Statistical Distributions

Giacomo Babudri

December 2023

1 Introduction

Statistical distributions provide a method of simulating the variations that occur in nature, by modeling uncertainty and variability in data. There are two main categories of distribution: ***discrete*** and ***continuous***. When sampled a continuous distribution returns a non-integer value, whereas a discrete distribution will return an integer.

If you flip a coin or roll a dice there are a finite set of possible outcomes. These finite outcomes define a discrete distribution. Other examples could be the result of a test where the result is pass or fail, or the nature of a part when categorized by Part Number or Type.

A continuous distribution has an uncountable number of possible outcomes. The duration of a journey will follow a continuous distribution - each incidence of a journey will take a marginally different time - or the time taken to perform a manual operation will be continuous. In each of these cases the times may be very similar, but when measured to several decimal places then differences will be seen.

2 Discrete

The most common discrete probability distributions include binomial and Poisson.

2.1 Binomial Distribution

A binomial probability distribution is one in which there is only a probability of two outcomes. In this distribution, data are collected in one of two forms after repetitive trials and classified into either success or failure. It generally has a finite set of just two possible outcomes, such as zero or one.

Example: Flipping a coin gives you the list Heads, Tails.

Functions:

- Probability Mass Function (PMF)

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1}$$

- Cumulative Distribution Function (CDF)

$$F(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \tag{2}$$

Application:

- *Genetics*: It can represent the probability of obtaining a certain number of offspring with a specific trait.

- *Fincance*: It can model the probability of a certain number of successful trades in a series of independent transactions.

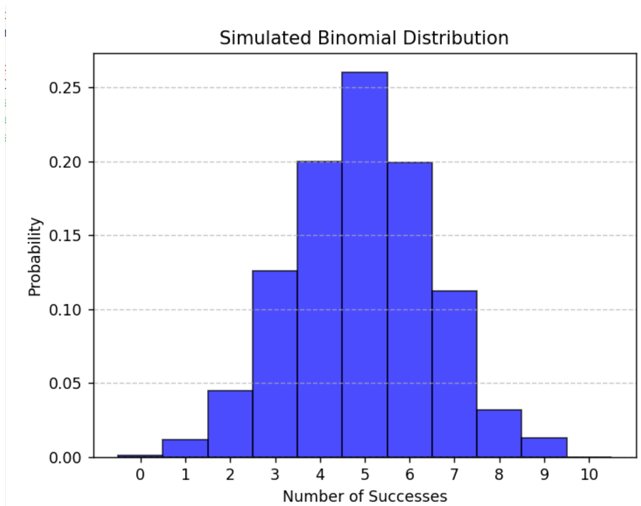

Figure 1: Simulation of a Binomial Distribution

2.2 Poisson Distribution

A Poisson distribution is a discrete probability distribution, meaning that it gives the probability of a discrete (i.e., countable) outcome. For Poisson distributions, the discrete outcome is the number of times an event occurs, represented by k.

You can use a Poisson distribution to predict or explain the number of events occurring within a given interval of time or space. “Events” could be anything from disease cases to customer purchases to meteor strikes. The interval can be any specific amount of time or space, such as 10 days or 5 square inches.

The two features that characterized a poisson distribution are:
Individual events happen at random and independently. That is, the probability of one event doesn’t affect the probability of another event.
You know the mean number of events occurring within a given interval of time or space. This number is called (lambda), and it is assumed to be constant.

Functions

- Probability Mass Function (PMF):

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{3}$$

- Cumulative Distribution Function (CDF):

$$F(X \leq k) = e^{-\lambda}\sum_{i=0}^{k}\frac{\lambda^i}{i!} \tag{4}$$

Application:

- *Traffic Flow:* The Poisson Distribution can be applied to predict the probability of observing a specific number of cars during that time frame. This information is valuable for optimizing traffic signal timings or planning road maintenance.
- *Medical Events:* The Poisson Distribution can be used to model the probability of observing a certain number of rare events within a defined time period. This information aids in resource allocation, risk assessment, and patient safety protocols.
- *Network Security:* The Poisson Distribution can help model the frequency of these incidents over a specific time period. This information is crucial for developing effective cybersecurity strategies, allocating resources for monitoring, and identifying unusual patterns that may indicate a security breach.

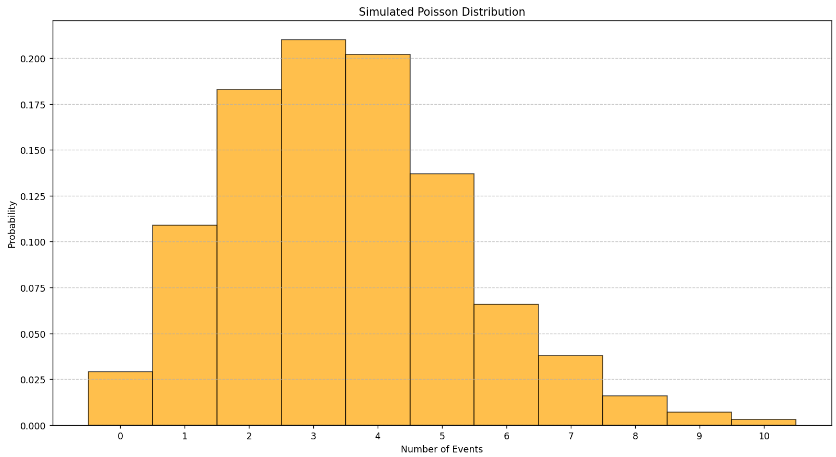

Figure 2: Simulation of poisson Distribution

3 Continuous

One of the main continuous distribution is the Normal Distribution

3.1 Normal Distribution

The Normal Distribution, also known as the Gaussian distribution, is a continuous probability distribution that is symmetrical and bell-shaped. It is widely used due to the Central Limit Theorem and the prevalence of naturally occurring phenomena exhibiting a normal distribution. The distribution is characterized by two parameters: the mean () and the standard deviation ().

Functions:

- Probability Density Function (PDF):

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

- Cumulative Distribution Function (CDF):

$$F(x|\mu,\sigma) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{6}$$

Application:

- *Financial Risk Management:* Portfolio managers and risk analysts use the Normal Distribution to model the distribution of returns. This aids in risk assessment, setting investment strategies, and estimating potential losses.

- *Physical Sciences:* Scientists use the Normal Distribution to analyze experimental results, assess measurement errors, and make statistical inferences. It provides a framework for understanding the variability in scientific observations.

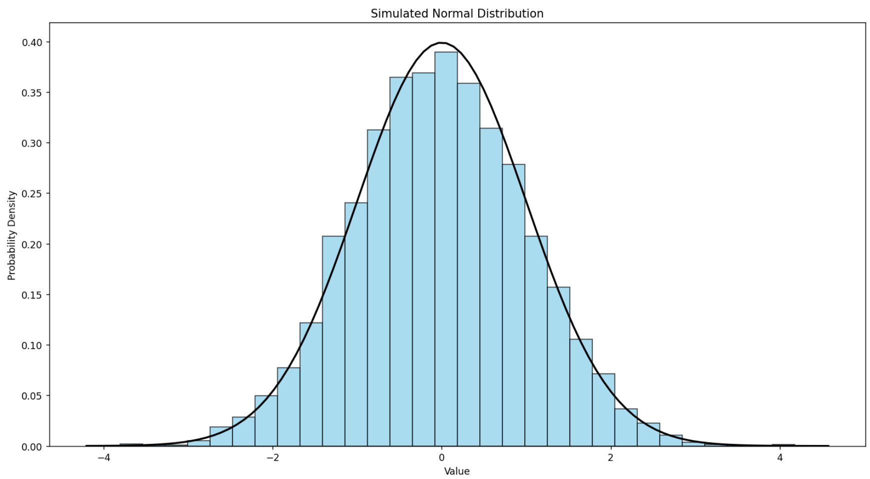

Figure 3: Simulation of Normal Distribution