HW4

Blake Conrad

06/30/2019

Contents

[Technical Report 3](#_Toc12816203)

[Introduction 3](#_Toc12816204)

[M/M/1 Queue 3](#_Toc12816205)

[Arrival 3](#_Toc12816206)

[Service 3](#_Toc12816207)

[Hypothesis Test: T-test 3](#_Toc12816208)

[Confidence Intervals 4](#_Toc12816209)

[Warmup, Batch Means, Regenerative Cycles, Antithetic Pairing 4](#_Toc12816210)

[G/G/1 Queue 5](#_Toc12816211)

[Hypothesis Test: T-test 6](#_Toc12816212)

[Confidence Intervals 6](#_Toc12816213)

[Conclusion 6](#_Toc12816214)

# Technical Report

## Introduction

In this project we have been asked to explore various aspects of simulation and output analysis. This is all using a 1 server queue. The following assumptions are made for the questions proposed throughout this analysis:

* For any arrival create a linear congruential generator for your random number generator. Let the prime numbers a=100801, c=103319 and m=193723. As a starting seed, use 50,001 for the arrivals. All simulations should start this way.
* For any service time create a congruential generator for your random number generator. Let the prime numbers a=7,000,313 and m=9,004,091. As a starting seed, use 94,907. All simulations should start this way.

We begin by assessing a M/M/1 Queue with a variety of statistical tests, performance metrics, and cycle detection techniques. Next, we go into a G/G/1 Queue to discuss similar aspects. After we complete both assessments, we explore the distinctions between these distributions by discussing their trade-offs, strengths, weaknesses, and the importance to variance in this trade-off and in simulation in general.

## M/M/1 Queue

We begin by creating an M/M/1 Queue with the following seed construction and rates for arrival and service distributions:

### Arrival

* For any arrival create a linear congruential generator for your random number generator. Let the prime numbers a=100801, c=103319 and m=193723. As a starting seed, use 50,001 for the arrivals. All simulations should start this way.
* Arrival rate: 1/3

### Service

* For any service time create a congruential generator for your random number generator. Let the prime numbers a=7,000,313 and m=9,004,091. As a starting seed, use 94,907. All simulations should start this way.
* Service rate: ½

We can quickly see that the rate we arrive is faster than the rate which we serve. In addition, there will likely exist a line because of the capacity constraint of 1 in the nature of M/M/1. We run a 30-replication simulation for 500 hours and acquire the following statistics on utilization.

### Hypothesis Test: T-test

We see the population, or theoretical mean, defined by formula:

.

We see from our sample Utilization from 30 replication average of 500-time stamp runs, results in:

.

We construct a statistical hypothesis test on this to determine if we have enough evidence to claim that the mean from the simulation matches that of the theory.

After running a T-test on this to determine the statistical significance we find the resulting: . This means that we do not suspect the mean of our simulation to be the same as the theoretical. This could be to a number of reasons. One, the seed could be off. Two, the random number generator might have launched a batch of tail-skewed numbers. Third, integrating different style generators for arrivals and departures could cause a distinction as well.

### Confidence Intervals

We next seek to know a confidence interval on the mean utilization. We also seek to validate our results by running multiple experiments and seeing how many new 30 replication simulations of 500 hours will produce means within the interval proposed 95% of the time. We set and run the test. The following is produced:

Repeat this entire process for 40 repetitions (1200 runs) and find all 95% confidence intervals, keeping track of mean utilization. We find that 96% of the time the theoretical mean was not captured in this interval. However, we do see 95% of the time is indeed in the interval proposed. We conclude that we have evidence to believe the interval produced does not contain the theoretical mean and our simulation is different enough to not be considered the same.

### Warmup, Batch Means, Regenerative Cycles, Antithetic Pairing

We repeat this procedure for three different circumstances to see if any change occurs. The following results are acquired:

* Warm Up (100 hours)
* Batch Means
* Regenerative Cycles

In addition to many other analytical findings to following, we see that only in the case of batch means do we see a result that is statistically significant with evidence in support of the true theoretical mean utilization for a M/M/1 Queue. The antithetic variates method reduces the variance of the simulation results. We run antithetic pairing on our original 30 replication simulation and see that the confidence intervals are very similar:

We see that the variance is reduced by about 10% after running antithetic pairing. In addition, the percentage of times that the result the true mean was found to be within the interval was brought to 67% of the time by introducing batch means. So we see moderate amounts of success from each of the techniques run on the original problem. After 1000 regenerative cycles and report the average time and confidence interval of the length of the regenerative cycle are listed below:

## G/G/1 Queue

1. Create a G/G/1 queue with an arrival distribution of 3/64 x2 between 0 and 4 and 0 else.

Run this for 30 replications each for 500 hours. You can either do this for utilization or expected number of people in line. Note it’s mean is 3. Let the service distribution be a normal (2,.25). You need to generate these and cannot have software generate them for you.

1. Use seeds 50001 for arrivals and 94907 for service times. Perform a t-Test to determine whether or not there is a statistical difference between the simulated data and 1 a. Also find the confidence interval.
2. Repeat this entire process for 40 repetitions and find all 95% confidence intervals. Determine how many of these 95% confidence intervals do not contain the theoretical mean from part 1. Observe that both 1 and 2 have the same average performance.
3. Use CRN as a VRT (revert to previous seeds). Compare the variances between 1a and 2a to determine if there is variance reduction.
4. Run this simulation for 1000 regenerative cycles and report the average time and confidence interval of the regenerative cycle.

We begin by creating an M/M/1 Queue with the following seed construction and rates for arrival and service distributions:

We begin by creating an G/G/1 Queue with the seed construction and rates for arrival and service distributions provided in the instructions.

We can quickly see that the rate we arrive is faster than the rate which we serve. In addition, there will likely exist a line because of the capacity constraint of 1 in the nature of G/G/1. We run a 30-replication simulation for 500 hours and acquire the following statistics on utilization.

### Hypothesis Test: T-test

We see the population, or theoretical mean, defined by formula:

.

We see from our sample Utilization from 30 replication average of 500-time stamp runs, results in:

.

We construct a statistical hypothesis test on this to determine if we have enough evidence to claim that the mean from the simulation matches that of the theory.

After running a T-test on this to determine the statistical significance we find the resulting: . This means that we do not suspect the mean of our simulation to be the same as the theoretical. This could be to several reasons. One, the seed could be off. Two, the random number generator might have launched a batch of tail-skewed numbers. Third, integrating different style generators for arrivals and departures could cause a distinction as well.

### Confidence Intervals

We next seek to know a confidence interval on the mean utilization. We also seek to validate our results by running multiple experiments and seeing how many new 30 replication simulations of 500 hours will produce means within the interval proposed 95% of the time. We set and run the test. The following is produced:

Repeat this entire process for 40 repetitions (1200 runs) and find all 95% confidence intervals, keeping track of mean utilization. We find that 96% of the time the theoretical mean was not captured in this interval. However, we do see 95% of the time is indeed in the interval proposed. We conclude that we have evidence to believe the interval produced does not contain the theoretical mean and our simulation is different enough to not be considered the same.

## Conclusion

Both M/M/1 and G/G/1 produce comparable results in terms of utilization. The data shows that the utilization is higher under a gamma distribution than an exponential. We see the variance of M/M/1 after 30 replications of length 500 hours produced: , where under the G/G/1: . Both the markov and gamma simulation models reject the null hypothesis of being the same as their respective theoretical mean utilization. This is justified in the variation of the models. The importance of variance is incredible, as it can cause a theoretical value to not match up with a simulated model. If the variances are minimized across each distribution, we would expect to see the simulated values of utilization to approach their theoretical counterpart. This could be done by using more variance reduction techniques, simulating more data, or finding a particularly good seed and random number generator. Variance in simulations is incredibly important, and can entirely change the output and confidence in what is analyzed overall for holistic decision making.