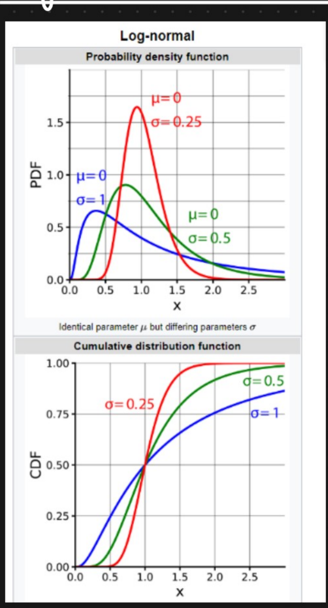
Agenda

Different types of Distribution

1. Normal / Gaussian Distribution ✓
2. Standard Normal Distribution ✓
3. Log Normal Distribution ✓
4. Power Law Distribution ✓
5. Bernoulli Distribution ✓
6. Binomial Distribution ✓
7. Poisson Distribution ✓
8. Uniform Distribution → Discrete, → Continuous ✓
9. Exponential Distribution ✓
10. F distribution ✓
11. Chi square distribution ✓
12. Hypothesis Testing }. ✓

(10)–(11): ANNOVA

① Log Normal Distribution {Continuous Random Variable}.

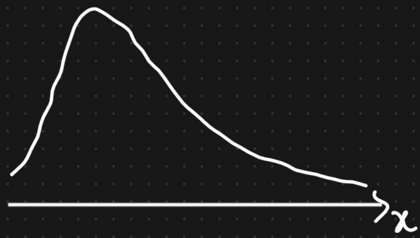


In probability theory, a log-normal (or lognormal) distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then Y = ln(X) has a normal distribution. Equivalently, if Y has a normal distribution, then the exponential function of Y, X = exp(Y), has a log-normal distribution.

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

$$Y \approx \ln(x) \Rightarrow \text{Normal Distribution}(\mu, \sigma^2).$$

ln = natural log ($\log_e$)



Age

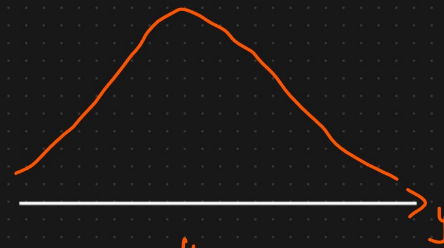
Logarithm Transformation

$\Rightarrow$

$\boxed{\ln(x)}$

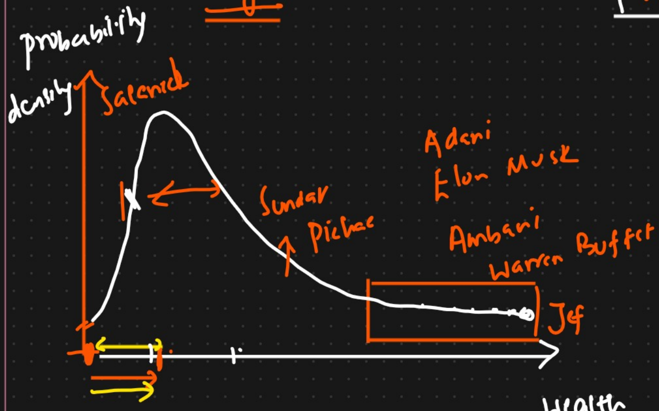
$\Leftarrow$

$\boxed{X = \exp(y)}$



$\Downarrow$

Model will get Trained efficient



DATA

1. Wealth distribution of the world
2. Salaried distribution in a Company
3. People writing length of comments
4. User spending time in read Articles

Wealth

$\Downarrow$

Notation $\quad \text{LogNormal}(\mu, \sigma^2)$

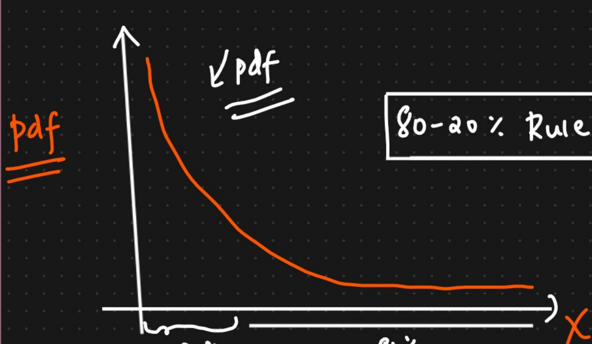
Parameters $\quad \mu \in (-\infty, +\infty)$

$\sigma > 0$

Pdf :

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

4. Power Law Distribution [Continuous Random Variable]



$\boxed{\text{80-20\% Rule}}$

In statistics, a power law is a functional relationship between two quantities, where a relative change in one quantity results in a proportional relative change in the other quantity, independent of the initial size of those quantities: one quantity varies as a power of another. For instance, considering the area of a square in terms of the length of its side, if the length is doubled, the area is multiplied by a factor of four

Eg: IPL Games

RCB

Project

1A 1PM 1BA

5-6 S/D

① 20% of the team is responsible for winning 80% of the match

② 80% of the sales in Amazon is derived from 20% of the product. Lead

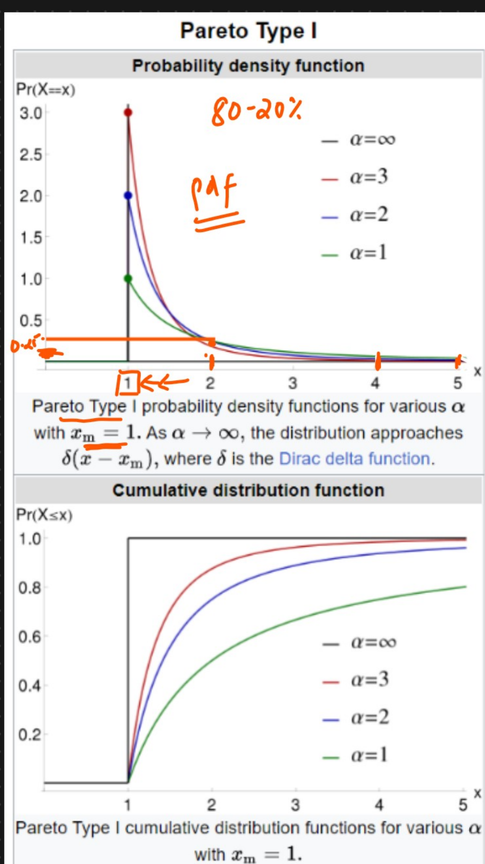
③ 80% of the wealth is distributed among 20% of the people.

④ 80% of the projection complete by 20% of team.

Types of Power Law Distribution

① Pareto Distribution ② Exponential Distribution

① Pareto Distribution [Continuous Random Variable]



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

pdf =

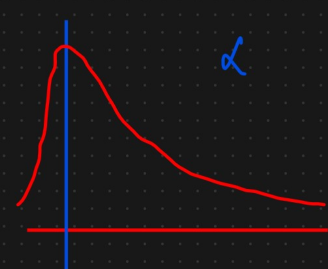
$$\overline{F}(x) = \Pr(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^{\alpha} & x \geq x_m, \\ 1 & x < x_m, \end{cases}$$

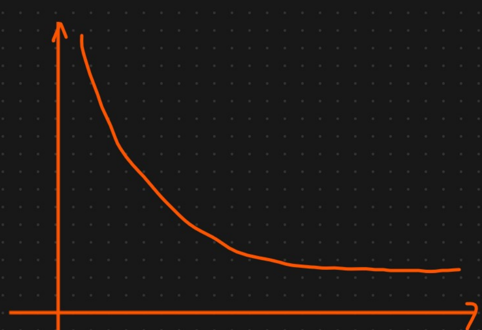
$$Pr(x \geq 4) = \left(\frac{1}{4}\right)^2 = \boxed{\frac{1}{16}}$$

$$Pr(X == 4) = \left(\frac{1}{2}\right)^2 = \frac{1}{4} = 0.25$$

$x_m = 2$

$$\left(\frac{1}{1}\right)^2 = 1$$





Box Cox

Transformation ⇒

⇓

α

(8) Bernoulli Distribution ÷ Outcome of the process is binary $\{1, 0\}$

$\{$Success, Failure$\}$.

Tossing a Fair Coin

$Pr(H) = 0.5 \Rightarrow p$

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

$\Rightarrow p + q = 1$

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

$= p(1-p)^0$

$= p$

$k \{1, 0\}$

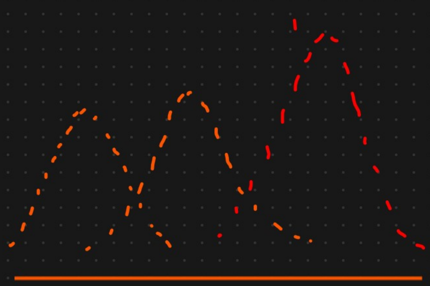
$$\begin{cases} p & \text{if } k=1 \\ 1-p=q. & \text{if } k=0 \end{cases}$$

pmf

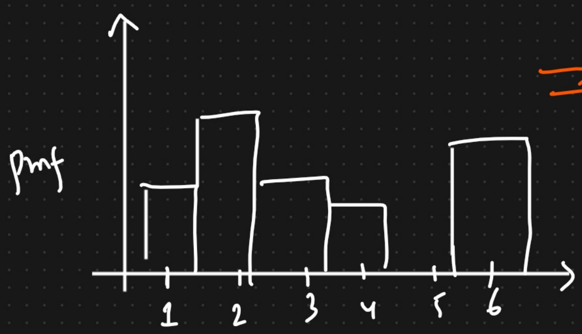
(9) Binomial Distribution → Combination of multiple Bernoulli Distribution

n, p k

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



(10) Poisson Distribution



No. of people visiting bank every hour

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

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

pdf

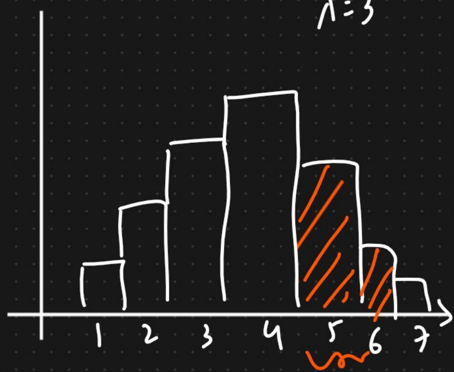
probability density

$f(x)$

pmf

Probability

$\lambda = 3$



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

$$= 0.101 = 10.1\%$$

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

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

4 Uniform Distribution

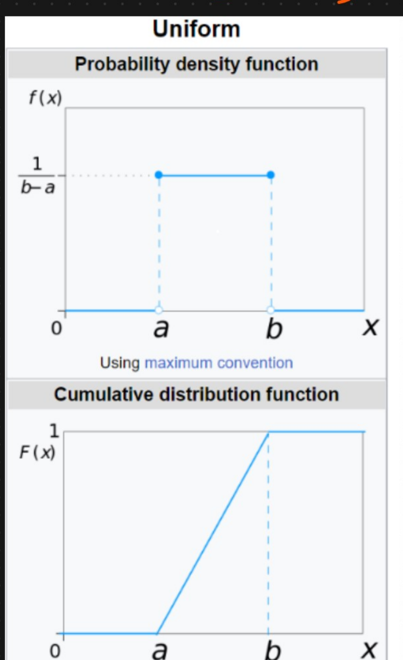
1 Continuous Uniform Distribution (pdf)

2 Discrete Uniform Distribution (pmf)

1 Continuous Uniform Distribution {Continuous Random Variable}

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

[15-30]   [Max, Min] ⇒ Interval



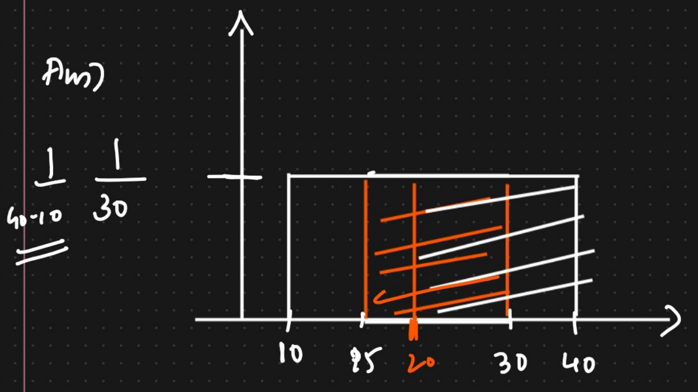
Notation: $U(a,b)$ $\boxed{b >> a}$

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

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

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

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



$x_1 = 15$

$x_2 = 30$

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

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

$$= \cancel{15}^{1} \times \frac{1}{\cancel{30}\,2} = 0.5$$

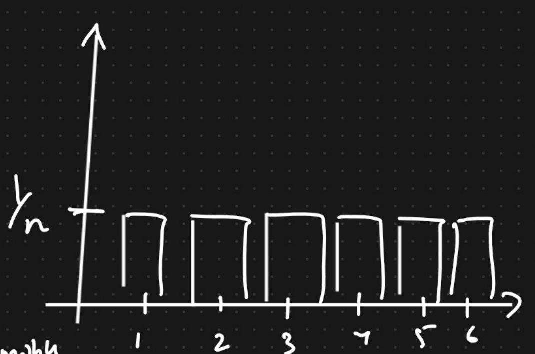
$$Pr(x \geq 20) = (40-20) * \frac{1}{30}$$

$$= \frac{2\cancel{0}}{3\cancel{0}} = 66.66\%$$

② Discrete Uniform Distribution {Discrete Random Variables}.

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

$Pr(1) = 1/6 \quad Pr(2) = 1/6$



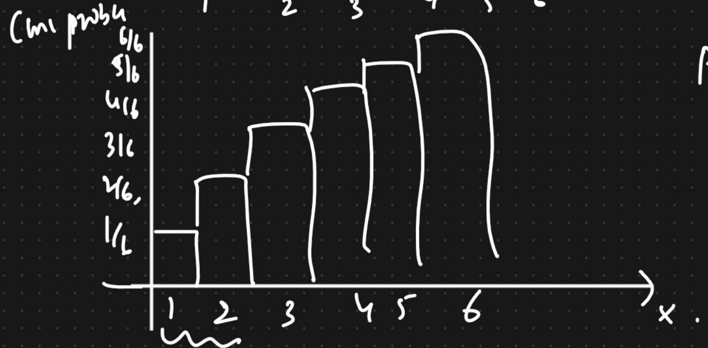
$n = b - a + 1$

$n = 6 - 1 + 1 = 6$

Notation $U(a, b)$

Parameters a, b with $b > a$

PMF $1/n$

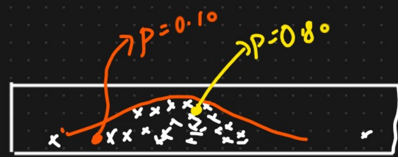


$Pr(X = 1 \text{ or } 2)$

④ <u>Hypothesis Testing</u> [Inferential Stats]

① <u>P value</u>

Out of 100 touches in their Space bar 10 times touching in that area.



<u>Hypothesis Testing</u>

<u>Person → Crime</u>

① Null Hypothesis H_0 – Person has not Committed crime.

Alternate Hypothesis H_1 – Person has Committed Crime

② <u>Experiments</u> : Proofs, DNA, finger prints, evidences ⇒ Judge ⇒ Person has Committed Crime

③ Reject the Null Hypothesis

OR

We fail to Reject H_0.

① Coin is fair or No through 100 experiments

H_0 = Coin is fair

H_1 : Coin is not fair

<u>Experiment</u> ;

C.I = 95%

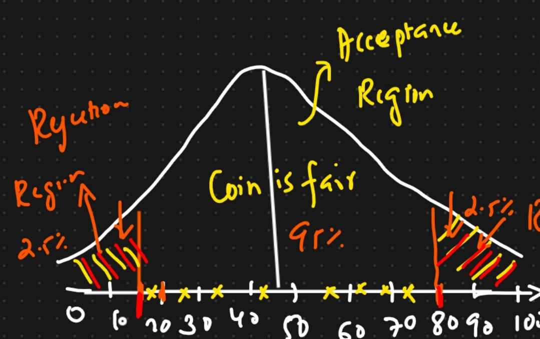
$\alpha = 1 - C.I = 0.05$

Significance Level

50 Head 50 Tails

60 Head 40 Tails

<u>70 Head</u> <u>30 Tail</u>



2.5% Rejection Region

No of Heads

$\boxed{\text{P value}} < 0.05$

$\Downarrow$

probability

value for

the Null hypothesis

To be True

$\Downarrow$

Confidence Interval

Assignment : Exploring Distribution