Part A – Theoretical Foundation

Statistical Distributions

A statistical distribution describes how the values of a random variable are spread (or distributed) across possible outcomes.

- Discrete Distributions
- Continuous Distributions

In Data Science & Statistics, distributions help model real-world data and make predictions.

Q-Q Plot

A **Q-Q (Quantile-Quantile) Plot** is a graphical tool used to check if a dataset follows a theoretical distribution (usually Normal).

Many statistical models assume **normality**. Q-Q plots validate that assumption.

- Straight line → Good fit
- Curved pattern → Not normal

Discrete and Continuous Distributions

Discrete Distribution	Continuous Distribution
Countable outcomes	Infinite values in interval
Uses PMF	Uses PDF
Example: Coin toss	Example: Height, time
Probability at exact point > 0	Probability at exact point = 0
Sum of probabilities = 1	Area under curve = 1

Bernoulli Distribution

A Bernoulli distribution is the simplest probability distribution.

It describes a random experiment with only two possible outcomes:

- Success (1) with probability p
- Failure (0) with probability 1-p

Binomial Distribution

The Binomial distribution is an extension of Bernoulli.

It models the probability of getting k successes in n independent Bernoulli trials.

Log-Normal Distribution

A **Log-Normal Distribution** is a distribution where a random variable **X** is log-normally distributed if its natural logarithm **ln(X)** follows a Normal distribution.

Key Characteristics:

- Positively skewed
- Mean > Median > Mode
- Only defined for X > 0
- Arises from multiplicative growth

Power Law Distribution

A Power Law distribution describes situations where:

Small events are very common
Large events are rare but extremely significant

Where:

- $\alpha > 1$ is the power-law exponent
- Larger α → faster decay
- Heavy tail distribution

Box-Cox Transform

Box-Cox is a **power transformation** used to:

- Stabilize variance
- Reduce skewness
- Make data more Normal-like

- Data must be strictly positive

Poisson Distribution

The Poisson distribution models the probability of a given number of events happening in a fixed interval of time, space, or area, if:

- Events occur independently.
- Events occur at a constant average rate (λ).
- Two or more events do not happen simultaneously.

Example:
If a call center gets 5 calls per hour (λ=5),
Probability of exactly 3 calls:

$$P(X=3)=0.1404$$

Z-score Probability

Score measures how many standard deviations a value is from the mean.

Interpretation:

- Z > 0 → Above mean
- Z < 0 → Below mean

PDF and CDF

PDF	CDF
Probability Density Function	Cumulative Distribution Function
Shows likelihood shape	Shows accumulated probability
Area under curve = 1	Increases from 0 to 1
Probability at exact point = 0	Gives cumulative probability