Statistical Methods for Data Science

DATA7202

Semester 1, 2024

Assignment 4 (Weight: 25%)

Assignment 4 is due on Wednesday 05.06.24 16:00.

Please answer the questions below. For theoretical questions, you should present rigorous proofs and appropriate explanations. Your report should be visually appealing and all questions should be answered in the order of their appearance. For programming questions, you should present your analysis of data using Python, Matlab, or R, as a short report, clearly answering the objectives and justifying the modeling (and hence statistical analysis) choices you make, as well as discussing your conclusions. Do not include excessive amounts of output in your reports. All the code should be copied into the appendix and the sources should be packaged separately and submitted on the blackboard in a zipped folder with the name:

```
"student_last_name.student_first_name.student_id.zip".
```

For example, suppose that the student name is John Smith and the student ID is 123456789. Then, the zipped file name will be `John.Smith.123456789.zip`.

1. **[100 Marks (see details below)]** Regular and priority packages arrive to a post office according to a Poisson process with rate $\lambda = 4$ (4 packages per unit time on average). It was noted that a package is regular with probability 0.9. There are 9 post-office workers and each package should be processed by a post-office employee. The processing time is exponentially distributed with mean 2 (that is, 0.5 packages can be processed on average per unit time). If all workers are busy with package processing, a newly arrived regular (or priority) package enters a *regular* (or a *priority*) waiting queue. As soon as a worker finishes a package processing, she either takes a new package from the waiting queues or remains idle if there are no packages waiting. Provided that there are waiting packages in both the regular and the priority queues, the worker will fetch a package from a priority queue.

   Perform a Discrete-Event Simulation study in Matlab, Python, or R, to answer the following question.

   > Assuming that the queues are empty at time 0, we would like to find the proportion of time that all workers are busy during the day from $t = 0$ to $T = 2000$. That is, we would like to simulate the process for 2000 unit times. Run the simulation and report the proportion along with 95% confidence interval.

   Perform a Discrete-Event Simulation study to answer the following question.

(a) **[15 Marks]** Give the problem summary and describe the project objective.

(b) **[15 Marks]** Give a specification of variables used in the simulation study. In addition, show a diagram that describes the project dynamics.

(c) **[40 Marks]** Results and Analysis. Using tables and figures, present a clear outcome of your study. Present the corresponding confidence intervals.

(d) **[20 Marks]** Formulate your conclusions.

(e) **[10 Marks]** Appendix. Include all code files used. Explain their interaction and provide a clear and well-commented code.