Modelling Queues

Dr Sarah Marshall
Department of Information Systems and Operations Management
University of Auckland Business School, New Zealand
sarah.marshall@auckland.ac.nz

4th December 2023
Statistics Teachers Day
University of Auckland

Introduction

Tēnā koutou katoa	Greetings to you all
Ko Ingarangi a Kōtirana te whakapaparanga mai	English and Scottish is my ancestry
engari	however,
Ko Kaitaia te whenua tupu	Kaitaia is where I was born
Nō Napier au	I am from Napier
Kei Tāmaki Makaurau au e noho ana	I am living in Auckland
He pouako matua au i He Manga Tauhokohoko	I am a Senior Lecturer at University of Auckland Business Schoo
Ko Sarah Marshall au	My name is Sarah Marshall
Tēnā tātou katoa	Greetings to one and all

A bit about me ...

- Grew up in Napier
- Studied Conjoint BSc/BCA and MSc at Victoria University of Wellington
- Studied PhD at University of Edinburgh, UK
- AUT 2014 - 2023
- UOA since October 2023
- I love using maths to model systems
- I teach **Business Analytics** and **Operations and Supply Chain Management**

Overview

- Introduction to modelling queues
- Task 1
- Task 2

Who has been stuck in this kind of queue?

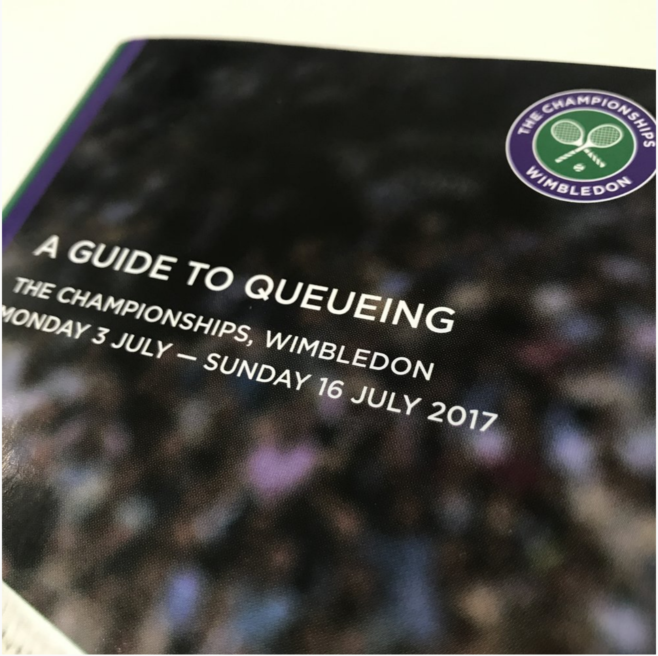
http://www.geograph.ie/photo/3414308

.... or this kind of queue?

Source: Sarah's Phone

.... or this kind of queue?

http://en.wikipedia.org/wiki/File:Waiting_in_line_at_a_food_store.JPG

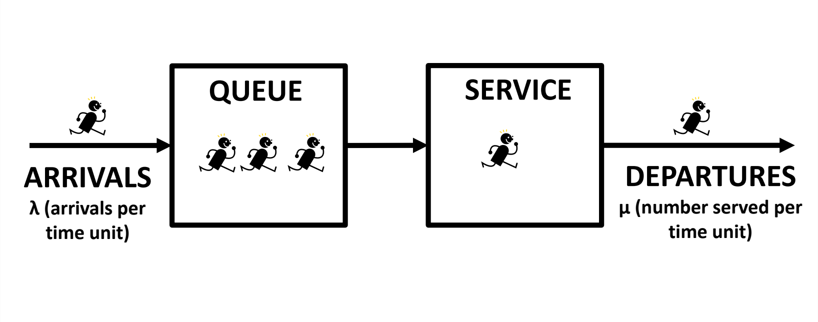


https://static1.squarespace.com/static/59d639183e00be6ee8e5633e/t/

How can we describe this queue?

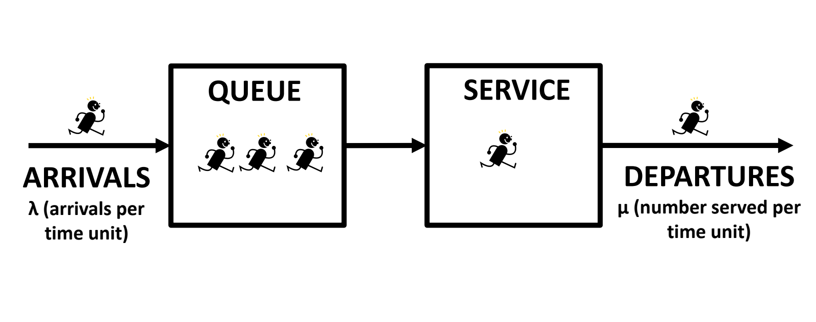
http://en.wikipedia.org/wiki/File:Waiting_in_line_at_a_food_store.JPG

Modelling a Queue



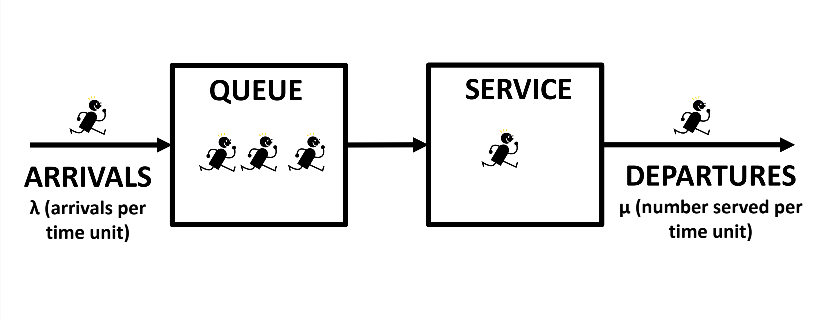
- Arrival Process
  - Probability distribution for time between customer arrivals (interarrival time)

Modelling a Queue



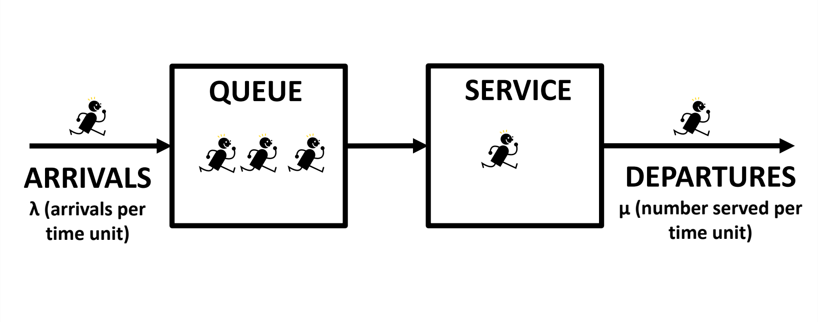
- Service Process
    - Number of Servers
    - Probability distribution for service time

Modelling a Queue



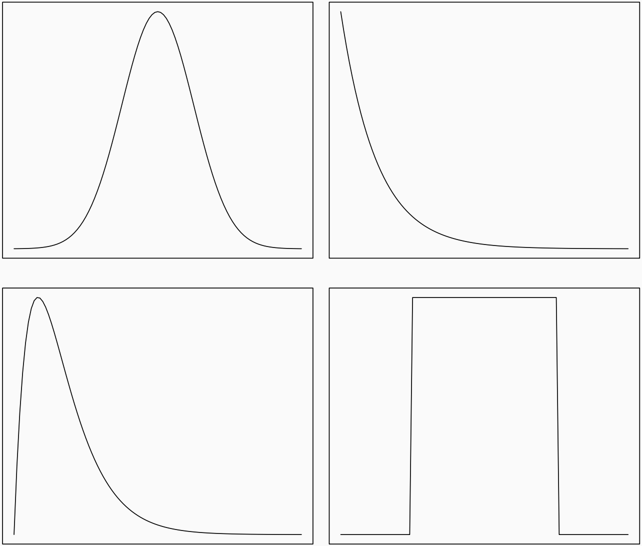
- Queueing Discipline
  - First-come-first-served
  - Last-come-first-served
  - Priority-based service
  - Service in random order

Modelling a Queue



- Type of queues (one vs several lines)

Probability distributions

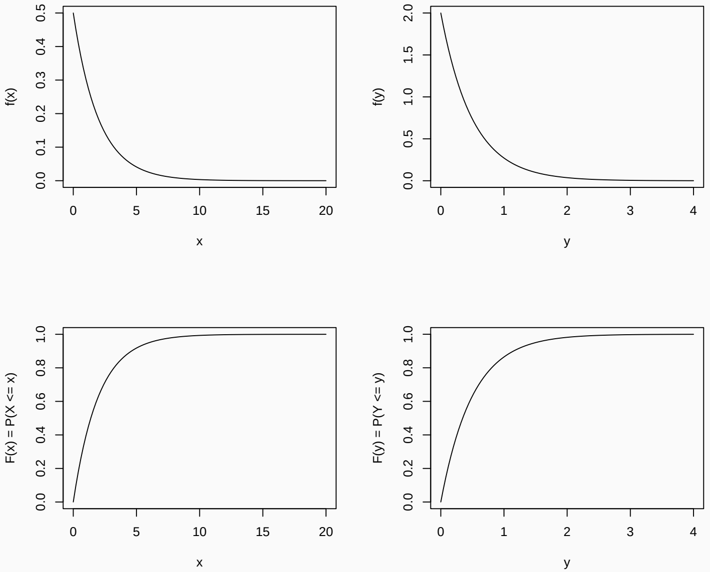


In queueing: need distributions with values that are ≥ 0.

M/M/1 Queue

- Arrival: interarrival time is exponentially distributed
- Service: service time is exponentially distributed
- Number of servers is 1

Exponential distribution



Measuring Performance of a Queue

- Number of customers in the:
  - Queue, L_q
  - Service, L_s
  - System, L
- Time spent in the:
  - Queue, W_q
  - Service, W_s
  - System, W

Questions of interest

- What happens if customers arrive faster than they are being served?
- If a queue has 3 servers, is it better to have individual queues or one queue?

Task 1: Supermarket Checkout Simulation

- Scan the QR code or go to `https://tinyurl.com/yytpyadt`
- Click "Run code" to run with the default parameters

Run code

```
```{r, echo = FALSE, results = "hide"}
lambda = 1                    # number of customers arriving per minute
mu = 2                        # number of customers served per minute
num_servers = 1               # number of checkout operators
simulation_length = 100       # number of minutes
```
```

- Two versions:
  - 1 queue with *n* servers
  - *n* queues with *n* servers

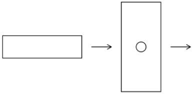
Task 1: Inspect the output

A single queue with n servers

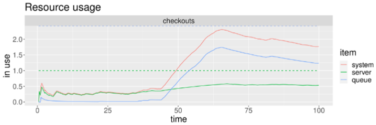
Summary of parameters

- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = one_line
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout



Number in the queue



Performance Measures

Name	Symbol	Simulation
Arrival Rate	lambda	1.0435720
Service Rate	mu	1.9405732
Number in System	L	1.7670723
Number in Service	L_s	0.5311114
Number in Queue	L_q	1.2359609
Time in System	W	1.7080247
Time in Service	W_s	0.5153117
Time in Queue	W_q	1.1927130

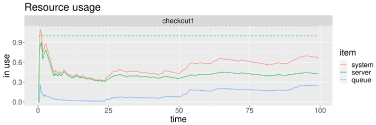
n queues with n servers

Summary of parameters

- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = individual_lines
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout

Number in the queue



Performance Measures

Name	Symbol	Simulation
Arrival Rate	lambda	1.0007467
Service Rate	mu	2.3036289
Number in System	L	0.6739753
Number in Service	L_s	0.4323573
Number in Queue	L_q	0.2416180
Time in System	W	0.6766882
Time in Service	W_s	0.4340977
Time in Queue	W_q	0.2425905

Task 1: Inspect the output

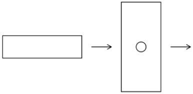
1 queue, n servers

A single queue with n servers

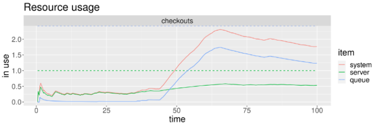
Summary of parameters

- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = one_line
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout



Number in the queue



Performance Measures

Name	Symbol	Simulation
Arrival Rate	lambda	1.0435720
Service Rate	mu	1.9405732
Number in System	L	1.7670723
Number in Service	L_s	0.5311114
Number in Queue	L_q	1.2359609
Time in System	W	1.7080247
Time in Service	W_s	0.5153117
Time in Queue	W_q	1.1927130

n queues, n servers

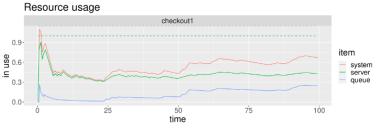
n queues with n servers

Summary of parameters

- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = individual_lines
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout

Number in the queue



Performance Measures

Name	Symbol	Simulation
Arrival Rate	lambda	1.0007467
Service Rate	mu	2.3036289
Number in System	L	0.6739753
Number in Service	L_s	0.4323573
Number in Queue	L_q	0.2416180
Time in System	W	0.6766882
Time in Service	W_s	0.4340977
Time in Queue	W_q	0.2425905

Task 1: Inspect the output

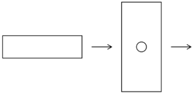
Summary of parameters

A single queue with n servers

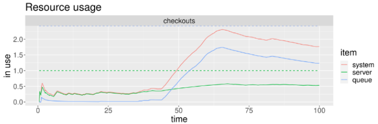
Summary of parameters

- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = one_line
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout



Number in the queue



Performance Measures

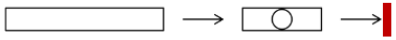
Name	Symbol	Simulation
Arrival Rate	lambda	1.0435720
Service Rate	mu	1.9405732
Number in System	L	1.7670723
Number in Service	L_s	0.5311114
Number in Queue	L_q	1.2359609
Time in System	W	1.7080247
Time in Service	W_s	0.5153117
Time in Queue	W_q	1.1927130

n queues with n servers

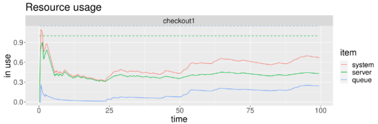
Summary of parameters

- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = individual_lines
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout



Number in the queue



Performance Measures

Name	Symbol	Simulation
Arrival Rate	lambda	1.0007467
Service Rate	mu	2.3036289
Number in System	L	0.6739753
Number in Service	L_s	0.4323573
Number in Queue	L_q	0.2416180
Time in System	W	0.6766882
Time in Service	W_s	0.4340977
Time in Queue	W_q	0.2425905

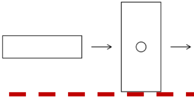
Task 1: Inspect the output

A single queue with n servers

Summary of parameters

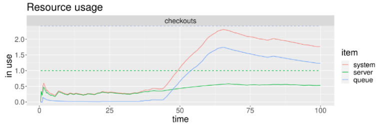
- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = one_line
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout



Number in the system, server and queue over time

Number in the queue



Performance Measures

Name	Symbol	Simulation
Arrival Rate	lambda	1.0435720
Service Rate	mu	1.9405732
Number in System	L	1.7670723
Number in Service	L_s	0.5311114
Number in Queue	L_q	1.2359609
Time in System	W	1.7080247
Time in Service	W_s	0.5153117
Time in Queue	W_q	1.1927130

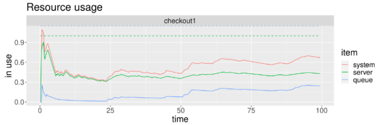
n queues with n servers

Summary of parameters

- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = individual_lines
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout

Number in the queue



Performance Measures

Name	Symbol	Simulation
Arrival Rate	lambda	1.0007467
Service Rate	mu	2.3038289
Number in System	L	0.6739753
Number in Service	L_s	0.4323573
Number in Queue	L_q	0.2416180
Time in System	W	0.6768882
Time in Service	W_s	0.4340977
Time in Queue	W_q	0.2425905

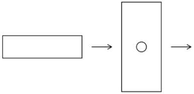
Task 1: Inspect the output

A single queue with n servers

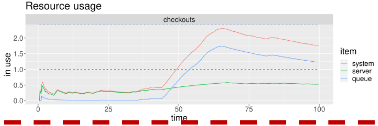
Summary of parameters

- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = one_line
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout



Number in the queue



Performance Measures

Name	Symbol	Simulation
Arrival Rate	lambda	1.0435720
Service Rate	mu	1.9405732
Number in System	L	1.7670723
Number in Service	L_s	0.5311114
Number in Queue	L_q	1.2359609
Time in System	W	1.7080247
Time in Service	W_s	0.5153117
Time in Queue	W_q	1.1927130

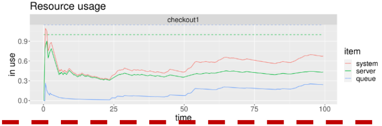
n queues with n servers

Summary of parameters

- Arrival rate, lambda = 1 per minute
- Service rate, mu = 2 per minute
- Number of servers, n = 1
- Type of queue = individual_lines
- Length of simulation = 100 min = 1.67 hours
- Traffic intensity, rho = 0.5

Queue Layout

Number in the queue



Performance Measures

Name	Symbol	Simulation
Arrival Rate	lambda	1.0007467
Service Rate	mu	2.3036289
Number in System	L	0.6739753
Number in Service	L_s	0.4323573
Number in Queue	L_q	0.2416180
Time in System	W	0.6766882
Time in Service	W_s	0.4340977
Time in Queue	W_q	0.2425905

Summary statistics at end of simulation

Task 1: Questions to investigate

1. Simulation length
   a. Change the code so the simulation runs for a longer time, e.g. `simulation_length = 10000`
   b. Run the simulation using the “run code” button and inspect the output
   c. Repeat a few times
2. Number of servers
   a. Change the number of servers `num_servers = 2`
   b. Run the simulation and inspect the output
   c. Repeat a few times

Task 1: Questions to investigate

3. What happens if customers arrive faster than they are served?
    a. Change the number of servers back to 1 `num_servers = 1`
    b. Change the values of lambda (arrival rate) and/or mu (service rate) so that lambda $>$ mu. e.g. `lambda = 3` and `mu = 2`
    c. Run the simulation and inspect the output
    d. Repeat a few times
    e. Conclusion: What happens to the number of people in the system if lambda $>$ mu?

Task 1: Questions to investigate

3. Which queuing configuration is best?
   Is it better to have 3 servers each with their own queue, or with a combined queue?
   a. Change the parameters as follows:
      ```
      lambda = 1
      mu = 2
      num_servers = 3
      ```
   b. Run the simulation and compare the output on the left and right
   c. Repeat a few times
   d. Conclusion: Which queueing configuration performs the best?

Task2: Designing your own supermarket checkout

- Scan the QR code or go to `https://tinyurl.com/38h2tzmb`

  - Two types of customers
    - Express (12 items or less)
      - Arrival rate 4 per minute
      - Service rate 1 per minute (service time = 1 min)
    - Regular (more than 12 items)
      - Arrival rate 1 per minute
      - Service rate 0.2 per minute (service time = 5 min)
  - Two types of checkouts
    - Self-checkout (12 items or less) cost = $10
    - Staffed-checkout (any number of items) cost = $100

Task2: Activity

Find a configuration of self and staffed checkouts which:

- Minimum cost, and must be $700 or less
- Average waiting time in the queue is less than the average service time
    - Express customers – average time in queue less than 1 minute
    - Regular customers – average time in queue less than 5 minutes

e.g. 7 staffed checkouts costs $700 so in budget, but the waiting time is too long.

- Make sure you run the simulation multiple times to ensure that your configuration consistently meets the waiting time requirements.

Conclusion

- Queueing is fun! (and is a great application of probability)
- Simulation is useful when theoretical results are not available and also for verifying theory
- Lots of cool applications beyond supermarket checkouts

Acknowledgments: Dr Anna Fergusson for setting the simulation up on the web

More info

Resources on github:
`https://github.com/sarahemarshall/queueing_is_fun_2023`

- slides
- full R source code (can run in RStudio)

Contact: Dr Sarah Marshall
Dept of Information Systems and Operations Management,
University of Auckland Business School
sarah.marshall@auckland.ac.nz