ELEN90054 Probability and Random Models
MATLAB Workshop 6 on Poisson Random Process Simulation;

week 11 = week of 14 May 2018

Prepare this workshop individually, in particular questions Q1,2,3,7a and 8a, before you come to your workshop session.

This workshop is worth 5% of the overall subject assessment and should be completed in pairs. Allocation will be random by your demonstrator, so you should not choose your project partner yourself. The next workshop will have a new allocation. Be aware that seeking or providing detailed assistance from/to people other than your workshop partner is collusion - see **http://academichonesty.unimelb.edu.au/plagiarism.html**.

Each group is expected to upload two files:

1. A pdf (scanned/typed) file containing their worked solutions;
   a. Only one member of the group needs to upload the pdf file.
   b. The naming convention of the file should include workshop number, day, time and the assigned group number, e.g. Workshop_2_Mon09_Gp5.pdf for group 5 in the Monday 0900 hrs workshop slot.
   c. You also need to **read and attach** the Engineering cover-sheet (see the ELEN90054 LMS site Workshops) signed by both of you.
2. A single zip file, containing all the required functions and **a single main script** that calls these functions to generate the required outputs as outlined in the workshop questions.
   a. Only one member of the group needs to upload the zip file.
   b. The following naming convention should strictly be followed: Workshop_2_Mon12_Gp5_Matlab.zip for group 5 in the Monday 1200 hrs workshop slot.

The workshop times are: Mon09, Mon12, Tue10, Tue17, Wed13, Wed17, Wed18, Wed19, Thu11, Thu19, Fri10, Fri13. The group numbers are randomly assigned by your demonstrator at the start of the workshop.

Both submissions should be made before the start of the next workshop of week 12. **This is a strict deadline.**

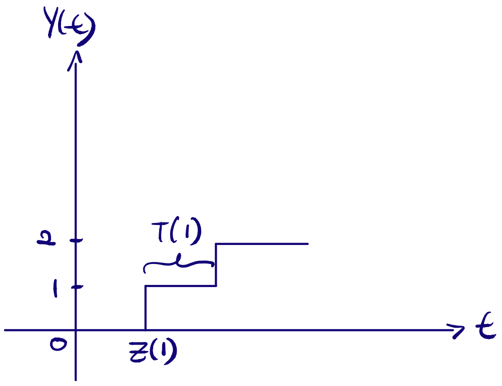
In special circumstances you can email the pdf/code to your demonstrator. For demonstrator email address information, see "staff info" on the ELEN90054 LMS site.

Note: The main purpose of this workshop is to investigate and simulate a random process, focusing on the Poisson process and related random sequences. Also to get you to experiment with a real life queuing system. Engineering applications can be found in packet networks where buffer overflow may occur when more packets arrive than the server can handle.

Poisson process

(total=36 marks + 4 on-time attendance marks)

This workshop is about University of Melbourne students filling up their water bottles at the drinking fountain at the North Entrance of the EE building. Our experiment consists of observing students filling up their bottles. We assume that we start this experiment at 10am, we call this $t = 0$. We also assume that students do not visit the drinking fountain twice. Our model involves a number of random processes, including:



- $Z(k)$ = arrival time of the k'th student
- $T(k)$ = time between the arrival of the $k-1$'th student and the k'th student (take $T(0) = 0$)
- $S(k)$ = time that it takes the k'th student to fill up his/her bottle (="service time")
- $W(k)$ = waiting time of the k'th student
- $Y(t)$ = the total number of students in the first t minutes (take $Y(0) = 0$).

The inter-arrival times are assumed to be independent and exponentially distributed with a mean of 4 minutes. Hence, $Y(t)$ is a Poisson process with an average arrival rate of λ arrivals/minute. You may use the following properties of the Poisson process:

1. The time between two subsequent visits has an Exponential(λ) distribution
2. In any interval of length τ, the number of visits has a Poisson($\lambda\tau$) distribution
3. For any two disjoint intervals, the number of visits in each are independent of each other

Questions

1. (1 mark) Explain that $\lambda = 0.25$ students.
2. (3 marks) Suppose that V_1 and V_2 are independent random variables with $V_1 \sim$ Poisson(μ_1) and $V_2 \sim$ Poisson(μ_2). By using moment generating functions (mgf's), show that

$$V_1 + V_2 \sim \text{Poisson}(\mu_1 + \mu_2).$$

3. a) (3 marks) Find an expression for the second order CDF of the random process $Y(t)$, taking the following into account:
   - as notation, use time instants t_1 and t_2, where $0 < t_1 < t_2$
   - formulate your answer in terms of the marginal (not joint) PMFs of two independent Poisson random variables V_1 and V_2 with suitably defined means. Clearly describe what V_1 and V_2 stand for.

   b) (2 marks) Let q be the joint probability that there are no more than 3 visits in the first 20 minutes and no more than 5 visits in the first 40 minutes. Use your answer of part a) to compute q.

Questions continue on the next page

4. (4 marks) Write a MATLAB function that returns a sample array of visit times up to $t = 60$ minutes, corresponding to one realisation (= sample path) of $Y(t)$ (i.e. the time of the first visit is stored at index 1, the time of the second visit at index 2, etc.). You may use the **exprnd** command. Using this function, plot two different realisations (= sample paths) of $Y(t)$ on the same axes (the commands **stairs** and **hold on** may be useful). **Clearly label the axes in your plot.**

5. Write another function that accepts the array of visit times of the previous question as an input, and returns the corresponding empirical values of:

   (a) (3 marks) The total number of visits by (i) $t = 20$ (ii) $t = 30$, and (iii) $t = 40$

   (b) (3 marks) The number of visits between (i) $t = 20$ and $t = 40$, (ii) $t = 30$ and $t = 50$, and (iii) $t = 40$ and $t = 60$.

6. (4 marks) Using 50,000 calls to the function from Q5:

   (a) Construct an empirical PMF of $Y(40)$ and plot it on the same axes as the true PMF

   (b) Compute an empirical value of the probability q, that was defined in Q3b, and compare it to the theoretical value from Q3b

   (c) Compute an empirical value of $E[Y(t)]$ for each of $t = 20$, $t = 30$ and $t = 40$

   (d) Compute an empirical value of $E[Y(t)Y(t + 10)]$ for $t = 20$ and for $t = 30$.

7. a) (1 mark) Is $Y(t)$ wide sense stationary (for $t \geq 0$)? Explain using theory.

   b) (2 marks) Compare your answers of part a) to the appropriate parts of Q6. Interpret and discuss your findings.

8. a) (3 marks) Find an expression for the autocorrelation function of $Y(t)$. Show your workings.

   b) (1 mark) Show that your expression of part a) is consistent with your empirical values from Q6.

9. It can be shown that the waiting time $W(k)$ is given by

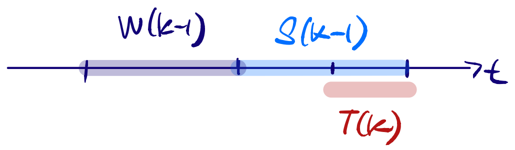
$$W(k) = \max\{W(k-1) + S(k-1) - T(k), 0\}.$$

   You may take $W(0) = 0$. Suppose that the service time $S(k)$ is distributed uniformly as $S(k) \sim \text{Uniform}(a, b)$ minutes.

   a) (4 marks) Assume that $a = 0.3$ and $b = 1$. Write a MATLAB function that returns a sample array of waiting times up to $n = 30$ students. Using 50,000 calls to this function, plot the average waiting time of n students as a function of n. **Clearly label the axes in your plot.**

   b) (2 marks) Repeat part a) for $a = 2$ and $b = 5$. Compare the resulting plot with the plot of part a) and comment on the difference. What does it say about the water fountain situation in these two cases?

End of MATLAB Workshop 6 Questions



Q1:

Base on the description of the problem, we know that the inter-arrival times are exponentially distributed with a mean of $4\frac{mins}{arrival}$. By dimensional analysis, this is equivalent to an average arrival rate of $\lambda = \frac{1}{4}\frac{arrival}{min}$.

Q2:

According to the definition of MGF, we can easily derive the MGF of random variables V_1, V_2 which are shown below:

$$\phi_{V_1}(s) = E[e^{SV_1}] = \mathrm{e}^{\mu_1(e^S-1)}$$

$$\phi_{V_2}(s) = E[e^{SV_2}] = \mathrm{e}^{\mu_2(e^S-1)}$$

$$\therefore \phi_{V_1+V_2}(s) = \phi_{V_1}(s) * \phi_{V_2}(s) = \mathrm{e}^{\mu_2(e^S-1)} * \mathrm{e}^{\mu_1(e^S-1)} = \mathrm{e}^{\mu_1 e^S-\mu_1+\mu_2 e^S-\mu_2}$$

$$= \mathrm{e}^{(\mu_1+\mu_2)*(e^S-1)}$$

So we can conclude that the combination of two independent Poisson random variables is also a Poisson random variable. That is, $V_1 + V_2 \sim \text{Poisson}(\mu_1 + \mu_2)$.

Q3:

(a) The second order CDF of the random process $\mathrm{Y(t)}$ are shown below: (where $0 < t_1 < t_2$)

$$\begin{aligned} F_{Y(t_1),Y(t_2)}(y_1,y_2) &= P[Y(t_1) \le y_1, Y(t_2) \le y_2\,] \\ &= P[Y(t_2) \le y_2 | Y(t_1) \le y_1\,] * \mathrm{P}[Y(t_1) \le y_1] \\ &= \mathrm{P}[Y(t_1) = 0] * P[Y(t_2 - t_1) \le y_2\,] + \mathrm{P}[Y(t_1) = 1] \\ &\quad * P[Y(t_2 - t_1) \le y_2 - 1\,] + \cdots + \mathrm{P}[Y(t_1) = y_1] \\ &\quad * P[Y(t_2 - t_1) \le y_2 - y_1\,] \end{aligned}$$

In this case we define two independent Poisson random variables V_1, V_2 to simplify above equation.

$$V_1 = Y(t_1)\ \sim Poisson(0.25t_1)$$

$$V_2 = Y(t_2 - t_1)\ \sim Poisson(0.25(t_2 - t_1)\,)$$

the CDF can be expressed by V_1, V_2.

$$\begin{aligned} &F_{Y(t_1),Y(t_2)}(y_1,y_2) \\ &= \mathrm{P}[V_1 = 0] * P[V_2 \le y_2\,] + \mathrm{P}[V_1 = 1] * P[V_2 \le y_2 - 1\,] + \cdots \\ &+ \mathrm{P}[V_1 = y_1] * P[V_2 \le y_2 - y_1\,] \\ &= \sum_{i=0}^{y_1}\sum_{j=0}^{y_2-i}\frac{(0.25t_1)^i}{i!}e^{-0.25t_1} * \frac{(0.25(t_2-t_1))^j}{j!}e^{-0.25(t_2-t_1)} \end{aligned}$$

(b) Under this circumstance, $\mathrm{q} = F_{Y(t_1),Y(t_2)}(y_1,y_2)$, where $t_1 = 20,\ t_2 = 40,\ y_1 = 3,\ y_2 = 5$.

$$\therefore V_1 \sim Poisson(0.25 * 20) = \text{Poisson}(5)$$

$$V_2 \sim Poisson(0.25 * (40 - 20)) = \text{Poisson}(5)$$

So we use the equation derive from (a), we can obtain the value of q.

$$q = F_{Y(20),Y(40)}(3,5) = \sum_{i=0}^{3}\sum_{j=0}^{5-i}\frac{(5)^i}{i!}e^{-5} * \frac{(5)^j}{j!}e^{-5}$$

$$= \frac{5^0}{0!}e^{-5} * \left(\sum_{k=0}^{5}\frac{5^k}{k!}e^{-5}\right) + \frac{5^1}{1!}e^{-5} * \left(\sum_{k=0}^{4}\frac{5^k}{k!}e^{-5}\right) + \frac{5^2}{2!}e^{-5}$$

$$* \left(\sum_{k=0}^{3}\frac{5^k}{k!}e^{-5}\right) + \frac{5^3}{3!}e^{-5} * \left(\sum_{k=0}^{2}\frac{5^k}{k!}e^{-5}\right)$$

Because the calculation is too large, we use MATLAB to calculate the value of q

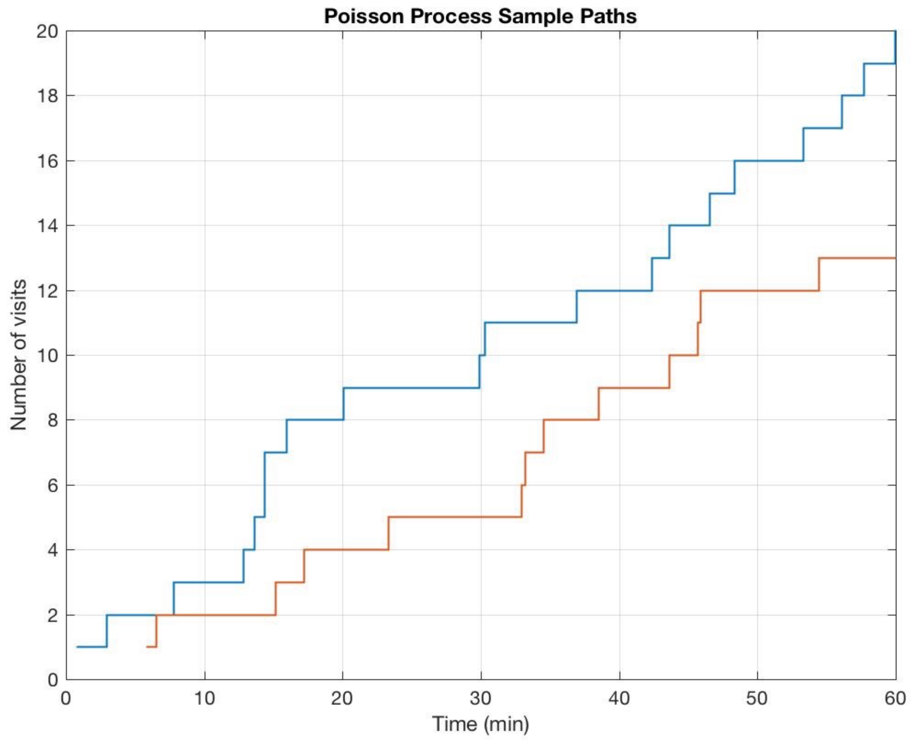
The code is as follows:

```
k5=0;
for v1=0:3
k1=(5^v1/factorial(v1))*exp(-5);
k3=0;
for v2=0:5-v1
k2=(5^v2/factorial(v2))*exp(-5);
k3=k2+k3;
end
k4=k3*k1;
k5=k4+k5;
end
disp(k5)
```

Then we can obtain the result is 0.0588.

4. (4 marks) Write a MATLAB function that returns a sample array of visit times up to $t = 60$ minutes, corresponding to one realisation (= sample path) of $Y(t)$ (i.e. the time of the first visit is stored at index 1, the time of the second visit at index 2, etc.). You may use the **exprnd** command. Using this function, plot two different realisations (= sample paths) of $Y(t)$ on the same axes (the commands **stairs** and **hold on** may be useful). **Clearly label the axes in your plot.**

Please find in attached MATLAB files the function **arrivalTime(lambda, tMax)** that returns a vector of arrival (visit) times up to **tMax**, given the parameter of the inter-arrival random variable T~exponential(**lambda**).



5. Write another function that accepts the array of visit times of the previous question as an input, and returns the corresponding empirical values of:

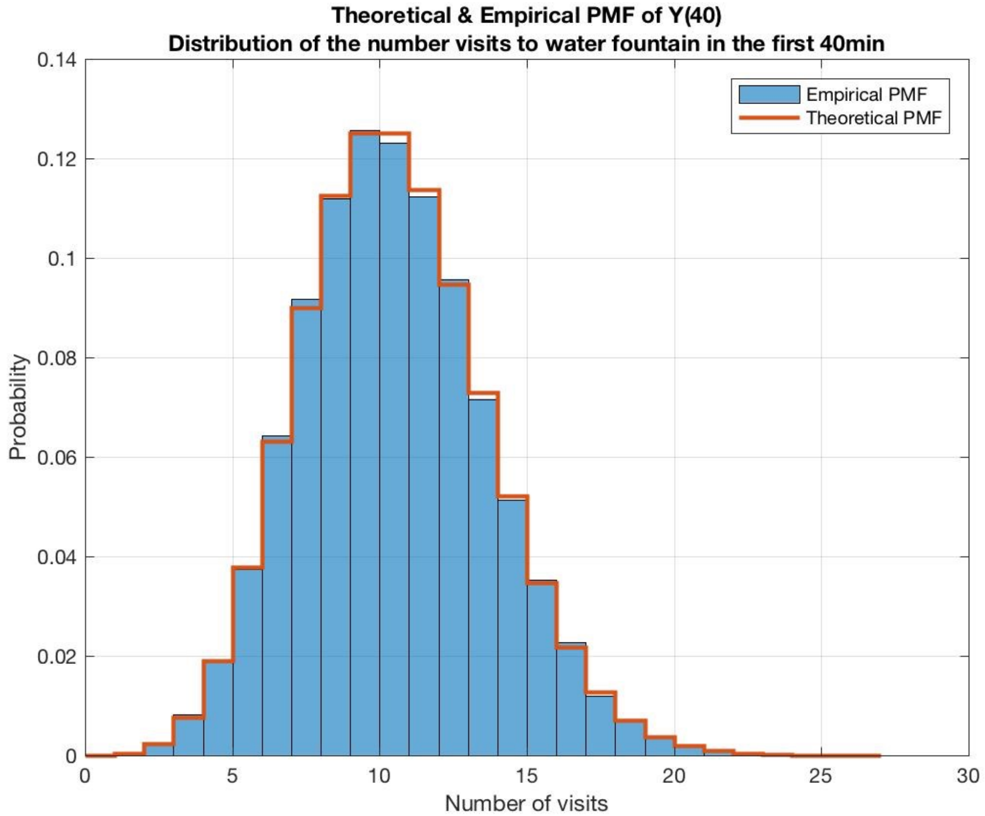
   (a) (3 marks) The total number of visits by (i) $t = 20$ (ii) $t = 30$, and (iii) $t = 40$

   (b) (3 marks) The number of visits between (i) $t = 20$ and $t = 40$, (ii) $t = 30$ and $t = 50$, and (iii) $t = 40$ and $t = 60$.

Please find in attached MATLAB files the function **arrivalNum(Z, t1, t2)** that returns the number of arrivals between time **t1** and **t2** in a vector of arrival times **Z**.
If **t2** is omitted, **arrival(Z, t1)** returns the number of arrivals between time 0 and **t1**.

6. (4 marks) Using 50,000 calls to the function from Q5:

(a) Construct an empirical PMF of $Y(40)$ and plot it on the same axes as the true PMF



(b) Compute an empirical value of the probability q, that was defined in Q3b, and compare it to the theoretical value from Q3b

(c) Compute an empirical value of $E[Y(t)]$ for each of $t = 20$, $t = 30$ and $t = 40$

(d) Compute an empirical value of $E[Y(t)Y(t + 10)]$ for $t = 20$ and for $t = 30$.

6b Empirical probability of Y(20) ≤ 3 AND Y(40) ≤ 5 is **5.93%**

6c Empirical expected value of Y(20) is **4.99**
Empirical expected value of Y(30) is **7.489**
Empirical expected value of Y(40) is **9.985**

6d Empirical expected value of Y(20)Y(20 + 10) is **42.401**
Empirical expected value of Y(30)Y(30 + 10) is **82.245**

7. a) (1 mark) Is $Y(t)$ wide sense stationary (for $t \geq 0$)? Explain using theory.

   b) (2 marks) Compare your answers of part a) to the appropriate parts of Q6. Interpret and discuss your findings.

7a

The Poisson Process Y(t) is **not** wide sense stationary for t ≥ 0 because its **mean is not constant**. In fact, its mean is **E[Y(t)] = lambda*t** which is dependant on the **t**.

Its auto-correlation function **R(t1, t2) = E[X(t1)X(t2)]** also **does not depend on the time difference t1 - t2 = tau**. If the auto-correlation function does depend on the time difference, then **E[X(t1)X(t1 + 10)]** should equal to **E[X(t2)X(t2 + 10)]**. But as observed from the above simulation results, this property does not hold.

7b

Empirical expected value of Y(20) is **4.99**
Empirical expected value of Y(30) is **7.489**
Empirical expected value of Y(40) is **9.985**
The above empirical expected values show that the mean of **Y(t) is not constant**, as it changes with different values of **t**.

Empirical expected value of Y(20)Y(20 + 10) is **42.401**
Empirical expected value of Y(30)Y(30 + 10) is **82.245**
The above empirical expected values are a function of the same time difference **t1 - t2 = 10**, but the **expected values are not not the same**. This shows that the auto-correlation function **R(t1, t2)** is changing in spite of the same time difference value. It is therefore not a function of the time difference and violates one of the criterions of wide sense stationarity.

Q8:

(a)

According the definition of autocorrelation function, we can derive the $\mathrm{R}(t_1, t_2)$ of $\mathrm{Y(t)}$ below: (we assume that $0 < t_1 < t_2$

$Y(t_1) \sim Poisson(0.25t_1)$, $\mathrm{Y}(t_2) \sim Poisson(0.25t_2)$

$$\mathrm{R}(t_1, t_2) = \mathrm{E}[Y(t_1) * \mathrm{Y}(t_2)\,] = \mathrm{E}\big[Y(t_1) * \big(Y(t_1) + \mathrm{Y}(t_2) - Y(t_1)\big)\big]$$

$$= \mathrm{E}\big[Y^2(t_1) + Y(t_1) * \big(\mathrm{Y}(t_2) - Y(t_1)\big)\big]$$

$$= \mathrm{E}[Y^2(t_1)] + \mathrm{E}\big[Y(t_1) * \big(\mathrm{Y}(t_2) - Y(t_1)\big)\big]$$

Because $\mathrm{Y}(t_2) - Y(t_1)$ is equivalent to $\mathrm{Y}(t_2 - t_1)$, furthermore $Y(t_1)$ and $\mathrm{Y}(t_2 - t_1)$ are independent, the equation can be expressed below:

$$\begin{aligned}\mathrm{R}(t_1, t_2) &= \mathrm{Var}[Y^2(t_1)] + \mathrm{E}^2[Y(t_1)] + E[Y(t_1)] * E[\mathrm{Y}(t_2 - t_1)] \\ &= 0.25t_1 + 0.25^2 {t_1}^2 + 0.25t_1 * 0.25(t_2 - t_1) \\ &= 0.25t_1 + 0.25^2 t_1 t_2\end{aligned}$$

(b)

For question 6, we have already get the empirical value of $\mathrm{E}[Y(20) * \mathrm{Y}(30)\,]$ and $\mathrm{E}[Y(30) * \mathrm{Y}(40)\,]$. Now we need to calculate the theoretical values by using the formula listed before shown below:

$$\mathrm{E}[Y(20) * \mathrm{Y}(30)\,] = \mathrm{R}(20, 30) = 0.25 * 20 + 0.25^2 * 20 * 30 = 42.5$$

$$\mathrm{E}[Y(30) * \mathrm{Y}(40)\,] = \mathrm{R}(30, 40) = 0.25 * 30 + 0.25^2 * 30 * 40 = 82.5$$

In previous question, $\mathrm{E}_{empirical}[Y(20) * \mathrm{Y}(30)\,] = 42.401$

$\mathrm{E}_{empirical}[Y(30) * \mathrm{Y}(40)\,] = 82.245$

From above we can find that the theoretical values are consistent with the empirical value.

9. It can be shown that the waiting time $W(k)$ is given by

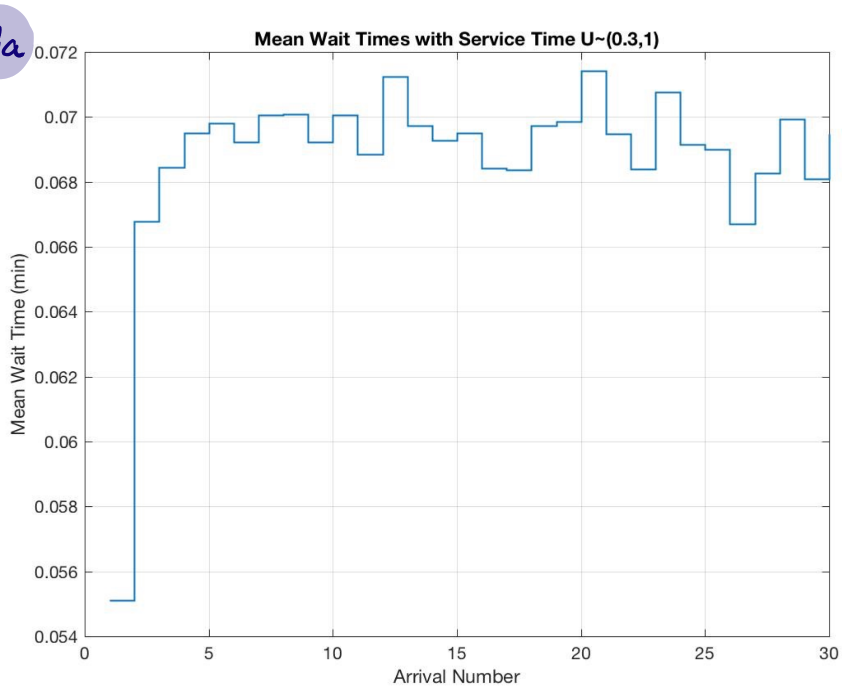
$$W(k) = \max\{W(k-1) + S(k-1) - T(k), 0\}.$$

You may take $W(0) = 0$. Suppose that the service time $S(k)$ is distributed uniformly as $S(k) \sim \text{Uniform}(a, b)$ minutes.

a) (4 marks) Assume that $a = 0.3$ and $b = 1$. Write a MATLAB function that returns a sample array of waiting times up to $n = 30$ students. Using 50,000 calls to this function, plot the average waiting time of n students as a function of n. **Clearly label the axes in your plot.**

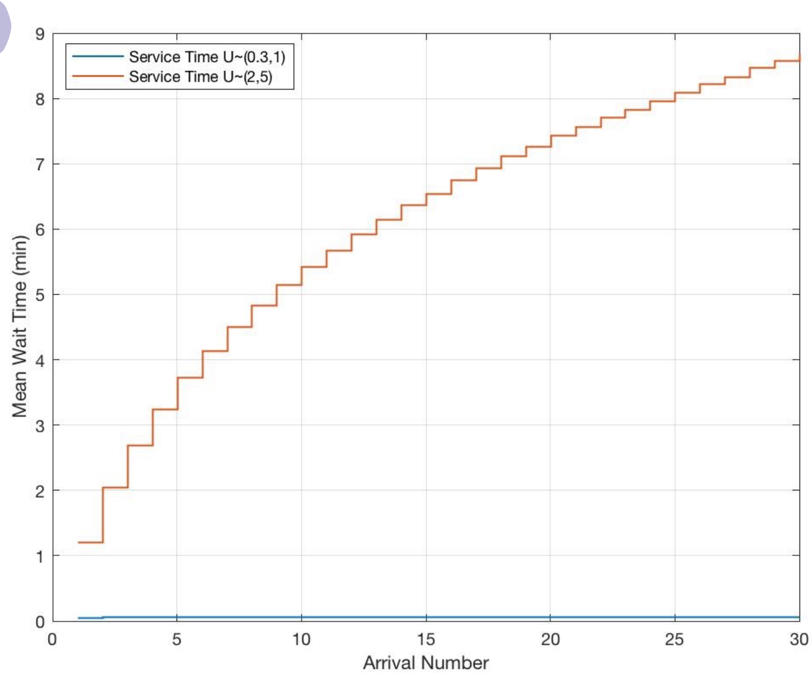
b) (2 marks) Repeat part a) for $a = 2$ and $b = 5$. Compare the resulting plot with the plot of part a) and comment on the difference. What does it say about the water fountain situation in these two cases?

9a



Please find in attached MATLAB files the function **waitTime(lambda, a, b, maxW)** that returns a vector containing **maxW** number of wait times generated with inter-arrival time T~exponential(**lambda**) random variable and service time S~uniform(**a**, **b**) random variable.

9b



The wait times generated by a service time of S~uniform(2, 5) is observed to be building up with every newly arrived student.

Students arriving at the water fountain in the first case with a service time of S~uniform(0.3, 1) is able to use the water fountain without having to wait too long.

Students arriving at the water fountain in the second case with service time of S~uniform(2, 5) however, will have to on average wait a certain amount of time, depending on where they are in the queue (arrival number).