Probability
&
Frequency Distribution

Prepared by
G.J.Rahul

Types of Probability Frequency Distribution

Discrete Probability Distribution: It represents the probability of outcomes for discrete random variables (i.e., those that have a countable number of possible outcomes).
– Examples include the binomial, Poisson, and geometric distributions.

Continuous Probability Distribution: It represents the probability of outcomes for continuous random variables (i.e., those that have an infinite number of possible outcomes).

Properties

- The probability of each outcome is between 0 and 1, inclusive.
- The sum of probabilities for all outcomes equals 1.

Graphical Representations

- **Histograms**: Especially useful for continuous data, histograms provide bars that show frequency of data in certain intervals.
- **Bar Graphs**: Ideal for discrete data, each bar represents a distinct value and its height represents its frequency.
- **Pie Charts**: Give a visual representation of each outcome's proportion to the whole.
- **Probability Mass Function (PMF)**: For discrete data, PMF gives the probability of each specific outcome.
- **Probability Density Function (PDF)**: For continuous data, PDF represents the likelihood of a value falling within a particular range.

- **Cumulative Distribution Function (CDF)**: Represents the probability that a random variable takes on a value less than or equal to x.

Uses of Frequency Distributions

- **Data Understanding:** At a glance, these distributions provide a clear picture of the data's distribution. For instance, is it skewed? Are there outliers?
- **Data Preprocessing:** Before applying machine learning algorithms, it's crucial to understand the dataset's distribution. This can help in normalizing the data, handling outliers, or even selecting the most appropriate algorithm.
- **Hypothesis Testing:** If you know the distribution of your data, you can determine which statistical tests are applicable. Some tests assume normal distribution, while others might be non-parametric.
- **Communication:** Visual distributions like histograms and bar charts derived from frequency distributions are fantastic for presentations and reports. They communicate complex data trends in a simple, digestible manner