Stochastic Simulation (MIE1613H) - Homework 3

Due: **March 14th**

- Submit your homework to Portal in PDF format by the above deadline. Late submissions are penalized.
- At the top of your homework include your name, student number, department, program and and e-mail address.
- You may discuss the assignment with other students, but each student must solve the problems, write the code and the solutions individually.
- **Code must be in Python 2.7. unless stated otherwise in the problem statement. You must include both the source code (including comments to make it easy to follow) and the output.**
- Full mark is given to answers that are correct and clearly explained. Write a brief and clear explanation of your solution for each problem.

Problem 1. For each of the following scenarios, state whether the estimator will be biased and explain why or why not.

(a) For an $M/G/1$ queue we are interested in estimating the probability that the tenth customer waits less than α time units when the system starts empty, i.e, $P(Y_{10} < \alpha)$, and we use

$$\frac{1}{n}\sum_{i=1}^{n} I(y_i < \alpha)$$

as our estimator, where $y_i, i = 1, \ldots, n$, are n iid samples of the waiting times of the tenth customer's waiting time Y_{10}.

(b) To estimate the **expected** steady-state waiting time of customers in the $M/G/1$ queue we simulate the queue n times and use

$$\frac{1}{n}\sum_{i=1}^{n} y_i,$$

where $y_i, i = 1, \ldots, n$ are the waiting times of the 10,000th customer to enter the system in each replication.

Problem 2. Exercise 16 from Chapter 4 of the textbook.

Problem 3. Exercise 16 from Chapter 6 of the textbook. **Note:** Data file (CallCounts.xls) is available on Portal.

Problem 4. Exercise 17 from Chapter 6 of the textbook. **Note:** For this question you may use Python or Excel. In either case, your need to explain your computations and report the data used to fit the regression model.)

Problem 5. Historical data for call durations in a bank call center are given in the (service-time_data.csv) file available on the portal.

(a) Plot a histogram of the data. (Make sure the size of the bins and the range of the axes are chosen such that the histogram visualizes the data properly.)

(b) Fit an Exponential, Lognormal, and Gamma distribution to the data by estimating the corresponding parameters using the MLE estimator. Discuss how well each distribution models the data. Use a Q-Q plot and a Kolmogorov–Smirnov (K-S) goodness of fit test to support your discussion.

(c) Fit a Gamma distribution to the data by matching the central moments. Are the parameters the same as the MLE estimator for Gamma?

(d) Write a program that generates samples from the interpolated empirical cdf (ecdf) of the data by first importing the data in Python and then writing a function that returns samples from the interpolated ecdf.

(e) Which of the above models would you recommend using in a simulation model; Exponential, Lognormal, Gamma, or empirical?