STATISTICS 1-2-2024

01 February 2024 18:30

Distribution Identification Structure

2 types of Problems → Business Problem/Statement
→ Data Analytics

Binomial Dist

(10) → (6)

Negative Binomial Dist

(6) → trial??

trial → ∞

90623
95123

Binary

A very interesting term

Tossing a Coin!

$P(H) = ?$

H, TH, TTH, TTTH, ...

$\frac{1}{2}$ $\frac{1}{2}\cdot\frac{1}{2}$ $\left(\frac{1}{2}\right)^3$ $\left(\frac{1}{2}\right)^4$ ....

$P(\text{Sum}) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots$

$= \frac{1/2}{1-1/2} = \frac{1/2}{1/2} = 1$

Independent Event

Const Prob of Success.. (Fair)

Discrete nature of Data

Limited outcomes..

Named (Product Quality Control)

Advanced Product Quality Control

Medical Trials (5 Days) [5]

Negative Binomial Distribution → Success, Failure

10 → 3

1000 → 10

[10] → 0.1

[10] → [5]

Negative → failure

→ When to apply???

(a) Total trials needed to reach success

$\frac{e^{-\lambda} \lambda^{x}}{x!}$

(b) Overdispersed data ..

Poisson ⇒ mean = Var, λ, $\sqrt{\lambda}$

Var > mean

(c) Modelling a success before stopping ..

More than 1/2/3 .... no of success are pre-decided or pre-registered ..

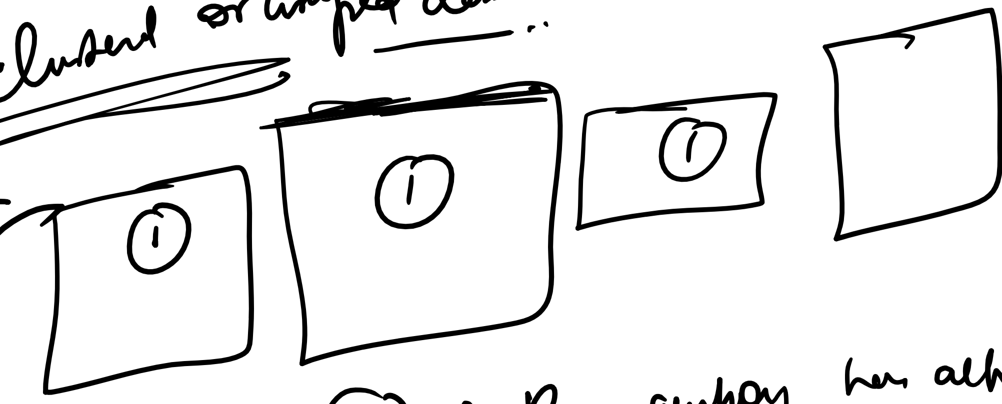
(d) Survival Analysis ..

→ death

... or Grouped data ..

(e) clustered or Grouped data ..

Size's are different



$Y = x f(x_i) + 8 x_j$ Every question has alter...

100 Questions

50/Something else

50x2 => 100

1) a b c) 3
2) a b c d e f 6
3) a b 2
4) a b c d 4
5

Marriage hall 5th

Only Paid members

Paid Employees

3.5 lakh f 2.5 lakh

3-4 lakh 5-6

CEO

India

Jun

Reg

Sub

56

15000

10

Long Tailed dist

Rare/extreme Case / very high or very low

5,72,893.8

Rare ...

-673.25

5,72,893.8

10 to 100

Bin

Trial fixed

Binary outcome

$p+q=1$

$np > npq$

np

npq

O D ✗

Right skewed

symmetric $p=q$

-ve Bin

∞

Non-negative integer counts of success

x/α

λ/α ✓

O D ✓ (var > mean)

+ve, -ve α, λ

Gamma

Poisson

$$\frac{e^{-\lambda} \lambda^{x}}{x!}$$

mean = λ = var

(i) Rare events

(ii) fixed interval..

(iii) Large Population

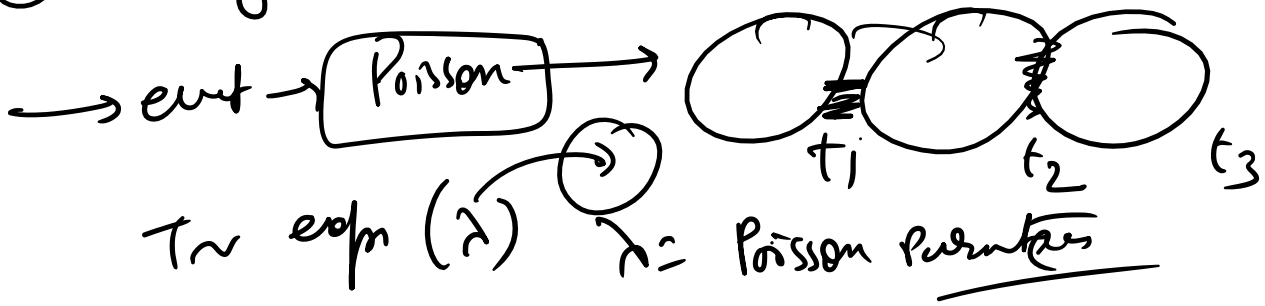
(iv) Approximation to Binomial

** $np < 10$

(iv)

Exponential distribution

(i) Waiting time between Poisson Events



$T \sim \exp(\lambda)$ λ = Poisson parameter

(ii) Memory Lesson

Conditional Prob of an event occurring in future given that it has not occurred until a certain Point

$$P(X > s+t \mid X > s) = P(X > t)$$

Random var Reporting time

$X \rightarrow$ time

$s, t \geq 0$

NORMAL Dist

CLT

Sampling of Sample mean of sufficiently large data (iid)

Dist → ND

$X_1, X_2 \ldots X_n$ → $\bar{X}$ → ND

$x_1, x_2 \ldots x_n$ $(\bar{X}) \rightarrow$ ND

$n \rightarrow \infty$

$\frac{\sigma}{\sqrt{n}}$

(3) Population dist is Symmetric

Single Parameter

ND

(4) Large Data $n > 30$

t test $n < 30$

$n = 30 \rightarrow$ ND / Z Test

Bin	Poi	Normal
n = finite	∞	×
Pr of success (p)	(p)	×
Trial indep	✓ ✓	×
OD ×	equal	×