Distributions

Continuous Data Distribution

- **Uniform Distribution**
  - Uniform distribution modes the random variable with equal probability for each & every value
  - Mathematical Representation
    - x ~ U (a , b)
      - x = Random Variable
      - U = Uniform Distribution
      - a = minimum value
      - b = maximum value
  - PDF (Probability Density Function)

    Uniform Distribution

    $$f(x) = \frac{1}{b-a} \text{ for } a \le x \le b$$

    Mean $\mu = \frac{a+b}{2}$

    Variance $\sigma^2 = \frac{(b-a)^2}{12}$
  - The probability a single continuous value is ZERO
- **Normal / Gaussian Distribution**
  - Normal distribution refers to distribution of continuous variable is symmetric with respect to mean
  - Mathematical Representation
    - x ~ N (μ , σ)
      - x = Random Variable
      - N = Normal Distribution
      - μ = Mean
      - σ = Standard Deviation
  - PDF (Probability Density Function)

    Normal Distribution Formula

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

    $\mu =$ mean of x

    $\sigma =$ standard deviation of x

    $\pi \approx 3.14159 \ldots$

    $e \approx 2.71828 \ldots$
- **Exponential Distribution**
  - Exponential distribution describes the time between two events
  - Mathematical Representation
    - x ~ E (λ)
      - x = Random Variable
      - E = Exponential Distribution
      - λ = Average rate of time
  - PDF (Probability Density Function)

    Exponential Distribution Formula

    $$f(x) = \begin{cases} \lambda e^{-\lambda * x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$
- **Log Normal Distribution**
  - The continuous probability distribution of a random variable whose logarithm is normally distributed is called a lognormal distribution
  - PDF (Probability Density Function)

    $$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)}$$
- **Pareto Distribution**
  - 80 : 20 Rule / Distribution
  - 80 % outcomes because of 20 % causes
  - Mathematical Representation
    - x ~ p (α)
      - x = Random variable
      - p = pareto distribution
      - α = shape parameter
  - PDF (Probability Density Function)

    $$pdf = f(x) = a\frac{x_m^a}{x^{a+1}}$$

Discrete Data Distribution

- **Bernoulli Distribution**
  - only one Trail
  - Two Possible Outcomes , success (k = 1) or failure (k = 0)
  - Mathematical Representation
    - x ~ B (1,P)
      - x - Random Variable
      - p - probability of success
      - (1 - p) - Probability of Failure
      - B - Bernoulli Distribution
      - 1 - no.of Trails
  - PMF (probability mass function)

    PMF of Bernoulli Distribution

    $$P(X = x) = \begin{cases} p \text{ for } x = 1 \\ 1-p \text{ for } x = 0 \end{cases}$$
- **Binomial Distribution**
  - For Number of Bernoulli trails (n)
  - Two Possible Outcomes , success (k = 1) or failure (k = 0)
  - Mathematical Representation
    - x ~ B (n , P)
      - x - Random Variable
      - p - probability of success
      - B = Binomial Distribution
      - n = no . of Trails
  - PMF (probability mass function)

    Binomial Distribution Formula

    $$b(x;n,P) = {}^nC_x . P^x . (1-P)^{n-x}$$

    $$P(x) = \frac{n!}{(n-x)!\,x!} . (p)^x . (q)^{n-x}$$
- **Poisson Distribution**
  - Poisson Distribution deals with average rate of time or space
  - Mathematical Representation
    - x ~ P (λ)
      - x = Random Variable
      - P = Poisson Distribution
      - λ = Average rate
  - PMF (probability mass function)

    $$P(X = x) = \frac{e^{-\lambda} . \lambda^x}{x!} \quad , x = 0,1,2,3 \ldots.$$

    $$, \lambda = 1,2,3, \ldots..$$
- **Multinomial Distribution**
  - N - Independent Trails
  - There are more then two possible outcomes (n > 2)
  - Mathematical Representation
    - x ~ M (n , pi)
      - x = Random Variable
      - m = multinomial Distribution
      - n - no . of Trails
      - pi = p1,p2,p3,....
  - PMF (probability mass function)

    $$P = \frac{n!}{(n_1!)(n_2!)\ldots(n_x!)} P_1^{n1} P_2^{n2} \ldots P_x^{nx}$$
- **Geometrical Distribution**
  - The number of trails is until you get success
  - The Possible outcomes are two Success or Failure
  - Mathematical Representation
    - x ~ G (n , p)
      - x = Random Varibale
      - G = Geometrical Distribution
      - n = no . of trails , until you get success
      - p = probability of success
  - PMF (probability mass function)

    Geometric Probability Distribution

    $P(X = n) = p(1-p)^{n-1}$

    $P(X > n) = (1-p)^n$

    Mean $\mu = \frac{1}{p}$

    Variance $\sigma^2 = \frac{1-p}{p^2}$

    p – probability of success

    n- number of first successful trial
  - CDF (Cumulative Distribution Function)
    - P(X ⩽ x) = 1 - (1 - p)**x
  - If the probability of success increases number of trails decreases