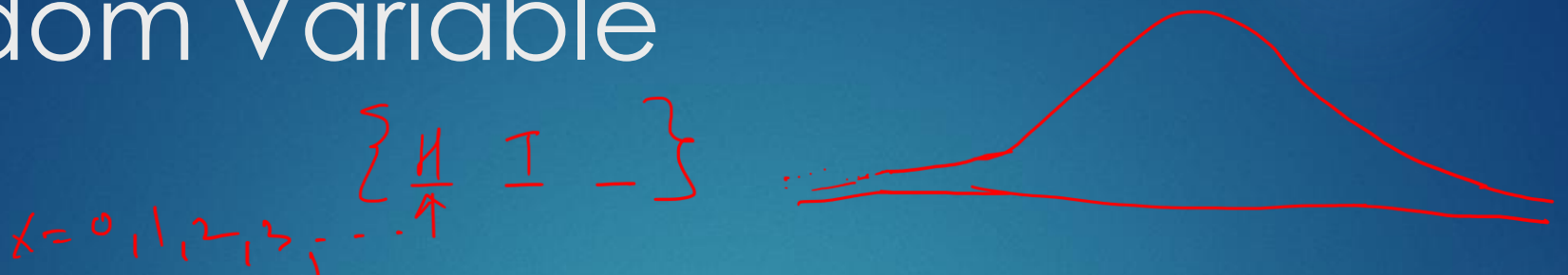
Module 1-S3: Fundamentals of Statistical Learning

POSTGRADUATE CERTIFICATE PROGRAM IN DIGITAL TRANSFORMATION STRATEGY & LEADERSHIP

NEENA PANDEY, IIM VISAKHAPATNAM

Random Variable



- A numerical measure of the outcome of an experiment
- May be discrete or continuous
- Examples?
  - No. of days that a supplier takes to supply raw material
  - No. of defects in a batch
  - Time between two customer arrivals
- In supply chain, most models are created with an assumption of continuous demand

Probability Distributions - Discrete

- Discrete Probability Function
  - Probability Distributions - Discrete
- Discrete Uniform Probability Distribution
- Discrete Uniform Probability Function

$$f(x) = 1/n$$

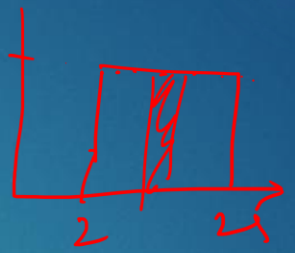
where: n = the number of values the random variable may assume

The values of the random variable are equally likely.

Probability Distributions - Discrete

- Binomial Probability Distribution
  - 'n' identical independent trials, two possible outcomes in each trial
  - The probability of both events does not change from trial to trial
  - Trials are independent
  - Example:
    - Showing an advertisement to 100 families, each either purchases a policy or not; Families are selected randomly (independent trials); Probability that at least 20 families buy the policy
    - Past experience – probability that any customer will buy from the store is 0.30. What is the probability that two out of the next three customers will buy

Probability Distributions - Continuous



- Continuous Uniform Probability Distribution
  - Probability is proportional to the interval's length

$$f(x) = 1/(b-a) \quad \text{for } a \le x \le b$$
$$= 0 \quad \text{elsewhere}$$

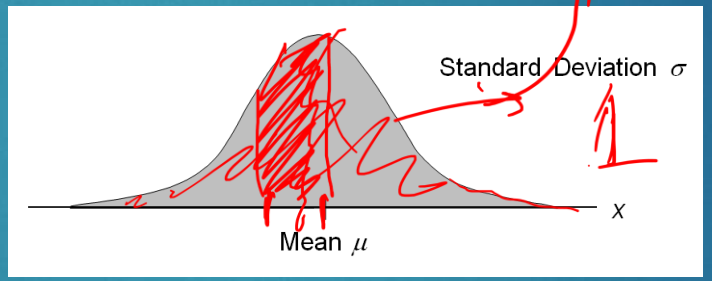
  - a, b: smallest, largest value the variable can assume
  - E.g., Flight time of an airplane traveling from Delhi to Mumbai
  - E.g., Battery life of iPad Mini is uniformly distributed between 8.5 and 12 hours – what's the probability that the battery life will be more than 11 hours?
  - Area as a measure of probability
    - Probability that x takes value between x1 and x2 = area under the curve f(x)

Probability Distributions – Continuous - Normal

- Normal Probability Density Function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

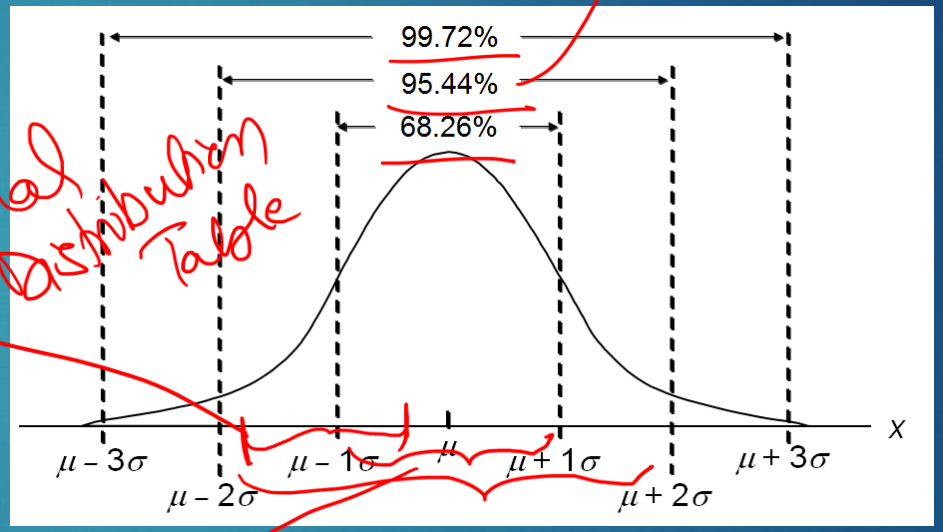
- Normal Distribution Curve:



- Total area under the curve = 1

Normal Probability Distribution

- **Empirical Rule**
- 68.26% of values of a normal random variable are within ±1 standard deviation of its mean.
- 95.44% of values of a normal random variable are within ±2 standard deviations of its mean.
- 99.72% of values of a normal random variable are within ±3 standard deviations of its mean.



Standard Normal Distribution