Stochastic Simulation (MIE1613H) - Homework 3

Due: March 21, 2021

- Submit your homework to Quercus in PDF format by the above 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 must include both the Python / R source code (including comments to make it easy to follow) and the output.
- You may discuss the assignment with other students, but each student must solve the problems, write the code and the solutions individually.
- 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. (20 Pts.) Exercise 16 from Chapter 4 of the textbook.

Problem 2. (15 Pts.) Exercise 16 from Chapter 6 of the textbook. **Note:** Data file (CallCounts.xls) is available on Quercus.

Problem 3. (15 Pts.) Exercise 17 from Chapter 6 of the textbook. **Note:** For this question you may use Python, R, or Excel, but make sure to explain your computations and report the data used to fit the regression.

Problem 4. Consider the discrete-time stochastic process $S_n = \sum_{i=1}^{n} X_i$, for $n = 1, 2, 3, \ldots$, where X_i's are iid random variables with CDF,

$$F(x) = 1 - \left(\frac{2}{x}\right)^3.$$

(a) (5 Pts.) Propose an inversion algorithm to generate samples of X_i's.

(b) (5 Pts.) Propose an unbiased estimator for $P(S_{10} > 40)$. Provide an estimate and a 95% CI using 10^6 samples.

Problem 5. In Ontario, when a patient requires an MRI image, the physician refers the patient to an imaging facility. The facility then schedules an appointment according to the urgency level of the patient. Consider a facility with a single MRI machine. Appointments are given at the beginning of each hour, starting 7AM and over the course of a 12 hour working day until 7PM. That is, the first appointment is at 7AM, the second at 8AM and so on. Patients are however not always punctual and may arrive before or after their scheduled appointment. Imaging time for each patient takes on average around 50 minutes but involves some variability. If there are still patients in the facility at the end of the day, the facility remains open until all patients with appointments on that day are served. Assume all patients show up for their appointments. Assume patients are served in order of arrival.

The attached data file (Problem_5_Data.xlsx) provides historical samples of imaging durations (time it took to conduct the imaging) and how much each patient's arrival deviated from the original appointment time. For example, a sample value of 3 in the "Deviation from appointment" column indicates that the patient was 3 minutes late for their appointment and a sample value of -5 indicates that they were 5 minutes early.

(a) (6 Pts.) Fit an appropriate distribution to the "Imaging Duration" and "Deviation from appointment" data. Justify your choice of the distributions.

(b) (4 Pts.) Explain, in words, how you would simulate arrivals driven by appointments, but subject to early or late arrivals according to the fitted distribution in Part (a).

(c) (10 Pts.) Develop a simulation model of the imaging facility using the distributions you obtained in part (a). Estimate (1) the expected average waiting time of patients during the day, and (2) the expected time it takes until all scheduled patients are served. Report a 95% CI for both estimates.

Problem 6. In your analysis of arrival data for a system that you are planning to simulate, you have estimated $Var(N(t))/E(N(t))$ to be approximately 1.6 for all $t \geq 0$. Since the estimate is significantly greater than 1, you have concluded that a non-homogeneous Poisson process is not a good choice and therefore you need a non-exponential base distribution with variance $\sigma_A^2 = 1.6$ to use with the inversion method for generating arrivals in your simulation model. As discussed in Exercise 35 of Chapter 6, one option for the base distribution is a balanced hyperexponential distribution; that is $\tilde{A}_n$ for $n = 2, 3, \ldots$, is exponentially distributed with rate λ_1 with probability p, and exponentially distributed with rate λ_2 with probability $(1-p)$. "Balanced" means that $p/\lambda_1 = (1-p)/\lambda_2$. Therefore, there are only two free parameters; λ_1 and p.

(a) (**5 Pts.**) Show that

$$p = \frac{1}{2}\left(1 + \sqrt{\frac{\sigma_A^2 - 1}{\sigma_A^2 + 1}}\right), \quad \lambda_1 = 2p,$$

achieves the desired arrival rate and variance. **HINT**: The first and second moments of the hyperexponential random variable are given by $p(1/\lambda_1) + (1-p)(1/\lambda_2)$ and $p(2/\lambda_1^2) + (1-p)(2/\lambda_2^2)$, respectivelly.

(b) (5 Pts.) Show that $G_e(t)$ or the distribution of the first inter-arrival time $\tilde{A}_1$ for the equilibrium renewal process is also hyperexponential but with $p = (1-p) = 1/2$, that is

$$G_e(t) = \frac{1}{2}(1 - e^{-\lambda_1 t}) + \frac{1}{2}(1 - e^{-\lambda_2 t}).$$

(c) (5 Pts.) Propose a method for generating samples from the hyperexponential distribution. **HINT**: Remember the defenition of the Hyperexponential distribution.

(d) (5 Pts.) Assuming $\Lambda(t) = e^t$, write a program in Python that generates customer arrival times for 10 time units. Report the arrival times for one realization.