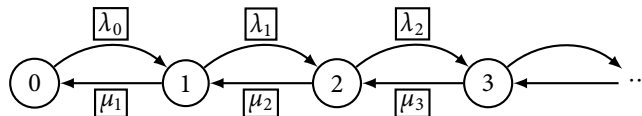# Lesson 20. The Birth-Death Process

1 The birth-death process

- Markov process with state space $\mathcal{M} = \{0, 1, 2, \dots\}$

    - State = number of customers in the <u>system</u>
    - 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:

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

    - λ_i is the **arrival rate** in state i
    - μ_i is the **service rate** in state i
    - 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:
    - Remaining time until next departure:
    - Remaining time until <u>something</u> 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.)

2 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 $\Rightarrow$ number of customer arrivals in nonoverlapping time intervals are likely to be independent

- "Superstore" $\Rightarrow$ system is large enough to accommodate all customers simultaneously

- Suppose the arrival rate is constant over time

$\Rightarrow$ Poisson process is a plausible model of customer arrivals

$\Rightarrow$ 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 <u>potential</u> customers arriving at a queueing system choose not to enter it

- Balking $\Rightarrow$ reduces arrival rate of <u>actual</u> 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

$\Rightarrow$ 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 $\Rightarrow$ arrival rate into the system is 0

- Although customers continue to arrive, there are no arrivals <u>from the perspective of the queueing system</u>

- Suppose:
    - customers arrive according to a Poisson process with rate λ
    - the capacity of the system is n customers

$\Rightarrow$ Arrival rate into the system in state i:

Case 4 (Small customer population). A fleet of automated guided vehicles (AGVs) moves parts between 10 work centers in a factory. A queueing model will help determine the number of AGVs needed.

- Work center $\longleftrightarrow$ customer, AGV $\longleftrightarrow$ server

- # potential customers is small relative to # customers that might be in queueing system
  $\Rightarrow$ the arrival rate depends on the state of the system

    - e.g. all work centers have a part to be moved $\Rightarrow$ all potential customers are in the system
      $\Rightarrow$ the arrival rate is 0

- Idea: model an individual customer's arrival in terms of the time it spends <u>outside</u> the system

- Suppose:

    - there are k potential customers (e.g. $k = 10$ in this case)

    - the time a customer spends outside the system $\sim$ Exponential(τ)

- i customers in the system $\Rightarrow$ [   ] customers outside the system

  $\Rightarrow$ first of these customers outside the system enters at a rate [   ]

  $\Rightarrow$ Arrival rate into the system in state i:

  [   ]

3   Formulating service rates of a birth-death process

Case 5 (Multiple identical servers). Parking is very limited at the University of Minnesota, 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 $\sim$ Exponential(μ)

- i cars in the parking lot $\Rightarrow$ first of these cars leaves at a rate [   ]

  $\Rightarrow$ Arrival rate into the system in state i:

  [   ]

Case 6 (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 $\sim$ Exponential(β)
  - the service time $\sim$ Exponential(μ)
  - s identical servers

- $i > s$ customers in system

$\Rightarrow$ [   ] customers receiving service and [   ] who might renege

$\Rightarrow$ Arrival rate into the system in state i:

4 Steady-state probabilities of a birth-death process

- Recall the steady-state equations:
$$\mathbf{0} = \pi^\top \mathbf{G} \qquad 1 = \pi^\top \mathbf{1}$$

- Rewrite the steady-state equations as scalar equations:

- Solve recursively in terms of π_0:

4

- Define

$$d_0 = 1 \qquad d_j = \prod_{i=1}^{j} \frac{\lambda_{i-1}}{\mu_i} \quad \text{for } j = 1, 2, \ldots$$

- Simplify and solve for π_0:

- Finally, solve for all the steady-state probabilities $\pi_0, \pi_1, \pi_2, \ldots$:

- Note: the steady-state probabilities exist provided that $0 < \sum_{j=0}^{\infty} d_j < \infty$

5 Next time...

- Using the steady-state probabilities to compute different performance measures

6 If we have time...

Example 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 $j/5$ when there are $j \leq 5$ customers already in the system, and 1 when there are $j > 5$ 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.

a. 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 $i = 0, 1, 2, \ldots$?).

b. Over the long run, how many customers are in the shop? (i.e. what is the probability there are 0 customers in the shop? 1? 2? etc.)