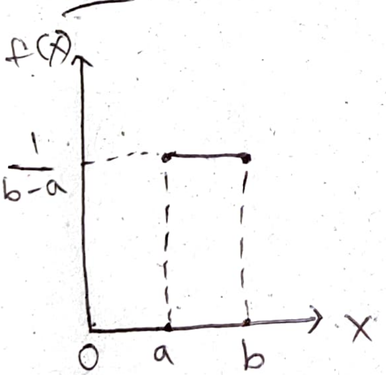
Uniform distribution.

- i) Continuous Uniform Distribution (pdf)
- ii) Discrete Uniform Distribution (pmf)

i) Continuous Uniform Distribution (cont. r.v)

- It is also called ~~regular~~ rectangular distribution.
- The distribution describes an experiment where there is an arbitrary outcome that lies between certain bounds. ~~The~~
- The bounds are defined by the parameters, a and b, which are the minimum and maximum values.
- The interval can be either be closed eg : [a, b] or open eg (a, b)

PDF

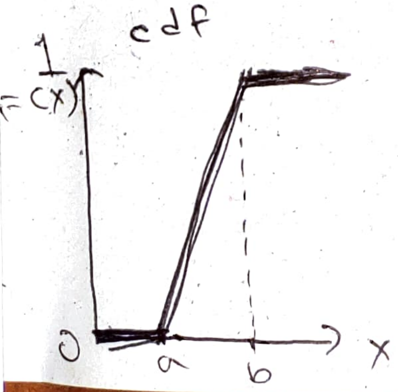


Notation : U(a, b)

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

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

cdf



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

- Mean $= \frac{1}{2}(a+b)$
- Variance $= \frac{1}{12}(b-a)^2$
- median $= \frac{1}{2}(a+b)$

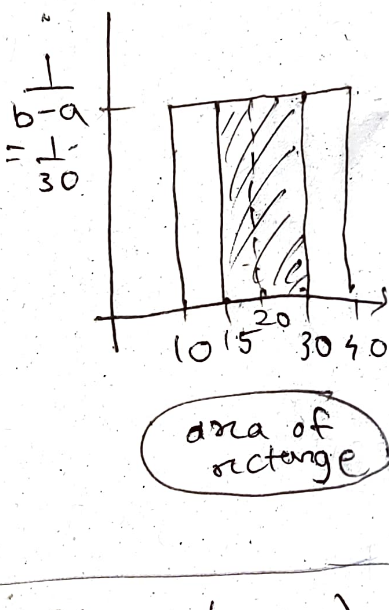
eg → The number of candies sold daily at a shop is uniformly distributed with a minimum of 10 and maximum of 40.

- Prob of daily sales to fall betn 15 and 30?

① $P(15 \le x \le 30) = (30-15) \times \frac{1}{30} = \boxed{0.5}$

length

breadth



② $Pr(x \ge 20) = (40-20) \times \frac{1}{30}$

$= 0.66 \Rightarrow 66.66\%$

area of rectange

ii) Discrete Uniform Distribution (discrete r.v.)

In probability and statistics, discrete uniform distributn is a symmetric distribution where finite no. of values are equally likely to be observed.

If there are 'n' values, then, prob of each value is $(1/n)$.

It is known, finite number of outcomes, equally likely to happen.

eg → rolling a die : $\{1, 2, 3, 4, 5, 6\}$

$P(1) = \frac{1}{6}$, $P(2) = \frac{1}{6}$, ..., $P(6) = \frac{1}{6}$

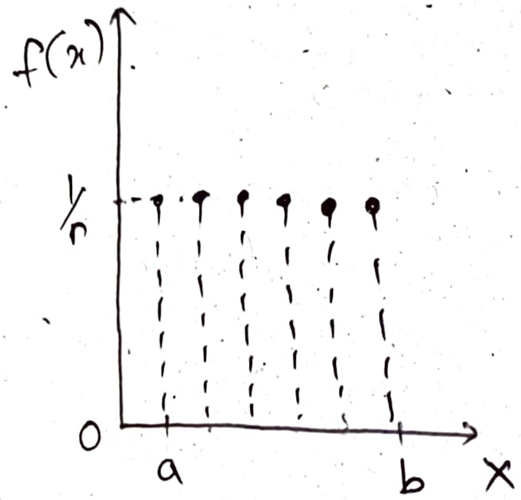
$\boxed{\therefore a = 1}$ $\boxed{b = 6}$

$\boxed{\begin{aligned} n &= b - a + 1 \\ &= 6 \end{aligned}}$

<u>pdf</u>

<u>Notatn:</u> $U(a, b)$

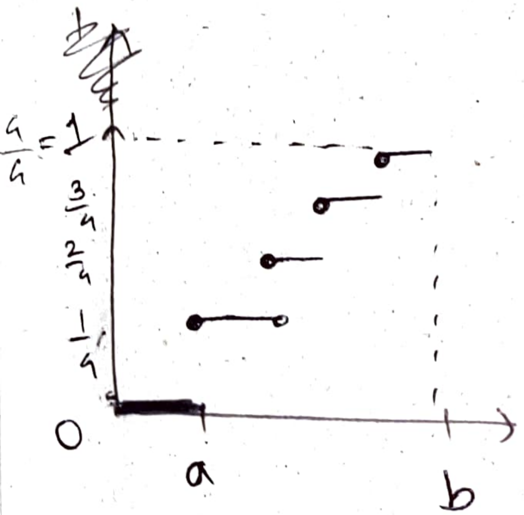
<u>Parameters:</u> a and b $b \geq a$.



$\boxed{\text{pmf} : \frac{1}{n}}$

- $n = 6$

Mean
median $\Big\}\ \frac{(a+b)}{2}$



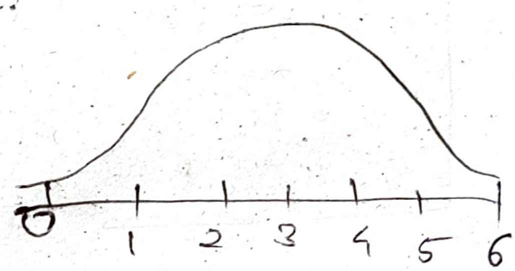
for $n = 4$

standard normal distribution. (Z-score)

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

$\therefore \mu = 3 \qquad \sigma = 1.414 \approx 1$ (assume 1)

- If we transform a distribution in such a way that we get a distribution with mean = 0 & std dev = 1, then that distribution is called as standard normal variate.



Formula:

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

$\{1, 2, 3, 4, 5\} \Longrightarrow \{-2, -1, 0, 1, 2\}$

mean = 0

$\sigma = 1$

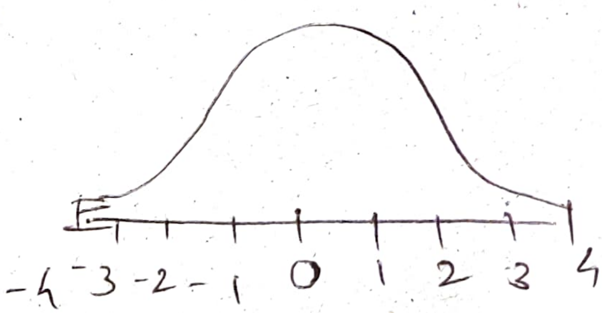
1) $\frac{1-3}{1} = -2$

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

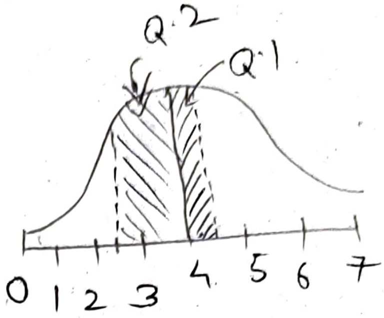
3) $\frac{3-3}{1} = 0$

4) $\frac{4-3}{1} = 1$

5) $\frac{5-3}{1} = 2$



say $\mu = 4$ & $\sigma = 1$



1) How many std dev 4.25 is away from the mean?

$\therefore x_i = 4.25$

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

2) How many std dev 2.5 away from mean?

$x_i = 2.5$ $\therefore$ z-score $= \cancel{4.25}\ \frac{2.5 - 4}{1} = -1.5$

eg → **Dataset**

Age (yrs)	Weight (kg)	Height (cm)	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

above values are of different range. To bring them in same range, we use $\boxed{\text{standardizatn}}$

Standardizatn is applying z-score on each feature

$$\boxed{\text{z-score} = \frac{x_i - \mu_{age}}{\sigma}}$$

same for wt, ht, salary.

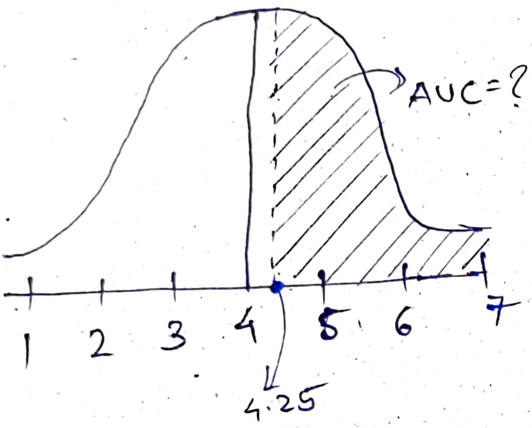
Standardizatn is used on a large scale in ML.

Problem statement on z-score: [z table]

$X = \{1, 2, 3, 4, 5, 6, 7\}$

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

$\mu = 4 \quad \sigma = 1$



* What percentage of scores ~~fall~~ fall above 4.25?

$x_i = 4.25$
$\mu = 4$
$\sigma = 1$

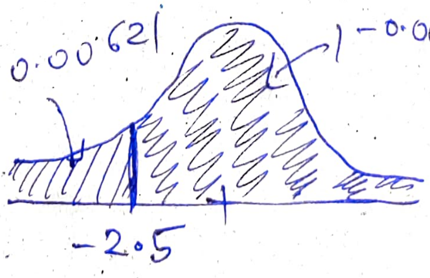
→ z-score = ~~4.25~~ $\frac{x_i - \mu}{\sigma} = \boxed{0.25}$

Hence, 4.25 is 0.25 std dev away from mean.
To find AUC, use z-table

z-score → how many std dev away from mean

$\boxed{\text{negative z-table}}$

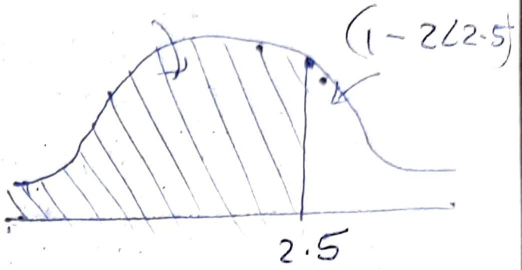
$P(z < -2.5)$
$= \boxed{0.00621}$



$1 - 0.00621$ →

Positive z-table

$P(z < 2.5) =$



(continuing pre. sum).

z-score = 0.25

$P(z < 0.25) = 4.25$

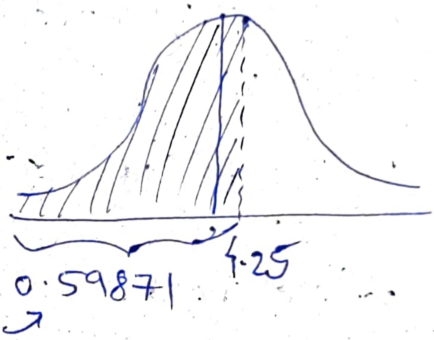
But we want area above 4.25

$\therefore P(z > 0.25)$

$= 1 - 0.59871$

$= 0.4013$

$= \boxed{40.13\%}$



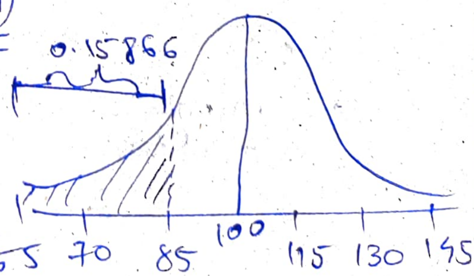
④ In India, avg IQ is 100 with $\sigma = 15$. What is the % of populatn expected to have an IQ lower than 85

$\rightarrow \mu = 100 \quad \sigma = 15$

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

$\therefore P(z < -1)$

$= \boxed{0.15866}$



IQ > 85 $\rightarrow$ $1 - 0.15866$

$= \boxed{84.13\%}$

Central Limit Theorem : It relies on the concept of a sampling distribution, which is the prob. distribution of a statistic for a large number of samples taken from a population.

CLT states that sampling distribution of the mean will always be normally distributed as long as the sample size is large enough. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution the sampling distribution of the mean will be normal.

$X \approx N(\mu, \sigma)$

n = samplesize (any value)

$S_1 = \{x_1, x_2, x_3, ----, x_n\} = \bar{x}_1$

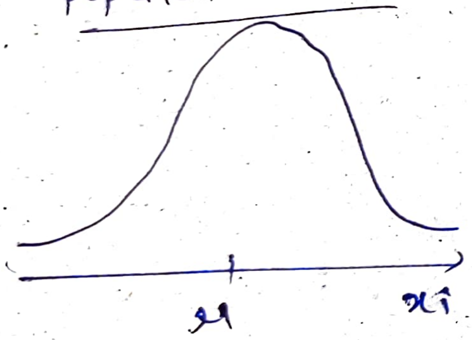
$S_2 = \{x_2, x_3, ----, x_n\} = \bar{x}_2$

$S_3 = \{ \qquad \bar{x}_3$

$\vdots$

$S_m \qquad \bar{x}_m$

populatn DATA



If we take this sample having mean $\bar{x}_1, \bar{x}_2, \bar{x}_3, ---\bar{x}_m$. & plot, we will get gaussian curve.

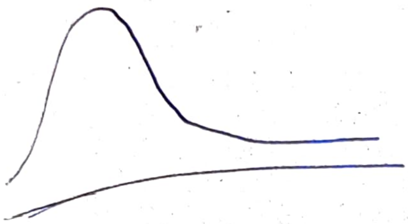
$\Downarrow$

Sampling distribution of mean

gaussian/normal dist

② $X \not\approx N(\mu, \sigma)$

$\boxed{n \geq 30}$ → sample size



$S_1 \rightarrow \bar{x}_1$

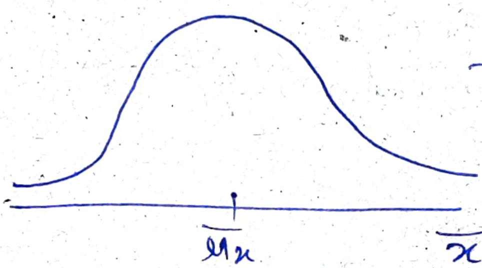
$S_2 \rightarrow \bar{x}_2$

$\vdots$

$\bar{x}_m$

$\Downarrow$ $\boxed{\text{CLT}}$

This will also follow gaussian/normal dist



$\rightarrow X \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Note: mean of sampling distribution will be same as population mean & std dev will be $\frac{\sigma}{\sqrt{n}}$

~~sampl~~