Concept of Random Variable:

Goals of Messi in a club Football Matches

Match No	Goal Number
01	1
02	3
03	2
04	2
05	0
06	1

Here for the same variable Messi Goals number are different for different matches. Also the ~~got~~ goal values don't have a relation with each other. This concept is random variable concept.

There are two types of Random variables.

→ 1) Discrete Random Variable

→ 2) Continuous Random Variable.

1) Discrete Random Variable: (Examples)

1) Number of Heads after tossing a coin 10 times.

2) Goals of Messi in World cup

3) Number of children in each section of class 6.

Continuous Random Variable: (Examples)

- → Heights of the students in a class
- → ~~No~~ Time remaining for the next bus to come
- → Family members consumption of coke from a 2L bottle.

Probability Distribution on Discrete Random Variables:

Let X be a discrete random variable which denotes the values $x_0, x_1, x_2, \dots x_n$

$$P(X = x_i) = f(x_i) = f_i \qquad x_i = \{x_0, x_1, x_2 \dots x_n\}$$

The value of f_i depends on x_i ie $i \rightarrow \{x_0, x_1, x_2 \dots x_n\}$

This Function, f_i called the ~~Pot~~ Probability Mass Function (PMF).

The set of ordered pair (x_i, f_i) is called the discrete probability distribution of X.

> Ordered pair $(a,b) = (u,v)$ iff, $a = u$ and $b = v$

This distribution represent as below:

X:	x_0,	x_1	x_2	$\dots\dots$	x_n
f_i:	f_0	f_1	f_2	$\dots\dots$	f_n

Let's take an example. Suppose we toss two coin. at a time.

Sample Space (S) = { HH, HT, TH, TT }

X is a random variable which says,

X : No. of head ~~appair~~ appears, for every coin toss. 2 times.

$X(HH) = 2$, $X(HT) = 1$, $X(TH) = 1$, $X(TT) = 0$

Range of X is $\{0, 1, 2\}$

$P(X=0) \rightarrow$ Means No Head $= \frac{1}{4}$ $\{P(X=0) = f_0\}$

$P(X=1) \rightarrow$ Means 1 Head $= \frac{2}{4} = \frac{1}{2}$ $\{P(X=1) = f_1\}$

$P(X=2) \rightarrow$ Means 2 Head $= \frac{1}{4}$ $\{P(X=2) = f_2\}$

Let's represent the distribution →

X :	0	1	2
(PMF) → f_i :	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Properties :

1) $f_i \geq 0 \rightarrow (\frac{1}{4} \geq 0, \frac{1}{2} \geq 0, \frac{1}{4} \geq 0)$
2) $\sum f_i = 1 \rightarrow (\frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1)$

Binomial Distribution:

I have noted down the concept well in PW-Skills part.

Question: You tossed a coin 3 times. What is the probability of getting 2 Heads?

In Binomial Distribution, PMF = ~~^nC_k~~

$$P(x) = {}^nC_x \, p^x (1-p)^{n-x}$$

n = number of trials

P : Probability of success

$n = 3$

$P_1 = \frac{1}{2}$

↳ The probability of getting Head for each toss

x : Number of success

Means number of heads $= 2$

$$\therefore P(x) = {}^3C_x \left(\tfrac{1}{2}\right)^x \left(1-\tfrac{1}{2}\right)^{3-x}$$

$$= {}^3C_x \left(\tfrac{1}{2}\right)^x \left(1-\tfrac{1}{2}\right)^{3-x}$$

$$\therefore P(2) = {}^3C_2 \left(\tfrac{1}{2}\right)^2 (1/2)^{3-2}$$

$$= {}^3C_2 \times \frac{1}{4} \times \frac{1}{2}$$

$$= \frac{3}{8}$$

Question 2: In a factory that produces light bulbs. It is known that 10% of the bulbs are defective. A inspector randomly selects 15 light bulbs from a recent batch. What is the probability that 3 of the 15 selected bulbs are defective.

→ Here the probability of success (finding a defective bulb) = 10% = 0.10

PMF Function, $P(X=K) = {}^{n}C_{K}\, P^{K}\, (1-P)^{n-K}$

Here, $n = 15$ (number of trials)

$K = 3$ (number of success) → Getting a deffective bulb

$P = 0.10$ (Probability of defective)

$\therefore P(X=K) = {}^{15}C_{3}\,(0.10)^{3}\,(1-0.10)^{15-3}$

$= 0.332$

Ans

Question 3:

A website adminstrator monitors the traffic to their site and observe that the avg click rate for a particular ad is 20%. If the add is shown to 100 people, What is the probability that it would be cliched by 15 people ?

Here, $n = 100$
$p = 0.2$
$K = 15$

$\because$ PMF $\rightarrow$ $P(X=K) = {}^{n}C_{K}\ p^{K} (1-p)^{n-K}$

$= {}^{100}C_{15}\ (0.2)^{15} (1-0.2)^{100-15}$

$= 0.201$

Ans

Bernouli Distribution:

This concept also noted down well in the PW-sikills part.

Question 1: In a series of basketball free throws, a player makes a successful shot (dentos by 1) with a probability of 0.75. What is the probability that the player makes a successful shots in a sequence of a free throws.

Here, $n = 1$
$K = 1$
$P = 0.75$

$P(X=K) = {}^{n}C_{K}\ p^{K} (1-p)^{n-K}$

$= {}^{1}C_{1}\ p^{1} (1-p)^{0}$

$= p$

$= 0.75$ Ans

Question 2:

In a deck of 20 playing cards, 6 cards are red and the rest are black. If a card is drawn at random from the deck, where red is considered a success. What is the probability of drawing a red card.

$$p = \frac{6}{20}$$

$$= \frac{3}{10} = 0.3$$

For Bernoulli Distribution

$$P(x=K) = {}^{n}C_{K} \, p^{K} (1-p)^{n-K}$$

Now, $\left.\begin{matrix} n=1 \\ K=1 \end{matrix}\right\}$ For Bernoulli

$$\therefore P(x=K) = 1 \times p^{1} \times (1-p)^{1-1}$$

$$= p \rightarrow \text{For success}$$

$q = (1-p) \rightarrow$ For not success

Example of Probability Distribution: (How we represent it)

Suppose, you have two dice. You need to show the probability distribution of the sum of the two dice after throwing.

Sum of the dices after throwing can be →

$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

Let's make the dice table →

	1	2	3	4	5	6	~~7 8 9 10~~
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	

$P(2) = \frac{1}{36}$

$P(3) = \frac{2}{36} = \frac{1}{18}$

$P(4) = \frac{3}{36} = \frac{1}{12}$

$P(5) = \frac{4}{36} = \frac{1}{6}$

$P(6) = \frac{5}{36}$

$P(7) = \frac{6}{36}$

$P(8) = \frac{5}{36}$

$P(9) = \frac{4}{36}$

$P(10) = \frac{3}{36}$

$P(11) = \frac{2}{36}$

$P(12) = \frac{1}{36}$

So, Representation of probability distribution

Dice (sum) → Possible outcomes

	2	3	4	5	6	7	8	9	10	11	12
P(sum)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

But the sample space was really less for the previous exam for which we ~~can~~ could draw the tables and calculate the probability.

But that can't be done if we wanted to get the probability of each sum for 10000, 1M Dice. right?

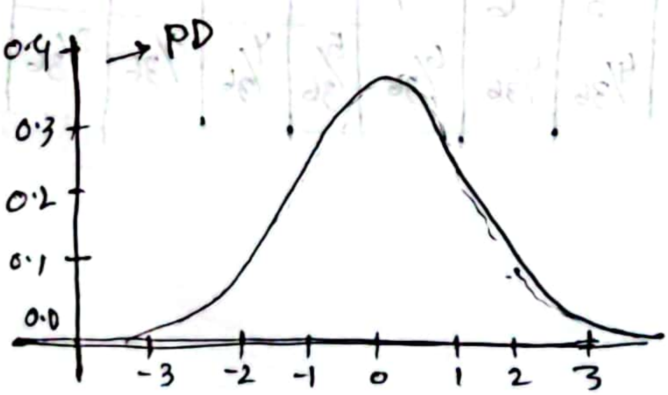
So, what we can do is we can create Probability distribution functions (PMF, PDF, CDF) using which we can plot graph and watch the probability distribution.

(I have already noted down PMF, PDF, CDF concepts in PW-Skills Section.)

For getting the probability of each sum from rolling 2 dice,

$$PMF = \begin{cases} 1/36 \text{ when } x \in \{2,12\} \\ 2/36 \text{ when } x \in \{3,11\} \\ \vdots \\ 0 \text{ otherwise.} \end{cases}$$

As PDF is little bit complex, I would note it again here briefly.



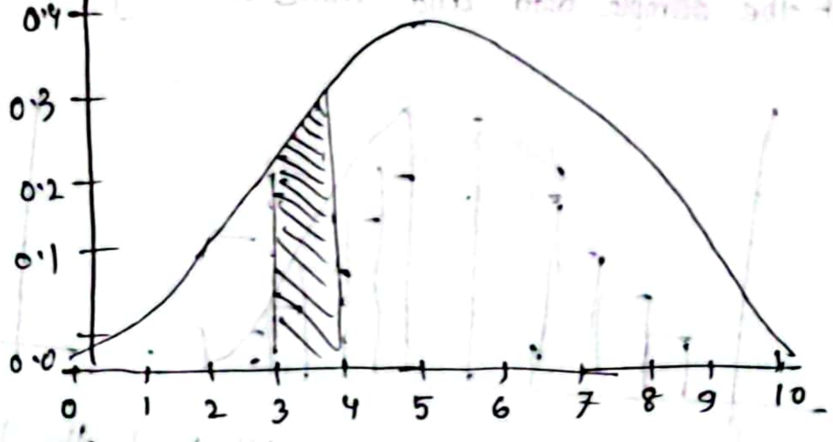
PD = Probability Density.

in PMF graph, y-axis we have probability

But in PDF, in Y-axis we have probability Density.

What is probability Density? What is the resean to use it?

→ In CDF, we use continuous Random variables. Suppose, we created a PDF where in x axis the cgpa are given. Now suppose, I want to find whose cgpa is 3.874. Between cgpa 3 to 4 there can be infinite numbers. The count is so big the probability of finding only one value in that range is 0 (almost 0). So that's why we can't use probability in Y axis for PDF. We have to use Probability density.



using PDF we can't say the probability if 3.874 instead we can find the probability of having a value between 3 to 4 by calculating the area.

$$\int_3^4 f(x)\,dx$$

↳ This integration will provide the value of probability for 3-4.

Now, the rectangle area that we got, we can make it more thinner to get the single value probabilities like 3.784

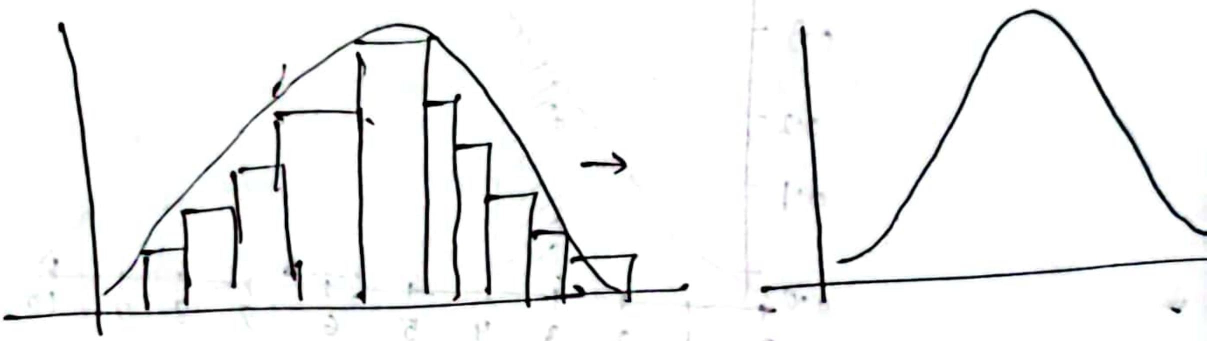
How to estimate Density:

There are two types of Density estimation:

1) Parametric 2) Non Parametric

Parametric Distribution:

First we plot the sample data and make a histogram.



Suppose this is the histogram. Then we will check the histogram looks like which distribution. Here, we can see that the histogram looks like Normal Distribution. Now we do everything according to Normal Distribution. We will get the μ (Mean) and σ (std) from the available data. Now from the sample data, we have to estimate the population Mean and population std (σ).

Then we have to use the PDF equation to calculate probability density.

$$PDF = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

for each sample data value.

Non Parametric Technique:

When after plotting the histogram, the graph is not similar to any probability Distribution. Then we apply Non parametric Technique.

Plot CDF for PDF:

If you Integrate PDF, you can get CDF (Integration)

~~Ifo~~ If you Differentiate CDF, you get PDF (Diffienciation).

CDF for PDF is noted down on PW-Skills section.

Uniform Distribution:

The concept is noted down in PW-Skills section well.

Some Examples:

① The height of a person randomely selected from a group of individuals whose height range from 5'6" to 5'10" would follow a continuous uniform distribution.

② The time taken for a machine to produce a product, where the production time ranges from 5 to 10 minutes, would follow a continuous uniform distribution.

③ The distance that a randomly selected car travels on a tank of gas, where the distance ranges from 300 to 400 miles, would follow a continuous uniform distribution.

Problem 1: Buses in a city are scheduled to arrive at a particular bus stop ~~ver~~ every 15 minutes. With the first bus arriving at exactly 8 AM. If a passenger arrives at the bus stop at a random time between 8 AM to 9 AM, What is the probability that they will wait less than 5 minutes for the next bus?

Here, $a = 0$ min

$b = 60$ min [Because Range 9 AM − 8 AM = 60 min]

$x_2 = 5$ min

$x_1 = 0$ min

$$Pr(0 \le x \le 5) = (x_2 - x_1)\frac{1}{b-a}$$

$$= \frac{5}{60} = \frac{1}{12}$$

Ans

Problem 2: In a parking lot, the available parking spaces are uniformly distributed througout the lot. The lot has a total of 200 spaces. On average 40% of the spaces are occupied. What is the probability that the next car arriving, finds an available parking space?

Here, Unoccupied spaces = $(100-40) = 60\%$

Number of unoccuppied spaces $= 0.6 \times 200$

$= 120$

$\therefore$ Probability of finding space $= \frac{120}{200}$

$= 0.6 = 60\%$ Ans

Problem 3: An internet router's download speed, uniformly distributed between 50 mbps and 100 mbps. Calculate the probability that a random user experiences a dowload speed at least 70 mbps?

Here, $b = 100$

$a = 50$

$x_2 = 100$

$x_1 = 70$

[because at least 70 mbps means minimum must be 70 mbps So max can be 100 mbps]

$\therefore$ Pr(speed at least 70 mbps) $= 100-70 \times \frac{1}{100-50}$

$= \frac{30}{50} = 0.6 = 60\%$.

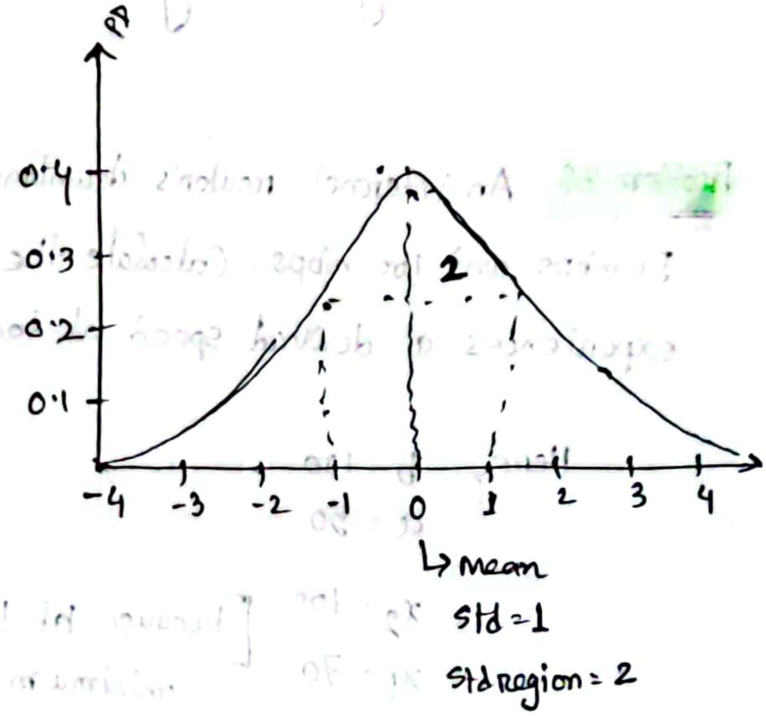
Ans

Normal Distribution: Normal distribution is noted in PW-Skills section.

But the Standard Normal Distribution part is not there.

Here is the explanation of that;

Standard Normal Distribution:



Parameters : $\mu = 0, \sigma = 1$

Notation : $x \sim N(0, 1^2)$

$$f_x(x) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-0}{1^2}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

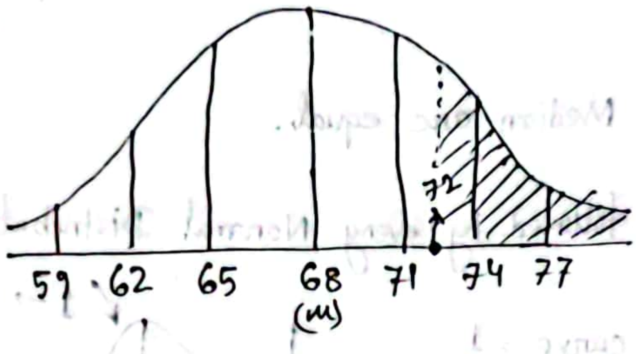
Standardization: To convert any normal distribution to the standard form

$$Z = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$$

standardization is very crucial to compare variables of different magnitudes.

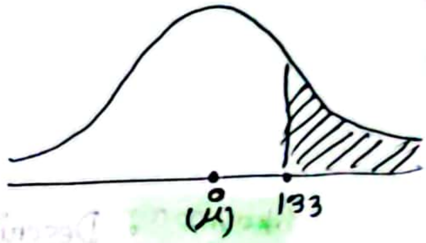
Problem: Suppose, the heights of adult males in a certain population follow a normal distribution with a mean of 68 inch, std of 3 inches. What is the probability that a randomly selected adult male from this population taller than 72 inches?

$x \sim N(68, 3)$



Standardized value of 72 : $Z = \frac{72-68}{3}$ $\left[\frac{x-\mu}{\sigma}\right]$

$= 1.33$



Using the z score in z score table, we found ~~0.988~~ 0.90824, this probability (~~987~~ 90.8%)

So, in the population ~~9.8~~ 90.8% people are shorter or equal to 72 inches.

$\therefore$ Taller people $= (1 - 0.90824)$

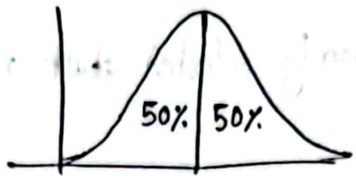
$= 0.09176$

$= 9.2\%$

The **advantage** of converting normal distribution to standard normal distribution is because then we can find the zscore and using the zscore and z table we can find the desired outcome.

Problem: Properties of Normal Distribution:

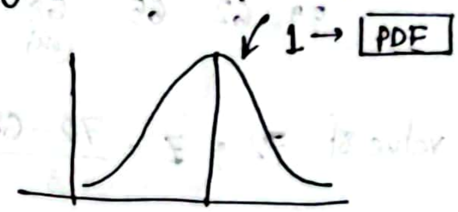
1) Symmentricity →



2) Mean, Mode and Median are equal.

3) Emperical rules followed by every Normal Distribution

4) Area under the curve = 1



Skewness: Describes the degree to which a dataset deviates from the normal distribution. (More noted on PW-skills section)

Moments in Statistics:

1st Moment → Mean

2nd Moment → Variance

3rd Moment → Skewness

4th Moment → Kurtosis.

Formula of sample skew: →

$$\frac{n}{(n-1)(n-2)} \sum \left(\frac{x - \bar{x}}{s} \right)^3 \rightarrow \text{(3rd moment)}$$