Queueing System of the Wellington Railway Station Ticket Booth

Group 1: Chad Kakau, Gale Bueno

Contents

Wellington Railway Station: an observation of a queueing system in action

Introduction

Queuing is a feature of contemporary living that we often encounter. Queuing arises whenever a system must be used for service by several jobs or customers, whether it be at restaurants or through the use of modern technology. Both the arrival and the service timings are assumed to be at random and independent. The study of queuing is significant because it provides both a theoretical framework for the service that we would anticipate from a system (e.g. Erlang, 1909) and how the system is constructed to provide some defined level of service. Simple queuing systems are characterized by naming (a) the arrival pattern, (b) the service mechanism, and (c) the queue discipline (Kendall, 1953;). The objective of this study is to observe a queuing system in action and prepare theoretical and simulation models that are used to estimate the impact of adjustments in the queue arrangement.

Erlang, A. K. 1909. “The theory of probabilities and telephone conversations”. Nyt Tidsskrift for Matematik. Vol, 20 (1909), p 33.

Kendall, D. G. (1953), “Stochastic processes occurring in the theory of queues and their analysis by the method of the embedded Markov chain”. The Annals of Mathematic Statistics. 24(3): p338-354.

Queuing system in observation

Group 1 collected data from Wellington Train Station the terminal station for Wellington city and the southern-most terminal of passenger trains in the North Island of New Zealand. The queuing system in observation is the train station passenger ticketing booths. The ticketing booths provide sales and services for the public and passengers on the trains. The queuing system is located adjacent to a supermarket, cafe, bar, and is en-route between the passenger platforms and the Wellington City main bus terminal. Customers can enter the queuing system:

- after disembarking a passenger train, or
- as pedestrian traffic from several street level and underground walkways

Methodology

The project adopted the following methodology for completing the research project:
1. secure permission to access and observe the system
2. observe and collect data on-site
3. convert data for analysis
4. clean and analyse data to establish baseline parameters
5. determine a best-fit distribution for the observed data 6. develop simulations based on theoretical, best-fit, and empirical models
7. report results of the research as presentation and in written report.

Permission to observe the system in action

Permission to collect data at the Wellington Railway station was granted by Kiwi Rail, with initial observation from the 11th to the 13th of April 2022. Data collection dates were extended, with permission, in order to collect enough data. Permission was granted on condition that health and safety guidelines were met. The project team provided a COVID-19 risk management plan as a further condition of the permission. Final date of observation was on the 27th of April 2022.

The Wellington train station system in observation

The Wellington train station ticketing booths form a M/M/C system, where C (the number of service channels) ranges between 3 and 4. This system has a first-come, first-serve (FCFS) discipline, where servers pull customers from one queue. Data was collected over five sessions, 3 morning sessions (7:00 am to 8:30am) and 2 afternoon sessions (4:30 pm - 5:00 pm) across five separate dates. Each observation records the time in the system which includes:

- time of arrival in the system,
- time of entering service,and
- and time of departing service,

For an immediate arrival to the system with no queue present, the service time is calculated (based on service start = arrival in system, and end of service = exit from system). In the presence of a queue, time of arrival in the system was explicitly observed, in addition to service start time (departure from the queue), and end of service time. To make collection of data easier, the lane of service was included in the collection. This allows observers to collect data quicker during peak hours. The lane number is essential, as the end of service time of a customer, will be the start time of the next customer who is served in the same lane, when a queue exists. Data was recorded using pen and paper.

One morning session failed to correctly collect data based on a M/M/C system, therefore was discarded from the final data, resulting in four sessions. A total of 403 observations were obtained. The data collected was manually transcribed and input into a csv file. Inter-arrival (time between each arrival in seconds) and service times (time spent being served in seconds) is calculated.

Initial data analysis to establish baseline parameters

The collected data was analysed from the csv file into R Studio, and from these the mean inter-arrival rate (λ) and service rate (μ) are calculated. Inter-arrival rate is calculated as the sum of all inter-arrival times divided by the total number of inter-arrival times. Service time is calculated as the mean time a customer spends in service (i.e. does not include queuing time). The following sections include R-code and outputs, although not all code-chunks are output in the knitted document.

The raw data had two outlier observations which were removed, leaving the following structure:

Inter-arrival Time

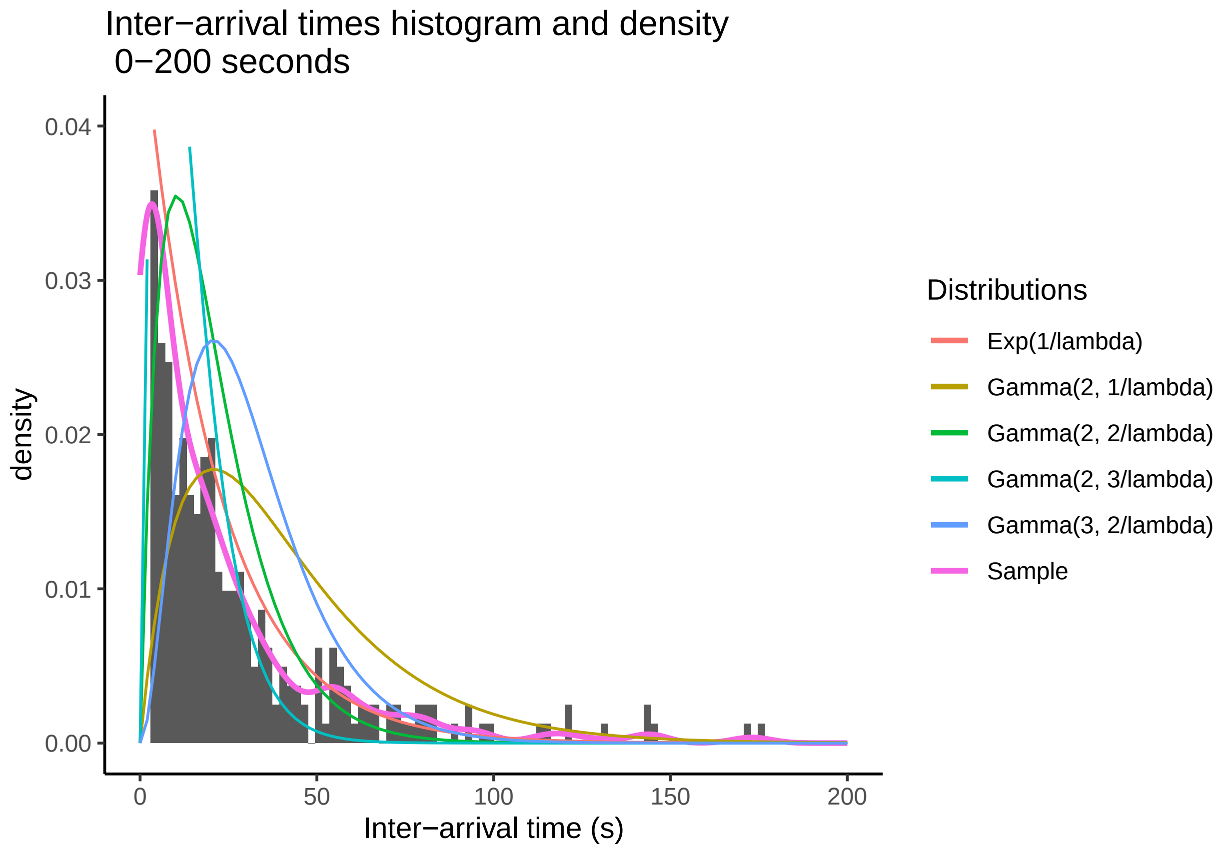
Inter-arrival rate is calculated as the sum of all inter-arrival times divided by the number of inter-arrival times:

```r
# estimate parameters for density plot, assuming exponential distribution
lamb <- mean(train.data$inter.arr.sec)
lamb
```

```
## [1] 20.73317
```

Plotting the inter arrival times

The inter-arrival times are plotted here, showing a histogram of the data, density function of the actual data and the theoretical exponential distribution, assuming $\lambda = 20.7331671$.



Goodness of fit

The inter-arrival time visually looks like a good fit to the exponential distribution, so we will begin by running a goodness of fit test for the data and set our rejection level to $\alpha = 0.05$. The chi-squared goodness of fit test was conducted with data sorted into 100 bins.

Results of the chi-square test:

Goodness of fit tests were conducted on five potential distributions, with the results presented in more detail, but summarised in the below table.

Exponential(1/23), Gamma(2, 1/23, Gamma(2, 2/23), Gamma(2, 3/23, Gamma(3, 2/23, 0.069, 0.08, 0.973, 0.973, 0.973

Testing for $Exp(\frac{1}{\lambda})$ distribution: H_0 : The data is consistent with the $Exp(\frac{1}{\lambda})$ distribution
H_1 : The data is not consistent with the $Exp(\frac{1}{\lambda})$ distribution

Since $p = 0.0693151$ is greater than 0.05 we fail to reject the null hypothesis and conclude that the inter-arrival time is consistent with data from an exponential distribution with the $Exp(\frac{1}{\lambda})$ distribution.

Testing for $Gamma(2, \frac{1}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(2, \frac{1}{\lambda})$ distribution H_1 : The data is not consistent with the $Gamma(2, \frac{1}{\lambda})$ distribution

Since $p = 0.0796664$, is greater than 0.05 we fail to reject the null hypothesis and conclude that the inter-arrival time is consistent with the $Gamma(2, \frac{1}{\lambda})$ distribution.

Testing for $Gamma(2, \frac{2}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(2, \frac{2}{\lambda})$ distribution H_1 : The data is not consistent with the $Gamma(2, \frac{2}{\lambda})$ distribution

Since $p = 0.9729881$, is greater than 0.05 we fail to reject the null hypothesis and conclude that the inter-arrival time is consistent with the $Gamma(2, \frac{2}{\lambda})$ distribution.

Testing for $Gamma(2, \frac{3}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(2, \frac{3}{\lambda})$ distribution H_1 : The data is not consistent with the $Gamma(2, \frac{3}{\lambda})$ distribution

Since $p = 7.5161203 \times 10^{-6}$, is less than 0.05 we reject the null hypothesis and conclude that the inter-arrival time is not consistent with the $Gamma(2, \frac{3}{\lambda})$ distribution.

Testing for $Gamma(3, \frac{2}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(3, \frac{2}{\lambda})$ distribution H_1 : The data is not consistent with the $Gamma(3, \frac{2}{\lambda})$ distribution

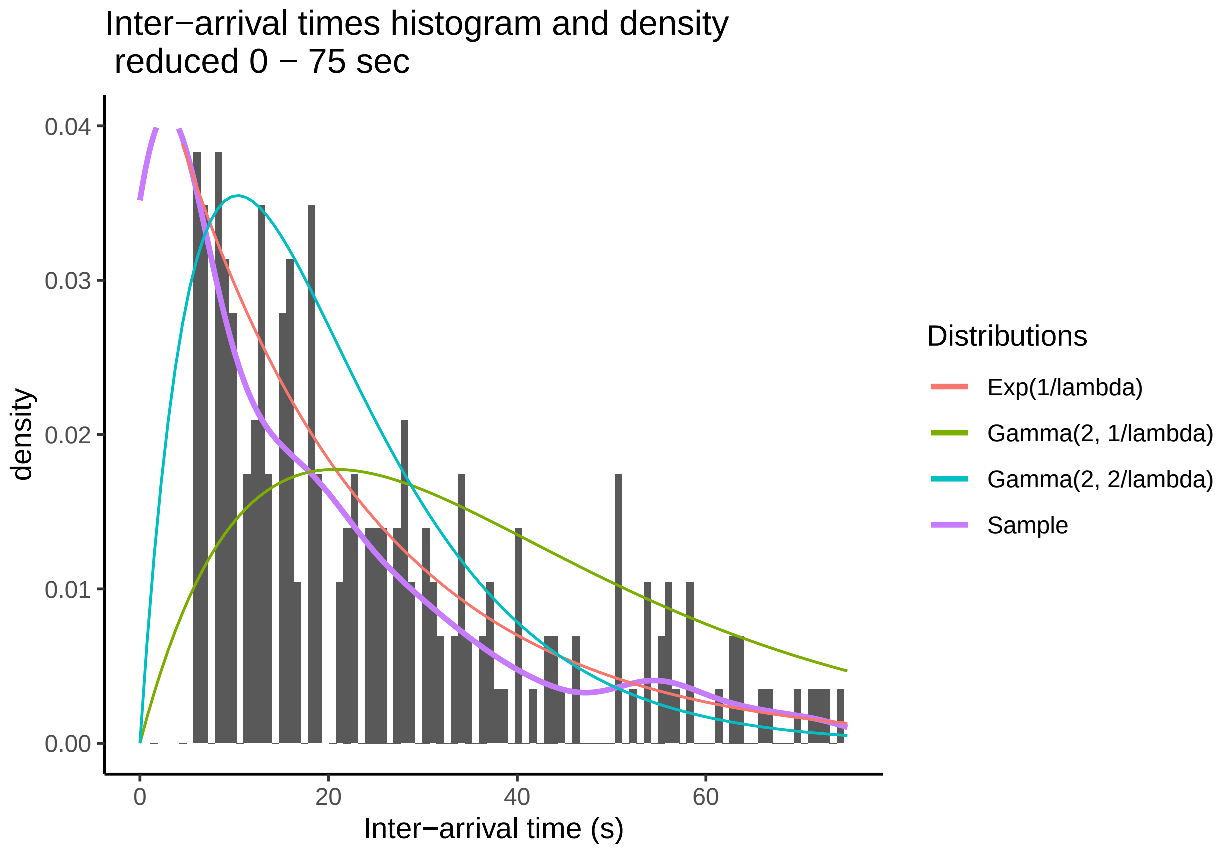
Since $p = 0.9729881$, is greater than 0.05 we fail to reject the null hypothesis and conclude that the inter-arrival time is not consistent with the $Gamma(3, \frac{2}{\lambda})$ distribution.

The highest p-value is 0.973. Since the goodness of fit test provided three high-scoring distributions, we select one arbitrarily:

$Gamma(2, \frac{2}{\lambda})$

Fitting a distribution

Initial visual inspection of the plot shows that the data may fit a Gamma distribution with shape = 2, rate = $\frac{2}{\lambda}$. This distribution is equivalent to an Erlang distribution with shape = 2, rate = $\frac{2}{\lambda}$.



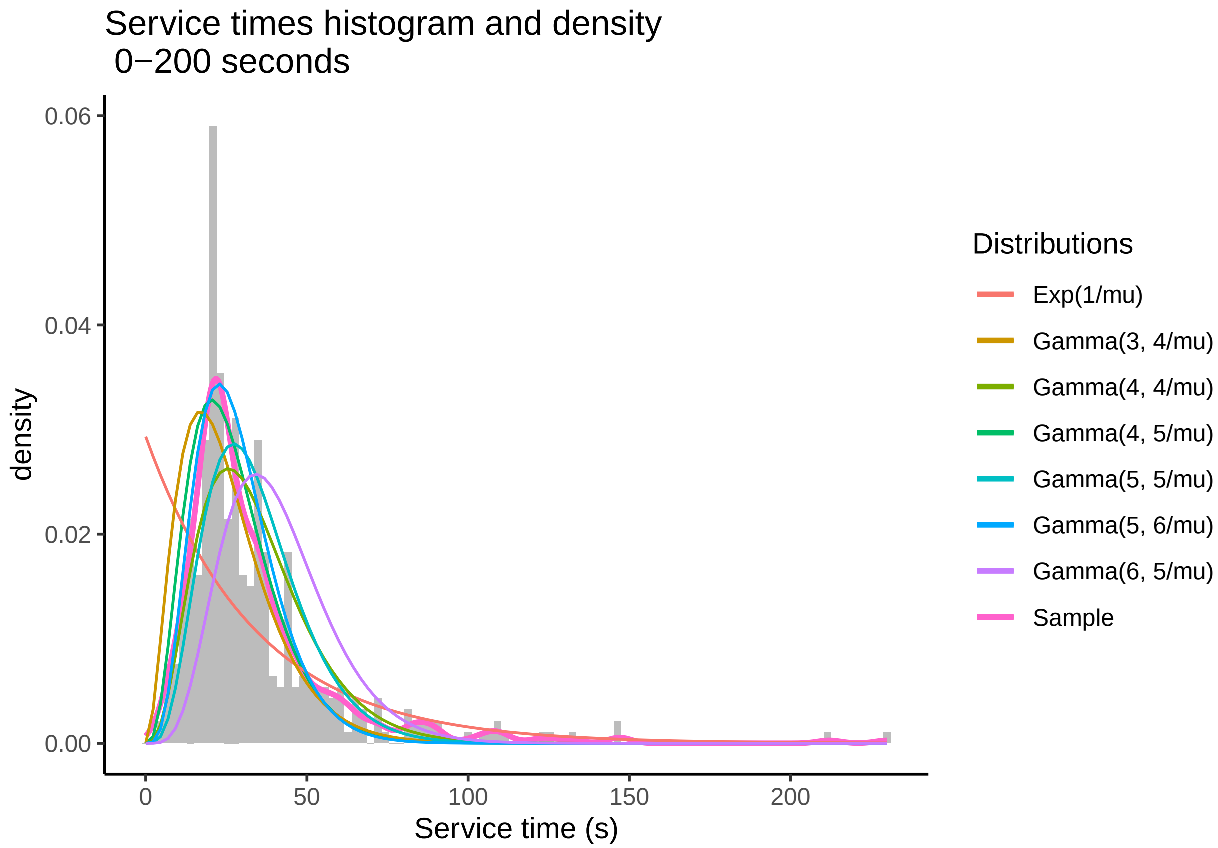
Service Time

The service time is calculated as the mean time a customer spends in service (i.e. does not include queuing time).

```
# calculate mean service time as mu
mu <- mean(train.data$Serv_time_sec)
mu
```

```
## [1] 34.09975
```

The service time is plotted here, showing a histogram of the service times, density plot of the actual data and theoretical density of the exponential distribution with $\mu = 34.0997506$.



Visually the service times look like they could be from a gamma distribution but it's difficult to fit a plot of the gamma density function to the scale of the histogram.

Goodness of fit

The service time visually looks like a possible fit to the exponential distribution, so we will begin by running a goodness of fit test for the data and set our rejection level to $\alpha = 0.05$.

H_0 : The data is consistent with the Exponential distribution
H_1 : The data is not consistent with the Exponential distribution

Results of the chi-square test:

A total of seven potential distributions were tested, with detailed results presented below, and summarised here:

Exp(1/23), Gamma(3, 4/34), Gamma(4, 4/34), Gamma(4, 5/34), Gamma(5, 5/34), Gamma(5, 6/34), Gamma(6, 5/34), 0.997, 0.003, 0.579, 0, 0.625, 0.544, 0

Testing for $Exp(\frac{1}{\mu})$ distribution: H_0 : The data is consistent with the $Exp(\frac{1}{\mu})$ distribution
H_1 : The data is not consistent with the $Exp(\frac{1}{\mu})$ distribution

Since $p = 0.9974454$ is greater than 0.05 we fail to reject the null hypothesis and conclude that the inter-arrival time is consistent with data from an exponential distribution with the $Exp(\frac{1}{\mu})$ distribution.

Testing for $Gamma(3, \frac{4}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(3, \frac{4}{\lambda})$ distribution
H_1 : The data is not consistent with the $Gamma(3, \frac{4}{\lambda})$ distribution

Since $p = 0.00279$, is less than 0.05 we reject the null hypothesis and conclude that the inter-arrival time is not consistent with the $Gamma(3, \frac{4}{\lambda})$ distribution.

Testing for $Gamma(4, \frac{4}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(4, \frac{4}{\lambda})$ distribution
H_1 : The data is not consistent with the $Gamma(4, \frac{4}{\lambda})$ distribution

Since $p = 0.5790907$, is greater than 0.05 we fail to reject the null hypothesis and conclude that the inter-arrival time is consistent with the $Gamma(4, \frac{4}{\lambda})$ distribution.

Testing for $Gamma(4, \frac{5}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(4, \frac{5}{\lambda})$ distribution
H_1 : The data is not consistent with the $Gamma(4, \frac{5}{\lambda})$ distribution

Since $p = 1.7898888 \times 10^{-5}$, is less than 0.05 we reject the null hypothesis and conclude that the inter-arrival time is not consistent with the $Gamma(4, \frac{5}{\lambda})$ distribution.

Testing for $Gamma(5, \frac{5}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(5, \frac{5}{\lambda})$ distribution
H_1 : The data is not consistent with the $Gamma(5, \frac{5}{\lambda})$ distribution

Since $p = 0.6253453$, is greater than 0.05 we fail to reject the null hypothesis and conclude that the inter-arrival time is consistent with the $Gamma(5, \frac{5}{\lambda})$ distribution.

Testing for $Gamma(5, \frac{6}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(5, \frac{6}{\lambda})$ distribution
H_1 : The data is not consistent with the $Gamma(5, \frac{6}{\lambda})$ distribution

Since $p = 0.5439745$, is greater than 0.05 we fail to reject the null hypothesis and conclude that the inter-arrival time is consistent with the $Gamma(5, \frac{6}{\lambda})$ distribution.

Testing for $Gamma(6, \frac{5}{\lambda})$ distribution: H_0 : The data is consistent with the $Gamma(6, \frac{5}{\lambda})$ distribution
H_1 : The data is not consistent with the $Gamma(6, \frac{5}{\lambda})$ distribution

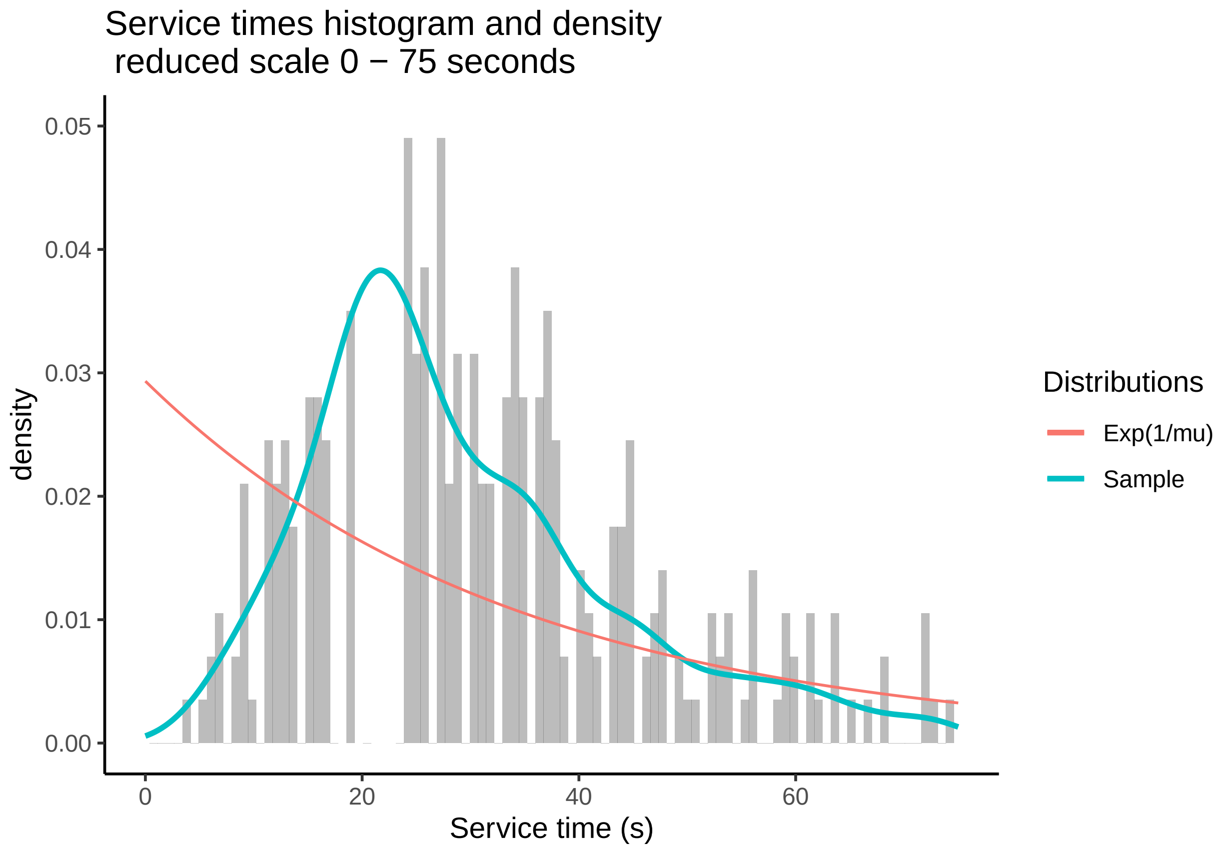
Since $p = 6.3339983 \times 10^{-6}$, is less than 0.05 we reject the null hypothesis and conclude that the inter-arrival time is not consistent with the $Gamma(6, \frac{5}{\lambda})$ distribution.

The highest p-value is 0.9974454

so we select the distribution:

$Exp(\frac{1}{mu})$

Plotting the $Exp(\frac{1}{mu})$ distribution against the data



Best-fit distributions

Based on this initial analysis we conclude that our M/M/4 queuing system has the following distributions:

1. Inter-arrival times: $Gamma(2, \frac{2}{\lambda} == Erlang(2, \frac{2}{\lambda}$
2. Service times: $Exp(\frac{1}{\mu}$

Simulating the system against M/M/4, Best-fit, and Empirical models

Using the parameters estimated from the empirical data we can calculate baseline performance measures as follows. Given:

$\lambda = \frac{1}{23}$ or one customer arriving every 23 seconds
$\mu = \frac{1}{34}$ or each customer taking 34 seconds in service $c = 4$ servers, we calculate baseline performance measures as:
$\frac{\lambda}{\mu} = \frac{34}{23} = 1.4783$
$\rho = \frac{\lambda}{c\mu} = 0.3699$

Average number of customers in the system: L

$$L = L_q + \frac{\lambda}{\mu}$$
$$where$$
$$L_q = \frac{(\frac{\lambda}{\mu})^{c+1}}{(c-1)!(c-\frac{\lambda}{\mu})^2}$$
$$= \frac{(1.4783)^5}{6(4-1.4783)^2}$$
$$= \frac{7.060}{38.1538}$$
$$= 0.1850$$
$$so$$
$$L = 0.1850 + 1.4783$$
$$= 1.6633$$

Average time in the system: W

$$W = \frac{L}{\lambda}$$
$$= \frac{1.6633}{\frac{1}{23}}$$
$$= 38.2559$$

Average server utilisation: B

$$B = \rho$$
$$= \frac{\lambda}{c\mu}$$
$$= 0.3699$$

Resulting in estimates for our baseline performance measures:

L = 1.66 customers in the system W = 38 seconds in the system B = 37% server utilisation

We will use these baseline performance measures to compare against simulations based on three different models. Each simulation consists and is generated using SimPy:

$N = 10000$ variates per simulation
$r = 50$ replications
$c = 4$
$\lambda = \frac{1}{23}$
$\mu = \frac{1}{34}$

Model 1: M/M/4 - $Exp(\frac{1}{\lambda}), Exp(\frac{1}{\mu})$

Model 1 establishes a simulation using the parameters estimated from the observed data:

- $\lambda = \frac{1}{23}$
- $\mu = \frac{1}{34}$

```r
base_L <- 1.66
base_W <- 38
base_B <- .37

m1_sim <- source_python("MM4_simulation.py")
```

Comparing the baseline values for M/M/4:

$L = 1.66$ customers in the system on average $W = 38$ seconds for a customer in the system $B = 0.37$ utilisation

With the simulation outputs for M1

$L = 1.23$ customers in the system, is slightly lower than the estimate. $W = 34.93$ seconds for a customer in the system, is lower than the estimate. $B = 0.89$ utilisation of servers, is much higher, more than twice the estimate.

Model 2: Best-fit distribution - $Gamma(2, \frac{2}{\lambda})$, $Exp(\frac{1}{\mu})$

Model 2 develops a simulation based on the best-fit distribution established earlier:

- inter-arrival times: $Gamma(2, \frac{2}{\lambda})$
- service times: $Exp(\frac{1}{\mu})$

```r
m2_sim <- source_python("Best_fit_simulation.py")
```

Comparing the baseline estimates:

$L = 1.66$ customers in the system on average $W = 38$ seconds for a customer in the system $B = 0.37$ utilisation

With the simulation outputs for M2 (Best-fit distribution)

$L = 0.37$ customers in the system, is much lower, less than a quarter of the estimate. $W = 34$ seconds for a customer in the system, is lower than the estimate. $B = 0.33$ utilisation of servers, is very close to the estimate.

Model 3 - Empirical distribution based on random variates sampled from observed data

Model 3 develops a simulatio based on the observed inter-arrival times and service times.

```r
m3_sim <- source_python("Empirical_simulation.py")
```

Comparing the baseline estimates:

$L = 1.66$ customers in the system on average $W = 38$ seconds for a customer in the system $B = 0.37$ utilisation

With the simulation outputs for M3 (Empirical distribution)

$L = 0.16$ customers in the system, is much lower, less than a tenth of the estimate. $W = 35$ seconds for a customer in the system, is lower than the estimate. $B = 0.07$ utilisation of servers, is much lower, around a fifth of the estimate.

Discussion of the simulation results

None of the models simulated was consistent across all three performance measures The results of the simulations, while not exactly the same as the theoretic estimates, are not so different that they warrant being immediately discounted. In fact, considering the area of study - simulation of stochastic models - simulating systems that are subject to high-levels of unpredictability at the level of individual observation, it is reassuring that the simulations deviated from the estimates. Had any particular model performed very closely to the estimate, the question would become whether the simulation truly offered the level of unpredictable variance that is necessary for an effective stochastic simulation.

. . . which indicates simulations can reproduce realworld variance Each model produced some performance measures that were close to baseline estimates, but no single model performed well across all three measures. Each model used different underlying distributions to draw variates for insertion into the simulation, so there should definitely be difference between the models. M1 and M2 shared the same service time distribution $Exp(\frac{1}{\mu})$ so time in the system and utilisation rate were predictably similar. In the absence of significant queueing, both of these measures are driven by service time. But it is clear that the distribution chosen to establish a model is very important and in fact drives much of the performance, especially in a simple system such as this.

Is the variance a sign of a good simulator, or a not so great implementation Comparing each model against baseline was useful, but with low experience and exposure to the simulation process and software, the research team questioned whether there were errors in implementation, calculation of parameters, or if the variance in the system represented real world unpredictability. Considering each performance measure is estimated from 500,000 individual observations, the latter seems likely. . . but only in so far that the simulations are driven by the selected parameters.

The simulation tools used are effective One of the initial questions posed as part of the research was whether the available tools were sufficient and the researcher concluded that the tools are indeed sufficient. Variance between and within models, (as evidenced by varying performance measures across 50 replications), demonstrates that it is possible to simulate a stochastic system. One of the key learnings though, is that theory provides a useful framework to give shape to the world and to help formulate, not just the answers, but more importantly, a way to ask the right questions.

Conclusion

The research team has reached three key conclusions:

1. the tools developed during the course of this project are very well suited to observing and simulating real-world queuing systems
2. there is tangible value in being able to share work in an online space to help reduce the barriers of time and space; and to reduce inefficiencies and duplication of effort, but
3. this needs to be tempered with regular contact and engagement between collaborators, because it is possible that development in isolation inhibits collective goals.

During the research we learn't about ourselves, the challenges of conducting real-world research (even for something as simple as standing and counting people arriving and departing from a public system), about the nuances of two programming languages and a number of functional packages within those languages, about working with reticent stakeholders, and about pushing the boundaries of our own development. It was definitely challenging, but looking at the culmination of the work, and recalling the systemic build-up and execution of tasks, it has also been incredibly rewarding.

This research and supporting documentation (including code, data, printouts, and future upgrades) is available online (https://github.com/ctkakau/train-station-queueing) . The researchers encourage future investigation of the Wellington train station ticketing counter, and welcome any input for improving this research or supporting material.