Guidelines: This problem set is due at 10:30 on Monday 26 March 2018 and will not be accepted late. You may not use for reference any source other than your textbook, lecture slides from this class, your own personal notes from class, and R help files. You may not seek or receive assistance from any person other than me. Failure to observe these rules will be treated as an Honor Code violation.

All solutions, including mathematics, must be typeset and submitted as a PDF (I recommend LaTeX). Include appropriate figures, data tables, etc. to support your discussion and conclusions. Strive to create a well-written, self-contained, research-article-like solution set. Include all source code with your solution set.

Problem 1: Use discrete-event simulation to simulate the performance of a single-server service node with a FIFO queue discipline and a *finite* service node capacity. Assume that:

- the arrival process has *exponential*(μ) interarrival times with average interarrival time $\mu = 2$;
- the service process (per job) consists of a number of service tasks equal to 1 + η, where η is a *geometric*(0.9) variate (see §3.1); and
- the time for each of the service tasks within a service process is, indepently for each task, a *uniform*(0.1, 0.2) random variate.

For this work, I suggest that you start with your next-event implementation of ssq, and use the streams capability in `vexp`, `vgeom`, and `vunif` from the `simEd` library.

(a) Based on 1 000 000 processed jobs, construct a table of the estimated steady-state probability of rejection for service node capacities of 1, 2, 3, 4, 5, and 6. (Note that a service node capacity of 1 corresponds to a server only with no queue.)

(b) Construct a similar table if the time per task is changed to be *uniform*(0.1, 0.3).

(c) Provide appropriate histograms to compare the two different service models. Discuss the models and comment on their appropriateness for a single-server queuing system.

(d) Comment on how the probability of rejection depends on the service process.

(e) Discuss what you did to convince yourself that your results are correct.

Problem 2: In collecting statistics from simulation output, you often focus on the sample mean $\bar{x}$ and sample variance s^2, which correspond to the first two population *moments*[1]: the population mean μ and the population variance σ^2. The third (central) moment is the population skewness, often labeled γ. The sample skewness is given by the following two-pass equation:

$$q^3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3 .$$

(a) Derive the one-pass version of this equation. (b) Provide an R function that, given a vector of data as a parameter, computes and returns as an R list the sample mean, sample standard deviation, and the sample statistic $q = \sqrt[3]{q^3}$. Your implementation must compute all three statistics using only a single pass through the data (non-Welford), and you may not use the `mean()` nor `sd()` functions available by default in R, nor may you use any non-standard R packages. (c) Devise and conduct an experiment to compare the histogram and

[1] Technically, the mean is the first *raw* moment and the variance is the second *central* (i.e., about the mean) moment.

sample skewness of many samples drawn from each of three different distributions having common mean: *uniform*, *exponential*, and *Erlang*. (d) Comment.

Note: The continuous random variable X is *Erlang*(n, β) if and only if

$$X = X_1 + X_2 + \cdots + X_n$$

where $X_1, X_2, \ldots, X_n$ is an *iid* sequence of *exponential*(β) random variables; the associated mean and standard deviation are $\mu = n\beta$ and $\sigma = \sqrt{n}\,\beta$ respectively. Alternatively, when the *gamma* distribution has integer shape parameter, the distribution is *Erlang*.

Problem 3: A test is compiled by selecting 12 different questions, at random and without replacement, from a well-publicized list of 120 questions. After studying this list you are able to classify all 120 questions into two classes, I and II. Class I questions are those about which you feel confident; the remaining questions define class II. Assume that your grade probability, conditioned on the class of the problems, is

	A	B	C	D	F
class I	0.6	0.3	0.1	0.0	0.0
class II	0.0	0.1	0.4	0.4	0.1

In other words, if a question is class I, you have probability 0.6 of earning an A, probability 0.3 of earning a B, and so forth. Each test question is graded on an A = 4, B = 3, C = 2, D = 1, F = 0 scale and a score of 36 or better is required to pass the test. (a) If there are 90 class I questions in the list, use Monte Carlo simulation to generate a histogram of scores. (b) Based on this histogram what is the probability that you will pass the test? (c) Provide a sense of the uncertainty of your result.

Problem 4: Provide brief descriptions of at least two projects of interest to you for an agent-based simulation final project. Provide citations of the sources for your ideas. As a couple of launching points, you can consider the external links at the bottom of the Wikipedia page on “agent-based model”, and the proceedings from recent years of the Winter Simulation Conference (`https://informs-sim.org/`).