

DSL201: Statistical Programming

Dr. Anil Kumar Sao

Assignment 2

Deadline: 13/02/24 11:59 pm

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_Asg2).
 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 and 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.
-

1. a) Generate a dataset representing the number of customer arrivals per hour in a store, assuming a Poisson distribution with a mean arrival rate of 5. Use Python to create a histogram to visualize the distribution. Calculate the mean and variance of the generated dataset and explain how they relate to the parameters of the Poisson distribution.
b) Generate a bivariate dataset with two features (e.g., height and weight) assuming a multivariate Gaussian distribution with given means $\mu_1 = 65$, $\mu_2 = 150$ and covariance matrix Σ $\begin{bmatrix} 25 & 10 \\ 10 & 36 \end{bmatrix}$. Visualize the data using a scatter plot in Python. Discuss the patterns in the plot and how they correspond to the parameters of the multivariate Gaussian distribution.
c) Generate a dataset for a univariate scenario (e.g., sales per day) and a bivariate scenario (e.g., temperature and sales) using both Poisson and multivariate Gaussian distributions. Create visualizations for each dataset using appropriate plots for each distribution (e.g., histogram for Poisson, scatter plot for multivariate Gaussian). Compare and contrast the visualizations, explaining how the characteristics of each distribution are reflected in the plots.

Note: Implement all functions from scratch

2. a) Write a Python code to generate 1000 random numbers from a Poisson distribution with a mean of 3.

- b) Transform the generated Poisson numbers to their squares and create a histogram to visualize the distribution of squared values.
- c) Now, write a Python code to generate 100 pairs of random numbers from a bivariate Gaussian distribution with means $\mu_1 = 2$, $\mu_2 = 3$ and covariance matrix $\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}$.
- d) Transform each pair of Gaussian numbers to their squares (both elements of the pair) and create a 3D scatter plot to visualize the distribution of squared values.
- e) Compare the histograms from the Poisson distribution and the scatter plot from the multivariate Gaussian distribution in terms of the shape and spread of the squared values.
- f) Discuss any insights gained from the visualizations and transformations, considering the characteristics of each distribution.

Note: Implement all functions from scratch