Analysis of Hospital Bed Requirements Using Discrete Event Simulation and Mathematical Modeling

Dincer Atasoy
Igdir University
Igdir, Turkey
dincer.atasoy@igdir.edu.tr

Abstract

Using SimPy and Discrete Event Simulation we have observed the different model responses of a system consisting of a hospital and people getting sick/healing under different initial conditions.

In our model, each independent person can get sick at an exponential rate. The hospital’s capacity is limited and accepts only a limited amount of sick people. When the hospital gets full, people are sent home to heal. This is simulated under different conditions and the simulation results have demonstrated that the model reaches the steady-state in every scenario. These cause the outputs of the simulation to be similar to each other even when the simulation is run with different initial conditions. The results are compared with the Machine Repair Problem and Erlang-Loss System and the model is validated.

1. Introduction

Analyzing large real systems is complex and it is impossible to observe the system under different conditions. Discrete Event Simulation [6] helps people understand the behavior and mechanism of a real system by simulating it under certain conditions. The simulation displays each step of the model and creates conclusions about the system. In a simulation entities and processes interact with each other, causing different results. This way, the real-time behavior of the system can be modeled as events occur.

This study simulates a real-time system consisting of a hospital and people getting sick/healing. The simulation demonstrates different outputs of the model as the hospital status/simulation time changes. In this study, the model outputs and responses are measured and discussed. The problem will be stated and the model will be explained in the next chapter. Following that, the numerical results will be shared and discussed in the scope of Queueing Systems [16] more precisely Machine Repair Problem [12] and Erlang-Loss System [5]. Finally, the conclusion of the study will be stated.

2. Related Work

Discrete event simulation is widely used for a lot of reasons and allows users to better understand its behavior and mechanism by replicating the behavior and mechanism of a genuine system under specific conditions. Unlike continuous event simulation, it suits making computations efficiently, and there are works that compare discrete event simulation with continuous event simulation in certain aspects [3].

Until now, there has been research [11] about using discrete event simulation to understand the underlying structures of how diseases are spreading. However, the amount of research has dramatically increased in recent years due to the COVID-19 virus.

Recent work [4] introduces bbsim.hospital, a resource-planning tool for hospitals dealing with the COVID-19 epidemic. Another work [7] intends to assess the effects of modifying consultation start time and patient arrival on wait times and congestion in a dual practice outpatient clinic. At the same time, a work [13] seeks to estimate the appropriate number of machines and operators needed, as well as their placement at various workstations, based on the available resources and the rate of samples to be examined every day. Finally, the goal of recent work [8] is to propose a decision-aid tool for hospital management that will allow them to decide on the bed needs for a specific hospital or network of hospitals over a short-to-medium term horizon.

3. Methodology

We aim to model a system where there are $N = 1582$ people that each can get sick with an exponential rate of $\lambda = 1/300$ [patients/day]. People who are sick can go to the hospital to heal or they can stay home. The probability of a sick person going to the hospital for healing is 0.2.
whereas the probability of a sick person staying home is 0.8.

Our simulation is a process-based simulation where processes interact with each other as objects with their own states and variables. There is an environment under which each process occurs i.e. people get sick/heal. And processes communicate through yield and wait for each other. Also, there are several processes active throughout the simulation. There is a simulation i.e. environment time running at all times, waiting for the events.

If our model was to be simulated as an event-based simulation, then there would be a list of events and the simulation would change its states as events occur according to their arrival time. In this case, we would have to implement an event scheduler, so that when people arrive, the system state changes. Since there is an event queue in an event-based case, only one routine would be active at a time. But in our case, there are several processes active in a time period, they work in threads. Also, in our model, we check the availability of the hospital by using the communication of processes and objects. We can easily do this with SimPy [10]. In an event-based approach, there would be no need for SimPy since there are no processes interacting with each other. So, we would use normal Python loops and variables to control the number of people in the hospital and decide whether the system should accept patients to the hospital or not by occurring each event. In event-based, keeping the FEL would be easier, since each event is listed in the beginning.

In both approaches, the results won’t change, only the implementation would change. So, one can decide on which simulation type to use by deciding the desired implementation.

There are $K = \frac{N}{24} = 66$ beds in the hospital which means that only 66 people can be healed at the hospital. After the hospital is full, the patients arriving at the hospital start to get rejected. Those patients are sent home to heal. Since the healing process at the hospital takes less time, the healing duration of a sick person at home takes $r \sim U[1, 2]$ times longer. The unit of time is days.

There are 3 different exponential healing rates for 3 different types of patients:

1. $\mu_1$: Healing rate of the patient being healed at the hospital.
2. $\mu_2$: Healing rate of the patient healing at home.
3. $\mu_3$: Healing rate of the patient that is initially meant to be healed at the hospital but sent home due to full capacity. So s/he is obligated to heal at home.

The healing rates are:

$$\begin{align*}
\mu_1 &= \frac{1}{6} \\
\mu_2 &= \frac{1}{10} \\
\mu_3^{-1} &= \mu_3^{-1} \cdot r = 6 \cdot r & \text{where } r \sim U[1, 2]
\end{align*}$$

We have a system as described above and we are to generate a Discrete Event Simulation for it. We have used SimPy [10] to simulate the events and observe the results. Events can be getting sick and a healing process is done for each person in the system. Everyone in the system can get sick and once a person is healed s/he can get sick again. For the simulation to be consistent with reality, we choose people who get sick randomly. So, anyone can get sick in the system, no one is favored.

The simulation runs in a SimPy environment called env. In that environment, the resource is the hospital with a capacity of 66, since each sick person arriving at the hospital expects to be healed there. If the resource is occupied, i.e. if there are 66 people in the hospital, then next arrivals are sent home, and until someone leaves the hospital sick people coming to the hospital get rejected. The simulation run time is 1000, 10000, or 100000 units of time.

So there are two subsystems in our model: hospitals and people’s homes.

To start to run the simulation, one should set the until and hospital_status to the desired simulation run time and hospital status. This can be done by changing the commented out variables at the beginning of the code:

```python
# until=1000
cr=1000
# until=10000

# hospital_status = "empty" # state
# hospital_status = "half-full" # state
cr=10000

# RANDOM_SEED = 978

random_seed = 979
```

Then, run the simulation by running the code in .ipynb file.

Firstly, 1582 people are created for the simulation. Each person lives in the same environment, has unique IDs and information on where to heal, heal duration, and departure time.

Secondly, the hospital is filled according to the hospital_status. The system adds the necessary number of people to the hospital before the simulation starts to run.

Finally, the simulation starts in the environment by calling the generator of sick people (person_generator()) and runs until the simulation time is done. During the runtime of the simulation, several people get sick, get healed and get sick again, and so on. The simulation starts after one interarrival time.

```python
# until=10000
until=10000
# until=100000

# hospital_status = "empty" # state
# hospital_status = "half-full" # state

# RANDOM_SEED = 978
random_seed = 979
```
The person_generator(), first checks whether the system is full of sick people or not. If everyone in the system is sick, then it waits for one to heal by yielding the generator_reactivate to reactivate the system. It then chooses a random healthy person to get sick (again) using the list healthy_people. This person is removed from the list of healthy people and his/her interarrival time is calculated using $\lambda \cdot |\text{healthy people}|$ as an exponential distribution [9]. After that, this person is added to the will_arrive_people list including his/her arrival time and id.

The simulation now waits for the selected sick person to arrive, i.e. it waits for his/her arrival time. It removes them from the healthy people list and Future Event List. At this step, the simulation decides whether that person will go to the hospital or not. After that, this person is processed to be sick and sent to get_sick() method.

In get_sick(), the healing place of the patient is checked. This is the part of the model where a person gets sick and decides to go to the hospital or not.

If he/she decided to go to the hospital, then the capacity of the hospital is checked. This is done by checking the hospital_count where the resource capacity is checked.

If the hospital is not full, then the patient gets accepted to the hospital. A hospital bed is requested (hospital.request()) and yielded. The heal duration is calculated using $\mu_1$, random.expovariate($\mu_1$).

Else if the hospital is full, the person is sent home to heal. The heal duration is calculated using $\mu_3$.

If he/she decides to heal at home, then they start healing at home. The heal duration is calculated using $\mu_2$, random.expovariate($\mu_2$).

Each sick person is then prepared for healing. In prep_to_heal() the departure time of the person is calculated and the sick person is added to the sick people list. The data of that person is added to the table along with the FEL.

Finally, the person starts to heal. In heal(), the simulation waits for the heal duration of the person. At this point, if everyone in the simulation is sick, then the simulation reactivates the hospital since one patient will be leaving. That person is removed from the sick list and added to the healthy list. The data of the recovered person is added to the table.

In our simulation, the output is an Excel table containing the important DES values as events occur. It contains Simulation Time, Person Id, Event Type, Healing Place, Heal Time, Number of Sick People, Number in the Hospital, Future Event List. These Excel tables can be found in Section 4.

Explanation of the DES Table:

1. Person with id 1210 gets sick and he/she is to be healed in the hospital. The number of sick people and the number in the hospital increase by 1. The next arrival and the current person’s departure time are seen in FEL. The departure time = simulation time + healing duration. Also, the interarrival time for the current sick person is written in the table in Figure 1.

2. Person with id 479 arrives and will be healed at home. The number of sick people increases by 1 and the number in the hospital remains unchanged. The next arrival and the current person’s departure time are added to the FEL in Figure 2.

3. In these examples, Heal Time is NaN since that value is written in the table only for healing events, i.e. when a sick person departs.

The simulation also calculates the following responses:

- long-run probability of the hospital is empty,
- proportion of sick people healing in the hospital,
- average number of sick people in the population,
- average proportion of sick people in the population,
- standard deviation of sick people in the population,
- average number of occupied beds in the hospital,
- standard deviation of occupied beds in the hospital
- total average sickness time.

The simulation also plots the change in the number of sick people in the system and the number of people in the hospital over Simulation Time (See 4.1 and 4.2).

The analysis of the outputs, graphs, and responses can be found in Section 4.

4. Numerical Analysis

4.1. Analysis of the Plots

As one can see from the graphs, with both seeds and each simulation run time-hospital status pair, the steady-state mean values are approximately the same. However, the standard deviation of the values gets larger as simulation run time gets larger which is the expected behavior of the system.

Moreover, when the system starts with half-full/full hospital, the number of people in the hospital decreases at a fast rate and reaches a steady state. This fast fall in the number of people in the hospital is due to the fast healing rates of people in the hospital. And when the steady-state is reached, due to the fast healing of people in the hospital, the hospital doesn’t reach its full capacity again. When people come to the hospital, they heal faster than the people healing at home (See $\mu_1, \mu_2$). These and other responses will be validated and analyzed.
4.2. Validation of Model with help of Queueing Systems

We can validate our model with Queueing Systems [2]. Our model could be an approximation to Machine Repair Problem [15] because the arrival rate is changing with a number of healthy people throughout the simulation. Also, population size is limited. Then, model will be $M/M/N/\infty/N$ [14].

Also, we can model our hospital system as another model. Because it has $K$ beds and the person who couldn't get help from the hospital will not wait for the hospital, he/she will be healed at their home. Thus, the system will be $M/M/k/k$ Queue(Erlang-Loss System) [5].

Let’s first calculate the common values of both systems:

Let $\lambda = 1/300, N = 1582$.

There are three different $\mu$’s in our case, so we have to find the expected value of $\mu$ in order to validate Machine Repair Problem responses. However, the probability of occurring $\mu_1, \mu_2$ depends on the probability of a person being rejected by the hospital, and this probability is dependent on $\lambda_e$ (because the arrival rate of the hospital is $\lambda_e \cdot 0.2$). Therefore, we will look at the edge cases of the probability to calculate an interval for it. If people who need to go to the hospital get accepted by the hospital i.e. capacity is not full:
Let \( R \), the service rate in the hospital, be the arrival rate to the hospital. Then, let \( \lambda = \lambda(N - L) \), where \( L = \frac{1}{\mu} : (1582 - 50.06) = 5.106 \) for \( \mu < 0.102 \).

Case 2: \( \mu = 0.102 \)

1. \( \mu = \frac{1}{E[S_1]} = E[S_1] \cdot 0 + E[S_2] \cdot 0.8 + E[S_3] \cdot 0.2 = \frac{1}{0.2 + 10.03} = 0.102 \)

2. If people who need to go to the hospital get rejected by the hospital i.e. capacity is full: Let \( R \in U[1, 2] \), then since \( \mu_1 \) and \( R \) are independent:

   \[ E[S_3] = E[S_1] \cdot E[R] = 61.5 = 9 \]

Then, \( 0.102 \leq \mu \leq 0.113 \) so we can continue to calculate values for two edge cases of service rate which are 0.102 and 0.113

Case 1: \( \mu = 0.102 \)

\[
\begin{align*}
P_0 & = \left[ \frac{\sum (\pi(S_3)) \cdot (\frac{1}{\mu})^n + \sum (\pi(S_1)) \cdot (\frac{1}{\mu})^n \cdot (\frac{1}{\mu})^n}{\sum (\pi(S_1)) \cdot (\frac{1}{\mu})^n} \right]^{-1} \\
P_1 & = \left[ \frac{\sum (\pi(S_2)) \cdot (\frac{1}{\mu})^n}{\sum (\pi(S_2)) \cdot (\frac{1}{\mu})^n} \right]^{-1} \\
\Rightarrow P_0 & = 5.06 \cdot 10^{-23} \\
P_n & = N(N, n)(\lambda/\mu)^n \cdot P_0 \\
\Rightarrow P_n & = 562(N, n)(\lambda/\mu)^n \cdot 5.06 \cdot 10^{-23} \\
L_{0.102} & = \sum n \cdot P_n = 50.06 \\
\Rightarrow \lambda_c = \lambda(N - L) = \frac{\lambda}{\mu} : (1582 - 50.06) = 5.118 \\
\end{align*}
\]

Let’s now validate the hospital model:

Let \( \lambda_h \) be the arrival rate to the hospital. Then, for \( \lambda_{c, 0.102} = 5.106 \).
\[ \lambda_h = \lambda_{e,0.102} \cdot (0.2) = \lambda_{e,0.102}/5 = \frac{5.118}{5} \approx 1.02 \]  
(7)

for \( \lambda_{e,0.109} = 5.118 \)

\[ \lambda_h = \lambda_{e,0.109} \cdot (0.2) = \lambda_{e,0.109}/5 = \frac{5.118}{5} \approx 1.02 \]  
(8)

1. Long run probability of hospital being empty \( (P_{0,h}) \):

\[ P_{0,h} = \frac{1}{e^{\lambda_h/\mu_1}} = \frac{1}{e^{\lambda_{0.102}/\mu_1}} = 0.0022 \]  
(9)

This matches the simulation result for the long-run probability of the hospital being empty which was 0.0025. Then, the response is validated.

2. Average number of beds occupied in the hospital \( (L_h) \):

\[ W_h = \frac{1}{\mu_1} = 6 \]  
\[ L_h = \lambda_h \cdot W_h = 1.02 \cdot 6 = 6.12 \]  
(10)

This matches the simulation result for the average number of occupied beds in the hospital which was 6.21.

3. Possibility of people who go to the hospital being rejected \( (P_{66}) \):

\[ P_{66} = P_{0,h} \cdot \left( \frac{\lambda_h/\mu_1}{5} \right)^6 = 0.0022 \cdot \left( \frac{1.02}{5} \right)^6 \approx 0 \]  
(11)

Then, the possibility of people who go to the hospital being rejected is 0. Therefore, \( \mu_3 \) is not important in our system, it can be omitted.

In conclusion, \( \mu \) is taken as

\[ \mu = \frac{1}{E[S]} = \frac{1}{E[S_1] + E[S_2] + E[S_3]} = \frac{1}{6} = 0.1666 \]  
(12)

and the effective arrival rate is taken as:

\[ \lambda_e = \lambda(N - L) = \frac{1}{300} \cdot (1582 - 46.85) = 5.118 \]  
(13)

Let’s now validate the rest of the model using the Machine Repair Problem:

The theoretical/calculated values are compared with Figure 9.

1. The average number of sick people in the population \( (L) \) :
We have calculated above that:

\[ L = L_{0,109} = \sum_{n=0}^{1582} n \cdot P_n = 46.85 \]  \hspace{1cm} (14)

This matches the result of the simulation where the mean value of \( L \) was 47.16.

Hence, our model’s average number of sick people is consistent with the theoretical value.

2. Total average sickness time \((W)\):

\[ W = \frac{L}{\lambda_e} = \frac{46.85}{5.118} = 9.15 \]  \hspace{1cm} (15)

This matches the result of the simulation for the total average sickness time which was 9.18.

Hence, our model’s total average sickness time is consistent with the theoretical value.

3. The average proportion of sick people in the population \((\frac{L}{N})\):

Average proportion of sick people in the population = \( \frac{L}{N} = 46.85 \div 1582 = 0.0296 \)

This matches the result of the simulation for the average proportion of sick people in the population which was 0.0298.

Hence, our model’s average proportion of sick people in the population is consistent with the theoretical value.

4.3. Analysis of Model Responses

We combined all data gathered by different settings in Figure 9 to analyze it, and we will analyze it one by one according to different settings.

1. long-run probability of the hospital being empty: -

When we look at the values \( P_{0,h} \) of different settings, all values are small probabilities, and they have a mean 0.002584 (Figure 10). It shows us that in the long run hospital is converging to be empty.

The main reason for that is the arrival rate is much bigger than the service rate. Thus, the hospital is mostly not empty. \( \lambda_e = 5.118, \mu = 0.109, L = 46.85 \) (calculated in 4.2).

Also, we can see the same from Figure 10 that the mean value of the average number of sick people in the population is 47.16. So, it is a low possibility that the system will get close to being empty and so hospital as well.

The standard deviation of the probability is large when compared to the mean value (almost the same). That means the variance of the probability is big. The main reason for that is since the probability value is low, the fluctuation of probability is big in small simulation time. (It is clear in above line plots (3 and 4.) which have simulation time equal to 1000 time units.)

2. the proportion of sick people healing in the hospital:

That response is straightforward to analyze. It has a small standard deviation and is always close to the mean value 0.2. It is expected since the probability of a sick person getting treated in the hospital is given as 0.2. Also, they have the same values because, in all settings, the number of beds in the hospital is not getting full. Therefore, a person who needs to heal in the hospital always has a chance to heal in the hospital. Hence, the final value of response is exactly the same as the defining value of needing treatment in the hospital.

3. the average number of sick people in the population:

We have validated the response above in Section 4.2. The response is not fluctuating so much because the simulation never exceeds the total number of beds in the hospital, and for every setting, it is between 47 and 48, also has a very small standard variation. As a result, because of the service rate and arrival rate and also the number of beds in the hospital, the simulation is reaching its steady-state quickly in every setting. Besides, the simulation run-time is big enough for the simulation to reach its steady-state.

4. the average proportion of sick people in the population:

We have validated the response above in Section 4.2 and it has exactly the same explanation with the average number of sick people in the population.

5. the average number of occupied beds in the hospital

We have validated the response above in Section 4.2. The response is not changing so much with different settings Because of the reason we mentioned above.

6. total average sickness time:

We have validated the response above in Section 4.2. The total average sickness time is equal to service time in our model, and it is generated with exponential distributions \( \mu_1, \mu_2, \mu_3 \). However, as we analyzed in Section 4.2, \( \mu_3 \) is never the case in our model, and we can find the weighted average of service time. Moreover, when we take the multiplicative inverse of the service time, we get average sickness time and as we validated in Section 4.2, it is very close to the theoretical value.
4.4. Conclusion of Numerical Analysis

The simulation is not changing in different settings. Because of four reasons:

1. Simulation time is long enough to get our simulation to steady-state, even 1000 unit time is enough.

2. The total number of beds in the hospital (K) is big enough for every person who needs to heal in the hospital to be able to heal in the hospital.

3. Both arrival rate and service rate are small compared to the total simulation time. Therefore, we couldn’t observe different responses in different simulation time settings.

4. The ratio of arrival rate and service rate is not big enough to reach the total population size. Therefore, total population size is not a constraint for the simulation.

Even though the simulation settings were not so great to observe different simulation responses in different settings, it is validated nicely with help of the Queueing Systems [1].

5. Conclusion

The study’s goal was to create a model to observe the real-time behavior of a system in which each person can get sick and go to the hospital/stay home for healing. This model is simulated under different starting conditions and it is observed that overall results don’t fluctuate from each other. The system reaches its steady-state in each simulation time and it is impossible for the system to fill the hospital during the run time since the ratio of arrival rates and service rates are relatively small compared to the number of beds in the hospital. Under a smaller hospital capacity or smaller simulation run time or higher arrival rates and lower healing rates, the results of the study would be different. Under these conditions, the model responses in each initial hospital status would also look more different than in our study. In the future, one can model a similar system with different initial conditions to get a variety of model responses.

References

[1] Ivo Adan and Jacques Resing. Queueing theory, 2002. 8
[2] Arnold O Allen. Probability, statistics, and queueing theory. Academic press, 2014. 4
[3] Yaman Barlas and Onur Özgün. Queueing systems in a feedback environment: Continuous versus discrete-event simulation. Journal of Simulation, 12(2):144–161, 2018. 1
[4] Thomas Bartz-Beielstein, Frederik Rehbach, Olaf Mersmann, and Eva Bartz. Hospital capacity planning using discrete event simulation under special consideration of the covid-19 pandemic, 2020. 1
[5] Agner Krarup Erlang. The theory of probabilities and telephone conversations. Nyt. Tidsskr. Mat. Ser. B, 20:33–39, 1909. 1, 4
[6] G.S. Fishman. Discrete-Event Simulation: Modeling, Programming, and Analysis. Springer Series in Operations Research and Financial Engineering. Springer New York, 2013. 1
[7] Weng Hong Fun, Ee Hong Tan, Ruzelan Khalid, Sondi Sararaks, Kar Foong Tang, Iqbal Ab Rahim, Shakirah Md. Sharif, Suhana Jawahir, Raoul Muhammad Yusof Sibert, and Mohd Kamal Mohd Nawawi. Applying discrete event simulation to reduce patient wait times and crowding: The case of a specialist outpatient clinic with dual practice system. Healthcare, 10(2), 2022. 1
[8] Jules Le Lay, Vincent Augusto, Xiaolan Xie, Edgar Alfonso-Lizarazo, Bienvenu Bongue, Thomas Celarier, Regis Gonthier, and Malek Masmoudi. Impact of covid-19 epidemics on bed requirements in a healthcare center using data-driven discrete-event simulation. In 2020 Winter Simulation Conference (WSC), pages 771–781, 2020. 1
[9] Albert W Marshall and Ingram Olkin. A multivariate exponential distribution. Journal of the American Statistical Association, 62(317):30–44, 1967. 3
[10] Norm Matloff. Introduction to discrete-event simulation and the simpy language. Davis, CA. Dept of Computer Science. University of California at Davis. Retrieved on August 2(2009):1–33, 2008. 2
[11] M.L. Penn, T. Monks, A.A. Kazmierska, and M.R.A.R. Alkoheji. Towards generic modelling of hospital wards: Reuse and redevelopment of simple models. Journal of Simulation, 14(2):107–118, 2020. 1
[12] A Pósfalvi and J Sztrik. A numerical approach to the repairman problem with two different types of machine. Journal of the Operational Research Society, 40(9):797–803, 1989. 1
[13] Michael Saidani, Harrison Kim, and Jinju Kim. Designing optimal covid-19 testing stations locally: A discrete event simulation model applied on a university campus. PLOS ONE, 16(6):1–16, 06 2021. 1
[14] J Sztrik. On the finite-source g/m/r queue. European Journal of Operational Research, 20(2):261–268, 1985. 4
[15] J Sztrik. The¡ g/m/r/fifo¿ machine-interference model with state-dependent speeds. Journal of the Operational Research Society, 39(2):201–207, 1988. 4
[16] J.A. White. Analysis of Queueing Systems. Elsevier Science, 2012. 1