Stochastic Simulation (MIE1613H) - Homework 3

Due: March 22, 2023

- Submit your homework on Quercus as a single PDF file by the deadline. Late submissions are penalized 10% each day the homework is late.
- At the top of your homework include your name, student number, department, and program.
- You may discuss the assignment with other students in general terms, but each student must solve the problems, write the code and the solutions individually.
- **The simulation models must be programmed in Python. For the input modelling questions you may use R or Python.** You must include both the source code (including comments to make it easy to follow) and the output of the simulation in your submission.
- **It is very important to explain your answers and solution approach clearly and not just provide the code/results**. Full marks will only be given to correct solutions that are fully and clearly explained.

Problem 1. (25 Pts.) You are asked to make projections about cycle times for a semiconductor manufacturer who plans to open a new plant. Here "cycle time" means the time from product release until completion. The process that you will consider has a single diffusion step with sub-steps as indicated in the diagram.

Raw material for two products (C) and (D) will begin at the CLEAN step, and make multiple passes until the product completes processing. Movement within each process is handled by robots and takes very little time (treat as 0) relative to the processing steps. The movement time from release to diffusion is 15 minutes (1/4 hour).

The anticipated product mix is to have 60% of products to be of type (C) and 40% of type (D). Product (C) requires 5 passes and product D requires 3 passes.

The OXDIZE step is deterministic but differs by product type: it takes 2.7 hours for (C) and 2.0 hours for (D).

(a) The CLEAN and LOAD/UNLOAD steps do not differ by product type but are subject to uncertainty. Historical data is provided in (SemiconductorData.xls). Represent them using the distributions available in PythonSim, and justify your choice using the graphical and statistical methods discussed in the lecture.

Product will be released in cassettes at the rate of 1 cassette/hour, 7 days a week, 24-hours a day (this is to achieve a desired throughput of 1 cassette/hour). Product is moved and processed in single cassette loads.

The table below shows the number of machines that are planned for each fabrication step:

(b) Provide estimates of the long-run average cycle times for each product. Determine and justify a warm-up period and run length. You may use the replication-deletion approach. Provide appropriate confidence intervals for your estimates.

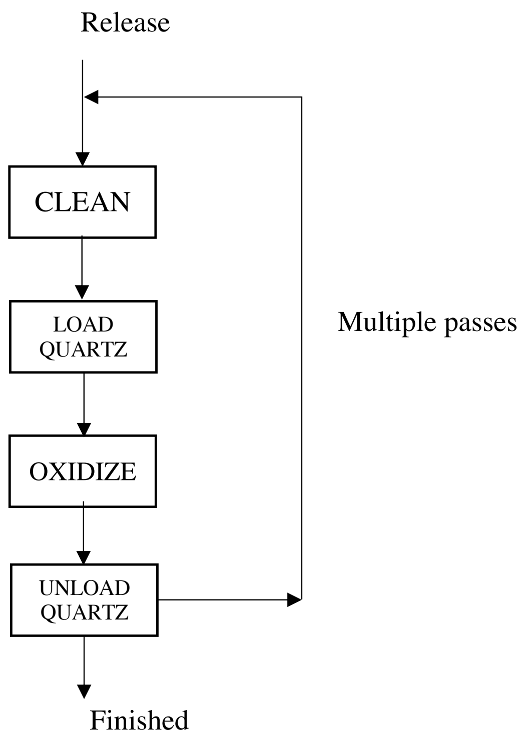

Figure 1: Diagram of the diffusion step.

Step	**Number of Machines**
CLEAN	9
LOAD QTZ	2
OXD	11
UNLOAD QTZ	2

Table 1: Number of available machines in each step.

Problem 2. (20 Pts.) An Automated external defibrillator (AED) can save a person's life in event of a cardiac arrest. To accelerate delivery, start ups have been developing drone technology to quickly deliver AEDs to the scene of the cardiac arrest in case of an emergency call. The AEDs will be maintained at various bases across the city to respond to calls in the area designated to each base.

You have been tasked to estimate the minimum number of drones at a certain base to ensure that with 95% probability there will be a drone available at the event of a cardiac arrest in the area covered by that base.

Assume that calls reporting cardiac arrests arrive according to a non-homogeneous Poisson process with (per minute) rate function,

$$\lambda(t) = \begin{cases} 4, & \text{7AM-12PM}, \\ 2, & \text{12PM-12AM}, \\ 1, & \text{12AM-7AM}. \end{cases}$$

throughout a day. Further, the total "drone busy time" i.e. from the time the call is received until the drone is back to the base and ready to dispatch again is exponentially distributed with mean 45 minutes. Assume at time 12AM, there are no calls (requests for AEDs) in the system.

(a) Explain, in words, your approach in determining the minimum number of drones at the base required to satisfy the provided service level. Specify the queueing model, corresponding parameters, and the performance measure you are estimating.

(b) Determine the minimum number of drones at the base using the inputs provided. Provide a 95% CI for your estimate.

Problem 3. (20 Pts.) The attached file (TSLA.csv) provides daily prices for Tesla's stock from March 2022 - March 2023.

(a) Fit a Geometric Brownian Motion (GBM) to the **Close** price by estimating the drift μ and volatility terms σ. Recall that for GBM the log-returns are independent Normal random variables, i.e.,

$$\log\left(\frac{S(t_i)}{S(t_{i-1})}\right) \sim N(\mu(t_i - t_{i-1}), \sigma(t_i - t_{i-1})),$$

for all $t_i, i = 1, \ldots, k$.

(b) Does the GBM fit the data well? Use your estimates from part (a) together with a Q-Q plot and statistical goodness of fit tests discussed in class to support your answer.

Problem 4. (15 Pts.) Normal distribution is sometimes not a good fit for log-returns. Cauchy distribution is another (heavy-tailed) distribution used to model the log-returns. The CDF of the Cauchy distribution is given by:

$$F(x) = \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2},$$

where x_0 and γ are parameters.

(a) Propose an inversion algorithm to generate samples of a Cauchy distribution with $x_0 = 0$ and $\gamma = 2$.

(b) Estimating the parameters of the Cauchy distribution using MLE is challenging. One approach to estimated the parameters using matching is to match median (1/2th quantile) and half of the inter-quartile range (difference between 1/4th and 3/4th quantiles) with their estimates from the data. Use this method to estimate the parameters using 1000 samples generated from part (a). How do they compare with the original parameters, i.e., $x_0 = 0$ and $\gamma = 2$?

Problem 5. (20 Pts.) Data on call counts to a call centre, by hour, for 1 month (31 days) are provided in CallCounts.xls. Let $N(t)$ represent the cumulative number of arrivals by time t. If the process is nonstationary Poisson, then $Var(N(t))/E(N(t)) = 1$ for all t, or stated differently $Var(N(t)) = \beta E(N(t))$ with $\beta = 1$. Since you have arrival count data, you can estimate $Var(N(t))$ and $E(N(t))$ at $t = 1, 2, \ldots, 8h$. Use these data to fit the regression model $Var(N(t_i)) = \beta E(N(t_i))$ and see if the estimated value of β supports the choice of a nonstationary Poisson arrival process. **Hints**: This is regression through the origin. Also, remember that $N(t_i)$ represents the total number of arrivals by time t_i.