


* It's the cumulative sum of all the Data points.

Probability Density Function:

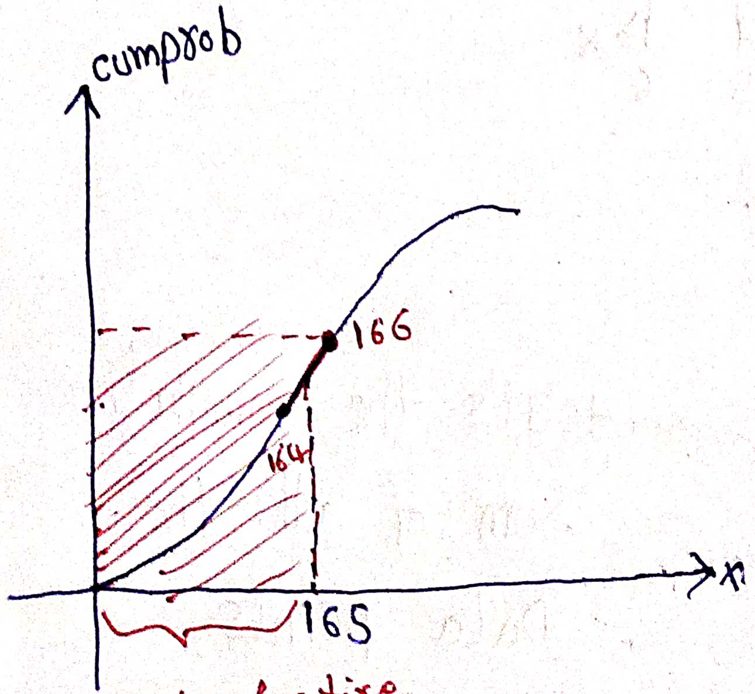
* used to represent continous Random variables.

Eg: Heights



50% of Entire Distribution

CDF:



50% of entire Distribution.

Bernoulli Distribution:

* It is the Distribution where two Discrete outcomes will be represented by p and $1-p$

eg: Tossing a coin

$P(H) = p = 0.50$

$\Rightarrow P(T) = 1-p \Rightarrow q$

Mean of Bernoulli Distribution : p

Variance : pq

std deviation : $\sqrt{pq}$

Binomial Distribution:

* It is same as Bernoulli, But two Discrete outcomes will be done For $\boxed{\text{'n'}}$ number of Times.

eg: Toss a coin 13 Times

Mean : np

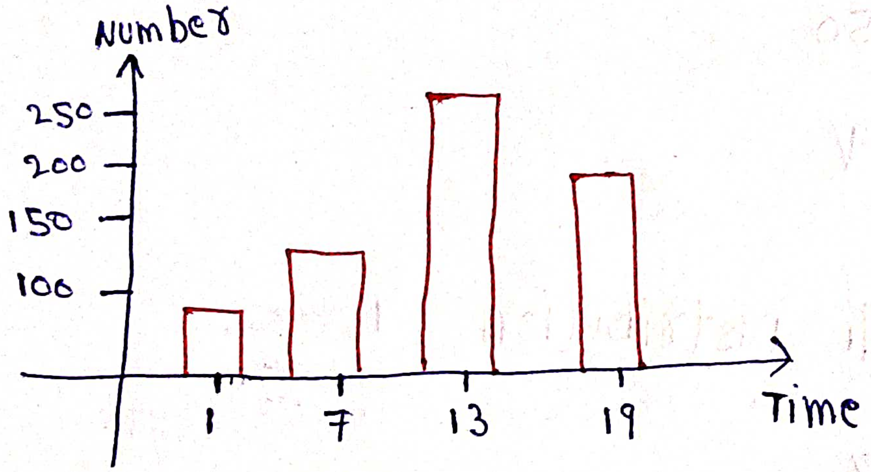
Variance : npq

std Deviation : $\sqrt{npq}$

Poisson Distribution:

* Describes the number of events occuring at a certain period of Time.

eg: No of people visiting Hospital every $\boxed{6}$ hours.



$\lambda = 135$

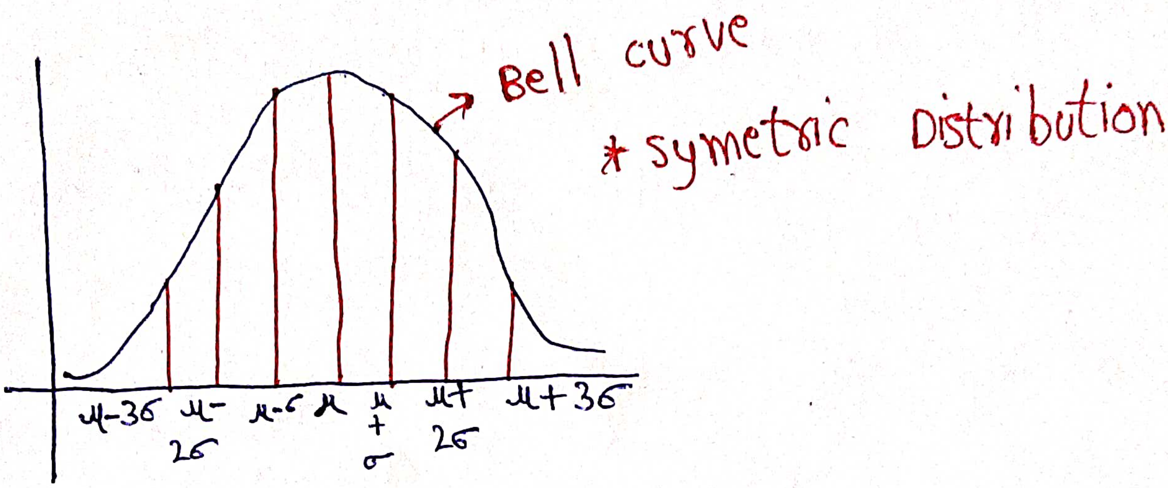
*(λ) Represents Expected no. of events at every Time Interval

Mean $\rightarrow \lambda t$

Variance $\rightarrow \lambda t$

* t = time Interval
  λ = Expected no. of Events

Normal (or) Gaussian Distribution:



Empirical Rule:

* 68% Data lies b/w $\mu-\sigma$ and $\mu+\sigma$
* 95% Data lies b/w $\mu-2\sigma$ and $\mu+2\sigma$
* 99.7% Data lies b/w $\mu-3\sigma$ and $\mu+3\sigma$

eg: Based on so much of Analysis, Researchers Found Height, weight, iris, etc... will mostly have guassian Distribution

uniform Distribution

continous uniform Distribution:

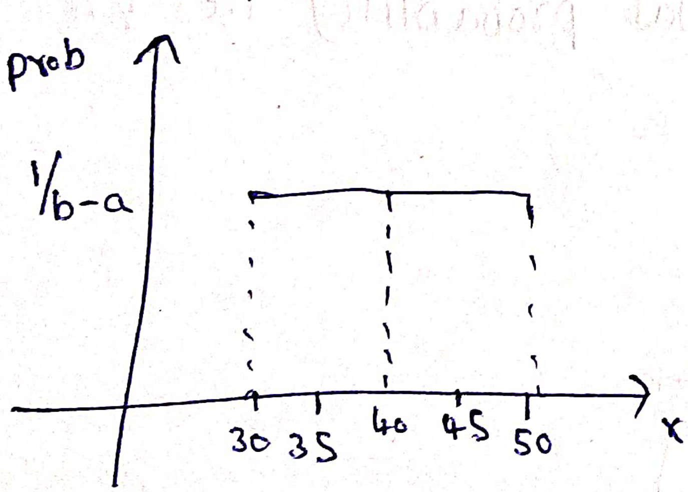
* In uniform Distribution, the probability of getting outcomes is equal
* In continous uniform Distribution, In between a specified Range the probability of getting outcomes is same.

$$pr = (x_2 - x_1) * \frac{1}{b-a}$$

a = lower bound

b = higher bount.

x_1, x_2 are datapoints.



$$pr(x \text{ b/w } 40 \,\&\, 45) = (\cancel{45}\, x_2 - x_1) * \frac{1}{b-a}$$

$$= (45 - 40) * \frac{1}{50-30}$$

$$= 5/20$$

$$= 1/4 \rightarrow 0.25 \rightarrow 25\%$$

$pr(x \text{ b/w } 30 \,\&\, 35) = (35-30) * \frac{1}{20}$

$= 0.25 \rightarrow \boxed{25\%}$

$pr(x \text{ b/w } 30 \,\&\, 45) = (45-30) * \frac{1}{20}$

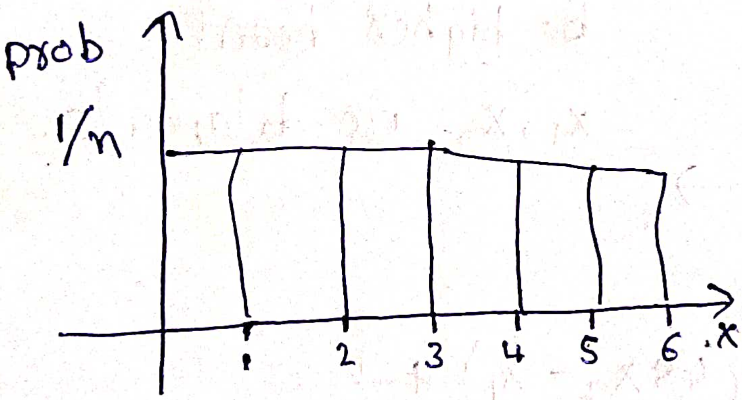
Mean : $a-b/2$ $= 15/20$

Variance : $(b-a)^2/12$ $= 0.75 \rightarrow \boxed{75\%}$

Discrete uniform Distribution:

* All the data has equal probability i.e, $1/n$

eg: Rolling a die

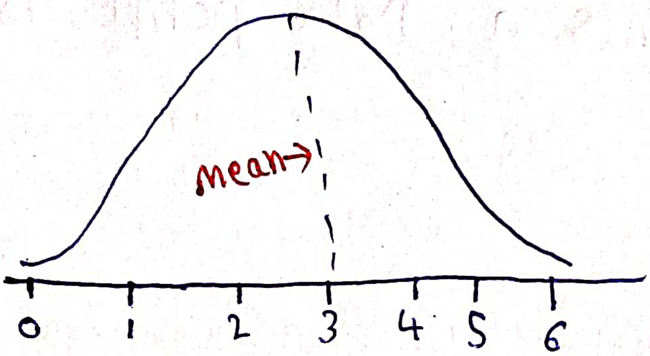


$pr(1) = 1/n \rightarrow 1/6$

Standard Normal Distribution:

* It is the process of converting Normal Distribution into -ve, zero and +ve Data

For eg this is Normal Distribution:



→ $\mu = 0$

$\sigma = 1$

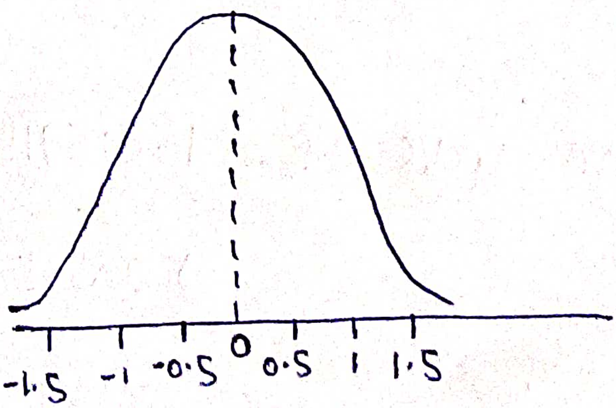
Let's Apply z score on all Data

$$z = \frac{x_i - \mu}{\sigma}$$

μ = mean = 3

σ = std Deviation = 2

data	z-score
0	-1.5
1	-1
2	-0.5
3	0
4	0.5
5	1
6	1.5



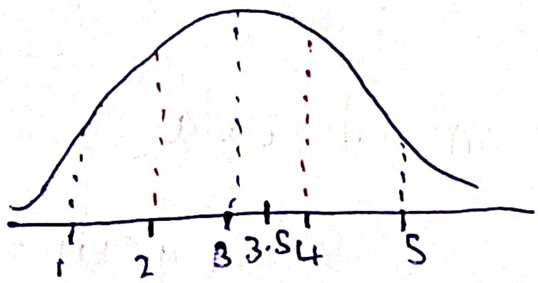
z - scores :

* z - score is used to standardize the Data.

* $z = \dfrac{x_i - \mu}{\sigma}$

* we can also releate 'z' with how many std deviation points a Data point is away from mean.

eg :



* Find pr of Region above 3.5 ⇒

$z = \dfrac{3.5 - 3}{1} \Rightarrow \boxed{0.5} \rightarrow$ Find it's value in z-Table

0.6915 } Its the region out of 1 Before 3.5

→ Region above 3.5 = 1 - 0.6915 ⇒ 0.3085

⇒ 30.85%