Lesson 12. An Introduction to Queueing Processes – The Birth-Death Process

1 Overview

- This lesson: a special type of Markov process – the **birth-death process**
- These processes model a variety of systems with **queues**

2 The Case of the Last Parking Space on Earth

Planning for construction of the proposed "Massive Mall" – the largest shopping mall and indoor golf course in the world – includes determining the amount of customer parking to provide. The developers of Massive Mall have told the planners to "give us enough parking for everybody!" The planners must decide what this is supposed to mean in practice, since the number of parking spaces must be some finite number.

Data from other similar malls indicate that the time between car arrivals is exponentially distributed with an expected rate of 1000 cars per hour. In addition, the time that a space is occupied is also exponentially distributed with a mean of 3 hours per car.

2.1 Markov process model

- Let's model this setting as a Markov process
- For the purpose of modeling and analysis, we (the planners) will pretend that parking is unlimited
- Then, we can use our analysis to determine how many spaces are sufficient to satisfy demand a large fraction of the time
- In addition, let's assume the car arrivals and car departures are independent and time-stationary
- State space:
- Transition rates for car arrivals:
- Transition rates for car departures (parking times):
    - Cars leave a single occupied parking space at a rate of cars per hour
    - Suppose there are i cars parked
    - Cars leave the parking lot at a rate of cars per hour
    - Therefore,

- Generator matrix:

2.2 Performance measures

- Suppose we find the steady-state probabilities $\pi_0, \pi_1, \pi_2, \ldots$
- We might be interested in the long-run expected number of cars in the parking lot:

- We might also be interested in the minimum number of parking spaces c^* that will accommodate all cars with probability $1 - \alpha$:

2.3 Is this a reasonable model?

- Car arrivals:
  - Independent?
  - Time-stationary?
  - Exponential interarrival times? (i.e. Poisson arrival process?)

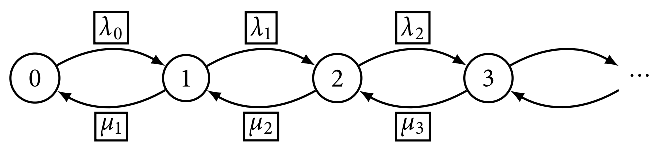
- Car departures (parking times):
  - Independent?
  - Time-stationary?
  - Exponential?

3 The birth-death process

- Markov process with state space $\mathcal{M} = \{0, 1, 2, \dots\}$
  - State = number of customers in the system
  - The **system** refers to all customers receiving service or waiting for service
  - The **queue** refers to only customers waiting for, but not yet receiving, service
- Generator matrix:

$$\mathbf{G} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

  - λ_i is the **arrival rate** in state i
  - μ_i is the **service rate** in state i (you can think of this as a **departure rate**)
  - These rates are measured in customers per unit time
- Transition rate diagram:



- Such a Markov process is called a **birth-death process**
  - State of the system can only increase/decrease by 1 at each transition
  - "Birth" or "death"
- Suppose the process is in state $i > 0$: there are i customers in the system
  - Remaining time until next arrival is exponentially distributed with rate
    - Expected remaining time until next arrival =
  - Remaining time until next departure is exponentially distributed with rate
    - Expected remaining time until next departure =
  - Remaining time until something happens is exponentially distributed with rate
    - Expected remaining time until something happens =
- The arrival rates λ_i and service rates μ_i are defined depending on the details of the queueing system (e.g. number of servers, limited queue capacity, etc.)

4 Formulating arrival rates of a birth-death process

Case 1 (Large customer population)**.** Kent Sporting Goods plans to open a "superstore" in a major city. A queueing analysis will be used to help determine staffing levels for the store.

- Major city with large customer population
  ⇒ number of customer arrivals in nonoverlapping time intervals are likely to be independent
- "Superstore" ⇒ system is large enough to accommodate all customers simultaneously
- Suppose the arrival rate is constant over time

⇒ Poisson process is a plausible model of customer arrivals

⇒ Arrival rate into the system in state i:

Case 2 (Balking)**.** The management of Sharon and LeRoy's Ice Cream has noticed that when potential customers find that the queue of waiting customers is too long, they sometimes go around the corner and buy ice cream at a grocery store. The management would like to incorporate this phenomenon into its staffing model.

- **Balking** occurs when potential customers arriving at a queueing system choose not to enter it
- Balking ⇒ reduces arrival rate of actual customers into system
- Let b_i = Pr{potential customer balks when i customers already in the system}
- Suppose potential customers arrive according to a Poisson process with rate λ

⇒ Arrival rate into the system in state i:

Case 3 (Limited capacity)**.** Customers visit the neighborhood hair stylist Fantastic Dan for haircuts. Dan's shop is small, so only 5 customers can wait inside. When Dan's shop is full, any customers that come by simply leave. Dan wants to investigate what happens if he expands his shop so more customers can wait inside.

- System capacity is reached ⇒ arrival rate into the system is 0
- Although customers continue to arrive, there are no arrivals from the perspective of the queueing system
- Suppose:
  - customers arrive according to a Poisson process with rate λ
  - the capacity of the system is n customers

⇒ Arrival rate into the system in state i:

5 Formulating service rates of a birth-death process

Case 4 (Multiple identical servers)**.** Parking is very limited at Simplexville University, so cars line up at the entrance to parking lots to wait for an available opening. The university would like to evaluate the effect of adding additional spaces to a particular lot.

- Car $\longleftrightarrow$ customer, parking space $\longleftrightarrow$ server
- Suppose:
  - there are s parking spaces in the lot
  - the time a car occupies a space ~ Exponential(μ)
- i cars in the parking lot $\Rightarrow$ first of these cars leaves at a rate

$\Rightarrow$ Service rate of the system in state i:

Case 5 (Reneging)**.** When customers call Fluttering Duck Airline's toll-free number to make reservations, they may be placed in a "hold" queue until an agent is available. Some customers will hang up if they are on hold too long. This phenomenon should be a part of Fluttering Duck Airline's capacity-planning models.

- **Reneging** occurs when customers in a queueing system choose to leave the system prior to receiving service
- Reneging $\Rightarrow$ increases service rate of the system
- Suppose:
  - the time a customer is willing to spend waiting in the queue prior to starting service ~ Exponential(β)
  - the service time ~ Exponential(μ)
  - s identical servers
- $i > s$ customers in system

  $\Rightarrow$ customers receiving service and who might renege

$\Rightarrow$ Service rate of the system in state i:

6 Next time...

- Computing steady-state probabilities and using them to compute different performance measures

7 Exercises

Problem 1 (Nelson 8.4, modified)**.** A small ice-cream shop competes with several other ice-cream shops in a busy mall. If there are too many customers already in line at the shop, then potential customers will go elsewhere. Potential customers arrive at a rate of 20 per hour. The probability that a customer will go elsewhere is $n/5$ when there are $n \le 5$ customers already in the system, and 1 when there are $n > 5$ customers already in the system. The server at the shop can serve customers at a rate of 10 per hour. Approximate the process of potential arrivals as Poisson, and the service times as exponentially distributed.

Model the process of customer arrivals and departures at this ice-cream shop as a birth-death process (i.e. what are λ_i and μ_i for $i = 0, 1, 2, \ldots$?).