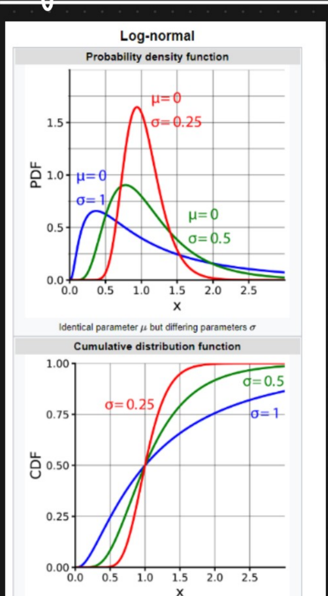
Agenda

Diffuent 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 ✓ } ANNOVA
12. Hypothesis Testing }. ✓

① Log Normal Distribution {Continuous Random Variable}.



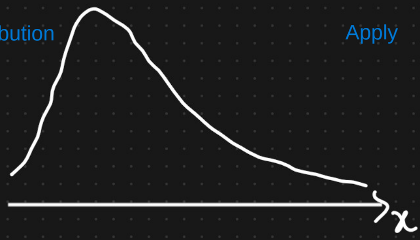
In probability theory, a log-normal (or lognormal) distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then Y = ln(X) has a normal distribution. Equivalently, if Y has a normal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

Logarithmic distribution (which are right skewed)



Apply Logarithm Transformation

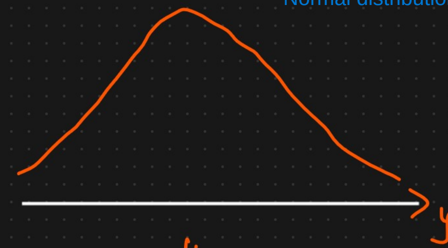
$\Rightarrow$

$\boxed{\ln(x)}$ (This is called natural log)

$\Leftarrow$

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

Normal distribution

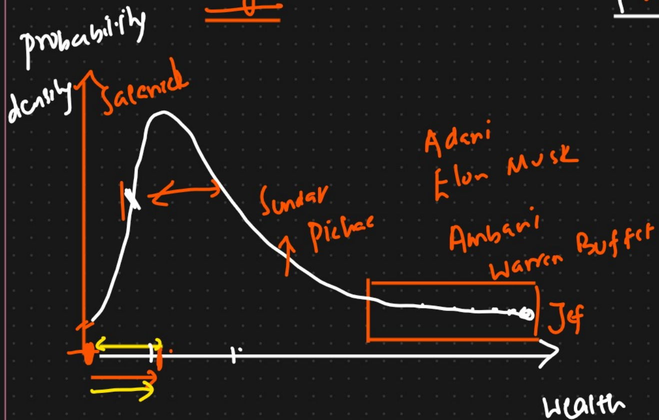


After doing log transformation

$\Downarrow$

Model will get Trained efficient

Age



Wealth

$\Downarrow$

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

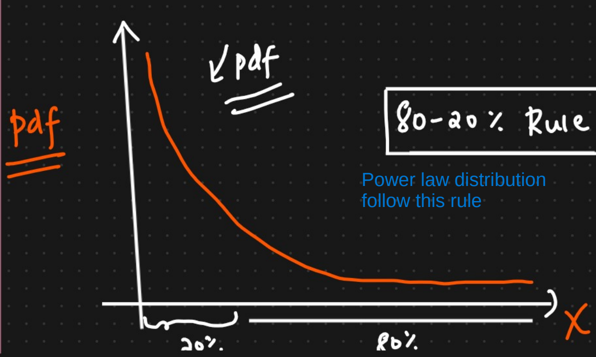
Notation: $\text{LogNormal}(\mu, \sigma^2)$

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

$\sigma > 0$

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

4. Power Law Distribution [Continuous Random Variable]



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

Power law distribution follow this 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/=

lead

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.

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

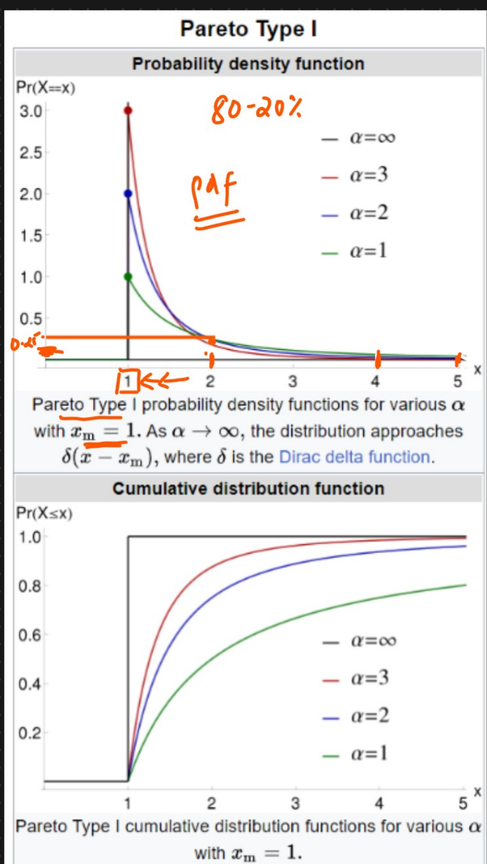
④ 80% of the projection complete by 20% of team.

Types of Power Law Distribution

① Pareto Distribution ② Exponential Distribution

① Pareto Distribution [Continuous Random Variable]

*extra-- for transforming log normal to pareto
put straight line at the max probability for log normal
then it will be pareto



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 $x_m = 1$.

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 =

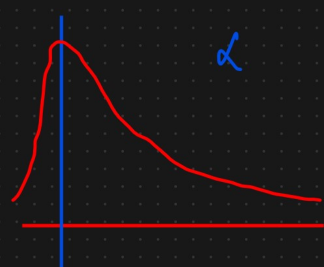
$$\overline{F}(x) = \Pr(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^{\alpha} & x \geq x_m, \\ 1 & x < x_m, \end{cases}$$

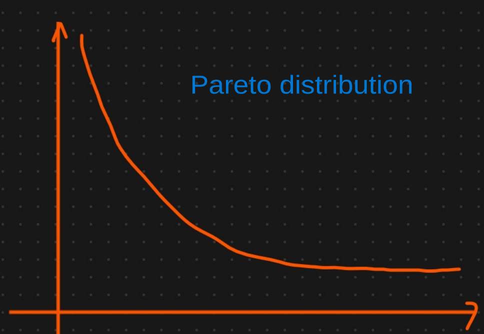
$$\Pr(x \geq 4) = \left(\frac{1}{4}\right)^2 = \boxed{\frac{1}{16}}$$

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

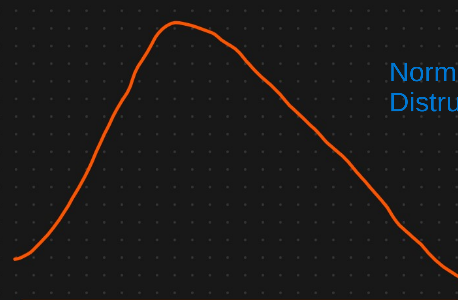
$x_m = 2$

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





Convert using
Box Cox
Transformation ⇒
⇓
α Alpha will be used here



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

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

It will follow PMF - probability mass function as outcome is discrete.

Tossing a Fair Coin

$Pr(H) = 0.5 \Rightarrow p$

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

$\Rightarrow p + q = \underline{\underline{1}}$

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

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

$= p$

$K \; \{1, 0\}$

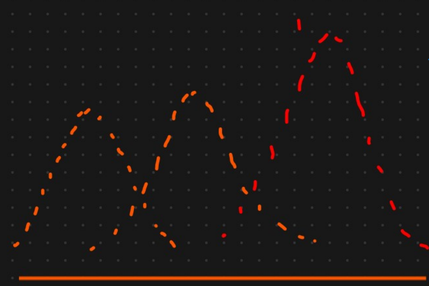
$$\begin{cases} p & \text{if } K=1 \\ 1-p=q. & \text{if } K=0 \end{cases}$$

$\underline{\underline{pmf}}$

9. Binomial Distribution → Combination of multiple Bernoulli Distribution

$\underline{\underline{n}}, \underline{\underline{p}}$ $\qquad$ $\underline{\underline{K}}$

$$pmf = {}^{n}C_{K} \; p^{K}(1-p)^{n-K}$$



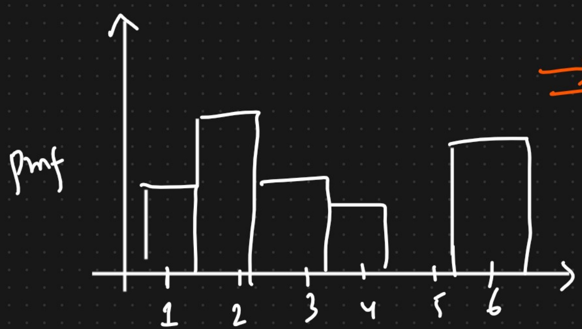
It is curve graph but still it is PMF that is why it is dotted.

10. Poisson Distribution

No. of operation happening at some specific interval of time.

No. of people visiting bank every hour

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



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

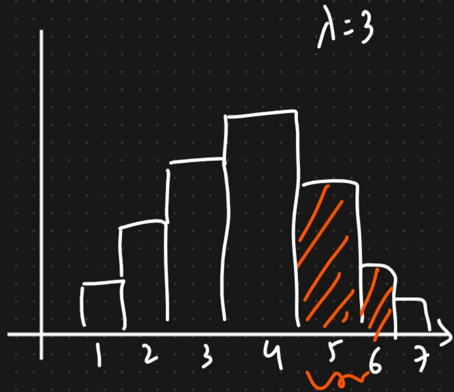
$\underline{\underline{pdf}}$

probability density

$f(x)$

$\underline{\underline{pmf}}$

Probability



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

$$= 0.101 = 10.1\%$$

$$Pr(X=5 \text{ or } 6) = Pr(x=5) + Pr(x=6)$$
$$= \frac{e^{-\lambda}\lambda^{5}}{5!} + \frac{e^{-\lambda}\lambda^{6}}{6!}$$

④ Uniform Distribution

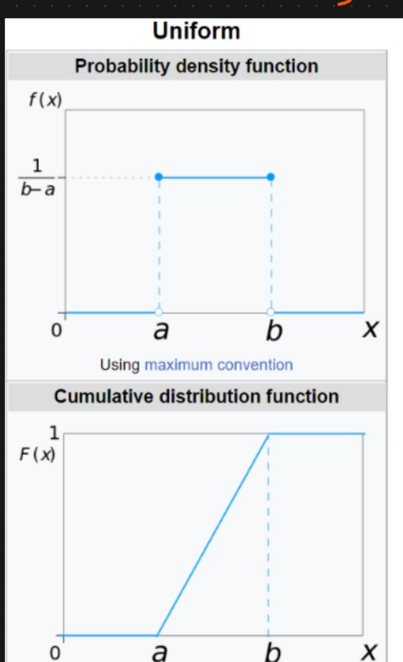
① Continuous Uniform Distribution (pdf)
② Discrete Uniform Distribution (pmf)

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
[MIN, MAX]

If dataset has any range where it is continuous then it is continuous uniform distribution



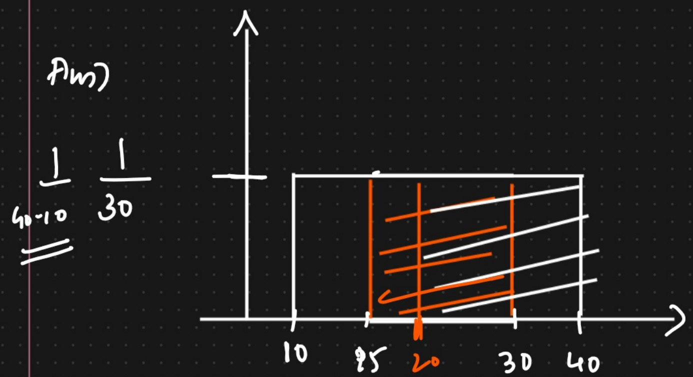
Notation : $U(a,b)$ $\boxed{b >> a}$

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

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

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.



$x_1 = 15$

$x_2 = 30$

$Pr(15 \leq x \leq 30)$ = length X breadth

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

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

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

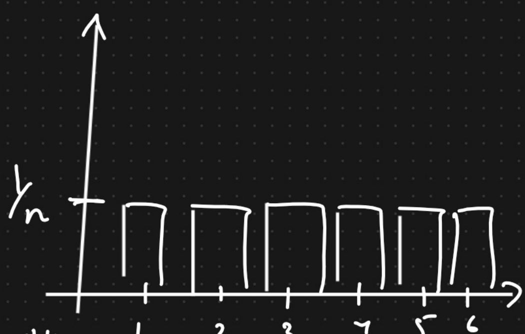
$= \frac{2\cancel{0}}{3\cancel{0}} = 66.66\%$

② <u>Discrete Uniform Distribution</u> {Discrete Random Variables}.

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

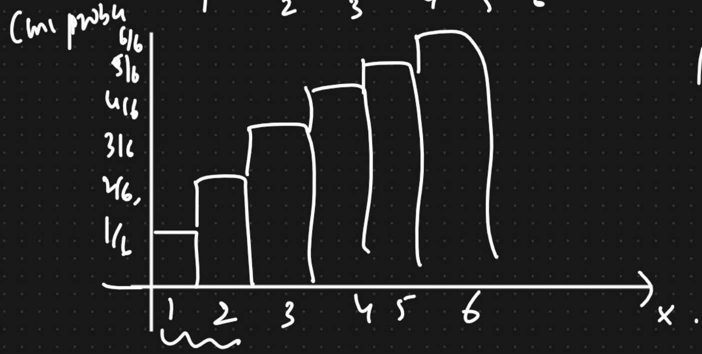
Here outcome of the graph comes as dotted graph for each data and data is not continuous. It is discrete random variables.

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



$n = \overset{max}{b} - \overset{min}{a} + 1$

$n = 6 - 1 + 1 = 6$



Notation $U(a, b)$

Parameters a, b with $b > a$

PMF $1/n$

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

F- distribution is used to compare
Variance between two groups

in CHI- square distribution when k value or
degree of freedom increase then
graph change from exponential to log normal
distribution but it never goes to Normal distribution

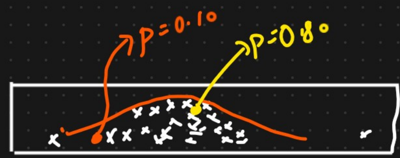
One key point is exponential, log normal and pareto are
CHI- square distribution.

④ Hypothesis Testing [Inferential Stats]

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

① P value

It is probability of null hypothesis
to be true



Hypothesis Testing

Person → Crime

Hypothesis testing is a powerful tool that can be used to make inferences about population parameters. However, it is important to remember that hypothesis testing is not a perfect science. There is always a chance of making a Type I error, which is rejecting the null hypothesis when it is actually true. There is also a chance of making a Type II error, which is failing to reject the null hypothesis when it is actually false.

The probability of making a Type I error is controlled by the significance level. The probability of making a Type II error depends on the size of the effect and the sample size.

When designing a hypothesis test, it is important to consider the Type I and Type II error rates and to choose a significance level that is appropriate for the situation.

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 has Committed Crime

③ Reject the Null Hypothesis

OR

We fail to Reject H_0. If judge will say that person has not committed a crime so we fail to reject null hypothesis.

① Coin is fair or No through 100 experiments

H_0 = Coin is fair

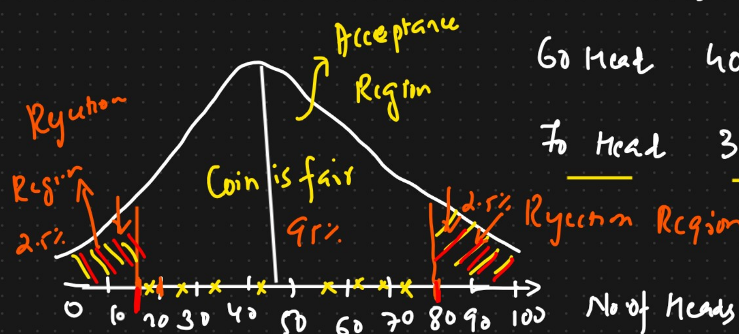
H_1 : Coin is not fair

Experiment :

50 head 50 Tails

60 Head 40 Tails

70 Head 30 Tail



C.I = 95%

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

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

if this condition become true then
we will be fail to rejet the null hypothesis.

probability
value for
the Null hypothesis
To be True

Confidence Interval

Basically means it is the probability of the value
which we have taken as to be true

Assignment : Exploring Distribution