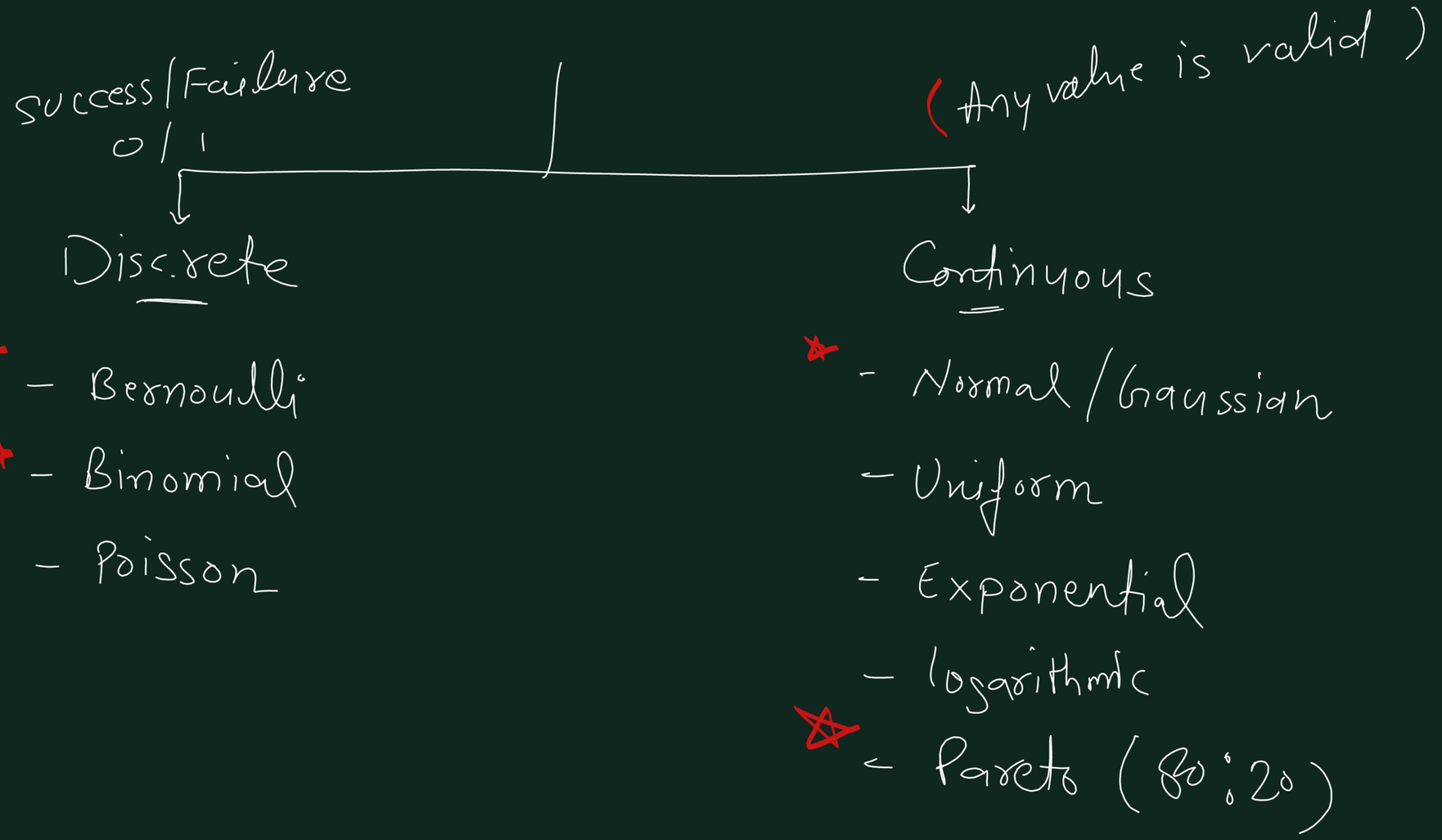
Probability Distribution



0/1

① Bernoulli

+ Binary Output
+ Single trial

$P(H) = 0.6 \rightarrow P(x=0)$

$P(\text{Not } H) = 1 - 0.6 \rightarrow P(x=1)$

$= 0.4$

② Binomial (multiple bernoulli)

+ Binary output
+ Multiple trial

n = no. of trial

p = Given probability

$X = ?$

$\Rightarrow$ P(H) in 10 trials, $P(H) = 0.5$

③ Normal/Gaussian : Mean ≈ Median ≈ Mode (Bell Curve)

(Symmetric Curve : ∩)

Continuous data can be understood : Mean (μ), Std (σ)

60% : $\mu \pm 1\sigma$: Boundary (LB, UB)

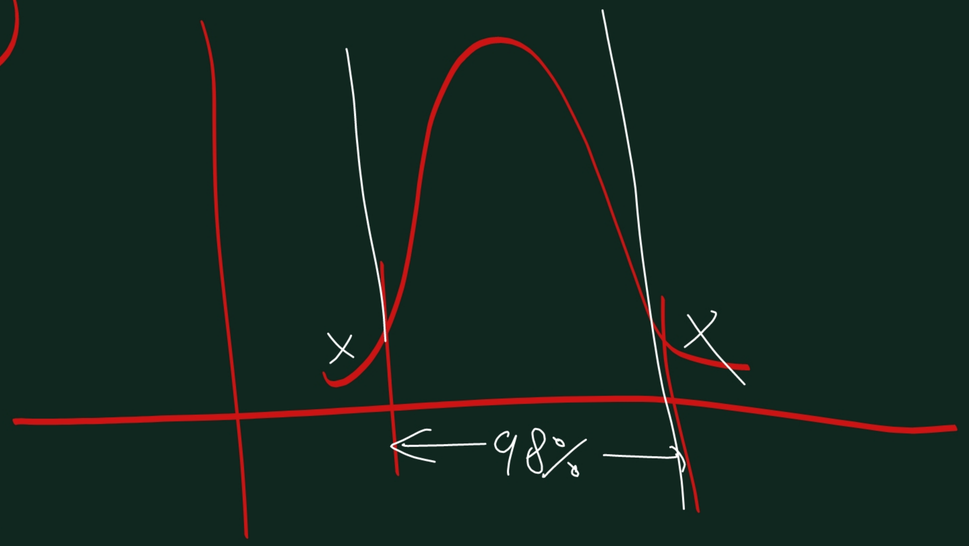
95% : $\mu \pm 2\sigma$

98% : $\mu \pm 3\sigma$

$\mu = 23, \sigma = 4.2$

LB = 23 − 2 * 4.2

UB = 23 + 2 * 4.2



④ Pareto (80 : 20)

↳ 20% data can explain 80% of distribution.

Eg: if you 10 members in a family
then only 2 members are decision maker.

(Probability of getting less 3 Heads)

Probability Distribution Function (scipy.stats)

① pmf() → Probability Mass Function → Discrete (=)

② pdf() → " Density " → Continuous (=)

③ cdf() → Cumulative Density Function → Discrete/Continuous

from scipy.stats import Bernoulli, Binom, Norm

Bernoulli.pmf( )

Binom.pmf( )

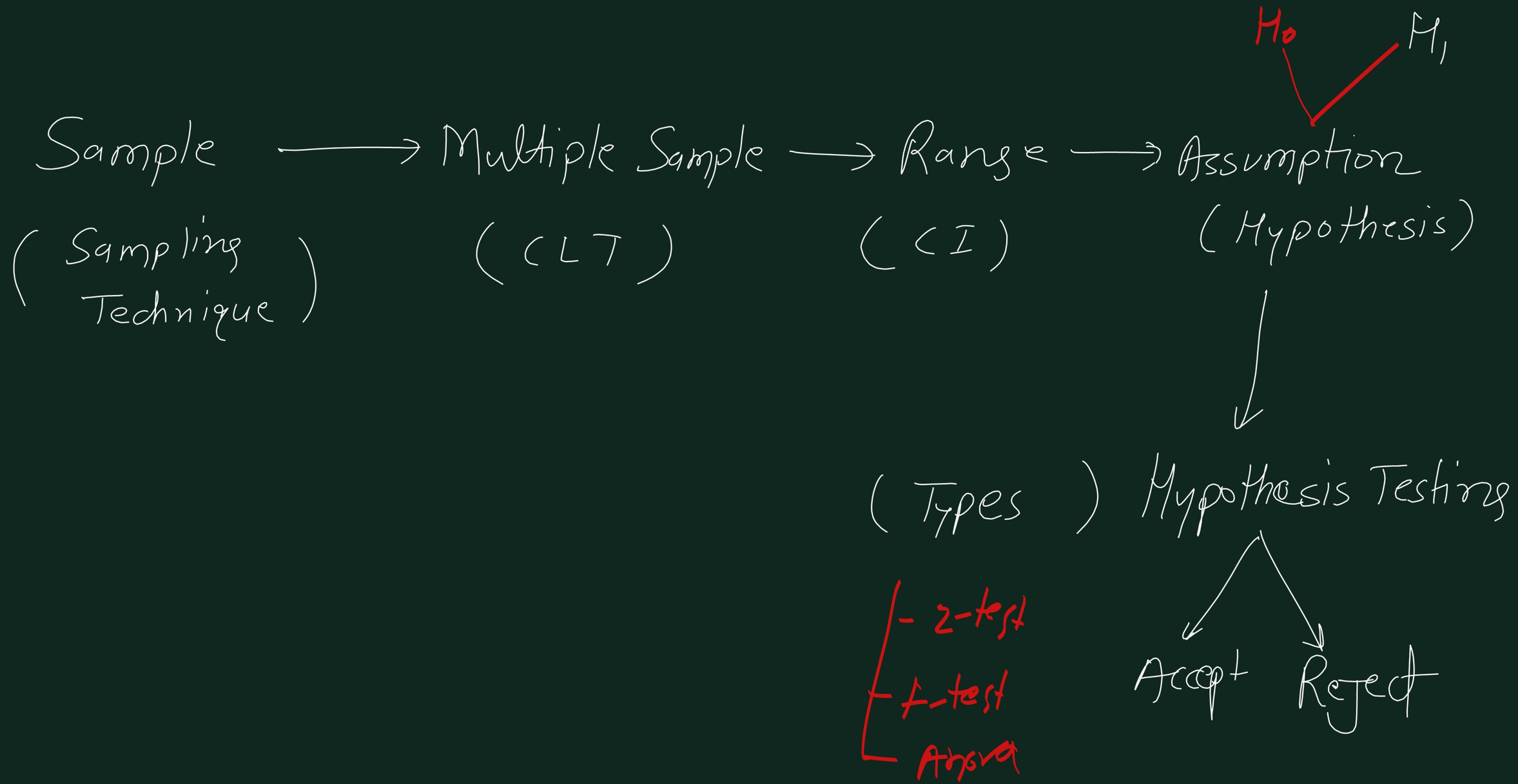
Norm.pdf( )

Inferential Stats

- Population (100% data)
- Sample (30% of data)

→ Help to conclude/infer about entire population/Data based on a sample/data.

Concluding
Predicting
Inferring


H_0
H_1
Sample
Multiple Sample
Range
Assumption
(Sampling Technique)
(CLT)
(CI)
(Hypothesis)
(Types)
Hypothesis Testing
- z-test
- t-test
- Anova
Accept
Reject