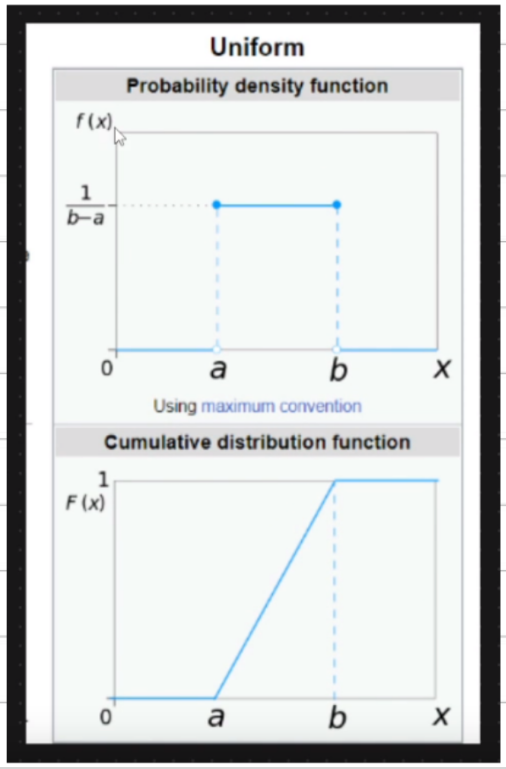
IV {Uniform Distribution}

there are 2 types of Uniform distribution ① Continous
② Discrete

① Continous Normal Distribution :- (for Continous Random Variables)

In probability theory and stastistics, the continous uniform distribution or rectangular distribution is a family of symmetric probability distributions. The distributions describes an experiment where there is an arbitrary outcome that lies b/w a certain bounds. The bounds are defined by parameters a and b, which are the minimum and maximum values. The interval can either be closed (eg. $[a,b]$) or open (eg. (a,b)).



notation: $\boxed{U(a,b)}$

parameters: $-\infty < a \leq b < \infty$

$$\boxed{PDF = \begin{cases} 1/(b-a) & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}}$$

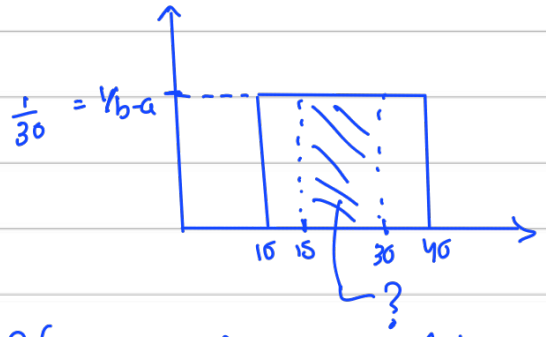
$$\boxed{CDF = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a,b] \\ 1 & \text{for } x > b \end{cases}}$$

$$\boxed{\text{mean} = \frac{1}{2}(a+b), \quad \text{median} = \frac{1}{2}(a+b), \quad \text{variance} = \frac{1}{12}(b-a)^2}$$

Ex:- No of candies sold daily at a shop is distributed with a min. of 10 and max. of 40.

i) what is the probability of daily sales to fall b/w 15 & 30?

Sol^n → as candies sold is uniformly distributed within 10 to 40



$P(15 \le x \le 30)$ = area of the shaded region

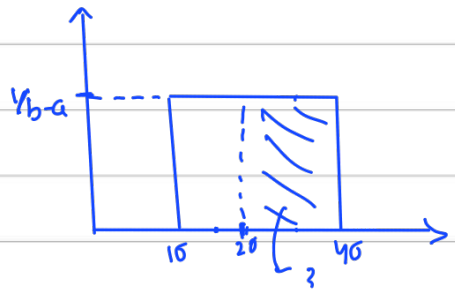
$= (30-15) \times \frac{1}{b-a}$

$= 15 \times \left(\frac{1}{30}\right)$

$= \underline{0.5}$

∴ 50% of daily sales will probably fall b/w this area

ii) $P(x \geq 20) = ?$



$P(x \geq 20) = (40-20) \times 1/30$

$= 20/30$

$= 0.\bar{6}$

$= 66.66\%$ of daily sales.

② Discrete Uniform Distribution: (for discrete random variables)

In probability theory and stastistics, the discrete uniform dist. is a symmetric probability distribution wherein a finite no. of values are equally likely to be observed; every one of n values has equal probability 1/n. Another way of saying "discrete uniform distribution" would be "a known, finite no. of outcomes equally likely to happen".

Ex:- Rolling a dice {1, 2, 3, 4, 5, 6}

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

where $a = 1$ & $b = 6$

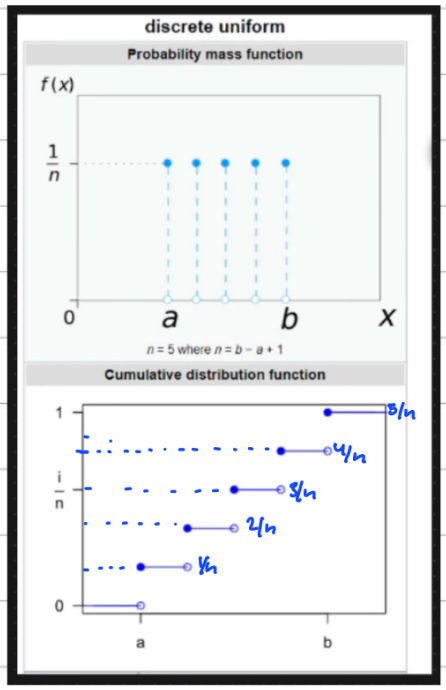
& $\boxed{n = b-a+1}$

Notations used: $U(a,b)$

parameters: a & b, $b > a$

$\boxed{PMF = 1/n}$

$\boxed{\text{Mean \& Median} = \frac{a+b}{2}}$



III) STANDARD NORMAL DISTRIBUTION :-

Let consider a random variable X such that

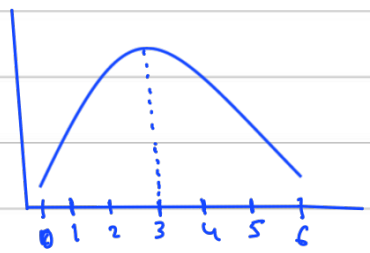
$$X = \{1, 2, 3, 4, 5\}$$

and suppose X is following a std. normal distribution such that

$$\text{mean} = \frac{\sum X}{n} = \frac{15}{5} = 3$$

$$\text{std deviation } (\sigma) = 1.414 \ (\approx 1 \Rightarrow \text{to make calculation easy})$$

So, the gaussian distribution is :-



$\Rightarrow \mu = 3$ & $\sigma = 1.414$

So if we convert the above distribution to a distribution where mean $= 0$ and std deviation $(\sigma) = 1$ then such type of distribution is c/a "<u>Standard Normal Distribution</u>".

Z-Score Method: it is used to convert a gaussian distribution to a normal distribution.

$$\boxed{Z \text{ score} = \frac{x_i - \mu}{\sigma}}$$

$\therefore$ for $x = \{1, 2, 3, 4, 5\}$

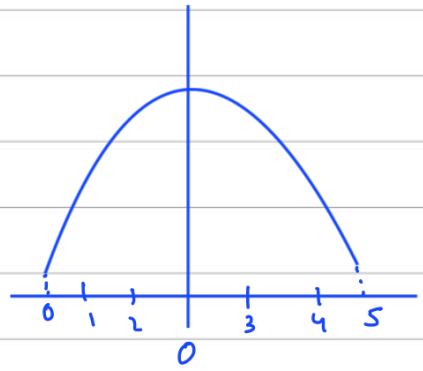
$Z(1) = \frac{1-3}{1} = -2$

$Z(2) = \quad = -1$

$Z(3) = \quad = 0$

$Z(4) = \quad = 1$

$Z(5) = \quad = 2$



Any distribution belongs to a standard normal distribution if its mean (μ) = 0 and $\sigma = 1$

i.e $\boxed{\text{dist}^n \in \text{Std. Normal dist (if } \mu = 0 \text{ \& } \sigma = 1)}$

Q. What is the significance of Z-score? Explain with example

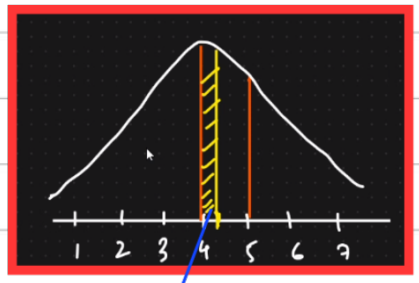
Sol^n → ① Z-score signifiece how much standard deviation is a value away from the mean.

Ex:

Suppose there is a deviation such that its $\mu = 4$ & $\sigma = 1$ then How many std. deviation 4.25 is away from the mean

Sol^n → calculating z-score

$$\text{z-score} = \frac{4.25 - 4}{1} = 0.25$$



$\therefore$ 4.25 is 0.25 std. deviation away from the mean

② Z-score can also be used to find the area under the curve b/w to deviations.

Q) Explain with an example that why we are using z-score and std. normal distribution?

Solⁿ →

Suppose we have a dataset with features: Ag, weight, height, salary

Eg: Dataset

Age (years)	Weight (kg)	Height (cms)	Salary (INR)
24	70	175	40K
25	60	160	50K
26	55	180	60K
27	40	130	30K
30	30	175	20K
31	25	180	70K

all the features have diff units.

But,

To solve a machine learning problem or create a ML model, 1 rule is to bring all the features to a same unit scale

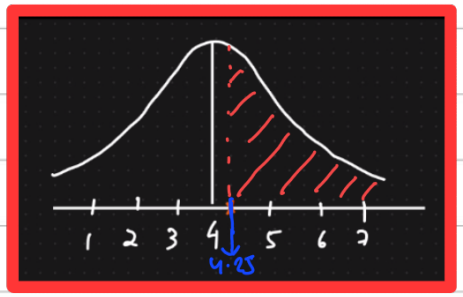
∴ we convert all these features by a technique c/a "Standardization (applying z-score method to all data-set)"

$$\boxed{Z(age) = \frac{x_i - \mu_{age}}{\sigma}}$$

Clustering Algo, Linear Regression, Logistic Regression are some example of algorithms that use standardization technique.

III) Z-SCORE APPLICATION & Z-TABLE :-

Q) for the following guassian distribution what %age of of data falls over 4.25?



1-Step - computing z-score

$$z\ score = \frac{x_i - \mu}{\sigma} = \frac{4.25 - 4}{1} = 0.25$$

∴ 4.25 stands around 0.25 std. deviation away from mean

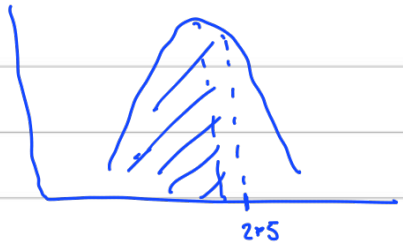
2nd Step: Using z-table to find to find are under the curve

two type of z-tables

① +ve z-table ② -ve z-table

z-table calculates area a left of the z score ∴

∴,



at 4.25 Left side area = 0.59871

to calculate for distribution above 4.25 :-

4.25 Right side area = 1 − 4.25 Left-side area

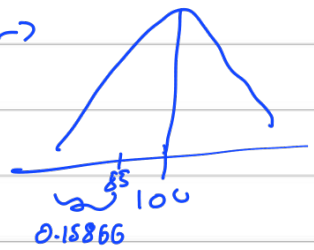
= 1 − 0.59871

= .40129 ≈ .4013

= 40.13% of data falls above 4.25

Qy In India avg IQ is 100, with a std. deviation of 15, What is the % of the population would you expect to have an IQ lower than 85?

Soln →



$\mu = 100$

$\sigma = 15$

$z \text{ score} = \frac{x_i - \mu}{\sigma} = \frac{85-100}{15} = -1$

according to z-table : 0.15866 ≈ 15.86% of IQs in india is below 85

∴ %age of people having IQ above 85 : 100 − 15.86 ≈ 84.14

★ find %age of ppl with b/w 75 ≤ IQ ≤ 100?

z score (75) = $\frac{75-100}{15} = \frac{-25}{15} = -1.66$ & ar = 0.4846 ≈ 48.46

z score (100) = $\frac{100-100}{15} = 0$ & ar = 50%

z score (100) − z score (75) = 1.54%

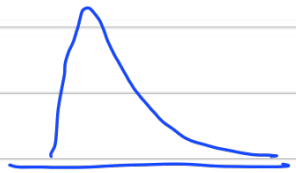
∴, 1.54% of ppl have IQ b/w 75 & 100.

IV) CENTERAL LIMIT THEOREM:-

Centeral limit theorem relies on concept of sampeling distribution, which is the probability distribution of a statistic for a large no. of samples taken from a population.

The Centeral Limit theorem says that the sampeling distribution of the mean will always be normally distributed, as long as the sample size is large enough. Regardless of weather the population has a normal, Possion, bionomial, or any other distribution, the sampling distribution of mean will always be std. normal

ex1: Suppose $X \not\approx N(\mu, \sigma)$ not a normal distribution

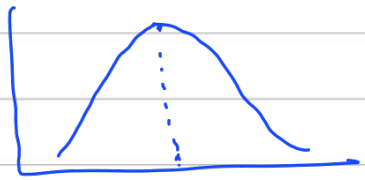


$\therefore$, creating $n \geq 30$ samples for non gaussion distibutions such that

$$S_1 \{x_{11}, x_{12}, x_{13} \ldots x_{1n}\} = \bar{x}_1$$
$$S_2 \{x_{21}, x_{22}, x_{23} \ldots x_{2n}\} = \bar{x}_2$$
$$\vdots$$
$$S_m \{x_{m1}, x_{m2} \ldots x_{mn}\} = \bar{x}_m$$

Now,

according to CLT the distribution of above sample means is always a normal i.e



★ mean of sample distribution remains same as population mean but the std deviation changes from σ to $\sigma/\sqrt{n}$, where $n \rightarrow$ Sample Size.

i.e $X \approx N(\mu, \sigma)$, $\mu_{new} = \mu$ & $\sigma_{new} = \frac{\sigma_{old}}{\sqrt{n}}$