ISyE 6644 – Spring 2023

Homework #1 — Due Tuesday, January 31

Below the term "`SIPmath Tools`" refers to the Excel add-in from `www.probabilitymanagement.org`. Follow the instructions on the syllabus to submit a single archive containing all and Simio files and a single PDF document containing all your derivations and experimental outcomes. Name the archive as `LastName-FirstName.zip`.

1. Consider the definite integral

$$\mu = \int_{-2}^{2} \frac{e^t}{1 + e^{t^2}}\, dt.$$

   (a) Use the Monte Carlo method discussed in class with the following 20 pseudo-uniform(0, 1) numbers

.43	.15	.54	.18	.93	.21	.64	.51	.72	.81
.87	.59	.98	.02	.28	.52	.51	.24	.77	.66

   to compute a point estimate and an approximate 95% confidence interval (CI) for μ. Use a scientific calculator with statistical functions and a table with quantiles for the t distribution.

   (b) Compute an estimate of the number of uniform(0, 1) observations that are required to obtain an approximate 95% CI for μ with half-length ≤ 0.02.

2. Estimate the integral in Problem 1 using an Excel spreadsheet with the `SIPmath Tools` or a programming language (Matlab, Python, R, C/C++, Java, etc.) and 100 uniform(0, 1) observations. The Excel function for uniform(0, 1) random numbers is `RAND()` and the function for π is `Pi()`. You can use the following Excel statistical functions: `AVERAGE(range)`, `STDEV.S(range)`, and `T.IVN(q,k)`. The latter function returns the $100q$th percentile of Student's t distribution with k degrees of freedom. For example, `T.INV(0.95,1000) = 1.646`.

   Use the "recalculate" key (`F9` key on a Windows machine or `Cmd-=` on a Mac) to repeat the experiment 100 times. Report the fraction of the 95% confidence intervals that contain the actual value of μ, which can be obtained from Wolfram Alpha.

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 $s = 1$ (start) to the sink $t = 9$ (end) 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. Also, all nodes can be labeled so that every arc (i, j) has $i < j$.

   A milestone is complete when all tasks on paths from the source to the respective node are complete. Hence the project is complete when all paths from s to t have been completed, and the project duration is the length of the longest $s \to t$ path. Assume temporarily that the task durations d_{ij} are fixed, and let $d(i)$ denote the completion time for milestone i (length of longest path from s to node i).

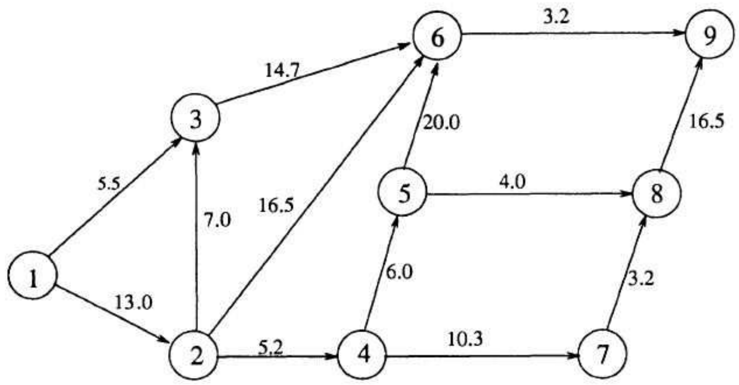

Figure 1: Stochastic activity network.

Let $\{\mathcal{P}_1, \ldots, \mathcal{P}_k\}$ is the set of $s \to t$ paths. Then 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 (i, j) in Figure 1 represent parameters of the distributions of the respective task durations. Specifically, the task durations are independent random variables with the following distributions:

- 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 (Python, R, C/C++, Java, etc.) or a spreadsheet with the SIPMath Tools to develop two simulation experiments based on 10^3 and 10^4 trials, respectively. At the end of each experiment 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. To compute the CIs, use the nonparametric method on slides 24–26 from the document `Error and Risk Estimation - Basics.pdf`. If you use the SIPMath Tools, you can simulate the duration of each arc and then each path in a single row using absolute references to the cells for the arcs in that path. In this case also compute a relative cumulative histogram for the project completion time.

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

Remark: If you use a programming script, let $P_j = \{i : (i, j) \in A\}$ be the set of nodes immediately preceding node j (e.g. $P_6 = \{2, 3, 5\}$). Then the project duration can be computed from the following recursion:

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

4. **[Extra credit (5 points) only if you submit answers to Problems 1–3.]** In this exercise we will use the variance cancellation methods on slides 29–31 from the document `Error and Risk Estimation - Basics.pdf`. A similar experiment is described in slides 50–55; feel free to use my file (with appropriate modifications). Let's simulate an M/G/1/FIFO queueing system with Poisson arrivals at a rate of $\lambda = 1/4$ per minute (mean of 4 minutes) and i.i.d. service times from the gamma distribution (see `https://en.wikipedia.org/wiki/Gamma_distribution`) with shape parameter $\alpha = 2$ and scale parameter $\beta = 1.5$. The mean and variance of the service times are equal to $\alpha\beta$ and $\alpha\beta^2$, respectively. The Excel formulas for generating the interarrival and service times are in the same slide set.

   First, compute the mean time an entity spends in the system in steady state. Recall that the mean entity delay (prior to service) can be obtained by Kingman's formula.

   Use a spreadsheet (or script) that simulates 2,200 entities based on a detailed table with the interarrival times, arrival times, times service starts, service times, and departure times. Start with an entity arrival to an empty system at time zero. Then conduct an experiment with 100 independent trials. Let $\bar{X}_i$ be the average time-in-system for entities 201 through 2,200 in trial i (disregard the "transient" data for the first 200 entities). Also let $\hat{x}_{0.9,i}$ be the point estimate of the 90th percentile of the time-in-system obtained from entities 201–2,200 in trial i. In the SIPMath Tools environment, you will create responses corresponding to the average time-in-system and the point estimate of the 90th percentile computed from the times-in-system for entities 201 to 2,200. The replicate values $\bar{X}_i$ and $\hat{x}_{0.9,i}$ can be found in the worksheet "PMTable".

   (a) Compute a point estimate and a 95% CI for $E(\bar{X}_i)$ using the variance cancellation method for the mean. Assuming that we have truncated a sufficient amount of transient data, this point estimate should be close to the mean (expected value) time in system per entity in steady state computed using Kingman's formula.
   (b) Obtain a histogram and a cumulative histogram for the random variable $\bar{X}_1$. (The subscript is irrelevant because the "replicate" averages $\bar{X}_i$ are i.i.d.) Does the shape of the histogram resemble the density of a known distribution? Name that distribution.
   (c) Use the "replicate" quantile estimates $\hat{x}_{0.9,i}$ and the variance cancellation method for quantiles to compute an approximate 95% CI for the 90th percentile of the time *an arbitrary* entity spends in the system in steady state. Is the latter point estimate close to the point estimate for the 90th percentile of the average time in system from part (b)?