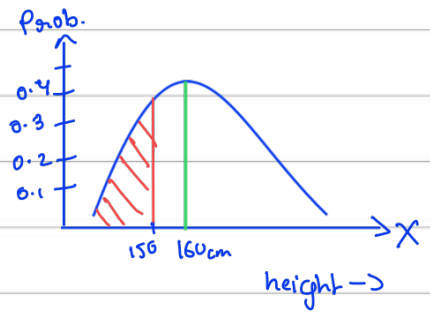
1) Probability Distribution and Density function

all PDF, PMF, CDF are part of Probability Distribution function

1.1) Probability Density Function :- (PDF)

It shows distribution of continous data.

Ex: Height of student



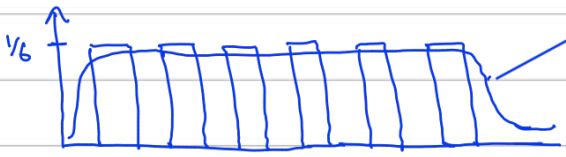
What is the probability for height less than 150?

Probability of having height < 150 is given by area of graph under 150

i.e $\boxed{P(x<150) = \text{Area under curve}}$

1.2) Probability Mass Function (PMF):- Distribution of a discrete Random variable

ex: Rolling of a dice $\{1, 2, 3, 4, 5, 6\}$



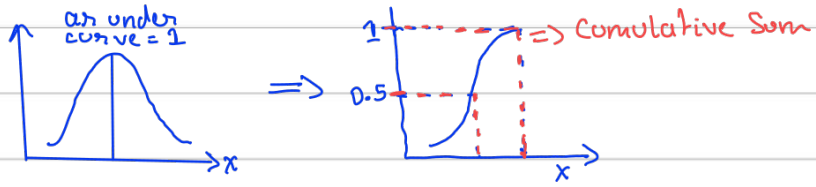
PMF: created by smoothening the histogram

Q) What is the the probability of $X < 4$?

Sol^n -> $P(x<4) = P(x=1) + P(x=2) + P(x=3) + P(x=4)$

$= (1/6) \times 4 = \boxed{2/3}$

1.3) Cumulative Distribution Function (CDF):- CDF is cumulative sum of every value wrt area under the curve of PDF.



II) Types of Probability Distribution :-

① Normal / Gaussian Distribution :- Bell curve
② Bernoulli Distribution (PMF) : 2 outcomes success or failure
③ Uniform Distribution
④ Log Normal Distribution (PDF)
⑤ Poisson Distribution Rule (PMF)
⑥ Power Law Distribution (PdF) : c/a 80 - 20% Rule
⑦ Binomial Distribution (PMF)

III) PDF vs CDF vs PMF

PDF → Probability density function
PMF → Probability mass function
CDF → Cumulative density function

① PMF :-

Requirement : Discrete random variable

PMF

Ex:- Rolling a dice

$\{1,2,3,4,5,6\}$ $P(X)(1) = 1/6$

$P(X)(2) = 1/6$

diagram:-

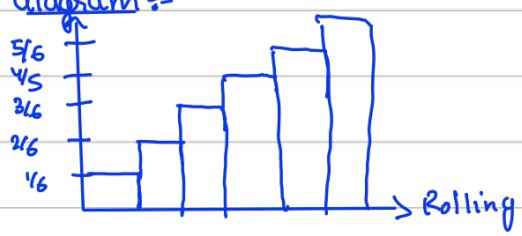
CDF

"

"

"

diagram :-



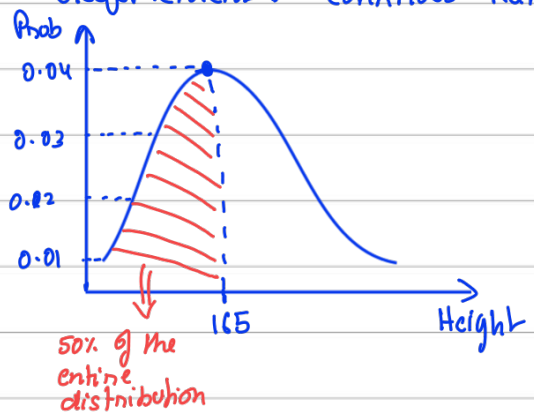
in CDF probabilities are cumulative-ly added

with CDF we can find the probability of lets say $P(x \leq 4)$

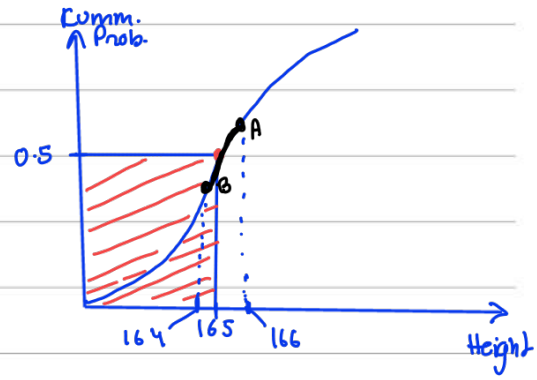
$P(x \leq 4) = P(1) + P(2) + P(3) + P(4)$
$= 4/5$

② **PDF**: Probability Density Function.

Requirement: Continous Random Variable.



CDF →



#finding gradient:-

Slope at pt A = 0.58

Slope at pt B = 0.50

$$\text{gradient} = \frac{0.58 - 0.50}{166 - 164} = \frac{0.08}{2} = 0.04 \quad (1)$$

therefore from equation (1) we conclude that Probability density is gradient descent of CDF

IV) Bernauli Distribution: [Binary Outcomes]

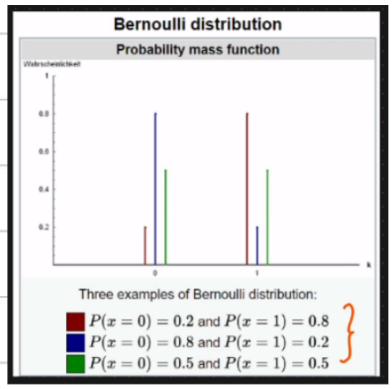
Bernauli Distrib^n and log normal distributions are not normal or gaussian distributions

In probability theory and Stastics, Bernaulli distribution, named after Swiss mathematician Jacob Bernaulli, is the discrette probability distribution of random variables which takes value 1 with probability p and value 0 with probability $q = 1 - p$. Less formally it can be thought as model for the set of possible outcomes of any single experiment that asks yes-no questions. Such questions leads to outcomes that are boolean-valued: a single bit whoose value is success/yes/true/one with probability p. and failure/no/false/zero with probability q.

Ex: ① tossing a fair coin. $\{\overset{H}{0}, \overset{T}{1}\}$:

$$P(T) = 0.5 = p$$

$$P(H) = 1 - 0.5 = 1 - p = q$$



② Weather a person will fail or pass?

Deriving probability mass function for Bernauli Distribution

for Bernaulli dist. R value is either 0 or 1

i.e, $k = 0$ or $k = 1$

so,

Pmf: probability of getting outcome $k = 0$ is give by:

$$P(x=k) = p^k (1-p)^{1-k}$$

→ for $k = 1$: $P(x=1) = p^1 (1-p)^0 = p$

→ for $k = 0$: $P(x=0) = p^0 (1-p)^{1-0} = (1-p) = q$

Hence its a Bernaulli distribution's pmf cuz we are getting p or q values for diff. scenarios.

PMF formula simplified

$$PMF = \begin{cases} q = 1-p & \text{if } k=0 \\ p & \text{if } k=1 \end{cases}$$

Calculating mean variance and std distribution for Bernauli

a). Mean:- mean of a Bernaulli distribution is always p.

mean given by :- $\boxed{E(k) = \sum_{i=1}^{k} k \cdot P(k)}$

ex: Suppose prob. of $P(k=1) = 0.6 = p$, then

$P(k=0) = 1 - 0.6 = 0.4 = q$

now using above eqn to find mean using above eqn

$$E(k=1) = \sum^{k} k \; P(k)$$

$$= (1 \times 0.6) + (0 \times 0.4)$$

$$= 0.6 = p$$

Hence, the mean of a Bernaulli Distribution is always p.

b) Median:

$$\text{Median} \begin{cases} 0 & \text{if } p < 1/2 \\ [0, 1] & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$$

c) Variance & std

$$\boxed{\begin{aligned} \text{variance} &= p(1-q) = pq \\ \text{std} &= \sqrt{pq} \end{aligned}}$$

IV Bionomial Distribution

for a process whoose outcomes are binary but the process is repeated n no. of times such type of distribution is called as Bionomial Distribution.

In probability theory and statistics, the bionomial distribution with parameter n and p ——— discrete probability distribution of the number of successes in a sequence of n independent experiments, each asking yes-no questions, and each with its own Boolean-valued outcome: (success with probability p) or failure (with prob. $q = 1-p$). A single Success/failure experiment is also called as Bernaulli trial or Bernaulli experiment, and a sequence of outcomes is called as Bernaulli process, for a single trial, i.e, $n=1$, the bionomial distribution is a Bernaulli distribution.

ex: Tossing a coin
{Bernaulli distribution}

$P(H) = 0.5 = 1-P = q \rightarrow 0$

$P(T) = 0.5 = P \longrightarrow 1$

Tossing a coin 10 times
{Bionomial distribution}

	1st time	2nd time	3rd time	...
$P(T) = P$		P	P	...
$P(H) = 1-P$		$1-P$	$1-P$	...

#So Bionomial Distribution can combine various Bernaulli distributions.

Parameters used:-

$n \in \{0,1,2,3 \ldots\}$: no. of Trials or Experiments

$p \in \{0 \text{ to } 1\}$: Success or probability of each trial

$q = 1-p$

PMF of Bionomial dist.:-

$$\boxed{P(k) = {}^nC_k \, p^k (1-p)^k}$$

, where $k\{0,1,2,\ldots\}$: no. of successes

Mean of Bionomial dist :- $\boxed{nP}$

Variance of Bionomial dist :- $\boxed{nPq}$

Std of Bionomial dist :- $\boxed{\sqrt{nPq}}$

VII Poisson Distribution :-

for : ① Discrete Distribution

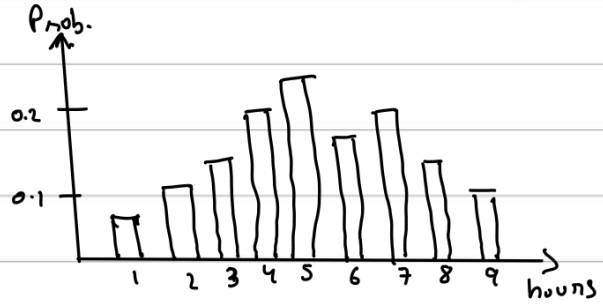
② Describes no. of events occuring in a fixed time interval.

ex: No. of people visiting hospital every hour

No. of people visiting banks " "

" " " " banks " "

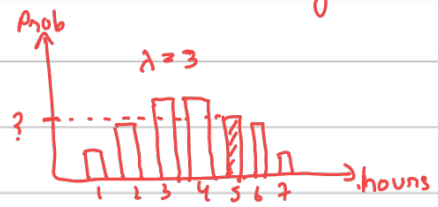
Distribution:



λ parameter: signifies expected no. of events to occur every hour

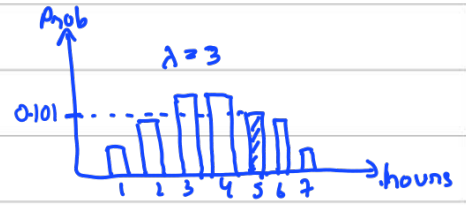
PMF for Poisson dist :- $\boxed{P(x) = \frac{e^{-\lambda} \lambda^{x}}{x!}}$

calculate the PMF for the following distribution with $\lambda = 3$.



Sol^n: $P(x=5) = \frac{e^{-\lambda} \lambda^{x}}{x!} = \frac{e^{-3} 3^{5}}{5!} = 0.101$

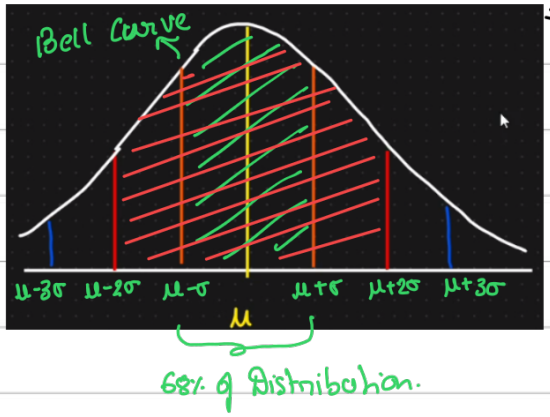
≈ 10.1%



Mean and Variance of Poisson dist :- $\boxed{E(x) = \mu = \lambda \times t}$

mean and variance are same in case of poisson distribution

VIII) Normal or Gaussian Distribution



⇒ Symmetric Distribution

also c/a "3 Sigma Rule"

Emperical Rule: within the 1st standard deviation from μ (i.e μ-σ to μ+σ) around $\boxed{68\%}$ of the data falls on this region and

$\boxed{95\%}$ of the distribution will be falling in the 2nd std. deviation and

$\boxed{99.7\%}$ of the distribution will be falling in the 3rd std deviation

QQ Plot: used to determine wether a distribution is Guassian/Normal distribution or not.

Q) What is the probability of a random variable to fall b/w

① $P(\mu-\sigma \le X \le \mu+\sigma) \longrightarrow \approx 68\%$

② $P(\mu-2\sigma \le X \le \mu+2\sigma) \longrightarrow \approx 95\%$

③ $P(\mu-3\sigma \le X \le \mu+3\sigma) \longrightarrow \approx 99.7\%$