Assignment #7 (Due April 30)

1. (**Computing Problem:** Telephone system) Consider a land-line telephone system with N (≥ 2) telephones connected to a switchboard by *lines* numbered 1, 2, ... , N. The switchboard has K ($< N$) *links* numbered 1, 2, ... , K, each of which can connect any two lines subject to the restriction that only one connection at any time can be made to a line. If more than one link is available and the called line is not in use, a placed call is connected (instantaneously) on the lowest-numbered available link. A placed call is immediately lost if either the called line is in use (*busy call*) or if no link is available (*blocked call*). (If a call is both busy and blocked, it is considered to be a blocked call.) Assume that the duration of successive calls placed at line i ($1 \leq i \leq N$) are uniformly distributed between 0 min. and 6 min., and that the successive times between the end of a call (placed or received) at line i to the next call placed at line i are i.i.d. as an exponential random variable with mean 6 minutes. (A lost call ends at the same instant that it is placed.) Whenever a call is placed at line i, the called line is equally likely to be any of the other lines. At time $t = 0$ there are no calls underway. Set $X(t) = \left(Z_1(t), \ldots, Z_N(t) \right)$, where $Z_i(t) = k$ if line i is connected (to some other line) on link k at time t and equals 0 if line i is not connected at time t. (This model can also be modified to represent other resource constrained pairwise-interaction systems.)

   a) Specify $\{X(t) : t \geq 0\}$ as a GSMP with event set $E = \left\{ e_1, e_2, \ldots, e_{N+K} \right\}$, where e_i = "call placed at line i" ($1 \leq i \leq N$) and e_{N+k} = "completion of call on link k" ($1 \leq k \leq K$).

   b) Write a program to simulate this GSMP with $N = 4$ lines and $K = 2$ links in order to compute a 95% confidence interval for each of the quantities in i) to iii) below. **Use an event list to schedule events.** [Code for implementing an event list in Python is posted on the class website. You may use this code in your own program.]

      i) The interquartile range for the amount of time that all links are simultaneously idle during the first 100 minutes. Use $m = 5$ sections of $k = 100$ observations each. Use the sectioning + jackknifing approach. [Hint: If your observations are arranged in an array `x`, you can write `x.sort` to sort the elements. To remove a section from an array `x`, you can set `y = copy(x)` and then set `y[a:b]=[ ]`, where `a` and `b` are appropriate indices.]

      ii) The long-run fraction of time that link 2 is busy. Use the regenerative method with regeneration points equal to successive entrances into the state $s = (0,0,0,0)$ and base your estimate on 500 cycles. Use the Taylor-Series approach for computing the ratio estimator and the confidence interval. Briefly explain why successive entrances to s indeed form a sequence of regeneration points.

      iii) Same performance characteristic as part (i), but now estimate the standard deviation (SD) instead of the interquartile range. Use the basic Taylor-series approach based on 500 runs. Based on the output of this simulation, give the approximate number of runs needed to estimate the SD to within $\pm 2\%$ with 95% probability.

c) Now suppose that the phone company is considering alternative per-minute calling rates. The foregoing model can be modified to capture the effect of price on demand by modeling the mean time M from the end of a call to the placement of the next call as a function of the per-minute calling rate r. Specifically, $M(r) = 6(r/0.05)^{1.3}$ for $r \geq 0.05$. Thus the higher the call rate, the fewer the number of calls placed. The three plans under consideration correspond to rates of 0.05, 0.20, and 0.60 dollars per minute. (Note that when a call is successfully placed, only the originator of the call is charged.) Modify your simulation program from Parts (a) and (b) to compute a point estimate of the expected total revenue to the phone company in the interval [0,100], based on a user-specified number of simulation replications. (You don't need to worry about confidence intervals, but you will need to compute the sample variance of the simulated revenue numbers over the simulation replications.) Then read Section 10.4.1 in the textbook, and apply the Dudewicz and Dalal (DD) procedure to select the best plan, such that you are indifferent between plans with total revenues that are within $0.10 of each other ($d^* = 0.10$ in the terminology of the textbook), and you want to select the truly best plan with a probability of at least 95%. Use $n_0 = 40$ replications for each plan under first-stage sampling. [Note: if you wish, you can perform the D&D calculations outside of your main simulation program, e.g., using a spreadsheet, R, Matlab, and so on.]

2. Law, Problem 11.13.

3. Law, Problem 11.17.

4. (Process monitoring) Consider a fabrication process that produces batches of computer chips at a rate of 3 batches per day. A batch is either "good" (G) or "bad" (B). A monitoring process periodically inspects the status of the fab process and makes adjustments as necessary. The monitoring rate is an adjustable parameter $\theta \in [0,1]$; here $\theta = 0$ means no monitoring and $\theta = 1$ means continuous monitoring. The cost of running the monitoring system at rate θ is $0.5\theta^2$. When the monitoring system is being run at rate θ, the sequence of quality ratings X_1, X_2, X_3 for the daily run of batches can be modeled as a discrete-time Markov chain with state transition matrix given by

$$P = \begin{matrix} & \begin{matrix} G & \quad B \end{matrix} \\ \begin{matrix} G \\ B \end{matrix} & \begin{pmatrix} \theta & 1-\theta \\ \theta^2 & 1-\theta^2 \end{pmatrix} \end{matrix}$$

Similarly, the initial distribution is given by $P\{X_1 = G\} = 1 - P\{X_1 = B\} = \theta$. The total (random) cost $c(N, \theta)$ incurred at the end of a day's operation consists of the cost of running the monitoring system plus a cost of $4N$, where N is the number of bad batches produced that day. Our goal is to use simulation to find a value of θ that minimizes the expected total daily cost.

a) Suppose that our first replication uses a value of θ_1, and we observe values $(X_1, X_2, X_3) = (G, B, G)$. Give a general expression for $L(h)$, the likelihood of observing this sequence of outcomes under parameter value $\theta + h$ *relative* to the likelihood of observing this sequence under parameter value θ.

CS 590M
Simulation
Peter J. Haas

Page 3 of 4
April 16, 2020
Spring Semester 2020

b) Suppose that our first replication uses a value of $\theta_1 = 0.5$, and we observe values as described in Part (a). Using your answer from Part (a), compute a numerical estimate Z_1 of $g'(\theta_1)$, where, in general, $g'(\theta)$ is the derivative of the function $g(\theta) = E_\theta\left[c(N,\theta)\right]$.

c) If we now execute a step of the Robbins-Monro Algorithm using a gain-parameter value of $a = 0.01$, what is the value of θ_2 ?

5. (Epidemiology) Consider an agent-based flu model in which a given large urban area is divided into many square regions called "quadrats." Under the flu model, each resident of the city moves along a random path that moves through the quadrats; residents behave in an i.i.d. manner. The probability that a given resident will enter a specified quadrat at any time is small, but there are many residents, so that the overall mean arrival rate γ (in units of persons per hour) to a given quadrat is reasonably large.

a) What is a plausible form for the probability distribution of W, the number of arrivals to a typical quadrat in a given two-hour period, e.g., 11am to 1pm? Briefly justify your answer.

b) We ran the simulation 5 times and then generated 10 bootstrap samples as follows:

		Mean
Original Data	5, 3, 2, 1, 9	4.0
Sample 1	3, 3, 1, 5, 2	2.8
Sample 2	5, 9, 1, 2, 1	3.6
Sample 3	1, 1, 5, 2, 1	2.0
Sample 4	3, 9, 9, 5, 2	5.6
Sample 5	2, 3, 1, 1, 5	2.4
Sample 6	9, 5, 5, 1, 3	4.6
Sample 7	5, 5, 2, 9, 3	4.8
Sample 8	3, 1, 1, 2, 9	3.2
Sample 9	9, 5, 9, 2, 2	5.4
Sample 10	2, 2, 2, 3, 9	3.6

Use the bootstrap pivot method to compute an 80% confidence interval for the mean. Note that this method is a simplified version of the algorithm given for nonlinear functions of means, with input data $X_1,\ldots,X_n$ instead of $(X_1,Y_1),\ldots,(X_n,Y_n)$ and $g(x,y)$ replaced by the identity function $g(x) = x$. [Hint: you can avoid doing 10 subtractions by delaying the subtraction operation as long as possible.]

c) Suppose that we want to estimate the mean 24 hour infection rate $\alpha = E[R]$ using i.i.d. simulation replications, and we have two simulation methods available. Method A produces an observation R_A in 20 seconds, on average. Method B produces on observation R_B in only 2 seconds on average, but the standard deviation of R_B is three times that of R_A. Which method is preferable? Justify your answer.

6. (**Extra credit:** Maintaining feasibility in Robbins-Monro) Consider the Robbins-Monro method in d dimensions. Suppose that the set of feasible solutions (assumed to be non-empty) is given by $\Theta = \{\theta : a \cdot \theta \leq b \text{ and } \theta \geq c\}$, where a and c are d-dimensional vectors, b is a real number, the symbol "$\cdot$" denotes the inner product, and for vectors x and y we say that $x \leq y$ if and only if $x_i \leq y_i$ for $i = 1, 2, \ldots, d$ (and similarly for $\geq$). Suppose that, for the nth iteration of R-M, we have that $a \cdot \theta_n > b$.

   a) First give an explicit formula for computing θ'_n, the perpendicular projection of θ_n onto the $(d-1)$-dimensional plane P defined by $a \cdot \theta_n = b$. [Hint: Recall that the vector a is perpendicular to P, and that the length of a is given by $\|a\| = (a \cdot a)^{1/2}$. Also recall the geometric definition of the inner product of two vectors: $a \cdot b = \|a\| \times \|b\| \times \cos\phi$, where ϕ is the angle between the vectors. Consider the right triangle formed by the three points θ_n, θ'_n, and y, where y is an arbitrary point on the plane P; see the figure below. First derive an expression for δ, the perpendicular distance from θ_n to the plane, so that $\theta'_n = \theta_n - \delta \cdot \dfrac{a}{\|a\|}$.]

   b) It may be the case that the projected point θ'_n fails to satisfy the lower bound constraint. In this case, give a procedure for modifying θ'_n to a new value θ_n^+ such that $a \cdot \theta_n^+ \leq b$ and also $\theta_n^+ \geq c$, so that θ_n^+ is feasible. [Hint: if θ satisfies $a \cdot \theta = b$ and if v satisfies $a \cdot v = 0$, then $\theta^{(r)} = \theta + rv$ satisfies $a \cdot \theta^{(r)} = b$ for any real number r.]