ISyE 6832 – Spring 2019

Homework #1 — Due Monday, February 4

1. Let U be uniformly distributed on [0, 1] and define the indicator function $1_A(x) = 1$ if $x \in A$ or 0 otherwise.

   (a) Show that the sequence $X_n = \sqrt{n}1_{(0,1/n)}(U)$ converges to 0 almost surely and in L^1, but not in L^2 (mean square).

   (b) Show that the sequence $X_n = 1_{[0,1/2+1/n)}(U)$ converges to $1_{[1/2,1]}(U)$ in distribution ($\Longrightarrow$), but not in probability.

   (c) For each $n = 1, 2, \ldots$ suppose that X_n is uniformly distributed with $P(X_n = \frac{k}{n}) = \frac{1}{n+1}$ for $k = 0, 1, \ldots, n$. Show that $X_n \Longrightarrow U$; so that continuous uniform distributions are limits of discrete ones.

2. Let $X_1, X_2, \ldots$ be i.i.d. random variables with the uniform distribution on [0, 1]. Show that

$$n[1 - \max\{X_1, X_2, \ldots, X_n\}] \Longrightarrow \text{expo}(\lambda = 1),$$

   where "expo" denotes the exponential distribution.

3. Consider the directed, acyclic stochastic activity network in Figure 1 below. The nodes represent milestones while the arcs represent tasks. Let A be the set of arcs. A path from the source $s = 1$ to the sink $t = 9$ is completed when the all tasks on the path are completed; hence, the completion time for a path is the sum of the task durations on that path.

   The project is complete when all $s \to t$ paths are completed; hence, the project duration is the length of the longest $s \to t$ path. Assume temporarily that the task durations d_{ij} are fixed, let $P_j = \{i : (i, j) \in A\}$ be the set of nodes immediately preceding node j (so that $P_6 = \{2, 3, 5\}$), and let $d(i)$ denote the completion time for milestone i (length of longest path from s to node i). Then the project duration can be computed from the following recursion:

   `set` $d(1) = 0$;
   `for` $j = 2, \ldots, t$: `set` $d(j) = \max_{i \in P_j}\{d(i) + d_{ij}\}$

   Alternatively, if $\{\mathcal{P}_1, \ldots, \mathcal{P}_k\}$ is the set of $s \to t$ paths, we can compute the project duration using path enumeration (an expensive task when the network is large):

$$d(t) = \max_{\ell=1\ldots,k} \sum_{(i,j)\in\mathcal{P}_\ell} d_{ij}.$$

   The labels μ_{ij} on the arcs represent parameters of the distributions of the respective task durations. Specifically, The task durations are independent random variables with the following distributions:

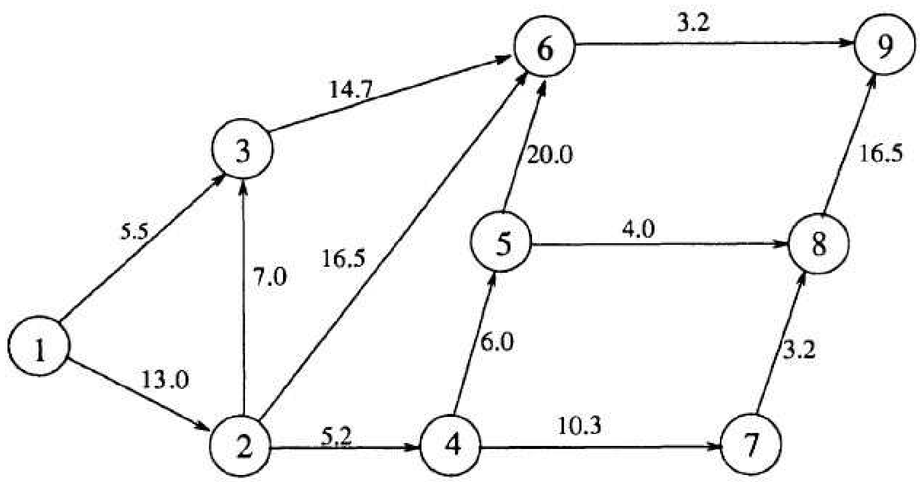

Figure 1: Stochastic activity network.

- The durations of tasks $(1, 2)$, $(1, 3)$, $(2, 4)$, $(6, 9)$, and $(7, 8)$ are $\max\{N(\mu_{ij}, \mu_{ij}/4), 0\}$, where $N(a, b)$ denotes a normal random variable with mean a and standard deviation b.
- The durations of the remaining tasks are exponentially distributed with mean μ_{ij}.

Use a programming language (Matlab, Python, R, C/C++, Java, etc.) or a spreadsheet with the SIPMath Modeler Tools to develop a simulation experiment based on k independent trails. Then use the values $k = 10^3, 10^4, 10^5$ to compute histograms for the project completion time as well as 95% confidence intervals (CIs) for the mean project completion time and the following percentiles of the project completion time: 5th, 25th, 50th (median), 75th, and 95th.

What happens as the number of trials increases? Please explain using a concise discussion.

4. Consider an $M/G/1$/FIFO queueing system with Poisson arrivals at rate 1 (per minute) and i.i.d. service times from the gamma distribution with shape parameter $\alpha = 0.8$ and scale parameter $\lambda = 1$; the mean of this distribution is $\alpha/\lambda = 0.8$ minutes. Let A_i be the time between arrivals $i - 1$ and i (ith interarrival time), let S_i be the service time of the ith entity, and let X_i be the delay in queue of the ith entity prior to service.

(a) Show that the sequence of delays can be computed by the following recursion

$$X_{i+1} = \max\{X_i + S_i - A_{i+1}, 0\}, \quad i = 1, 2, \ldots. \tag{1}$$

(b) Use Kingman's formula to compute the mean entity delay $\mu \equiv \lim_{i\to\infty} \mathrm{E}(X_i \mid X_1 = x_1)$ in steady state:

$$\mu = \frac{\rho}{1-\rho} \cdot \frac{c_A^2 + c_S^2}{2} \cdot ES,$$

where EA and ES are the the mean interarrival time and service time, respectively; $\rho = ES/EA$ is the traffic intensity; $\mathrm{Var}(A)$ and $\mathrm{Var}(S)$ are the variances of the interarrival and service times, respectively; and $c_A^2 = \mathrm{Var}(A)/(EA)^2$ and $c_S^2 = \mathrm{Var}(S)/(ES)^2$ are the squared coefficients of variation of the interarrival and service times, respectively. We will use the true mean delay to evaluate the effectiveness of the simulation methods in parts (c) and (d) below.

(c) Suppose that the system starts in the empty and idle state, hence $X_1 = 0$. Create a simulation experiment with $k = 100$ independent trials (replications), each generating $n = 10^3, 10^4, 10^5$ consecutive entity delays. Use the (pseudo-)uniform random number generator in your package of choice to generate the interarrival and service times. Then use recursion (1) to generate the delays.

For each value of n, generate k average entity delays $Y_j = n^{-1}\sum_{i=1}^{n} X_i,\ j = 1, \ldots, k$, and use the i.i.d. sample $\{Y_1, Y_2, \ldots, Y_k\}$ to compute an approximate 95% CI for μ using the formulas

$$\bar{Y}_k \pm t_{k-1,0.975}\frac{S_k}{\sqrt{k}}; \quad S_k^2 = \frac{1}{k-1}\sum_{j=1}^{k}(Y_j - \bar{Y}_k)^2.$$

Plot the above CIs relative to the steady-state mean μ and comment on their behavior as n increases. Does the fact that the delay process $\{X_i\}$ has a transient phase matter?

(d) Now make an experiment with 100 independent trials/replications. Let $n = 10^5$. In each trial start with an empty system ($X_1 = 0$), collect $n + 1000$ delays $X_1, X_2, \ldots, X_{n+1000}$ and compute the following 95% CI for μ:

$$\bar{X}_n \pm z_{0.975}\frac{S_n}{\sqrt{n}},$$

where

$$\bar{X}_n \equiv \frac{1}{n}\sum_{i=1001}^{n+1000} X_i$$

is the "truncated" average delay of entities $1001, \ldots, 1000 + n$ (to defuse the effects of the initial transient period) and

$$S_n^2 = \frac{1}{n-1}\sum_{i=1001}^{n+1000}(X_i - \bar{X}_n)^2$$

is the respective sample variance of the truncated sample. What percentage of the 100 CIs contain the true mean μ? Is this percentage surprising?