All students should read §3.1 of the textbook in preparation for this LGA as well as the next lecture.

The following numbered questions should be split across your group and the solutions discussed during the next lecture period.

i. You may use a pRNG provided by your language of choice (provided it is **not** `rand(3)`).

ii. You may need to implement your own $Uniform(a,b)$, $Equilikely(a,b)$, $Exponential(\mu)$ or $Geometric(p)$ routines (definitions 2.3.3, 2.3.4, 3.1.1, and 3.1.2 respectively).

Students should review the learning goals for the day, determine which are applicable to their questions and provide answers or commentary to their group members. When using the Internet to formulate answers (some questions may require this), keep track of **where** you find your information on the web. You may be asked for, and are expected to have (in Email-able form), URLs supporting your investigations.

1. Perform the following data collection and analysis:

   (a) i. Create $N = 10$ different datasets of $n = 1000$ points chosen from the $Uniform(0,1000)$ distribution, for clarity we'll notate one of these datasets as $\{x_i\}_{i=1}^{n}$.

   ii. Sort each of the these individual datasets, and call the the new ordering $\{a_i\}_{i=1}^{n}$ (think: **arrival times**). Then calculate the distance between each sequential (after sorting) pair of points (within each sorted dataset). To be clear: you are calculating the values $\{r_1 = a_1 - 0, r_2 = a_2 - a_1, \ldots, r_i = a_i - a_{i-1}\}$ (think: **interarrival times**).

   What is the statistical mean and standard deviation of this (collectively) $1000N$ sized population of derived values $\{a_*\}$?

   (b) Now, plot two **empirical CDFs**, one for the $1000N$ sized $\{a_*\}$ and one for the $1000N$ sized set of $\{r_*\}$. An emperical CDF is generated for a set of data $\{u_k\}_{k=1}^{K}$ by first sorting the data into $\{s_k\}_{k=1}^{K}$, and plotting the coordinate pairs $(s_1, \frac{1}{K}), (s_2, \frac{2}{K}), \ldots, (s_i, \frac{i}{K}), \ldots, (s_k, 1)$.

   You will compare these results to the findings of question 3 when we reconvene in the next lecture.

2. Return to your (or your group's at the time) solution to `lga-coding-ssqs.pdf` question 2 where Little's equations were used to calculate $\bar{q}$, $\bar{x}$ and $\bar{l}$. Change the code so that **instead** of reading arrival times and service times from a file, they are determined by using random variates:

   1. The **interarrival time** of each job comes from an $Exponential(1)$ random variate.

   2. The service time for each job comes from a $Uniform(0.5, 1.5)$ random variate.

   Run 1000 jobs through your simulation for **five** different random seeds and display the time domain statistics $\bar{q}$, $\bar{x}$ and $\bar{l}$ for your group to discuss at the next lecture.

3. Perform the following data collection and analysis:

   (a) Create $N = 10$ different datasets of $n = 1000$ tuples of points according to the following algorithm:

   ```
   procedure GenerateOneDataSet()
     r_1 ← Uniform(0,2)
     a_1 ← r_1
     i ← 2
     do (
         r_i ← Uniform(0,2)
         a_i ← a_{i−1} + r_i
         increment i
     ) while ( i ≤ 1000 )

     # two sequences of 1000 numbers each
     return {r_i}_{i=1}^{1000}, {a_i}_{i=1}^{1000}
   ```

   What is the statistical mean and standard deviation of this (collectively) $1000N$ sized population of $\{r_*\}$ data points?

   (b) Now, plot two **empirical CDFs**, one for the $1000N$ sized $\{a_*\}$ and one for the $1000N$ sized set of $\{r_*\}$. An emperical CDF is generated for a set of data $\{u_k\}_{k=1}^{K}$ by first sorting the data into $\{s_k\}_{k=1}^{K}$, and plotting the coordinate pairs $(s_1, \frac{1}{K}), (s_2, \frac{2}{K}), \ldots, (s_i, \frac{i}{K}), \ldots, (s_k, 1)$.

   You will compare these results to the findings of question 1 when we reconvene in the next lecture.

4. Return to your (or your group's at the time) solution to `lga-coding-ssqs.pdf` question 2 where Little's equations were used to calculate $\bar{q}$, $\bar{x}$ and $\bar{l}$. Change the code so that **instead** of reading arrival times and service times from a file, they are determined as such:

   1. Determine the job arrival times by generating 1000 random values from $Uniform(0, 1000)$ and then sorting them smallest to greatest. These are your $\{a_1, a_2, \ldots, a_{1000}\}$ arrival times.
   2. The service time for each job comes from a $Uniform(0.5, 1.5)$ random variate.

   Run these 1000 jobs through your simulation for **five** different random seeds and display the time domain statistics $\bar{q}$, $\bar{x}$ and $\bar{l}$ for your group to discuss during your next group time.

5. Return to your (or your group's at the time) solution to `lga-coding-sis.pdf` question 3 where the SIS simulation trace data file for demands was pushed through the simulation with various starting point offsets. These generated a range of values for the total (dependent) cost.

   Change the coded solution so that the 100 weekly demands are determined instead by:

   $$d_i = 17 + Geometric(0.93)$$

   Run the simulation with **10 different seeds** and tabulate the results for your group to review. The average and minimum of `sis1.dat` match that of the variate valued d_i of this assignment. What else might account for the total cost differences compared to `lga-coding-sis.pdf` question 3?