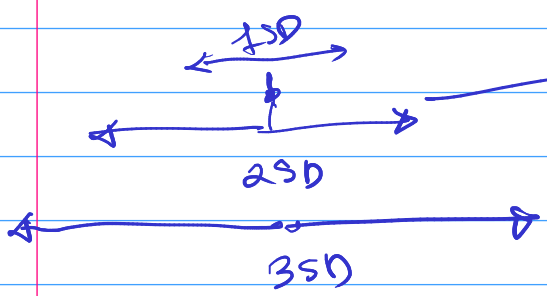
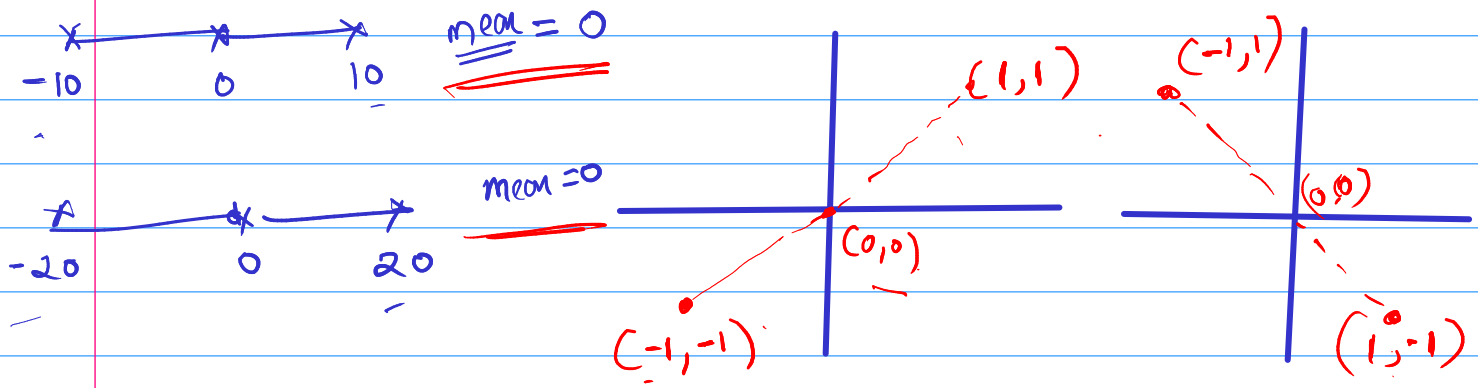
1. Measure of central Tendency → mean, median, mode

2. Measure of Dispersion → Range — (max-min), variance → (σ^2) = , SD → $\sqrt{\sigma}$, Covariance



Covariance

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



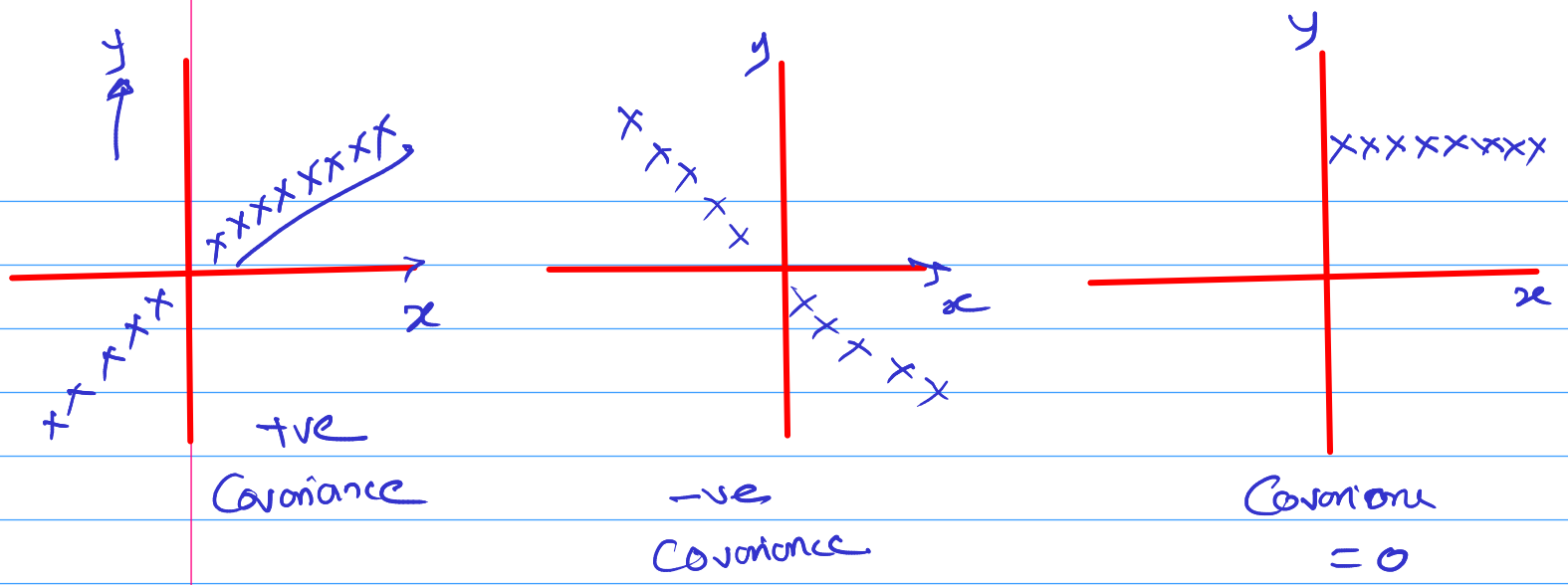
$$\sigma = \frac{1^2 + 0^2 + 1^2}{3} = 2/3$$

$$\sigma^2 = \frac{1^2 + 0^2 + 1^2}{3} = 2/3$$

Covariance is a statistical measure that describes the degree to which two variables are linearly related. It measures how much two variables change together, such that when one variable increases, does the other variable also increase, or does it decrease?

$$\text{min} = Q_1 - 1.5(IQR)$$

$$\text{max} = Q_3 + 1.5(IQR) \quad IQR = Q_3 - Q_1$$



Disadvantages of using Covariance

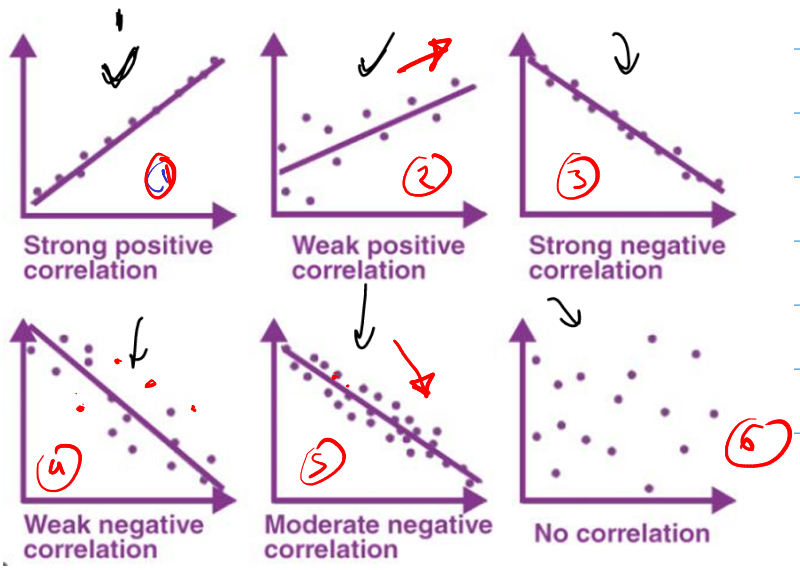
One limitation of covariance is that it does not tell us about the strength of the relationship between two variables, since the magnitude of covariance is affected by the scale of the variables.

Correlation

Correlation refers to a statistical relationship between two or more variables. Specifically, it measures the degree to which two variables are related and how they tend to change together.

Correlation is often measured using a statistical tool called the correlation coefficient, which ranges from -1 to 1. A correlation coefficient of -1 indicates a perfect negative correlation, a correlation coefficient of 0 indicates no correlation, and a correlation coefficient of 1 indicates a perfect positive correlation.

-1 to 1 → Range



Probability

Variable (Algebic Variable)

$x + 10 = 5$

$x = 5$ → assignment operator (=)

Random Variable) – Toss

$X = \{H, T\}$

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

Discrete (RV) (Whole No.)

Age = 20, 30, 31

no. of children = 1, 0, 2, 3,

Continuous (RV)

$X = \{0, 10\}$

0 0.1, 0.11, 0.23, 0.49, 10

9.999.

Probability :– How likely something is to happen.

Toss a Coin = Sample = $\{H, T\}$ – Total = 2

$P(H) = 1/2$ $P(T) = 1/2$

Probability Distribution

1. What are Probability Distributions?

A probability distribution is a list of all of the possible outcomes of a random variable along with their corresponding probability values.

Coin Toss	H (1)	T (0)
probability	1/2	1/2

Dice	1	2	3	4	5	6
probability	1/6	1/6	1/6	1/6	1/6	1/6

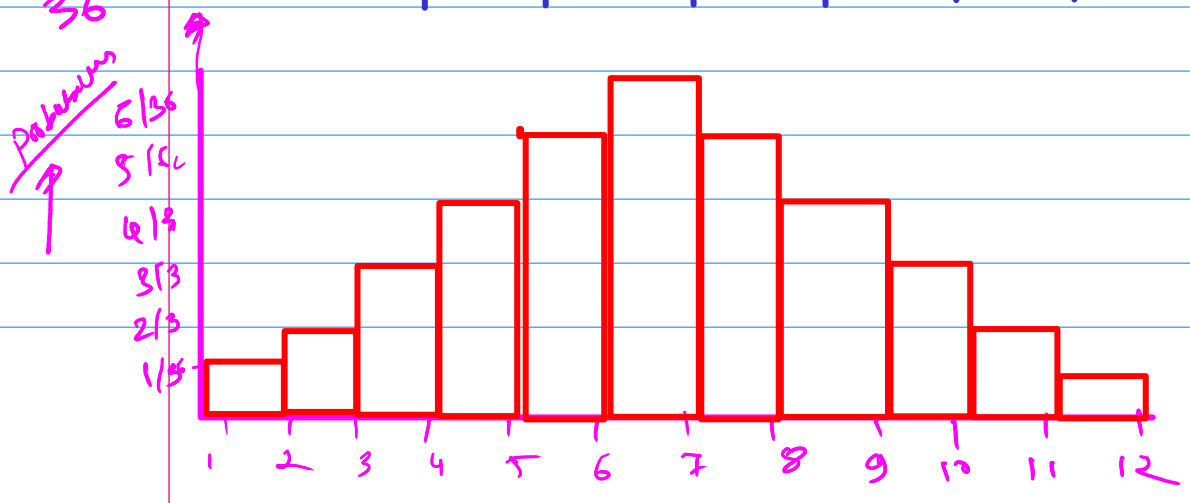
2-Dice Thrown

Dice 2 \ Dice 1	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

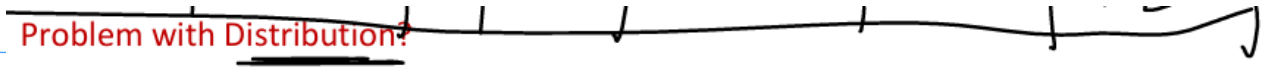
6 = 6 = 6 = 6 = 6 = 6 = 36

Probability Distribution

2 = 1/36
3 = 2/36
4 = 3/36
5 = 4/36
6 = 5/36
7 = 6/36
8 = 5/36
9 = 4/36
10 = 3/36
11 = 2/36
12 = 1/36



eg = Rolling a 10 dice together.



Problem with Distribution?

In many scenarios, the number of outcomes can be much larger and hence a table would be tedious to write down. Worse still, the number of possible outcomes could be infinite, in which case, good luck writing a table for that.

Example - Height of people, Rolling 10 dice together

Solution - Function?

What if we use a mathematical function to model the relationship between outcome and probability?

Probability Distribution Function

Why are Probability Distributions important?

- Gives an idea about the shape/distribution of the data.
- And if our data follows a famous distribution then we automatically know a lot about the data.

PDF

A probability distribution function (~~PDF~~) is a mathematical function that describes the probability of obtaining different values of a random variable in a particular probability distribution.

Probability Mass Function (PMF)

CMF

Cumulative mass function

Probability Density Function. (PDF)

CDF

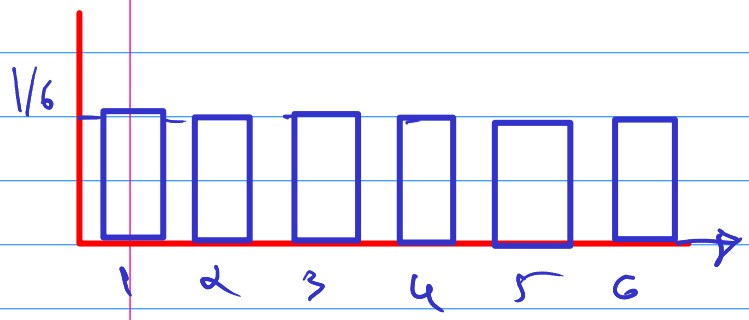
Cumulative Distribution Function

PMF → discrete random Variable

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

1/6

$$Y = \begin{cases} 1/6 & \text{if } x \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases}$$



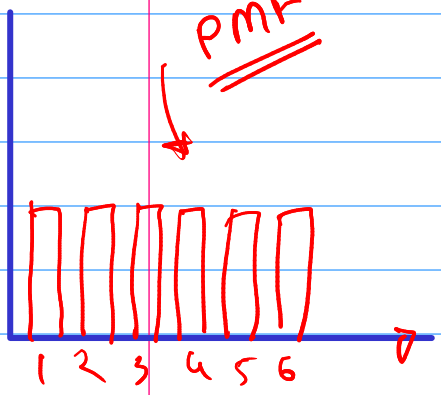
$P_0(x \leq 4)$

$= P(1) + P(2) + P(3) + P(4)$

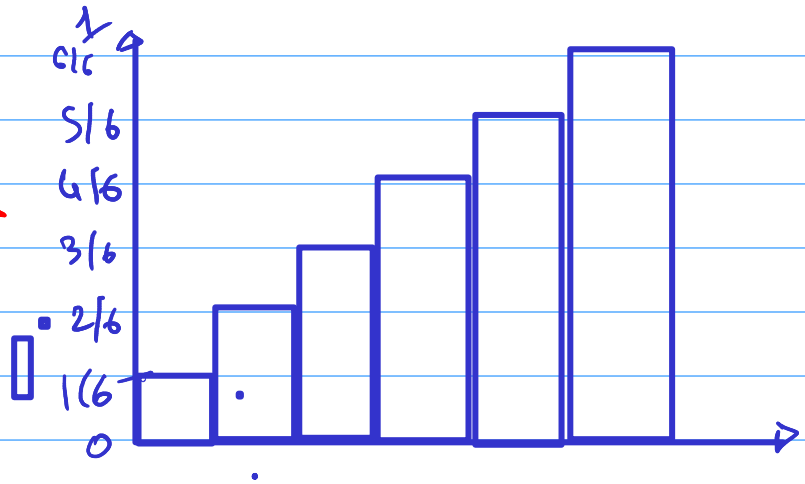
$= 1/6 + 1/6 + 1/6 + 1/6$

$P(x \leq 8)$

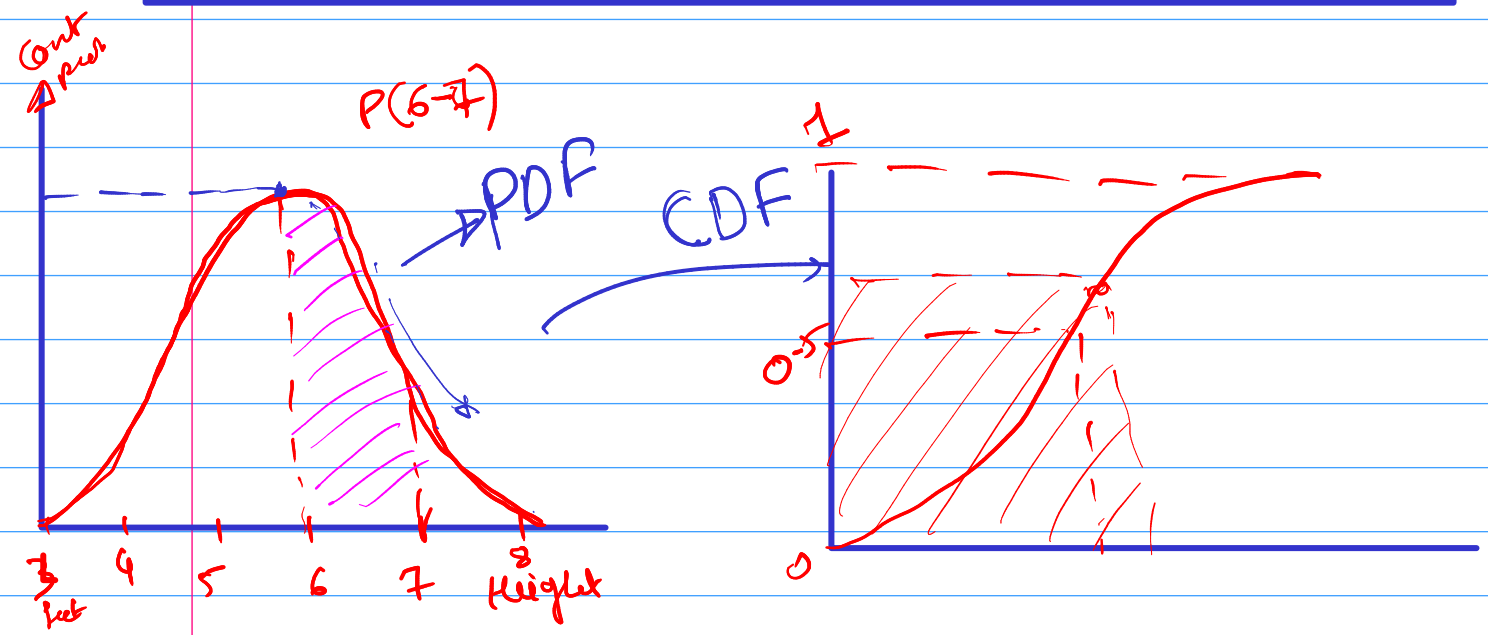
CDF (cumulative Distribut funct)



CMF
CDF



Probability Density function - Continuous Random variable



➤ Types of Probability Distribution: -

1. Normal or Gaussian Distribution
2. Bernoulli Distribution
3. Uniform Distribution
4. Poisson Distribution
5. Binomial Distribution
6. Log-Normal Distribution