DS201/DSL253: Statistical Programming

Assignment 06

20/03/2025

Instructions for Submission: You can submit your solution as a Jupyter Notebook/Matlab file with comments and discussions on the results obtained in each step.

1. Follow Standard Report Format: Include sections like Introduction, Data, Methodology, Results, Discussion, and Conclusion.
2. File Naming Convention: Adhere to the specified naming convention for each file you submit (e.g., RollNumber FirstName Asg1).
3. Refrain from using zip files. If necessary, submit multiple files.
4. Include comments in the code explaining the logic and any assumptions made.
5. Include References: Cite any external sources or references used in your assignment.
6. Code Quality: Ensure your code follows best practices, is well-organized, and avoid plagiarism as a plagiarism check will be conducted.
7. Be aware that late submissions are not permitted; ensure timely submission.
8. Coding can be done in any language.

Question 1

You are a data scientist working for a smartphone manufacturer. The company claims that the battery life of their latest model follows a normal distribution with a mean (μ) of 20 hours and a standard deviation (σ) of 2 hours. However, you suspect that the actual battery life might differ due to manufacturing variability. To investigate this, you decide to simulate and analyze battery life data.

(i) Write a function to simulate n_1 smartphone battery life measurements based on the claimed distribution ($\mu = 20$, $\sigma = 2$). Use Maximum Likelihood Estimation (MLE) to estimate the mean and standard deviation of the battery life from your simulated data. Plot a histogram of the simulated data and overlay the normal PDF using your estimated parameters. Experiment with different values of n_1 (e.g., 10, 100, 1000) and n_2 (e.g., 100, 1000) and observe how the estimates change.

(ii) Repeat the simulation n_2 times (e.g., 1000 trials) and plot a histogram of the estimated means and standard deviations. Mark the true values ($\mu = 20$, $\sigma = 2$) on the same plot. Experiment with different values of n_1 (e.g., 10, 100, 1000) and n_2 (e.g., 100, 1000) and observe how the estimates change. Are the estimators for μ and σ biased or unbiased? Discuss your observations and give suggestions.

Question 2

You are working on a project to measure the temperature of a chemical reaction using a sensor. The true temperature (X) follows a normal distribution with a mean (μ) of 50°C and a standard deviation (σ) of 5°C. However, the sensor introduces some random noise (η) due to calibration issues, where η is uniformly distributed between -1°C and 1°C. The measured temperature is $Y = X + \eta$.

(i) Simulate n_1 temperature measurements (Y) by adding the sensor noise to the true temperature (X). Use MLE to estimate the mean and standard deviation of the true temperature (X) from the noisy measurements (Y). Plot a histogram of the noisy measurements and overlay the normal PDF using your estimated parameters.

(ii) Repeat this experiment n_2 times and plot a histogram of the estimated means and standard deviations. Mark the true values ($\mu = 50$, $\sigma = 5$) on the same plot. Compare your results with part (a). How does the sensor noise affect your ability to estimate the true temperature? Are the estimators still unbiased? Discuss your findings.

Question 3

You are an analyst at a financial firm studying the daily returns of a high-risk stock. Unlike normal stocks, the returns of this stock follow a **t-distribution** with heavier tails, meaning extreme gains or losses are more likely. You want to estimate the mean (μ) and standard deviation (σ) of the stock returns.

(i) Simulate n_1 daily stock returns using a t-distribution with a mean (μ) of 0.1% and a standard deviation (σ) of 2%. Use MLE to estimate the mean and standard deviation of the stock returns. Plot a histogram of the simulated returns and overlay the t-distribution PDF using your estimated parameters.

(ii) Now, suppose the stock returns are affected by market noise (η), where η is uniformly distributed between -0.5% and 0.5%. Repeat the experiment with the noisy returns ($Y = X + \eta$) and compare your results with part (a). How does the t-distribution and noise affect your estimates? Are the estimators biased? Discuss your observations.