Assignment: Statistical Modelling of Stock Returns using MLE and Hypothesis Testing

1. Objective

The purpose of this assignment is to analyse real-world financial data by modelling the distribution of stock returns using statistical techniques. The assignment involves:

1. Estimating distribution parameters using Maximum Likelihood Estimation (MLE).
2. Hypothesis testing to assess the goodness-of-fit for different distributions.
3. Identifying the most plausible distributional model for stock returns.
4. Constructing confidence intervals and evaluating return anomalies.

2. Data Acquisition and Preprocessing

2.1 Choose any publicly listed stock from a major exchange (preferably NSE).

2.2 Download at least one year of daily closing price data for the selected stock. You may use APIs such as yfinance, nsepy, or work with a pre-downloaded CSV file.

2.3 Compute the daily log returns

3. Maximum Likelihood Estimation (MLE) of Distribution Parameters

3.1 Consider the following 10 probability distributions as candidates for modeling the stock returns:

1. Normal distribution
2. Student's t-distribution
3. Laplace distribution
4. Exponential distribution
5. Gamma distribution
6. Beta distribution
7. Cauchy distribution
8. Logistic distribution
9. Weibull distribution
10. Log-normal distribution

3.2 Choose any five of the above distributions and estimate their parameters using Maximum Likelihood Estimation based on your computed daily returns.

3.3 For each distribution, clearly specify the parameters you are estimating (e.g., location, scale, shape) and include the corresponding estimation code and output.

4. Hypothesis Testing:

4.1 For each of the five selected distributions, formulate a null hypothesis:

H_0: The daily stock returns follow this distribution.

4.2 Perform goodness-of-fit tests to assess each hypothesis. Recommended tests include Kolmogorov-Smirnov (KS) test

4.3 Record the p-values and test statistics for each distribution.

4.4 Rank the five distributions in terms of their goodness-of-fit. A higher p-value suggests better fit under the null hypothesis.

4.5 Clearly state which distribution best fits the stock returns based on your analysis.

5. Confidence Interval Analysis for Returns

5.1 Assuming the best-fitting distribution identified in Section 4 is valid, compute a 95% confidence interval for the mean of daily returns.

5.2 Use the appropriate distribution-specific formula for constructing the confidence interval.

5.3 Plot the daily return time series, and mark the dates where returns fall outside the 95% confidence interval. These points represent statistically unusual returns.

5.4 Briefly comment on possible market events or conditions that could have caused these outliers, if evident from the data or timeline.

6. Submission Guidelines

6.1 Submit a Jupyter notebook or Python script containing:

- Clean and well-commented code for all steps.
- Output from parameter estimation and hypothesis testing.
- All required plots: return histogram with fitted distributions, p-value comparisons, and the return time series with CI bounds.

6.2 Include a 1-page written summary (PDF or Latex) with answers to the following:

- Which distribution best models the returns?
- What are the estimated parameters?
- What are the test results and p-values?

- What do the confidence interval outliers indicate?