Unit I

Univariate Statistics :

 Common Univariate Probability Distributions: Univariate statistics involve analyzing a single variable at a time. In probability theory and statistics, several common univariate probability distributions describe the likelihood of different values occurring for a particular random variable:

- Normal Distribution: The most well-known probability distribution, often referred to as the "bell curve." It is characterized by its symmetric, bell-shaped curve and is widely used in statistical modeling due to its prevalence in many natural phenomena.

- Binomial Distribution: Describes the probability of getting a certain number of successes in a fixed number of independent Bernoulli trials (experiments with two possible outcomes, such as success or failure).

- Poisson Distribution: Models the number of events that occur within a fixed interval of time or space, assuming that events occur independently and at a constant rate.

- Exponential Distribution: Describes the time between events in a Poisson process (a process where events occur randomly and independently at a constant average rate).

- Uniform Distribution: Assumes that all values in a given range have an equal probability of occurring.

These probability distributions are fundamental in statistical modeling, hypothesis testing, and making predictions based on data.

In data analytics, descriptive statistics play a crucial role in summarizing and understanding the characteristics of data. Univariate frequencies and data visualization techniques help analysts gain insights into individual variables, while univariate statistics and probability distributions assist in making probabilistic statements and modeling data for further analysis.
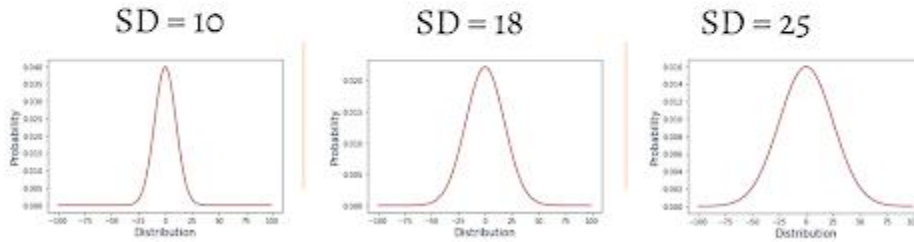
The 4 Most Common Distributions

There are many different probability distributions out there; some of them are more common than others. Let's look at the four most commonly used distributions in data science.

1. NORMAL DISTRIBUTION

NORMAL DISTRIBUTION

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

SD = 10   SD = 18   SD = 25

Gaussian distribution (normal distribution) is famous for its bell-like shape, and it's one of the most commonly used distributions in data science. Many real-life phenomena follow normal distribution, such as peoples' height, the size of things produced by machines, errors in measurements, blood pressure and grades on a test.
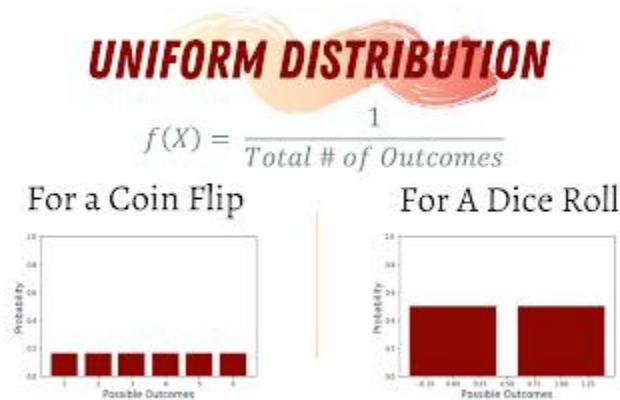
2.BINOMIAL DISTRIBUTION

BINOMIAL DISTRIBUTION

$$b(x, n, P) = {}_nC_x * P^x * (1 - P)^{n-x}$$

p = 0.2   p = 0.5   p = 0.9
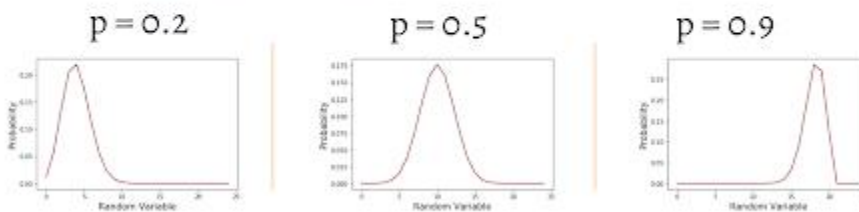
A binomial experiment is a statistical experiment, where a binomial random variable is the number of successes (x) in repeated trials of a binomial experiment (n). The probability distribution of a binomial random variable is called a binomial distribution.

3. UNIFORM DISTRIBUTION

A uniform distribution, also called a rectangular distribution, is a probability distribution that has a constant probability, such as flipping a coin or rolling dice. This distribution has two types. The most common type in elementary statistics is the continuous uniform distribution (which forms the shape of a rectangle). The second type is the discrete uniform distribution. Here's how to generate a uniform distribution in Python.
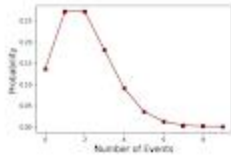
4. POISSON DISTRIBUTION

Poisson distribution can be found in many phenomena, such as congenital disabilities and genetic mutations, car accidents, meteor showers, traffic flow and the number of typing errors on a page.
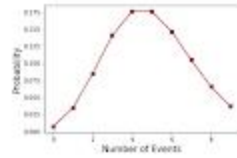
POISSON DISTRIBUTION

$$p(X) = \frac{e^{-\mu}\mu^x}{x!}$$

μ = 2

μ = 5

μ = 10