Statistical Implementations in Python

Part 2: Probability Distributions

Created by : Felice Benita

Probability Distributions

A probability distribution is a mathematical function that describes how the values of a random variable are distributed or, in other words, the likelihood of each possible outcome. It provides a framework for understanding the possible values a variable can take and how frequently they are expected to occur.

Created by Felice Benita

Types of Probability Distributions

Probability distributions can be broadly categorized into discrete and continuous:

- Discrete Probability Distributions: These are used for variables that can take on specific, separate values (often integers). Examples include the binomial and Poisson distributions. For instance, the number of heads in a series of coin flips is a discrete variable.
- Continuous Probability Distributions: These apply to variables that can take on an infinite number of values within a range. Examples include the normal (Gaussian) and exponential distributions. For instance, the height of people in a population is a continuous variable.

Created by Felice Benita

EXAMPLE

Using dataset of **E-commerce Product Sales** with data columns: Product_ID, Product_Category, Units_Sold, Revenue, Discount_Percentage, Return_Rate, Customer_Rating, Days_in_Inventory, Sales_Channel (online, in-store).

→ Find the probability distributions of this data set.

[2]:
```
# Reading data
df = pd.read_csv('File Dirr/ecommerce_product_sales.csv')
df.head()
```

[2]:

	Product_ID	Product_Category	Units_Sold	Revenue	Discount_Percentage	Return_Rate	Customer_Rating	Days_in_Inventory	Sales_Channel
0	P0001	Beauty	39	207.04	50	0.12	2.7	91	In-Store
1	P0002	Books	44	12.62	20	0.04	2.9	141	In-Store
2	P0003	Home & Kitchen	43	293.55	25	0.20	4.7	289	Online
3	P0004	Books	38	130.29	0	0.22	3.6	218	Online
4	P0005	Books	40	227.88	15	0.17	2.2	77	Online

Here's the full code: [Click Here]

Created by Felice Benita

Normal Distribution

The normal distribution (or Gaussian distribution) is a continuous probability distribution that is symmetric and bell-shaped. It is one of the most widely used probability distributions in statistics due to its natural appearance in various fields, such as biology, economics, and more.

68-95-99.7 Rule

- Approximately 68% of data falls within one standard deviation of the mean.
- About 95% falls within two standard deviations.
- Around 99.7% lies within three standard deviations.

Symmetry

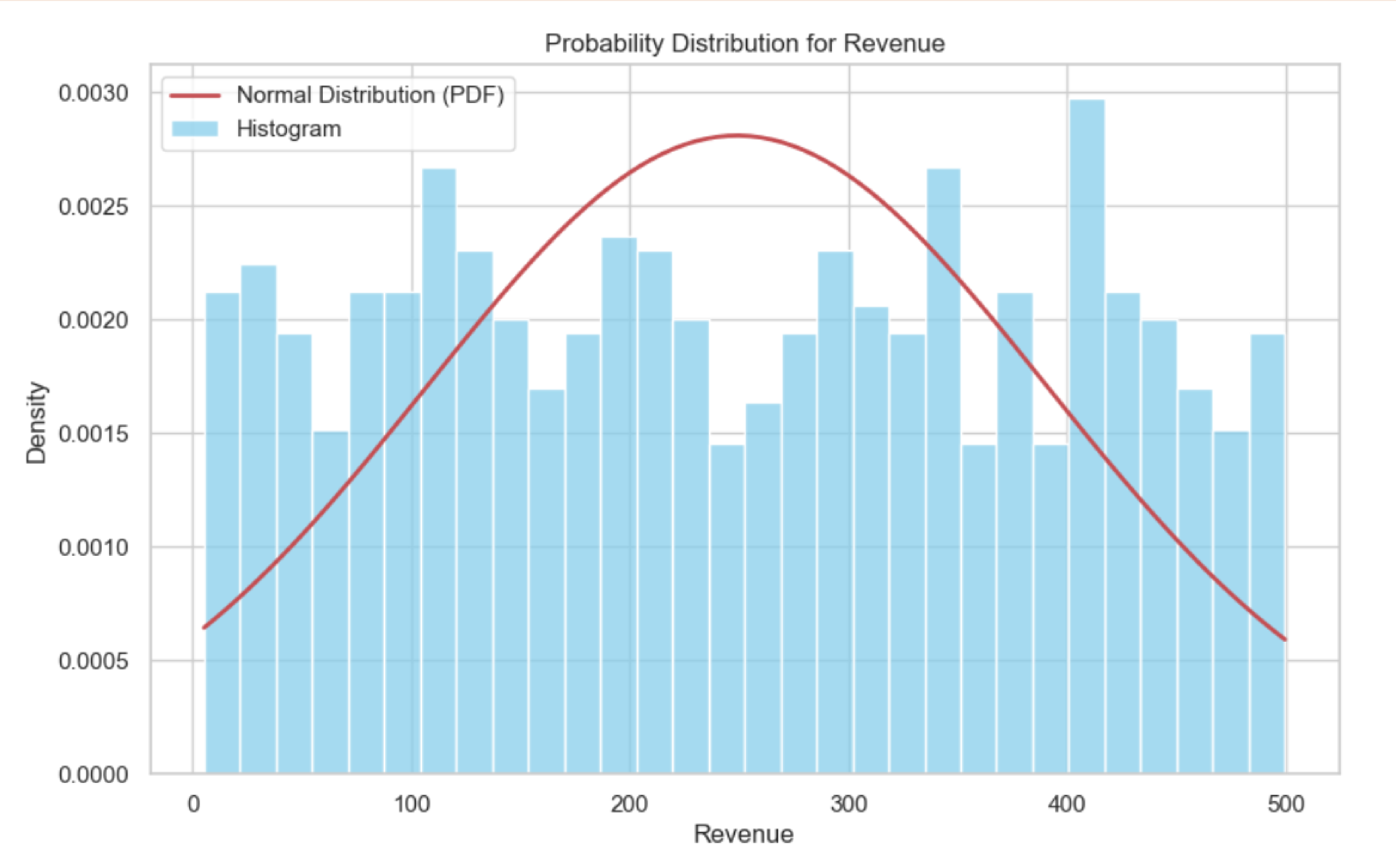
The left and right sides of the curve are mirror images of each other, making the distribution symmetrical around the mean.

Bell-Shaped Curve

The curve peaks at the mean and tails off symmetrically toward the extremes.

Created by Felice Benita

Normal Distribution



Created by Felice Benita

Binomial Distribution

The Binomial Distribution is a discrete probability distribution that describes the number of successes in a fixed number of independent trials, each with the same probability of success. It's commonly used in scenarios where there are only two possible outcomes for each trial, such as "success" and "failure."

Key Features of Binomial Distribution:

Fixed Number of Trials (n): The number of trials must be predetermined.

Two Possible Outcomes: Each trial results in one of two outcomes, often termed as "success" (often denoted by kk) and "failure."

Constant Probability of Success (p): The probability of success remains constant for each trial.

Independence of Trials: The outcome of one trial does not affect the outcome of another.

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

- $P(X = k)$ is the probability of getting k successes.
- $\binom{n}{k}$ (read as "n choose k") is the binomial coefficient, calculated as $\frac{n!}{k!(n-k)!}$.
- p is the probability of success on each trial.
- $(1-p)$ is the probability of failure.
- n is the total number of trials.
- k is the number of successes.

Created by Felice Benita

Poisson Distribution

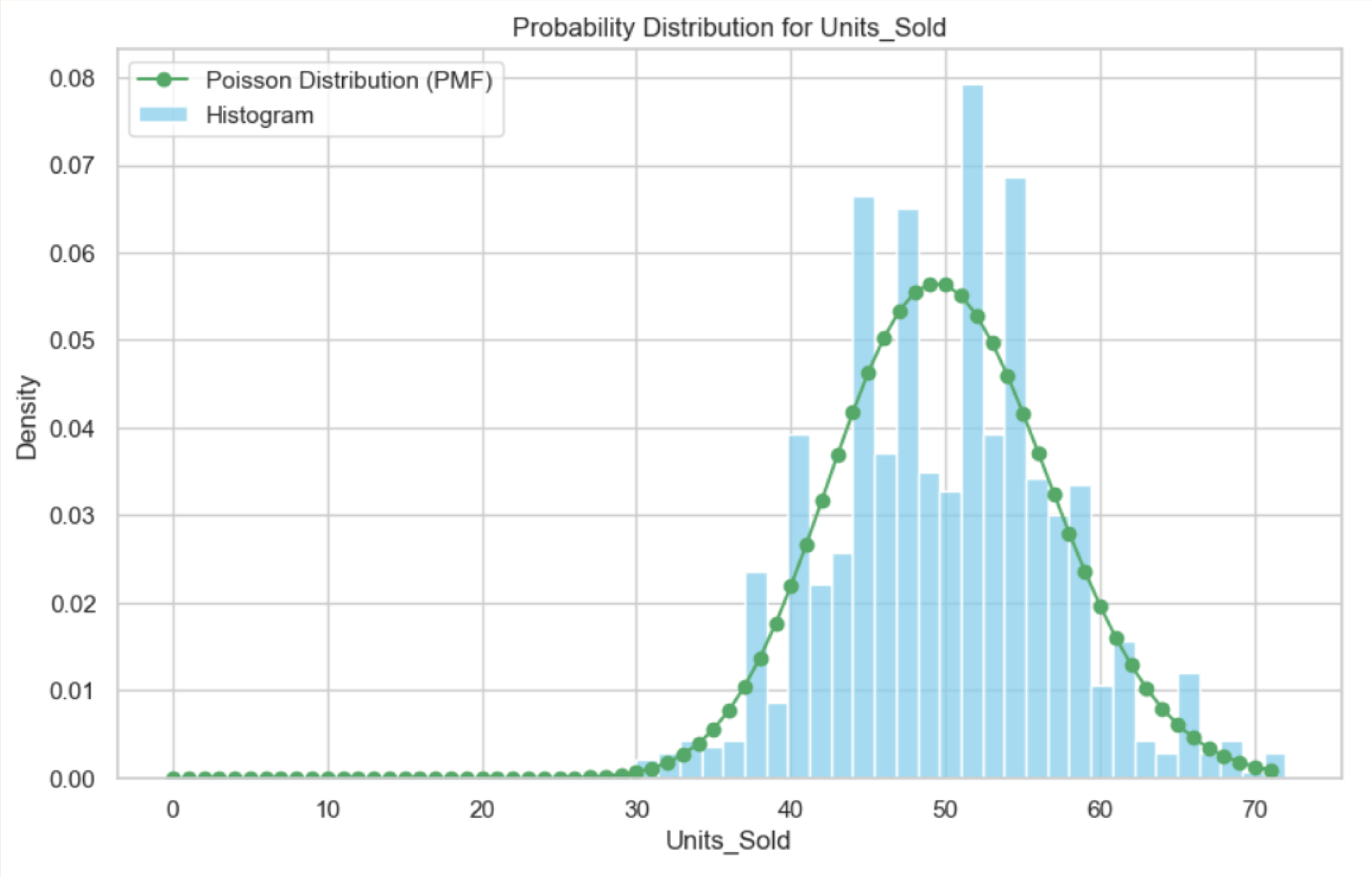
The Poisson Distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space, given that these events happen with a known constant mean rate and are independent of the time since the last event. It's often used to model the number of times an event occurs within a specified period or area when these events are rare or infrequent.

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

- $P(X = k)$ is the probability of k events occurring.
- e is Euler's number (approximately equal to 2.71828).
- λ is the average number of events in the interval.
- k is the number of events.

Created by Felice Benita

Poisson Distribution



Created by Felice Benita

Exponential Distribution

The Exponential Distribution is a continuous probability distribution that describes the time between events in a Poisson process, which is a process in which events occur continuously and independently at a constant average rate. It is often used to model the time until the next event occurs.

<u>Key Features of Exponential Distribution:</u>

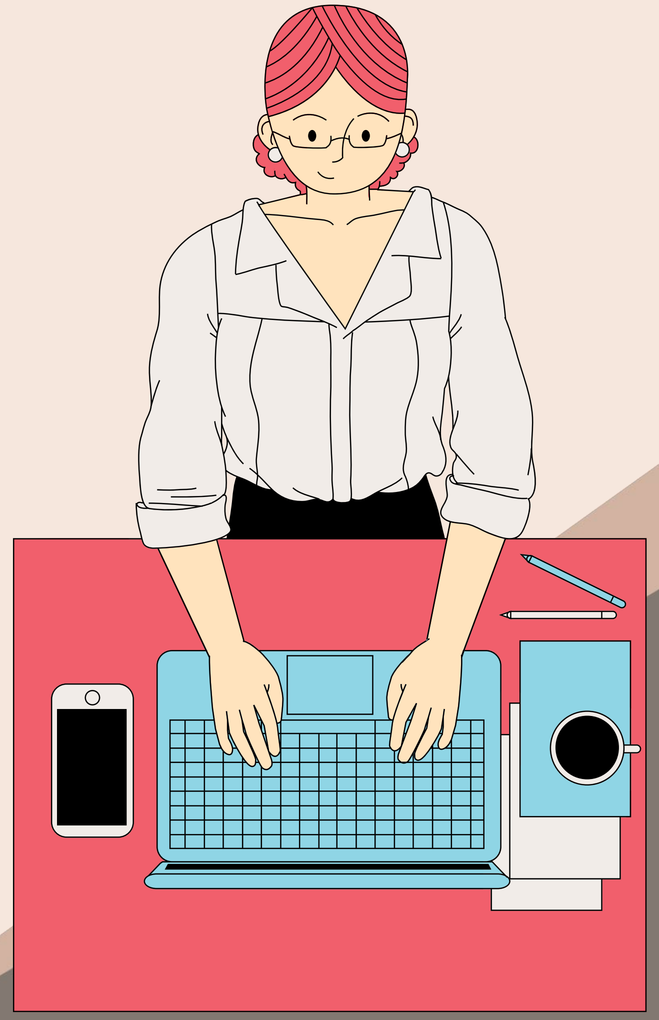
Memoryless Property: The exponential distribution is memoryless, meaning that the probability of an event occurring in the future is independent of how much time has already passed.

Continuous: Unlike the Poisson distribution, which is discrete, the exponential distribution deals with continuous data.

Rate Parameter (λ): The distribution is characterized by its rate parameter $\lambda\lambda$, which is the average number of events per unit time. The mean time between events is $1/\lambda$.

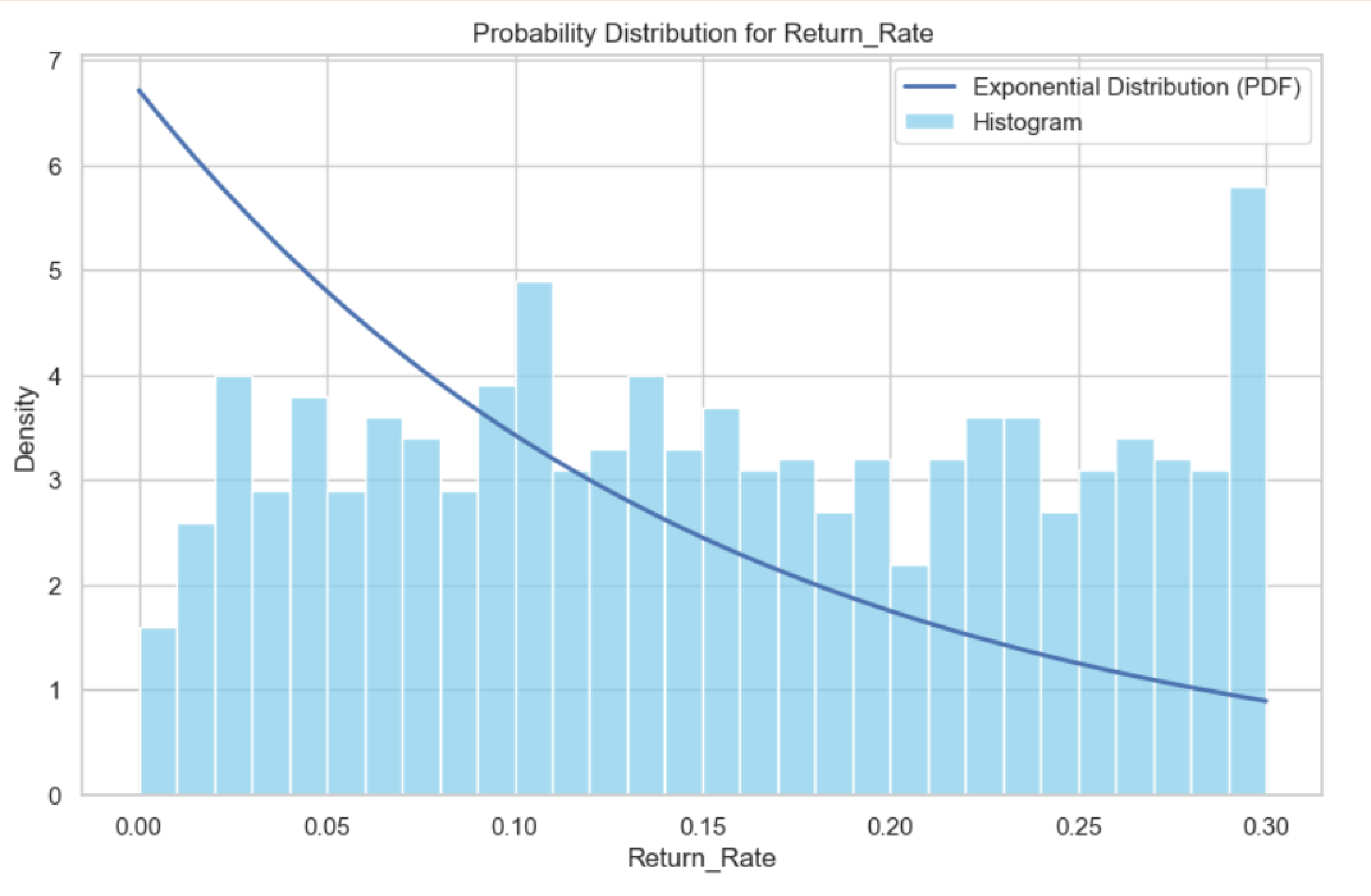
$$f(x;\lambda) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

- $f(x;\lambda)$ is the probability density function.
- λ is the rate parameter.
- x is the time until the next event occurs.
- e is Euler's number (approximately 2.71828).



Created by Felice Benita

Exponential Distribution



Created by Felice Benita

Conclusion

Probability distributions enable us to quantify uncertainty and make informed predictions about future events based on past data. Understanding the characteristics, applications, and limitations of various distributions is essential for effective data analysis and decision-making in numerous fields.

Created By : Felice Benita

THANK YOU

Created By : Felice Benita