Data Distributions

A probability distribution describes how the values of a random variable are spread or distributed.

It tells us the likelihood of different outcomes and is a fundamental concept in statistics and data science.

Distributions can be:

- Discrete: Possible outcomes are countable (e.g., coin toss, dice rolls).
- Continuous: Possible outcomes form a continuous range (e.g., height, temperature).

1. Normal Distribution

A continuous, symmetric distribution shaped like a bell curve.

In a normal distribution:

- Mean = Median = Mode
- Most of the data is clustered around the mean.

- **Properties**

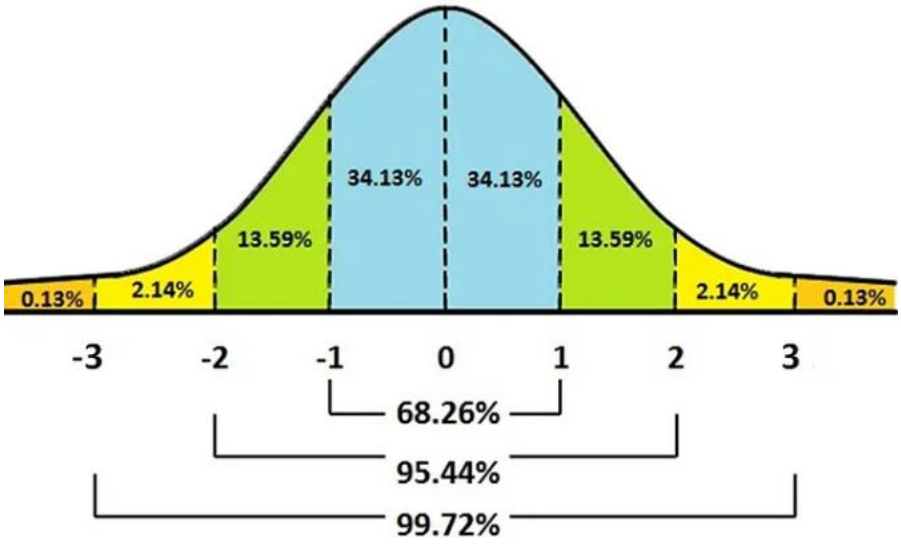
Symmetrical around the mean.

The 68-95-99.7 Rule:

- ~68% of values within 1 standard deviation (σ).
- ~95% within 2σ.
- ~99.7% within 3σ.

- **Example**
  - Human height distribution.
  - IQ scores.

- **Visual Example**



2. Uniform Distribution

A distribution where all outcomes are equally likely.

Can be discrete or continuous.

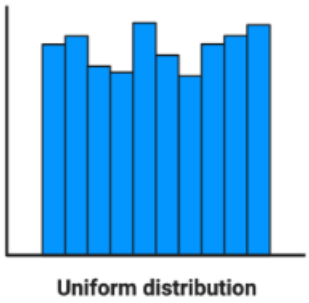
- **Properties**

Flat shape (no peak).

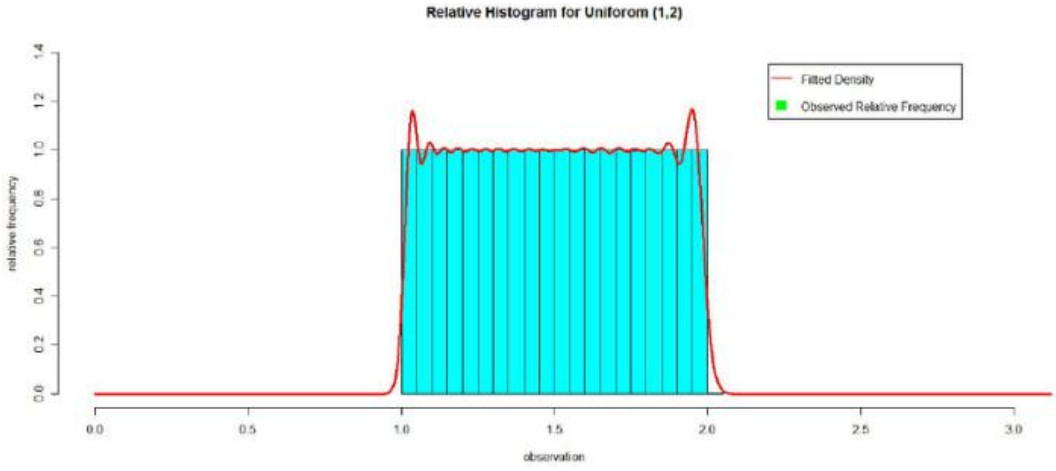
Probability of each outcome = 1 / (number of outcomes).

- **Example**
  - Rolling a fair die (discrete uniform).
  - Random number generator between 0 and 1 (continuous uniform).

- **Visual Example**



Uniform distribution



3. Binomial Distribution

A discrete distribution that gives the probability of getting exactly k successes in n independent trials, where each trial has two possible outcomes: success or failure.

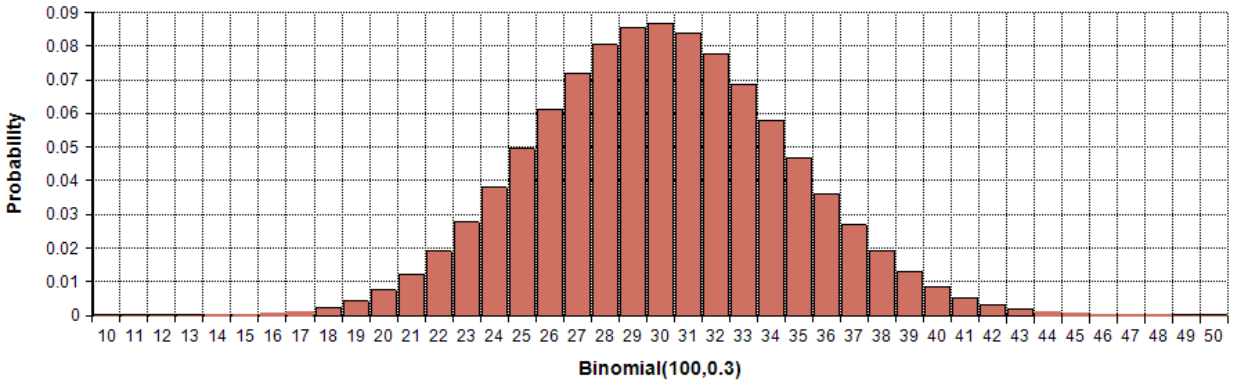
- **Formula**

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

where
n = the number of trials (or the number being sampled)
x = the number of successes desired
p = probability of getting a success in one trial
$q = 1 - p$ = the probability of getting a failure in one trial

- **Example**
  - Number of heads in 10-coin tosses (p=0.5).
  - Number of defective items in a batch.

- **Visual Example**



4. Poisson Distribution

A discrete distribution that describes the probability of a given number of events occurring in a fixed interval of time or space.

- **Formula**

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where

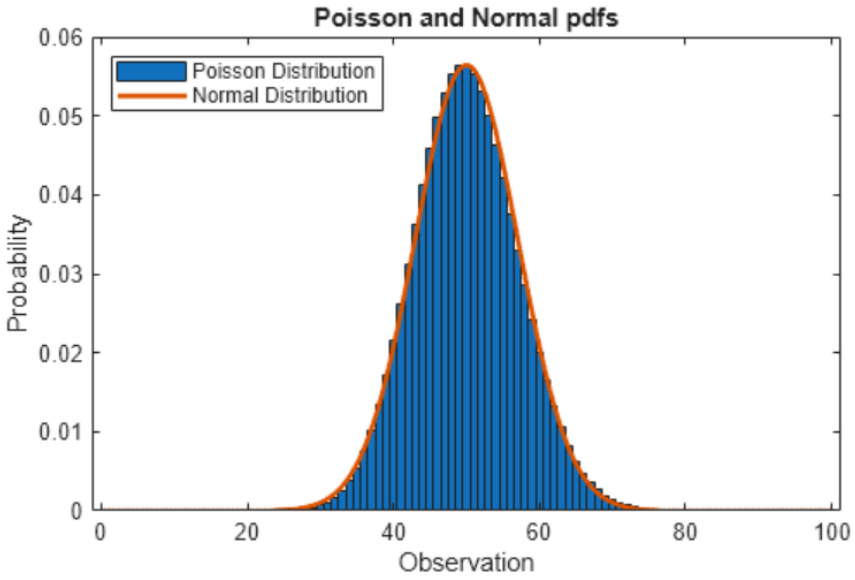
x = 0, 1, 2, 3, …

λ = mean number of occurrences in the interval

e = Euler's constant ≈ 2.71828

- **Example**
  - Number of customer arrivals at a store per hour.
  - Number of emails received per day.

- **Visual Example**



5. Comparison

Distribution	Type	Shape	Example
Normal	Continuous	Bell curve	Human height
Uniform	Both	Flat	Rolling a die
Binomial	Discrete	Varies with n, p	Coin tosses
Poisson	Discrete	Skewed, depends on λ	Customer arrivals

6. Common Mistakes

- Confusing discrete vs continuous distributions.
- Assuming all data is normally distributed.
- Using binomial when trials are not independent.
- Misinterpreting λ in Poisson (it's the average rate, not fixed).