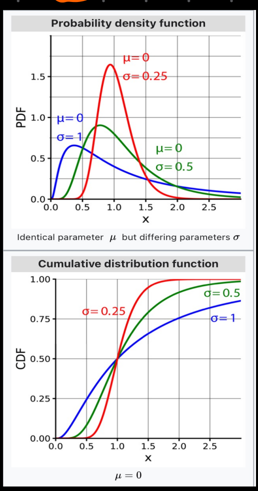
Statistics-3

Types of Distribution:

1) Normal | Gaussian Distribution
2) Standard Normal
3) log Normal
4) Power Law
5) Bernoulli
6) Binomial
7) Poisson Distribution
8) Uniform Distribution { Discrete / continous
9) Exponential Distribution
10) F Distribution
11) CHI square Distribution
12) Hypothesis testing.

* Log Normal Distribution: {continous Random variable}



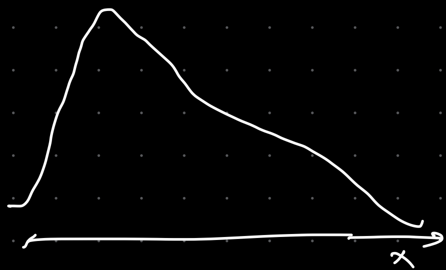
* In Probability theory, a log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed.

* Thus, if the random variable X is log-normally distributed, then $Y = \ln(X)$ has a normal distribution. Equivalently, if Y has a normal distribution, then the exponential function of Y, $X = \exp(Y)$, has a log-normal distribution.

$X \approx \text{lognormal}(\mu, \sigma^2)$

$Y \approx \ln(X) \longrightarrow$ Normal distribution (μ, σ^2)

ln = natural log ($\log_e$)

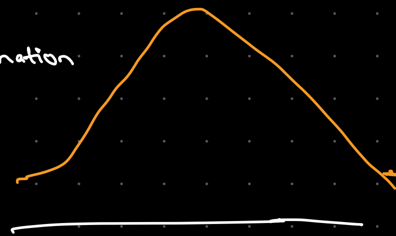


Logarithm Transformation

$\Rightarrow$

$\ln(x)$

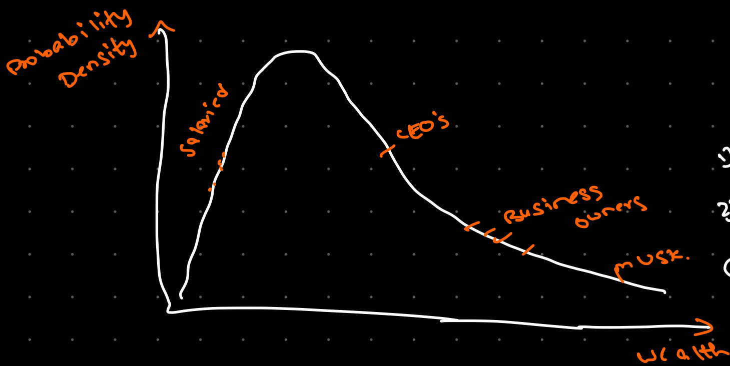
$\Leftarrow$



$X = \exp(y) \Rightarrow e^y$

$\Downarrow$

model will get trained properly



Data

1) wealth distribution of world

2) salaries distribution in company

3) People writing length of comments

Notation: $\text{lognormal}(\mu, \sigma^2)$

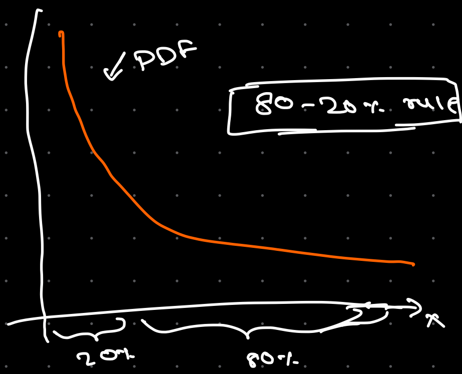
Parameters: $\mu \in (-\infty, +\infty)$

$\sigma > 0$

pdf:
$$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)$$

Power law Distribution

(continuous Random variable)



* In statistics, a power law is a functional relationship between two quantities where a relative change in one quality results in a proportional relative change in the other quantity, independent of initial size of those quantities, one quantity varies as a power of another.

For instance, considering the area of the square in terms of length of its side, if the length is doubled, the area is multiplied by a factor of 4.

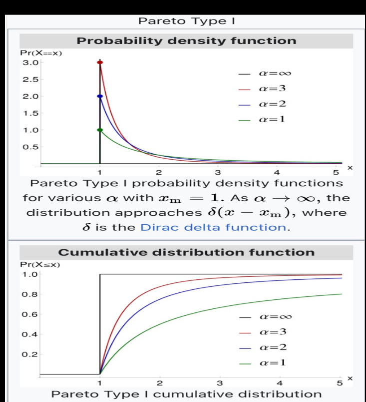
Ex:

1) 20% of the team is responsible for 80% of match.
2) 80% of the sales in Amazon is derived from 20% of the products.
3) 80% of the wealth is distributed among 20% of people.
4) 80% of the project is completed by 20% of team.

Types of power law Distribution

① Pareto distribution ② Exponential distribution.

Pareto distribution (continuous Random variable).



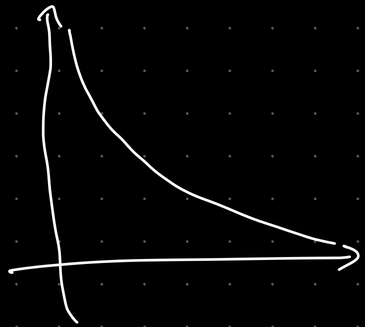
Pareto Type I probability density functions for various α with $x_m = 1$. As $\alpha \to \infty$, the distribution approaches $\delta(x - x_m)$, where δ is the Dirac delta function.

Pareto Type I cumulative distribution functions for various α with $x_m = 1$.

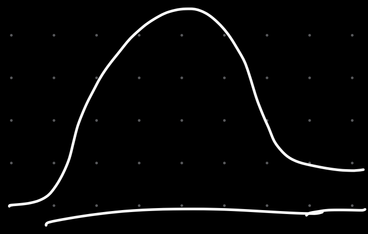
Parameters	$x_m > 0$ scale (real) $\alpha > 0$ shape (real)

* If X is a random variable with a Pareto (Type I) distribution, then the probability that X is greater than some number x, ie, the survival function (also called tail function) is given by,

$$\bar{F}(x) = \Pr(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^\alpha & x \geq x_m \\ 1 & x < x_m \end{cases}$$



Box-cox
Transformation
$\Rightarrow$
$\Downarrow$
α



Bernoulli Distribution:

outcome of a process is binary $\{1, 0\}$

tossing a coin

$Pr(H) = 0.5 \Rightarrow p$

$Pr(T) = 0.5 \Rightarrow 1-p = q$

$\boxed{p+q=1}$

$\boxed{Pdf = p^k (1-p)^{1-k}}$

$= p(1-p)^0$

$= p$

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

pmf

* Binomial Distribution

Binomial distribution is n times Bernoulli Distribution

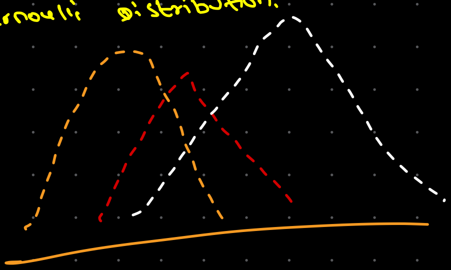
$\Rightarrow$ Binomial = n x Bernoulli

or set of Bernoulli distribution

* Combination of multiple Bernoulli distribution.

$n, p \qquad k$

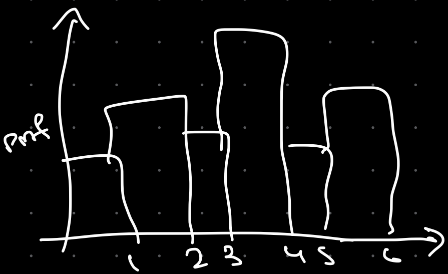
$$Pmf = {}^nC_k \, p^k (1-p)^{n-k}$$



Poisson Distribution

No. of people visiting bank every hour

$\boxed{\lambda = 3}$ → Expected no. of people to come at that specific time interval



PDF — probability density

pmf — probabilities

pmf $\quad Pr(x == 5) = \dfrac{e^{-\lambda} \lambda^{x}}{x!}$

$\lambda = 3$

$$Pr(x=5) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-3} 3^5}{5!}$$

$$= 0.101 = 10.1\%$$

$$Pr(x=5 \text{ or } 6) = Pr(x=5) + Pr(x=6)$$

$$= \frac{e^{-\lambda}\lambda^5}{5!} + \frac{e^{-\lambda}\lambda^6}{6!}$$

* Uniform Distribution

Continuous uniform Distribution (PDF)

Discrete uniform Distribution (PMF)

Continuous uniform Distribution

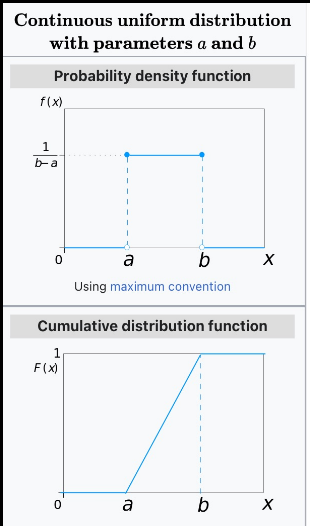
The number of candies sold daily at a shop is uniformly distributed

[15 - 30] $\quad$ {min, max] ⇒ Interval.

Notation : $U(a, b)$

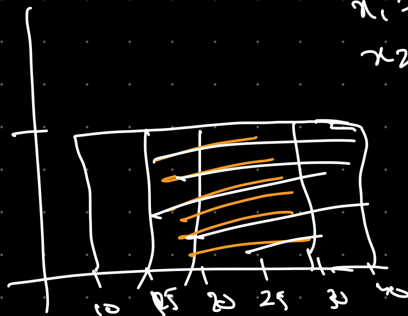
Parameters: $-\infty < a < b < \infty$

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



Ex: The number of candies sold daily at a shop is uniformly distributed with a maximum of 40 and a min of 10.

i, Probability of daily sales to fall between 15 and 30.



$x_1 = 15$

$x_2 = 30$ $\Pr(15 \le x \le 30)$

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

$= 15 \times \frac{1}{30}$

$= 0.5$

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

$= \frac{20}{30} = 0.66\%$

② Discrete uniform Distribution {Discrete Random variable}

Rolling a dice $= \{1,2,3,4,5,6\}$

$\Pr(1) = \frac{1}{6}$ $\Pr(2) = \frac{1}{6}$

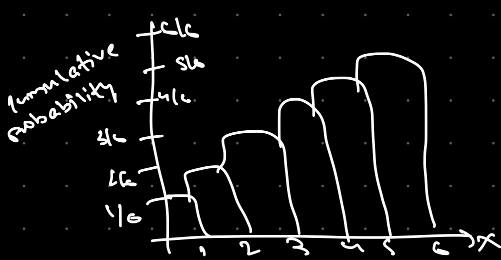
$n = b - a + 1$

$n = 6 - 1 + 1 = 6$

notation $= U(a,b)$

Parameters $= a, b$ with $b > a$

PMF $= 1/n$



$\Pr(x = 1 \text{ or } 2)$

Hypothesis testing (Inferential statistics)

① P-value

Hypothesis testing   Person → crime

① Null hypothesis (H_0) → Person has not committed crime

Alternate hypothesis (H_1) → Person has committed crime.

② Experiments:-

Proofs, DNA, finger prints, evidences

↳ Judge → person committed crime.

③ Reject the null hypothesis

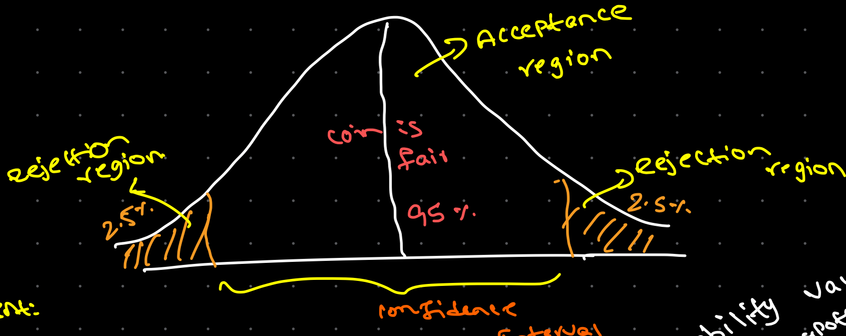
or

we fail to reject null hypothesis.

Ex: coin is fair or not through 100 experiments

H_0 → coin is fair

H_1 → coin is not fair



Experiment:

Confidence Interval = 95%

$\alpha = 1 - CI$
$= 0.05$

P.value < 0.05 → probability value for the null hypothesis to be true.