IERG3050 Simulation and Statistical Analysis (Spring 2021)
Assignment 4

Instructions:

1. Do your own work. You are welcome to discuss the problems with your fellow classmates. Sharing ideas are great, and do write your own programs and explanations.
2. All work should be submitted onto the blackboard before the due date.
3. You are advised to submit a zip or rar containing all necessary files. In the submission, you should have the following items
   a. A pdf file explaining all your files and your answers to each part.
   b. A python program, assignment4_basic.py, refers to the work in the basic part.
   c. Three more python programs, assignment4_taskX.py, refer to task X in the extended parts (X ranges from 1 to 6).
4. Do type/write your work neatly. If we cannot read your work, we cannot grade your work.
5. If the submitted programs cannot be run or are irrelevant to the problem (e.g. empty python programs), you will receive no scores for that part.
6. Assignment 4 is a project-like assignment, and will take 20% of the course total.
7. Do spend at least three days for the project-like assignment 4.
8. Due date: 27th Apr, 2021 (Tuesday) 23:59

Basic part: Simulation for a canteen (50%)

Simulation Background:

In operating a canteen, customers arrive at the canteen and order food. Dining time of a customer is defined as the time that the customer stays inside the canteen. When the canteen is full, customers need to wait outside in a single queue. The waiting time of a customer is defined as the time that the customer waits outside the canteen. Total service time is the sum of the dining time and the waiting time. Before the simulation, N customers are generated. The simulation will begin with the N customers. The simulation will not generate more customers during the simulation.

In the canteen, there are different sizes of tables. Customers will pick the table randomly. The size of the tables is stored in a list of integers. The floor plan of the canteen is fixed throughout the simulation. It means that the canteen will not change the floor plan during the simulation.

The canteen opens from 11:00 am to 2:00 pm daily, while customers arrive at the canteen from 11:00 am to 12:59 nn. When customers get into the canteen before 2:00 pm, they can finish their meal after 2:00 pm. When the time is 2:00 pm, customers waiting outside the canteen cannot have their lunch in this canteen. If a customer cannot get in the canteen for consecutive k days, the customer will never come back to this canteen.

Due to a pandemic of a flu, canteen owner and customers pay very high attention to the flu. When someone dines with a carrier, a person having flu, in the same table at the same time, the person will be infected by the flu. At the end of a day, everyone is subjected to test whether he/she has the flu or not. If the person does not have the flu, the test will be negative. If the person has the flu, the test has a p chance (Bernoulli trial) to be positive.

Once the person is tested positive, the person will be sent to the hospital immediately. On the other hand, the canteen will also blacklist the patients.

Write a simulation for a canteen during lunch hour for D days. Submit your program as assignment4_basic.py.

Notations for the simulations:

- Number of days in the simulation: D, $D \geq 1$
- Number of the customer: N, $N \geq 1$
- Number of customers at the beginning on day i: n_i, $n_i \leq N$
- Number of tables: T_size, T_size ≥ 1
- Arrival time for customer c at day i: a_{ci}, $0 \leq a_{ci} < 120$
- Dining time for customer c: d_c
- Parameter for Bernoulli trial in testing positive with flu in the simulation: p, $0 \leq p \leq 1$
- Parameter for consecutive days in removing customer in the simulation: $k \geq 2$

Part A: Initial settings

There are two ways to initialize the system, which are initialized by file and initialized by random numbers.

Initialization with file

1. The first argument refers to the name of the configuration file.
2. The configuration file is a two-column matrix with file extension csv.
3. An example of the configuration file is as follows.
   - D,100
   - N,100
   - T_size,12
   - T_detail,"2,2,2,2,2,2,4,4,4,4,4,4"
   - p,0.8
   - k,2
   - Customer_file,"customer.csv"
   - Result_file,"simulation_result.csv"
   - Summary_file,"simulation_summary.csv"
4. The file for the customers is a (D+2)xN table. The file name is specified in the configuration file.
   - Header is "customer_id,dining_time,carrier,arrival_time_day_1,...,arrival_time_day_D"
   - Dining time is an integer ranging from 30 to 100.
   - Carrier is a binary variable. 1 means the customer is a carrier. 0 means the customer is not a carrier.
   - Following D columns specify the arrival time of each customer on different days. Arrival time is an integer, the offset from 11:00 am. For example, 110 means that the customer arrives at 12:50 (= 11:00 am + 110 min).
5. The file for the canteen is a list of integers, specifying the size of each table. The first table is table 1 (one-based).

Initialization with random numbers

1. There are some default numbers.
    - D = 100
    - N = 100
    - T_size = 12
    - Size of tables = [2, 2, 2, 2, 2, 2, 4, 4, 4, 4, 4, 4]
    - Customer number is one-based. If i mod 10 == 1, then customer i is a carrier.
    - p = 0.8
    - k = 2
2. Set the random seed to 0
3. Generate N integers as min(30+expo(20), 120), corresponding to the dining time of N customers.

Part B: One day in the simulation

1. The simulation starts with empty-and-idle state.
2. The canteen only opens during 11:00 am and 2:00 pm.
3. At day i, you need to set the random seed to i. The first simulation day is day 1.
4. For initialization with random numbers, customers arrive uniformly during 11:00 am and 12:59 nn.
5. Set the random seed to 10000+i
6. When a customer arrive,
    - If the canteen is full, the customer waits outside in the queue (FCFS).
    - If the canteen is not full, the customer picks a random number X ~ U(1,n+1), where n is the number of tables). The customer will go to table X. If the table X is not full, the customer will dine at table X. If the table X is full, the customer will start linear search to find the next available table. For example, a customer picks table 3, but tables 3 - 5 are full, the next available table is 6, then the customer will dine at table 6. Furthermore, the next table of table n is table 1.
7. At 2 pm, the canteen will close. Customers in the queue cannot dine at the canteen. However, customers inside the canteen can continue their meal.
8. Suppose customer A has flu, while customer B does not have flu. If customer A and customer B dine at the same table at the same time, customer B will be infected by the flu. Customer B can spread the flu immediately after being infected by customer A.
9. Set the random seed to 20000+i
10. At the end of day, customers with flu have a chance p being blacklisted by the canteen (removed from the system).
11. Customers cannot dine at the canteen for consecutive k days and will never come back to the canteen again (removed from the system).

Part C: Output statistics

1. If the simulation is initialized by random numbers, you need to output the configuration file as config_seed0.csv and the customer information as customer_seed0.csv. The file format is the same as the initialization with file. If the arrival time for a customer is unavailable, the entry is -1.
2. Output the result of the simulation to either the specified output file in the configuration file or simulation_result.csv in the random setting.
   - Header is “Day,Time,Event,len(Queue),Number_dining,Number_finish”
   - Day is the day of simulation. It is one-based.
   - Time is the simulated time. The start of simulation of a day is 11:00.
   - Events can only be “Customer i arrives” or “Customer i dined at table t and departs”. Customers not having lunch in the canteen will only have the arrival event.
   - len(Queue) is the length of the queue at the simulated time.
   - Number_dining is the number of customers dining in the canteen at the simulated time.
   - Number_finish is the number of customers departed from the canteen at the simulated time.
3. Output the summary of the simulation to either the specified output file in the configuration file or simulation_summary.csv in the random setting.
   - Header is “Day,ni,Number_finish,Carrier,Carrier_removed,Hunger,Hunger_removed”
   - Day is the day of simulation. It is one-based.
   - ni is the number of customers at the beginning of the day.
   - Number_finish is the number of customers who have lunch in the canteen.
   - Carrier is the number of carriers at the end of the day.
   - Carrier_removed is the list of carriers blacklisted by the canteen at the end of the day.
   - Hunger is the number of customers unable to have lunch in the canteen.
   - Hunger_removed is the list of customers blacklisting the canteen.

   Hint: Do study the code provided in the lecture, “introduction to queueing system”. It is very useful to the basic part.

Extended parts (50%)

Task 0. Specifying your selected tasks (2%)

In the extended part, each task carries 16%. You are asked to select 3 tasks, write your programs and explain your work.
First, explicitly, state clearly which 3 tasks you have chosen.
Second, for each selected task, you need to write a program. The start of the task is the program in the basic part. You need to submit separate programs for each task.
Third, there are some questions in each task, you need to explain clearly how your program can solve the task, and show the results.

Task 1. Flu with incubation period (16%)

The incubation period is the number of days between when you are infected with flu and when you might see symptoms. The incubation period varies from different conditions.

There are some features added or changed in the simulation.

1. In the basic part, there is a member binary variable indicating whether the customer is inflected by the flu or not. In task 1, we use 2 variables. One refers to whether the customer is a carrier of the flu or not. Another refers to the length of the customer being infected by the flu.
2. When the length, described in step 1, is less than the incubation period, the customer cannot be detected by some tests. The customer will not be blacklisted by the canteen.
3. When the length, described in step 1, is the same as the incubation period, the customer will be detected as positive to certain tests. The customer will immediately be blacklisted by the canteen at the end of the day.
4. The length of the incubation period is a constant in one simulation run.

You are asked to study the relations between the length of the incubation period and the spread of the flu in the canteen. Describe your experiment settings with detail. Conduct the experiments and necessary analyses. Finally, give your conclusions.
You also need to submit your program as assignment4_task1.py.

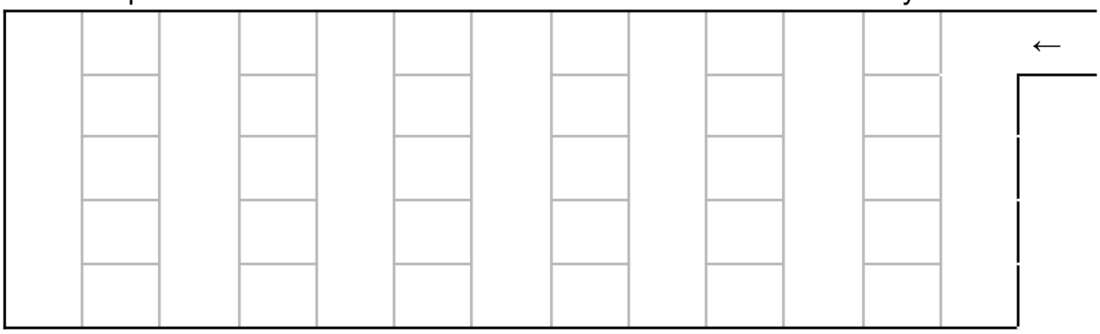
Task 2. Workflow representation and traffic generation (16%)

What are open class and closed class? Explain these terms using the simulation in the assignment as an example.

A theorem states that open representations are pessimal. Verify this theorem experimentally. Explain your experiment settings with detail. Conduct the experiments and necessary analyses. Finally, give your conclusions.
You also need to submit your program as assignment4_task2.py.

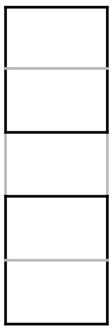
Task 3. Analysis on the canteen configurations (16%)

The floor plan of the canteen looks like below. Entrance is indicated by the arrow.

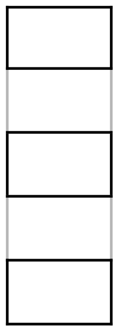


In the floor plan, each column can assign either two 4-person tables or three 2-person tables.

Two 4-person tables:



Three 2-person tables:



The profit of the canteen varies with the number of customers in a day linearly. Design the floor plan such that the profit of the canteen can be maximized in the simulation. Describe your experiment settings with detail. Conduct the experiments and necessary analyses. Finally, give your conclusions.
You also need to submit your program as assignment4_task3.py.

Task 4. Analysis on the total service time (16%)

To analyse the profit of a canteen, another direction is to study the total service time of the customers. If the total service time is smaller, then the canteen can earn more by serving more customers. In this extended part, we focus on the analysis on the relations between the distribution of the arrival rate and the total service time of the customers. You need to pay attention to the following changes:

1. In this extended part, you should not consider the flu.
2. Every day, the number of customers should be the same, and this number is up to your experiment design. The same set of customers will come to this canteen everyday.
3. The probability distribution of the arrival rate should be the same for one simulation.

4. The dining time for each customer should be the same across all simulations. For example, customer A takes 100 time unit to finish his meal on day 1 in a simulation, it means that customer A takes 100 time unit to finish his meal everyday in all simulations.
5. The customers will only arrive between 11:00 am and 12:59 nn. If the customer cannot get into the canteen before 2:00 pm, the customer cannot enjoy their lunch. In your analysis, you need to consider customers not having lunch and describe clearly how you handle these cases.

Besides the uniform distribution in the basic part, you need to compare at least 3 different probability distributions. Describe your experiment settings with detail. Conduct the experiments and necessary analyses. Finally, using two-stage sampling, give your conclusions, which probability distribution can give a smaller total service time.
You also need to submit your program as assignment4_task4.py.

Task 5. Transient states and steady states (16%)

In the basic part, what are transient states? What are steady states? Construct two experiments such that they convert to different steady states after 50 days in the simulation. Both simulations should initialize with empty-and-idle state. Describe your experiment settings with detail. Conduct the experiments and necessary analyses. Finally, give your conclusions.
You also need to submit your program as assignment4_task5.py.

Task 6. Sensitivity analysis (16%)

In the simulation, the output variable is the number of customers in a day. What are the input variables of the simulations? Design the experiments for determining all variables that the simulation is sensitive to. Describe your experiment settings with detail. Conduct the experiments and necessary analyses. Finally, give your conclusions.
You also need to submit your program as assignment4_task6.py.