Course: Applied Probability MW [2:15 – 3:30]

Chapter: [17]

Markov Chains

Section: [17.1]

What Is a Stochastic Process?

Introduction

Interest

Sometimes we are interested in *how a random variable changes over time*.

Stochastic Process

The study of how a random variable evolves over time includes *stochastic processes* (عمليّة عشوائيّة).

Markov Chain

An explanation of stochastic, a type of stochastic process known as a *Markov chain* is included.

Start

We begin by defining the concept of a stochastic process.

Stochastic Process

- Suppose we observe some characteristic of a system at discrete points in time (labeled 0, 1, 2, . . .).
- Let X_t be the value of the system characteristic at time t.
- In most situations, X_t is not known with certainty before time t and may be viewed as a random variable.
- A discrete-time stochastic process is simply a description of the relation between the random variables $X_0, X_1, X_2, \cdots$.

The Gambler's Ruin Example [1]

At time 0, I have \$2.

$X_0 = 2$

At times $1, 2, \cdots$, I play a game in which I bet (يراهن) \$1.

$X_1, X_2, X_3, \cdots$

With probability p, I win the game,

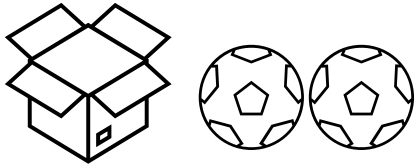
and with probability $1 - p$, I lose the game.

My goal is to increase my capital (رأس المال) to \$4, and as soon as I do, the game is over.

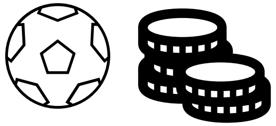
The game is also over if my capital is reduced to \$0.

Define X_t to be my capital position after the time t game (if any) is played.

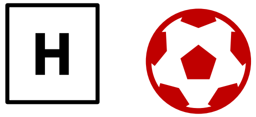
Choosing Balls Example [2]



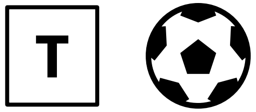
A box contains two unpainted balls at present.



We choose a ball at random and flip a coin.



If the chosen ball is unpainted and the coin comes up heads, we paint the chosen unpainted ball red.



if the chosen ball is unpainted and the coin comes up tails, we paint the chosen unpainted ball black.

If the ball has already been painted, then (whether heads or tails has been tossed) we change the color of the ball (from red to black or from black to red).

We define time t to be the time after the coin has been flipped for the t^{th} time and the chosen ball has been painted.

Choosing Balls Example [2]

We define time t to be the time after the coin has been flipped for the t^{th} time and the chosen ball has been painted.

The state at any time may be described by the vector

$$[\, u \quad r \quad b \,]$$

number of unpainted balls

number of red balls

number of black balls

$X_0 = [2 \quad 0 \quad 0] \;= 200$

After the first coin toss

$X_1 = [1 \quad 1 \quad 0]$ or $X_1 = [1 \quad 0 \quad 1]$

If $X_t = [0 \quad 2 \quad 0]$ then

$X_{t+1} = [0 \quad 1 \quad 1] \;= 011$

Continuous–Time Stochastic Process

- **Is simply** a stochastic process in which the state of the system can be viewed at any time, not just at discrete instants in time.
- **For example**, the number of people in a supermarket t minutes after the store opens for business may be viewed as a continuous-time stochastic process.

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [17]

Markov Chains

Section: [17.2]

What Is a Markov Chain?

Definition

A discrete–time stochastic process is a *Markov chain* if, for $t = 0,1,2,\cdots$ and all states,

$$P(X_{t+1} = i_{t+1} | X_t = i_t, X_{t-1} = i_{t-1}, \cdots, X_1 = i_1, X_0 = i_0) \\ = P(X_{t+1} = i_{t+1} | X_t = i_t) \tag{1}$$

- Equation (1) says that the probability distribution of the state at time $t + 1$ **depends on** the state at time t (i_t) and does not depend on the states the chain passed through on the way to it at time t.
- For all states i and j and all t, $P(X_{t+1} = j | X_t = i)$ is **independent** of t. This assumption allows us to write

$$P(X_{t+1} = j | X_t = i) = p_{ij} \tag{2}$$

Notes

1

p_{ij} is the probability that given the system is in state i at time t, it will be in a state j at time $t+1$.

2

If the system moves from state i during one period to state j during the next period, we say that a **transition** (انتقال) from i to j has occurred.

3

The p_{ij}'s are often referred to as the **transition probabilities** for the Markov chain.

4

- Equation (2) implies that the probability law relating the next period's state to the current state does not change and remains **stationary** over time.
- Any Markov chain that satisfies (2) is called a **stationary Markov chain**.

Notes

5

We also must define q_i to be the probability that the chain is in state i at the **time** $\mathbf{0}$; in other words, $P(X_0 = i) = q_i$.

6

We call the vector

$$\mathbf{q} = \begin{bmatrix} q_1 & q_2 & \cdots & q_s \end{bmatrix}$$

the **initial probability distribution for the Markov chain**.

7

In most applications, the transition probabilities are displayed as an $s \times s$ **transition probability matrix P**.

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{bmatrix} \begin{array}{l} \sum = 1 \\ \sum = 1 \\ \vdots \\ \sum = 1 \end{array}$$

$$0 \le p_{ij} \le 1$$

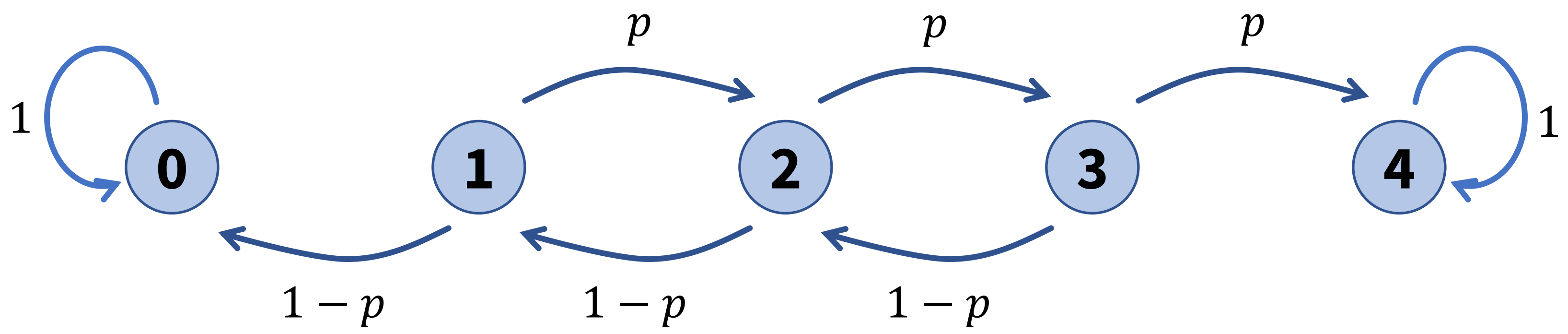
$$\sum_{j=1}^{s} p_{ij} = 1 \qquad \forall i = 1,2,\cdots,s$$

Example: The Gambler's Ruin (Continued)

Find the transition matrix for Example 1.

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

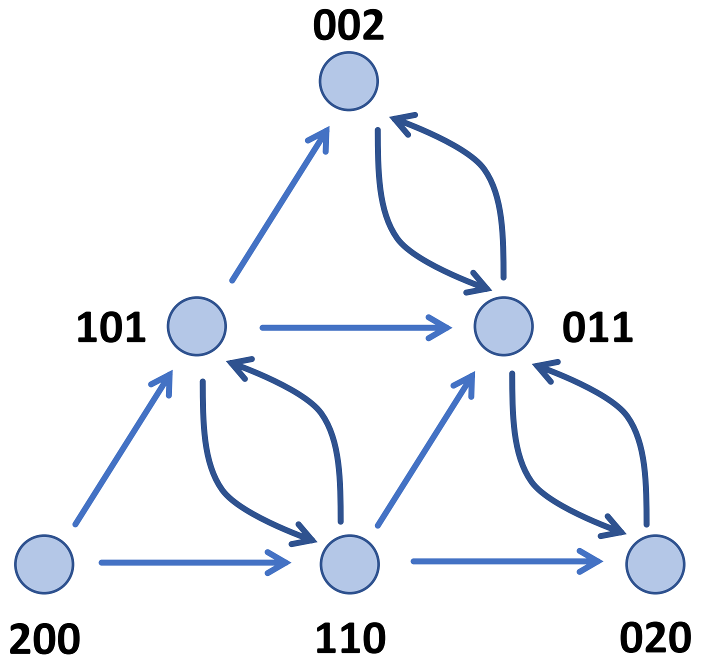
A transition matrix may be represented by a **graph** in which each node represents a state and arc (i, j) represents the transition probability p_{ij}.



Example: Choosing Balls (Continued)

Find the transition matrix for Example 2.

$$P = \begin{array}{c} \\ 200 \\ 110 \\ 101 \\ 011 \\ 020 \\ 002 \end{array} \begin{array}{c} \begin{array}{cccccc} 200 & 110 & 101 & 011 & 020 & 002 \end{array} \\ \begin{bmatrix} 0 & 0.50 & 0.50 & 0 & 0 & 0 \\ 0 & 0 & 0.50 & 0.25 & 0.25 & 0 \\ 0 & 0.50 & 0 & 0.25 & 0 & 0.25 \\ 0 & 0 & 0 & 0 & 0.50 & 0.50 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$



$$P(101|200) = P(\text{Unpaint} \cap \text{Tail}) = 1 \times 0.5 = 0.5$$

$$P(011|101) = P(\text{Unpaint} \cap \text{Head}) = 0.5 \times 0.5 = 0.25$$

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [17]

Markov Chains

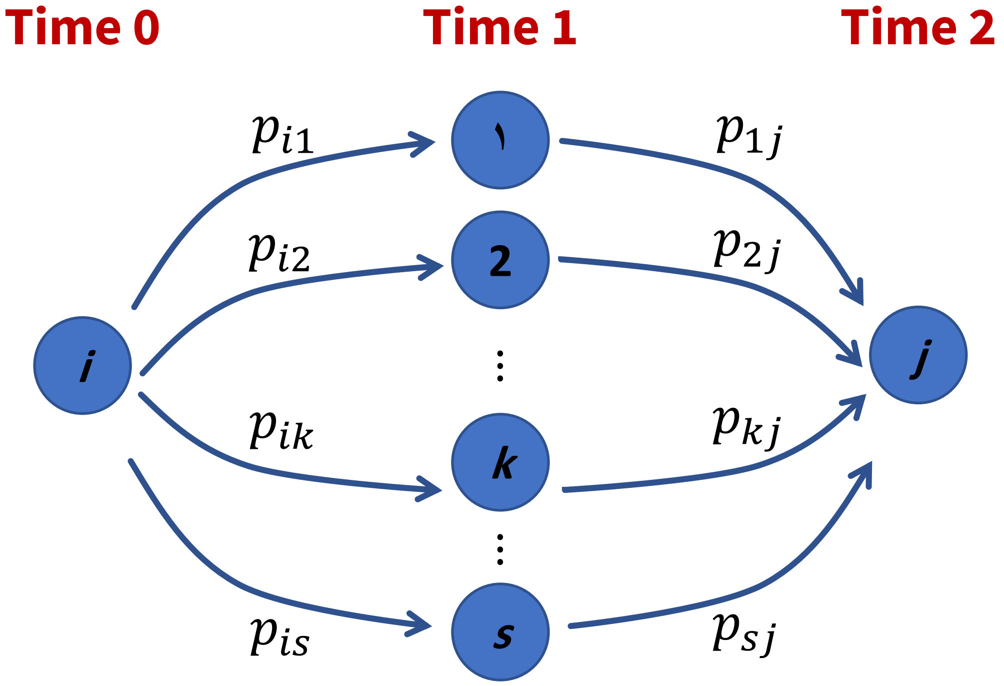
Section: [17.3]

n –Step Transition Probabilities

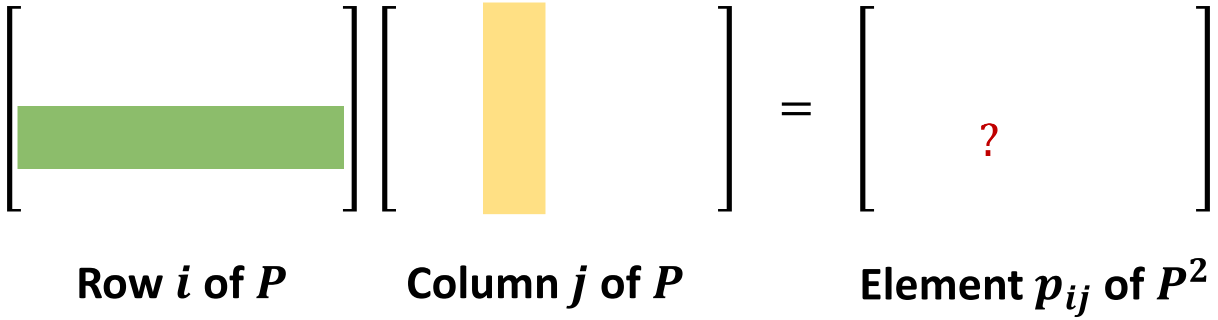
The Idea

A question of interest when studying a Markov chain is

If a Markov chain is in a state i at time m, what is the probability that n periods later than the Markov chain will be in state j?



$$P_{ij}(2) = p_{i1}p_{1j} + p_{i2}p_{2j} + \cdots p_{is}p_{sj}$$



For $n > 1$: $P_{ij}(n) = ij^{\text{th}}$ element of P^n.

The Idea

This probability will be independent of m, so we may write

$$\begin{aligned} P(\mathbf{X}_{m+n} = j | \mathbf{X}_m = i) &= P(\mathbf{X}_n = j | \mathbf{X}_0 = i) \\ &= P_{ij}(n) \end{aligned}$$

where $P_{ij}(n)$ is called the n −step probability of a transition from state i to state j.

$$P_{ij}(0) = \begin{cases} 1 & : \quad i = j \\ 0 & : \quad i \neq j \end{cases}$$

$$\begin{aligned} \mathbf{P}^n &= \mathbf{P}\,\mathbf{P}^{n-1} \\ &= \mathbf{P}^{n-1}\,\mathbf{P} \end{aligned}$$

$$\begin{aligned} \mathbf{P}^n &= \mathbf{P}^{n-m}\,\mathbf{P}^m \\ &= \mathbf{P}^m\,\mathbf{P}^{n-m} \quad ; \quad 0 < m < n \end{aligned}$$

The Cola Example

Suppose the entire cola industry produces only two colas. Given that a person last purchased cola 1, there is a 90% chance that her next purchase will be cola 1. Given that a person last purchased cola 2, there is an 80% chance that her next purchase will be cola 2.

1. If a person is currently a cola 2 purchaser, what is the probability that she will purchase cola 1 two purchases from now?

State 1 = person has last purchased cola 1

State 2 = person has last purchased cola 2

$$P = \begin{array}{cc} & \begin{array}{cc} \text{Cola 1} & \text{Cola 2} \end{array} \\ \begin{array}{c} \text{Cola 1} \\ \text{Cola 2} \end{array} & \begin{bmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{bmatrix} \end{array}$$

$$P(\mathbf{X}_2 = 1 | \mathbf{X}_0 = 2) = P_{21}(2) = 0.34$$

$$P^2 = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix}$$

The Cola Example

Suppose the entire cola industry produces only two colas. Given that a person last purchased cola 1, there is a 90% chance that her next purchase will be cola 1. Given that a person last purchased cola 2, there is an 80% chance that her next purchase will be cola 2.

2. If a person is currently a cola 1 purchaser, what is the probability that she will purchase cola 1 three purchases from now?

State 1 = person has last purchased cola 1

State 2 = person has last purchased cola 2

$$P = \begin{array}{c} \\ \text{Cola 1} \\ \text{Cola 2} \end{array} \begin{array}{c} \begin{array}{cc} \text{Cola 1} & \text{Cola 2} \end{array} \\ \begin{bmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{bmatrix} \end{array}$$

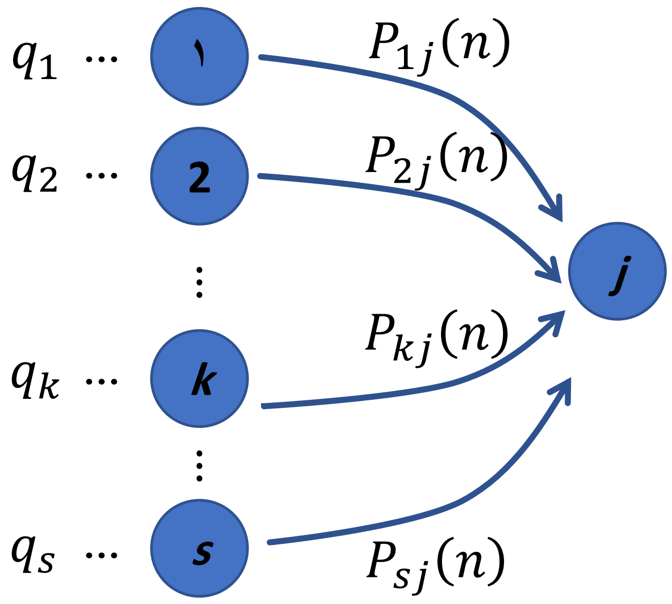
$$P(\mathbf{X}_3 = 1 | \mathbf{X}_0 = 1) = P_{11}(3) = 0.781$$

$$P^3 = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix} = \begin{bmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{bmatrix} \qquad P^2 = \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix}$$

Note

- In many situations, we do not know the state of the Markov chain at time 0.
- As defined in Section 17.2, let q_i be the probability that the chain is in state i at time 0.
- Then we can determine the probability that the system is in state i at time n by using the following formula:

$$\text{Probability of being in state } j \text{ at time } n = \sum_{i=1}^{s} q_i P_{ij}(n) = \begin{bmatrix} q_1 & q_2 & \cdots & q_s \end{bmatrix} \begin{bmatrix} \text{Column} \\ j \text{ of } P^n \end{bmatrix} = \mathbf{q} \begin{bmatrix} \text{Column} \\ j \text{ of } P^n \end{bmatrix}$$



The Cola Example (Continue)

Suppose the entire cola industry produces only two colas. Given that a person last purchased cola 1, there is a 90% chance that her next purchase will be cola 1. Given that a person last purchased cola 2, there is an 80% chance that her next purchase will be cola 2.

$$P = \begin{array}{c c} & \begin{array}{cc} \text{Cola 1} & \text{Cola 2} \end{array} \\ \begin{array}{c} \text{Cola 1} \\ \text{Cola 2} \end{array} & \begin{bmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{bmatrix} \end{array}$$

Suppose 60% of all people now drink cola 1, and 40% now drink cola 2. Three purchases from now, what fraction of all purchasers will be drinking cola 1?

$$P^3 = \begin{bmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{bmatrix}$$

$$\text{Prob.} = \begin{bmatrix} 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 0.781 \\ 0.438 \end{bmatrix} = 0.6438$$

$$\mathbf{q} = \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$

The Cola Example (For Large n)

Suppose the entire cola industry produces only two colas. Given that a person last purchased cola 1, there is a 90% chance that her next purchase will be cola 1. Given that a person last purchased cola 2, there is an 80% chance that her next purchase will be cola 2.

To illustrate the behavior of the n −step transition probabilities for large values of n, we have computed several of the n −step transition probabilities for the Cola example in the following table.

n-Step Transition Probabilities for Cola Drinkers

n	$P_{11}(n)$	$P_{12}(n)$	$P_{21}(n)$	$P_{22}(n)$
1	.90	.10	.20	.80
2	.83	.17	.34	.66
3	.78	.22	.44	.56
4	.75	.25	.51	.49
5	.72	.28	.56	.44
10	.68	.32	.65	.35
20	.67	.33	.67	.33
30	.67	.33	.67	.33
40	.67	.33	.67	.33

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [17]

Markov Chains

Section: [17.4]

Classification of States in a Markov Chain

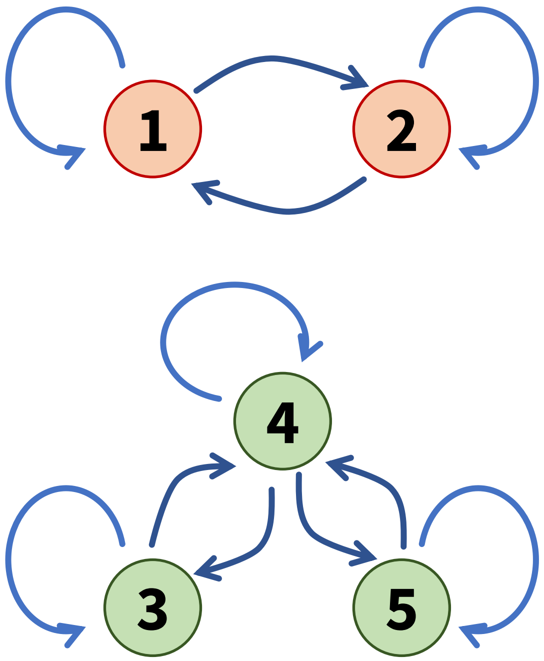
Classifications

We use the following transition matrix to illustrate the following definitions.

$$P = \begin{bmatrix} 0.4 & 0.6 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0 & 0.8 & 0.2 \end{bmatrix}$$

DEFINITION [1]

Given two states i and j, a **path** from i to j is a sequence of transitions that begins in i and ends in j, such that each transition in the sequence has a positive probability of occurring.

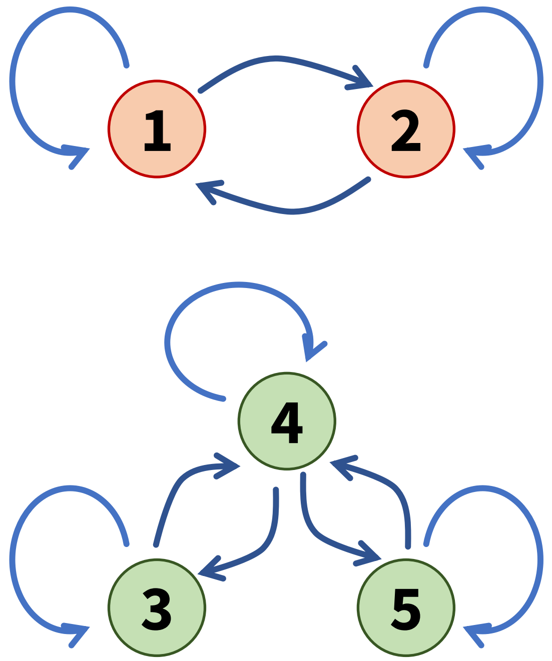


For example, $3-4-5$ is a path from 3 to 5.

Classifications

We use the following transition matrix to illustrate the following definitions.

$$P = \begin{bmatrix} 0.4 & 0.6 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0 & 0.8 & 0.2 \end{bmatrix}$$



DEFINITION [2]

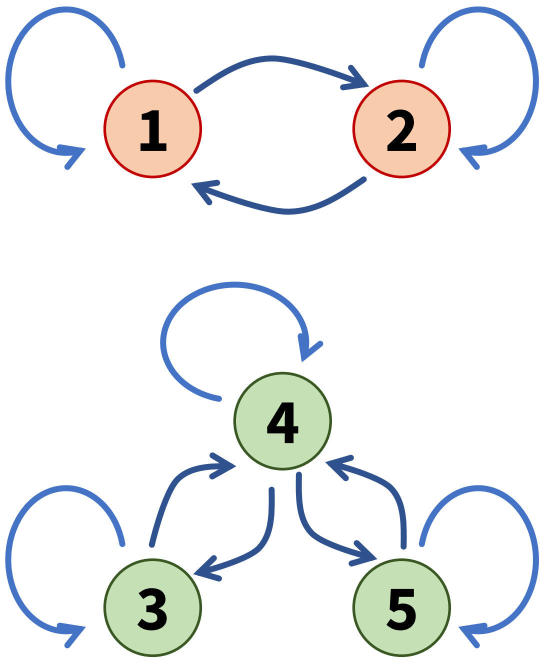
A state j is **reachable** from state i if there is a path leading from i to j.

For example, state 5 is reachable from state 3 via the path 3–4–5, but state 5 is not reachable from state 1 since there is no path from 1 to 5.

Classifications

We use the following transition matrix to illustrate the following definitions.

$$P = \begin{bmatrix} 0.4 & 0.6 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0 & 0.8 & 0.2 \end{bmatrix}$$



DEFINITION [3]

Two states i and j are said to **communicate** if j is reachable from i, and i is reachable from j.

For example, states 1 and 2 communicate since we can go from 1 to 2 and from 2 to 1.

Classifications

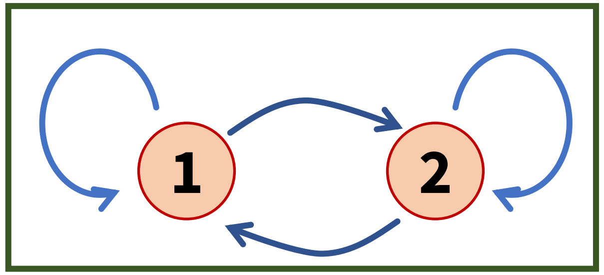
We use the following transition matrix to illustrate the following definitions.

$$P = \begin{bmatrix} 0.4 & 0.6 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0 & 0.8 & 0.2 \end{bmatrix}$$

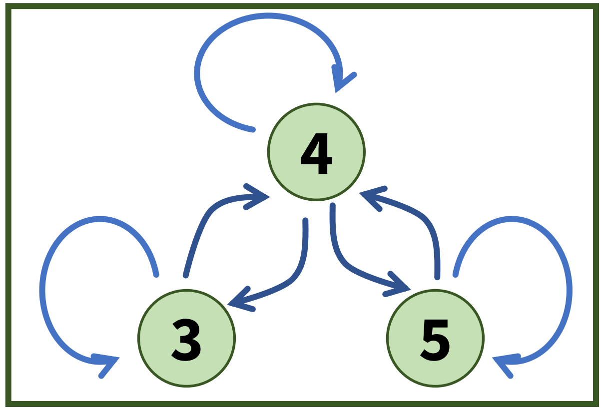
DEFINITION [4]

A set of states S in a Markov chain is a **closed set** if no state outside of S is reachable from any state in S.

For example, $S_1 = \{1, 2\}$ and $S_2 = \{3, 4, 5\}$ are both closed sets.



S_1



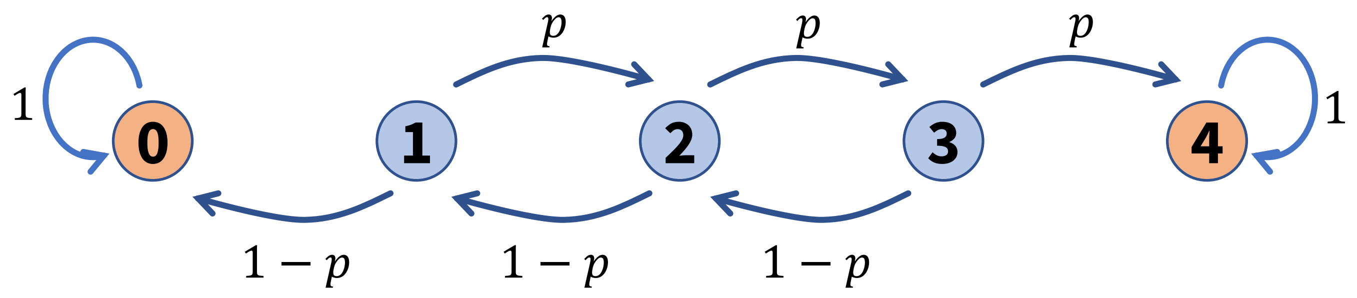
S_2

Classifications

DEFINITION [5]

A state i is an **absorbing** state if $p_{ii} = 1$.

- Whenever we enter an absorbing state, we never leave the state.
- **For example**, the gambler's ruin, states 0 and 4 are absorbing states.
- Of course, an absorbing state is a closed set containing only one state.

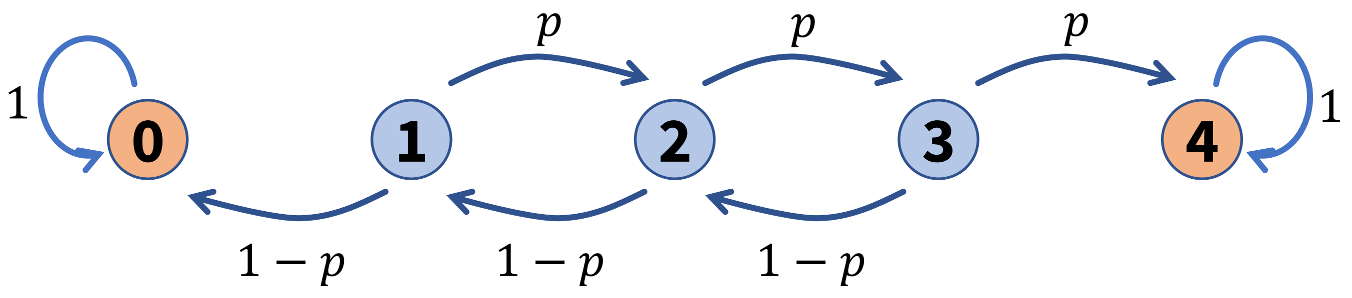


Classifications

DEFINITION [6]

A state i is a **transient** state if there exists a state j that is reachable from i, but the state i is not reachable from state j.

- In other words, a state i is transient if there is a way to leave state i that never returns to state i.
- **For example**, In the gambler's ruin example, states 1, 2, and 3 are transient states.
- After a large number of periods, the probability of being in any transient state i is zero.

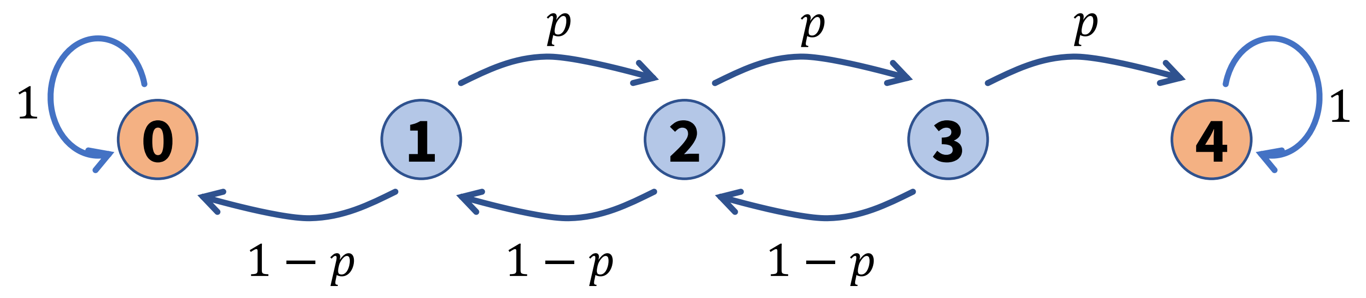


Classifications

DEFINITION [7]

If a state is not transient, it is called a **recurrent** state.

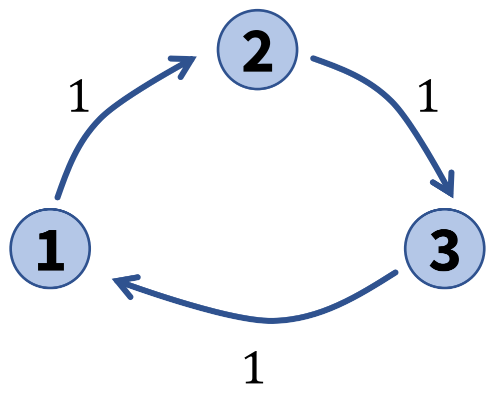
For example, In the gambler's ruin example, states 0 and 4 are recurrent states (and also called *absorbing* states).



Classifications

DEFINITION [8]

A state i is **periodic** with period $k > 1$ if k is the smallest number such that all paths leading from state i back to state i have a length that is a multiple of k. If a *recurrent* state is not periodic, it is referred to as **aperiodic**.



For the Markov chain with transition matrix $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

each state has period 3.

For example, if we begin in state 1, the only way to return to state 1 is to follow the path 1–2–3–1 for some number of times m. Hence, any return to state 1 will take $3m$ transitions, so state 1 has period 3. This means that $P_{11}(3m) = 1$.

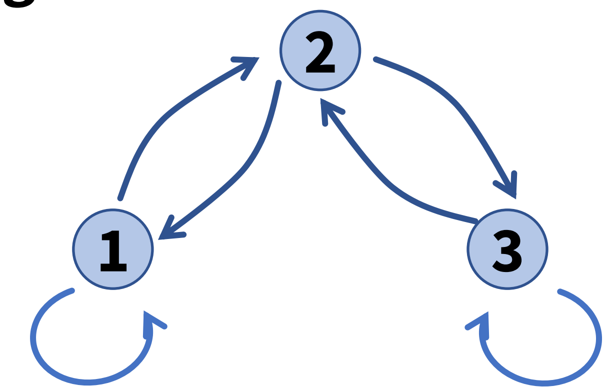
Classifications

DEFINITION [9]

If <u>all</u> states in a chain are *recurrent*, *aperiodic*, and *communicate* with each other, the chain is said to be **ergodic**.

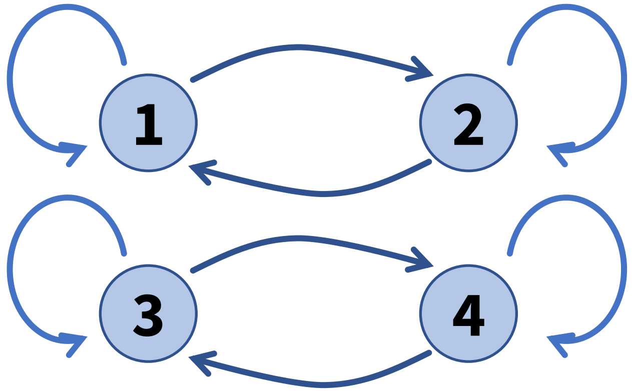
For example,

$$P_1 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$



Ergodic

$$P_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$



Not Ergodic

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [17]

Markov Chains

Section: [17.5]

Steady-State Probabilities and Mean First Passage Times

Introduction

- In our discussion of the cola example, we found that after a long time, the probability that a person's next cola purchase would be cola 1 approached .67 and .33 that it would be cola 2.

- These probabilities did not depend on whether the person was initially a cola 1 or a cola 2 drinker.

- In this section, we discuss the important concept of steady-state probabilities, which can be used to describe the long-run behavior of a Markov chain.

n-Step Transition Probabilities for Cola Drinkers

n	$P_{11}(n)$	$P_{12}(n)$	$P_{21}(n)$	$P_{22}(n)$
1	.90	.10	.20	.80
2	.83	.17	.34	.66
3	.78	.22	.44	.56
4	.75	.25	.51	.49
5	.72	.28	.56	.44
10	.68	.32	.65	.35
20	.67	.33	.67	.33
30	.67	.33	.67	.33
40	.67	.33	.67	.33

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 0.67 & 0.33 \\ 0.67 & 0.33 \end{bmatrix}$$

Theorem 1

Let P be the transition matrix for an s −state **ergodic** chain. Then there exists a vector $\pi = [\pi_1 \quad \pi_2 \quad \cdots \quad \pi_s]$ such that

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_s \\ \pi_1 & \pi_2 & \cdots & \pi_s \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_s \end{bmatrix}$$

- Recall that the ij^{th} element of P^n is $P_{ij}(n)$. Theorem 1 tells us that for any initial state i,

$$\lim_{n\to\infty} P_{ij}(n) = \pi_j$$

- Observe that for large n, P^n approaches a matrix with identical rows. This means that after a long time, the Markov chain settles down (تستقر), and (independent of the initial state i) there is a probability π_j that we are in state j.

Steady-State Distribution

The vector $\pi = [\pi_1 \quad \pi_2 \quad \cdots \quad \pi_s]$ is often called the **steady-state distribution**, or **equilibrium distribution**, for the Markov chain.

For a given chain with transition matrix P, how can we find the steady-state probability distribution?

- If n is large, solve the system $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$.
- Unfortunately, this system of equations has an infinite number of solutions.
- Note that for any n and any i,
  $$P_{i1}(n) + P_{i2}(n) + \cdots + P_{is}(n) = 1$$
- Letting n approach infinity, we obtain
  $$\pi_1 + \pi_2 + \cdots + \pi_s = 1$$
- To obtain unique values of the steady-state probabilities, replace any equation in $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$ by $\pi_1 + \pi_2 + \cdots + \pi_s = 1$.

The Cola Example

- To illustrate how to find the steady-state probabilities, we find the steady-state probabilities for the cola example.
- Recall that the transition matrix for the cola example was $P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$.

$$\pi = \pi\, P$$

$$\begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

$$\pi_1 = 0.9\pi_1 + 0.2\pi_2$$
$$\pi_2 = 0.1\pi_1 + 0.8\pi_2$$

Replacing the first equation with the condition $\pi_1 + \pi_2 = 1$, we obtain the system:

$$1 = \pi_1 + \pi_2$$
$$\pi_2 = 0.1\pi_1 + 0.8\pi_2$$

Solving for π_1 and π_2 we obtain $\pi_1 = \frac{2}{3}$ and $\pi_2 = \frac{1}{3}$.

Use of Steady-State Probabilities in Decision Making

In the Cola example, suppose that each customer makes one purchase of cola during any week (52 weeks = 1 year). Suppose there are 100 million cola customers. One selling unit of cola costs the company \$1 to produce and is sold for \$2. For \$500 million per year, an advertising firm (شركة إعلانات) guarantees to decrease from 10% to 5% the fraction of cola 1 customers who switch to cola 2 after a purchase. Should the company that makes cola 1 hire the advertising firm?

- At present, a fraction $\pi_1 = \frac{2}{3}$ of all purchases are cola 1 purchases.
- Each purchase of cola 1 earns the company \$1 profit.
- Since there are a total of 5200000000 cola purchases each year, the cola 1 company's current annual profit is $\frac{2}{3} \times 5200000000 = \$3466666667$

Use of Steady-State Probabilities in Decision Making

In the Cola example, suppose that each customer makes one purchase of cola during any week (52 weeks = 1 year). Suppose there are 100 million cola customers. One selling unit of cola costs the company \$1 to produce and is sold for \$2. For \$500 million per year, an advertising firm (شركة إعلانات) guarantees to decrease from 10% to 5% the fraction of cola 1 customers who switch to cola 2 after a purchase. Should the company that makes cola 1 hire the advertising firm?

- The advertising firm is offering to change the P matrix to

$$P^* = \begin{bmatrix} 0.95 & 0.05 \\ 0.20 & 0.80 \end{bmatrix}$$

- For P^*, the steady-state equations become

$$\pi_1 = 0.95\pi_1 + 0.20\pi_2$$
$$\pi_2 = 0.05\pi_1 + 0.80\pi_2$$

- Replacing the second equation by $\pi_1 + \pi_2 = 1$ and solving, we obtain $\pi_1 = 0.8$ and $\pi_2 = 0.2$.

Use of Steady-State Probabilities in Decision Making

In the Cola example, suppose that each customer makes one purchase of cola during any week (52 weeks = 1 year). Suppose there are 100 million cola customers. One selling unit of cola costs the company $1 to produce and is sold for $2. For $500 million per year, an advertising firm (شركة إعلانات) guarantees to decrease from 10% to 5% the fraction of cola 1 customers who switch to cola 2 after a purchase. Should the company that makes cola 1 hire the advertising firm?

- Now the cola 1 company's annual profit will be

$$(.80)(5{,}200{,}000{,}000) - 500{,}000{,}000 = \$3{,}660{,}000{,}000$$

- Hence, the cola 1 company should hire the ad agency.

Mean First Passage Times

- For an **ergodic** chain, let m_{ij} = expected number of transitions before we first reach state j, given that we are currently in state i.
- m_{ij} is called the **mean first passage time** from state i to state j.
- In the Cola Example, m_{12} would be the expected number of bottles of cola purchased by a person who just bought cola 1 before first buying a bottle of cola 2.
- By solving the following linear equations, we may find all the mean first passage times.

$$m_{ii} = \frac{1}{\pi_i} \quad , \qquad m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj}$$

Mean First Passage Times

- To illustrate the use of these equations, let's solve for the mean first passage times in the Cola Example.
- Recall that $P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$, $\pi_1 = \frac{2}{3}$ and $\pi_2 = \frac{1}{3}$.
- Then $m_{11} = \frac{1}{2/3} = 1.5$ and $m_{22} = \frac{1}{1/3} = 3$.
- Now,

$$
\begin{aligned}
m_{12} &= 1 + p_{11}m_{12} = 1 + 0.9m_{12} \Rightarrow m_{12} = 10 \\
m_{21} &= 1 + p_{22}m_{21} = 1 + 0.8m_{12} \Rightarrow m_{21} = 5
\end{aligned}
$$

- This means, for example, that a person who last drank cola 1 will drink an average of ten bottles of soda before switching to cola 2.

Exercise [8]

Three balls are divided between two containers. During each period a ball is randomly chosen and switched to the other container.

a) Find (in the steady state) the fraction of the time that a container will contain 0, 1, 2, or 3 balls.

$$P = \begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

$$\pi_0 = \frac{1}{3}\pi_1$$
$$\pi_1 = \pi_0 + \frac{2}{3}\pi_2$$
$$\pi_2 = \frac{2}{3}\pi_1 + \pi_3$$
$$\pi_3 = \frac{1}{3}\pi_2$$
$$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3$$

$$\pi_0 = \frac{1}{8}$$
$$\pi_1 = \frac{3}{8}$$
$$\pi_2 = \frac{3}{8}$$
$$\pi_3 = \frac{1}{8}$$

Exercise [8]

Three balls are divided between two containers. During each period a ball is randomly chosen and switched to the other container.

b) If container 1 contains no balls, on the average how many periods will go by before it again contains no balls?

$$m_{00} = \frac{1}{\pi_0} = \frac{1}{1/8} = 8$$

$$\pi_0 = \frac{1}{8}$$
$$\pi_1 = \frac{3}{8}$$
$$\pi_2 = \frac{3}{8}$$
$$\pi_3 = \frac{1}{8}$$

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [17]

Markov Chains

Section: [17.6]

Absorbing Chains

Absorbing States and Absorbing Chains

- Many interesting applications of Markov chains involve chains in which some of the states are absorbing and the rest are transient states.
- Such a chain is called an **absorbing chain**.
- Consider an absorbing Markov chain: If we begin in a transient state, then eventually we are sure to leave the transient state and end up in one of the absorbing states.
- A state in a Markov chain is called an **absorbing state** if, once the state is entered, it is impossible to leave.

$$P = \begin{array}{c} A \\ B \\ C \end{array} \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix}$$

State A is absorbing state.

Absorbing States and Absorbing Chains

- A state in a Markov chain is **absorbing** if and only if the row of the transition matrix corresponding to the state has a 1 on the main diagonal and 0's elsewhere.
- The presence of an absorbing state in a transition matrix does not guarantee that the powers of the matrix approach a limiting matrix.

Definition A Markov chain is an **absorbing chain** if:

1. There is at least one absorbing state; and
2. It is possible to go from each nonabsorbing state to at least one absorbing state in a finite number of steps.

Standard Form

- For any absorbing chain, one might want to know certain things.
  1. If the chain begins in a given transient state, and before we reach an absorbing state, what is the expected number of times that each state will be entered? How many periods do we expect to spend in a given transient state before absorption takes place?
  2. If a chain begins in a given transient state, what is the probability that we end up in each absorbing state?
- To answer these questions, we need to write the transition matrix with the states listed in the following order: transient states first, then absorbing states.

Standard Form

$$P = \left[\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array}\right]$$

- $\boldsymbol{I}$ is an $m \times m$ identity matrix reflecting the fact that we can never leave an absorbing state.
- $\boldsymbol{Q}$ is an $(s - m) \times (s - m)$ matrix that represents transitions between transient states.
- $\boldsymbol{R}$ is an $(s - m) \times m$ matrix representing transitions from transient states to absorbing states.
- $\mathbf{0}$ is an $m \times (s - m)$ matrix consisting entirely of zeros. This reflects the fact that it is impossible to go from an absorbing state to a transient state.

Example

Consider the following transition matrix:

$$\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} .2 & .3 & .4 & .1 \\ 0 & 1 & 0 & 0 \\ .5 & .3 & 0 & .2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

The matrix **P** can be rearranged and partitioned as:

$$\mathbf{P}^* = \begin{array}{c} \\ 1 \\ 3 \\ 2 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 3 & 2 & 4 \end{array} \\ \left(\begin{array}{cc|cc} .2 & .4 & .3 & .1 \\ .5 & 0 & .3 & .2 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \end{array}$$

$$Q = \begin{bmatrix} 0.2 & 0.4 \\ 0.5 & 0.0 \end{bmatrix}$$

$$R = \begin{bmatrix} 0.3 & 0.1 \\ 0.3 & 0.2 \end{bmatrix}$$

Goal

Given the definition of $\boldsymbol{R}$ and $\boldsymbol{Q}$ and the *unit column vector* $\mathbf{1}$ (of all 1 elements),

it can be shown that

- Expected time in state j starting in state i = element (i, j) of $(\boldsymbol{I} - \boldsymbol{Q})^{-1}$.
- Expected time to absorption = $(\boldsymbol{I} - \boldsymbol{Q})^{-1}\ \mathbf{1}$.
- Probability of absorption = $(\boldsymbol{I} - \boldsymbol{Q})^{-1}\boldsymbol{R}$.

Example: The Gambler's Ruin

In the Gambler's Ruin example let $p = 0.5$, then the transition matrix is

$$p = \begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

We write the matrix p in standard form

$$p = \begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 3 \\ 1 \\ 2 \\ 0 \\ 4 \end{array} & \begin{bmatrix} 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$p = \begin{array}{c c} & \begin{array}{ccccc} 3 & 1 & 2 & 0 & 4 \end{array} \\ \begin{array}{c} 3 \\ 1 \\ 2 \\ 0 \\ 4 \end{array} & \left[\begin{array}{ccc|cc} 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

Example: The Gambler's Ruin

$$p = \begin{array}{c c} & \begin{array}{ccccc} 3 & 1 & 2 & 0 & 4 \end{array} \\ \begin{array}{c} 3 \\ 1 \\ 2 \\ 0 \\ 4 \end{array} & \left[\begin{array}{ccc|cc} 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

$$\boldsymbol{Q} = \begin{array}{c c} & \begin{array}{ccc} 3 & 1 & 2 \end{array} \\ \begin{array}{c} 3 \\ 1 \\ 2 \end{array} & \begin{bmatrix} 0 & 0 & 0.5 \\ 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix} \end{array} \qquad \boldsymbol{R} = \begin{array}{c c} & \begin{array}{cc} 0 & 4 \end{array} \\ \begin{array}{c} 3 \\ 1 \\ 2 \end{array} & \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

$$\boldsymbol{I} - \boldsymbol{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0.5 \\ 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{bmatrix}$$

$$(\boldsymbol{I} - \boldsymbol{Q})^{-1} = \begin{bmatrix} 1.5 & 0.5 & 1 \\ 0.5 & 1.5 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

Example: The Gambler's Ruin

$$p = \begin{array}{c} \\ 3 \\ 1 \\ 2 \\ 0 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 3 & 1 & 2 & 0 & 4 \end{array} \\ \left[\begin{array}{ccc|cc} 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

$$(\boldsymbol{I}-\boldsymbol{Q})^{-1} = \begin{bmatrix} 1.5 & 0.5 & 1 \\ 0.5 & 1.5 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad \boldsymbol{R} = \begin{array}{c} \\ 3 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 4 \end{array} \\ \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

$$(\boldsymbol{I}-\boldsymbol{Q})^{-1}\mathbf{1} = \begin{array}{c} 3 \\ 1 \\ 2 \end{array} \begin{bmatrix} 1.5 & 0.5 & 1 \\ 0.5 & 1.5 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}$$

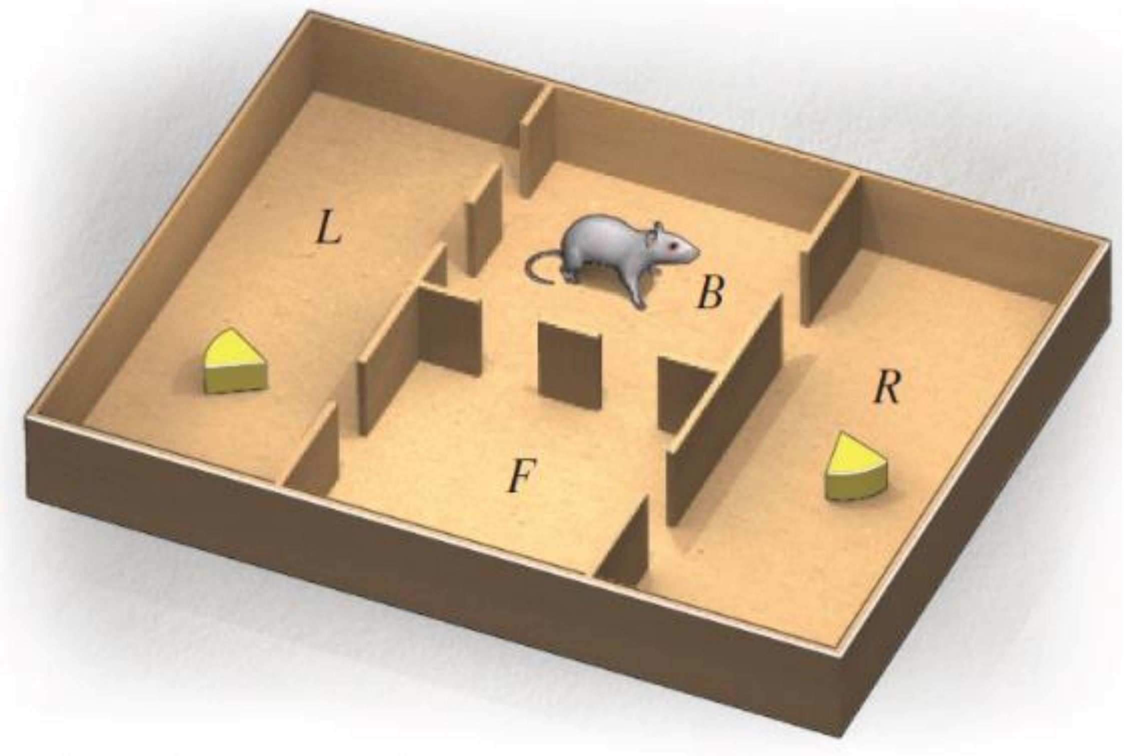
Expected time to absorption

$$(\boldsymbol{I}-\boldsymbol{Q})^{-1}\boldsymbol{R} = \begin{array}{c} 3 \\ 1 \\ 2 \end{array} \begin{bmatrix} 1.5 & 0.5 & 1 \\ 0.5 & 1.5 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.75 \\ 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix}$$

Probability of absorption

Psychology Example

A rat is placed in room *F* or room *B* of the maze shown in the figure. The rat wanders from room to room until it enters one of the rooms containing food, *L* or *R*. Assume that the rat chooses an exit from a room at random and that once it enters a room with food it never leaves.



a) **What is the long-run probability that a rat placed in room *B* ends up in room *R*?**

b) **What is the average number of exits that a rat placed in room *B* will choose until it finds food?**

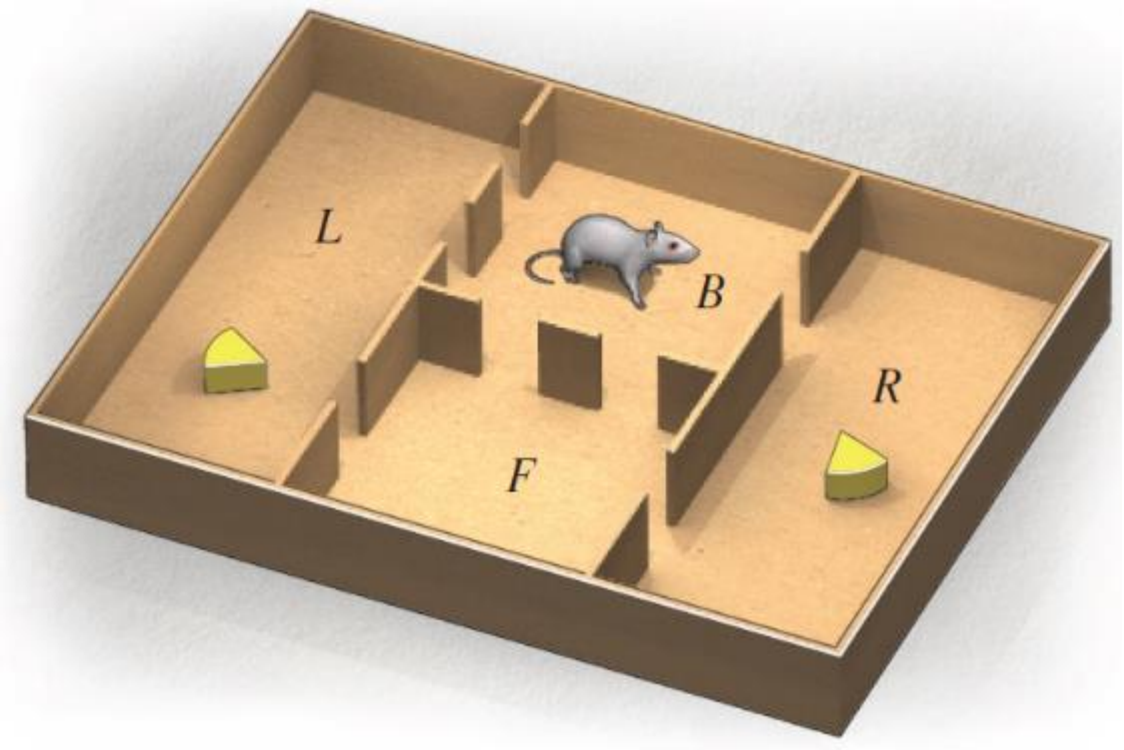
Psychology Example

$$P = \begin{array}{c} \\ B \\ F \\ L \\ R \end{array} \begin{array}{c} \begin{array}{cccc} B & F & L & R \end{array} \\ \left[\begin{array}{cc|cc} 0 & 0.40 & 0.40 & 0.20 \\ 0.50 & 0 & 0.25 & 0.25 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

$$Q = \begin{bmatrix} 0 & 0.4 \\ 0.5 & 0 \end{bmatrix} \qquad R = \begin{bmatrix} 0.40 & 0.20 \\ 0.25 & 0.25 \end{bmatrix}$$

$$(I - Q)^{-1} = \begin{bmatrix} 1.25 & 0.50 \\ 0.625 & 1.25 \end{bmatrix}$$

$$(I - Q)^{-1}R = \begin{bmatrix} 1.25 & 0.50 \\ 0.625 & 1.25 \end{bmatrix} \begin{bmatrix} 0.40 & 0.20 \\ 0.25 & 0.25 \end{bmatrix}$$
$$= \begin{bmatrix} 0.6250 & 0.3750 \\ 0.5625 & 0.4375 \end{bmatrix}$$



a) **What is the long-run probability that a rat placed in room *B* ends up in room *R*?**

b) **What is the average number of exits that a rat placed in room *B* will choose until it finds food?**

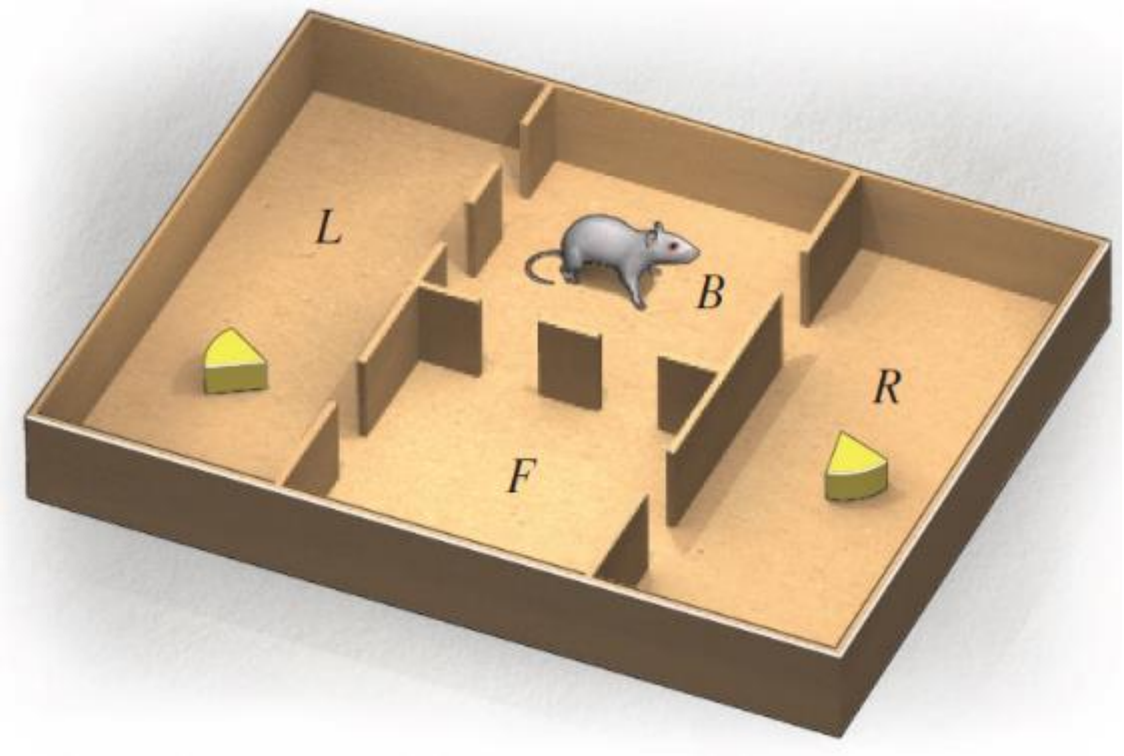
Psychology Example

$$P = \begin{array}{c c} & \begin{array}{cccc} B & F & L & R \end{array} \\ \begin{array}{c} B \\ F \\ L \\ R \end{array} & \left[\begin{array}{cc|cc} 0 & 0.40 & 0.40 & 0.20 \\ 0.50 & 0 & 0.25 & 0.25 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

$$Q = \begin{bmatrix} 0 & 0.4 \\ 0.5 & 0 \end{bmatrix} \qquad R = \begin{bmatrix} 0.40 & 0.20 \\ 0.25 & 0.25 \end{bmatrix}$$

$$(I - Q)^{-1} = \begin{bmatrix} 1.25 & 0.50 \\ 0.625 & 1.25 \end{bmatrix}$$

$$(I - Q)^{-1}\mathbf{1} = \begin{bmatrix} 1.25 & 0.50 \\ 0.625 & 1.25 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1.750 \\ 1.875 \end{bmatrix}$$



a) **What is the long-run probability that a rat placed in room *B* ends up in room *R*?**
b) **What is the average number of exits that a rat placed in room *B* will choose until it finds food?**

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [20]

Queuing Theory

Section: [20.1]

Some Queuing Terminology

Description

- Each of us has spent a great deal of time waiting in lines.
- To describe a *queuing system*, an **input process** and an **output process** must be specified.
- Examples of input and output processes are:

Situation	Input Process	Output Process
Bank	Customers arrive at bank	Tellers serve the customers
Pizza parlor	Request for pizza delivery are received	Pizza parlor send out truck to deliver pizzas

The Input or Arrival Process

- The input process is usually called the **arrival process**.
- Arrivals are called **customers**.
- We assume that *no more than one arrival* can occur at a given instant.
- If more than one arrival can occur at a given instant, we say that **bulk arrivals** (**وصول بالجملة**) are allowed.
- Models in which arrivals are drawn from a small population are called **finite source** models.
- If a customer arrives but fails to enter the system, we say that the **customer has balked** (**تم رفض العميل**).

The Output or Service Process

- To describe the output process of a queuing system, we usually specify a probability distribution – **the service time distribution** – which governs a customer's service time.
- We study two arrangements of servers: **servers in parallel** and **servers in series**.
- *Servers are in parallel* if all server provide the same type of service and a customer need only pass through one server to complete service.
- *Servers are in series* if a customer must pass through several servers before completing service.

Queue Discipline

- The **queue discipline** (ضبط الطوابير) describes the method used to determine the order in which customers are served.
- The most common queue discipline is the **FCFS** discipline (**first come, first served**), in which customers are served in the order of their arrival.
- Under the **LCFS** discipline (**last come, first served**), the most recent arrivals are the first to enter service.
- If the next customer to enter service is *randomly chosen* from those customers waiting for service it is referred to as the **SIRO** discipline (**service in random order**).

Queue Discipline

- Finally we consider **priority queuing disciplines**.
- A **priority discipline** classifies each arrival into one of *several categories*.
- Each category is then given a *priority level*, and within each priority level, customers enter service on an **FCFS** basis.
- Another factor that has an important effect on the behavior of a queuing system is the method that customers use to determine which line to join.

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [20]

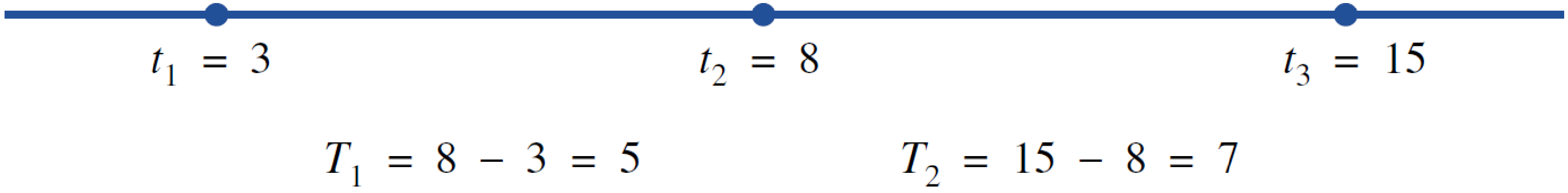
Queuing Theory

Section: [20.2]

Modeling Arrival and Service Processes

Modeling the Arrival Process

- As previously mentioned, we assume that at most one arrival can occur at a given instant of time.
- We define t_i to be the time at which the i^{th} customer arrives.
- For $i \geq 1$, we define $T_i = t_{i+1} - t_i$ to be the i^{th} **interarrival time**.



- In modeling the arrival process, we assume that the T_i's are *independent*, *continuous* **random variables** described by the random variable $\boldsymbol{A}$.

Modeling the Arrival Process

- The assumption that each T_i is continuous is usually a good approximation of *reality*.
- The assumption that each interarrival time is governed by the same random variable implies that the distribution of arrivals is independent of the time of day or the day of the week. This is the assumption of **stationary interarrival times**.
- Stationary interarrival times is often *unrealistic*, but we may often approximate reality by breaking the time of day into segments.

Modeling the Arrival Process

- We assume that $\boldsymbol{A}$ has a density function $a(t)$.
- Of course, a negative interarrival time is impossible. This allows us to write

$$P(\boldsymbol{A} \leq c) = \int_0^c a(t)dt \quad \text{and} \quad P(\boldsymbol{A} > c) = \int_c^\infty a(t)dt$$

- We define $\frac{1}{\lambda}$ to be the mean or average interarrival time (*will have units of hours per arrival*).

$$E(\boldsymbol{A}) = \frac{1}{\lambda} = \int_0^\infty ta(t)dt$$

- We define λ to be the **arrival rate**, which will have units of arrivals per hour.

Modeling the Arrival Process

- In most applications of queuing, an important question is how to choose $\boldsymbol{A}$ to reflect reality and still be computationally tractable?
- The most common choice for $\boldsymbol{A}$ is the **exponential distribution**.
- An exponential distribution with parameter λ has a density $a(t) = \lambda e^{-\lambda t}$.
- Using integration by parts, we have:

$$E(\boldsymbol{A}) = \frac{1}{\lambda}$$

$$\mathrm{var}(\boldsymbol{A}) = \frac{1}{\lambda^2}$$

$$P(\boldsymbol{A} \leq c) = 1 - e^{-\lambda c} \text{ and } P(\boldsymbol{A} > c) = e^{-\lambda c}$$

Modeling the Arrival Process

Example: In each of the following cases, determine the average arrival rate per hour, and the average interarrival time in hours.

a) One arrival occurs every 20 minutes.

$$\lambda = \frac{60}{20} = 3 \quad \text{arrivals/hour}$$

$$\frac{1}{\lambda} = \frac{1}{3} \text{ hour/arrival}$$

b) Number of arrivals in a 30-minute period is 10.

$$\lambda = 2 \times 10 = 20 \quad \text{arrivals/hour}$$

$$\frac{1}{\lambda} = \frac{1}{20} \text{ hour/arrival}$$

Modeling the Arrival Process

Lamma 1 (No-Memory Property): If $\boldsymbol{A}$ has an exponential distribution, then for all nonnegative values of t and h,

$$P(\boldsymbol{A} > t + h | \boldsymbol{A} > t) = P(\boldsymbol{A} > h)$$

Proof: The RHS $= P(\boldsymbol{A} > h) = e^{-\lambda h}$. Now,

$$\begin{aligned} \text{LHS} &= P(\boldsymbol{A} > t + h | \boldsymbol{A} > t) \\ &= \frac{P(\boldsymbol{A} > t + h \cap \boldsymbol{A} > t)}{P(\boldsymbol{A} > t)} \\ &= \frac{P(\boldsymbol{A} > t + h)}{P(\boldsymbol{A} > t)} \\ &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h} = \text{RHS} \end{aligned}$$

Modeling the Arrival Process

- The no-memory property of the exponential distribution is important, because it implies that if we want to know the probability distribution of the time until the next arrival, then it does not matter how long it has been since the last arrival.
- This means that to predict future arrival patterns, we need not keep track of how long it has been since the last arrival.

Modeling the Arrival Process

Example: The time between arrivals at the State Revenue Office is exponential with mean value .04 hour. The office opens at $8{:}00_{\text{A.M}}$.

1) Write the exponential distribution that describes the interarrival time.

$$\frac{1}{\lambda} = 0.04 \Rightarrow \lambda = 25$$
$$f(t) = 25e^{-25t}\ ;\ \ t > 0$$

2) Find the probability that no customers will arrive at the office by $8{:}15_{\text{ A.M}}$.

$$P\left(t > \frac{15}{60}\right) = P(t > 0.25) = e^{-25\times 0.25} \approx 0.00193$$

Using Excel `= 1 - EXPON.DIST(0.25, 25, TRUE)`

Modeling the Arrival Process

Example: The time between arrivals at the State Revenue Office is exponential with mean value .04 hour. The office opens at $8{:}00_{\text{A.M}}$.

3) It is now $8{:}35_{\text{ A.M}}$. The last customer entered the office at 8:26. What is the probability that the next customer will arrive before $8{:}38_{\text{ A.M}}$.?

$$P\left(t < \frac{3}{60}\right) = P(t < 0.05) = 1 - e^{-25\times 0.05} \approx 0.713$$

Using Excel `= EXPON.DIST(0.⬩5, 25, TRUE)`

4) What is the average number of arriving customers between 8:10 and $8{:}45_{\text{ A.M}}$?

$$25 \times \frac{45 - 10}{60} = 25 \times \frac{35}{60} \approx 14.58 \text{ arrivals}$$

Modeling the Arrival Process

Relation Between Poisson Distribution and Exponential Distribution

If interarrival times are exponential, the probability distribution of the number of arrivals occurring in any time interval of length t is given by the following important theorem.

Theorem 1

Interarrival times are exponential with parameter λ if and only if the number of arrivals to occur in an interval of length t follows a **Poisson distribution** with parameter λt.

Modeling the Arrival Process

- A discrete random variable $\boldsymbol{N}$ has a **Poisson distribution** with parameter λ if, for $n = 0,1,2,\cdots,$

$$P(\boldsymbol{N} = n) = \frac{e^{-\lambda}\lambda^n}{n!}$$

- If $\boldsymbol{N}$ is a Poisson random variable, it can be shown that $E(\boldsymbol{N}) = \mathrm{var}(\boldsymbol{N}) = \lambda$.
- If we define $\boldsymbol{N}_t$ to be the number of arrivals to occur during any time interval of length t, Theorem 1 states that

$$P(\boldsymbol{N}_t = n) = \frac{e^{-(\lambda t)}(\lambda t)^n}{n!} \quad \text{for } n = 0,1,2,\cdots$$

- Since $\boldsymbol{N}_t$ is Poisson with parameter λt, then $E(\boldsymbol{N}_t) = \mathrm{var}(\boldsymbol{N}_t) = \lambda t$.

Modeling the Arrival Process

Example: The number of cups of coffee ordered per hour at a coffeeshop follows a Poisson distribution, with an average of 30 cups per hour being ordered.

1. Find the probability that exactly 50 cups are ordered between 10 A.M. and 12 midday.

$$\lambda t = (30)(2) = 60$$

$$P(N_2 = 50) = \frac{e^{-60} \cdot 60^{50}}{50!} \approx 0.023271$$

Using Excel `= POISSON.DIST(50, 60, FALSE)`

Modeling the Arrival Process

Example: The number of cups of coffee ordered per hour at a coffeeshop follows a Poisson distribution, with an average of 30 cups per hour being ordered.

2. Find the mean and standard deviation of the number of coffee cups ordered between 9 A.M. and 1 P.M..

$$\text{mean} = \lambda t = (30)(4) = 120$$
$$\text{s.d} = \sqrt{120} \approx 10.95$$

Using Excel `= SQRT(120)`

Modeling the Arrival Process

What assumptions are required for interarrival times to be exponential?

1. if the arrival rate is stationary,
2. if bulk arrivals cannot occur,
3. if past arrivals do not affect future arrivals.

Theorem 2

If the *above assumptions* hold, then $\boldsymbol{N}_t$ follows a Poisson distribution with parameter λt, and interarrival times are exponential with parameter λ.

Modeling the Arrival Process

Notes:

1. For small Δt, the probability of one arrival occurring between times t and $\mathrm{t} + \Delta t$ is $\lambda\Delta t + o(\Delta t)$, where $o(\Delta t)$ refers to any quantity satisfying

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

2. The probability of no arrival during the interval between t and $\mathrm{t} + \Delta t$ is $1 - \lambda\Delta t + o(\Delta t)$.
3. The probability of more than one arrival occurring between t and $\mathrm{t} + \Delta t$ is $o(\Delta t)$.

Modeling the Arrival Process

Example: The number of cups of coffee ordered per hour at a coffeeshop follows a Poisson distribution, with an average of 30 cups per hour being ordered.

3. Find the probability that the time between two consecutive orders is between 1 and 3 minutes.

$$
\begin{aligned}
P\left(\frac{1}{60} \le t \le \frac{3}{60}\right) &= P\left(t \le \frac{3}{60}\right) - P\left(t \le \frac{1}{60}\right) \\
&= \left(1 - e^{-30\cdot\frac{3}{60}}\right) - \left(1 - e^{-30\cdot\frac{1}{60}}\right) \\
&= e^{-0.5} - e^{-1.5} \approx 0.383
\end{aligned}
$$

Using Excel

```
= EXPON.DIST(3/60,30,TRUE) - EXPON.DIST(1/60,30,TRUE)
```

Modeling the Arrival Process

The Erlang Distribution

1. If interarrival times do not appear to be exponential they are often modeled by an *Erlang distribution*.
2. An **Erlang distribution** is a continuous random variable (call it $\boldsymbol{T}$) whose density function $f(t)$ is specified by two parameters:
   - a rate parameters R,
   - and a shape parameters k (k must be a positive integer).

Modeling the Arrival Process

The Erlang Distribution

3. Given values of R and k, the Erlang density has the following probability density function:

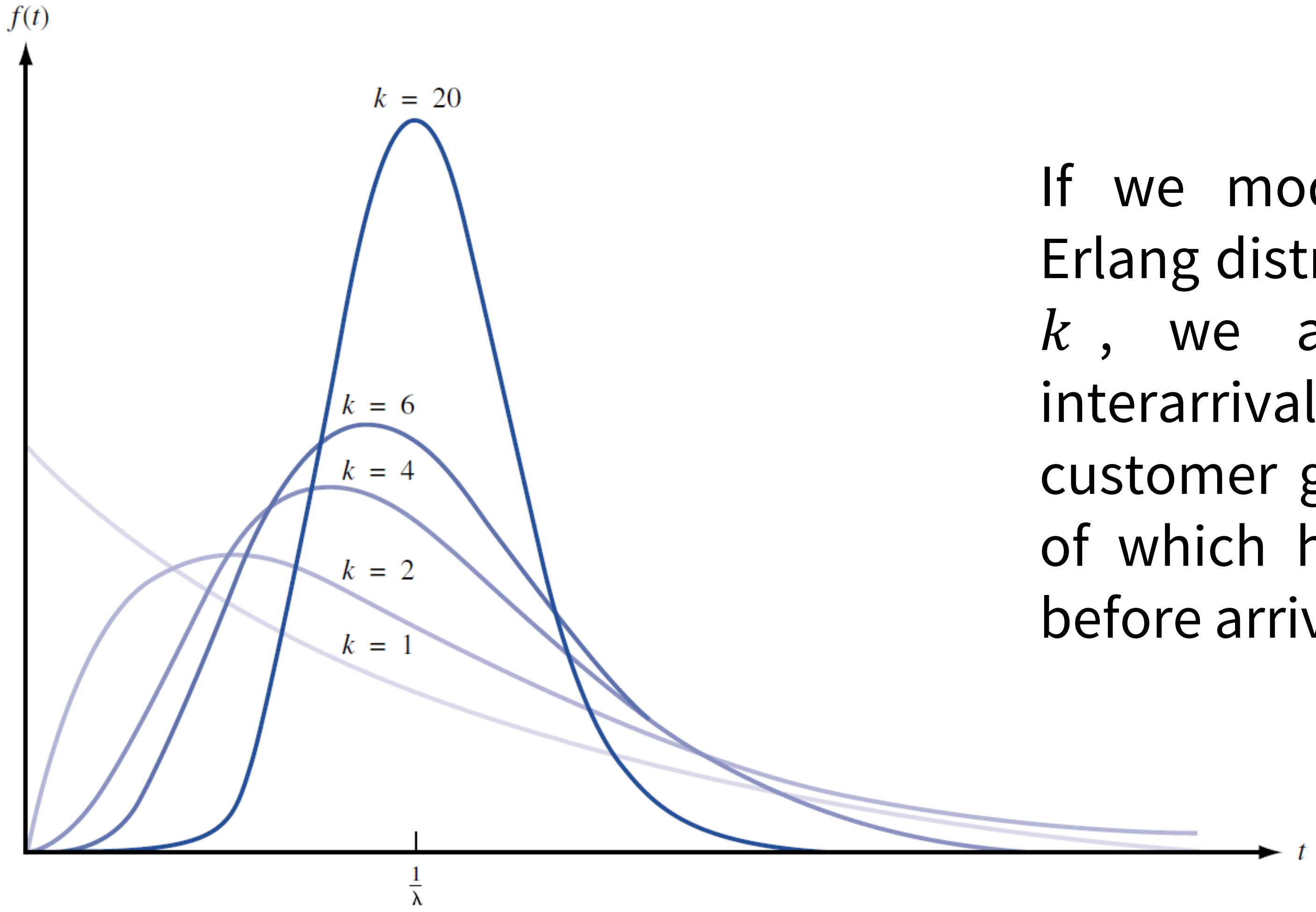
$$f(t) = \frac{R\,(Rt)^{k-1}e^{-Rt}}{(k-1)!}$$

4. Using integration by parts, we can show that if $\boldsymbol{T}$ is an Erlang distribution with rate parameter R and shape parameter k, then

$$E(\boldsymbol{T}) = \frac{k}{R} \quad \text{and} \quad \mathrm{var}(\boldsymbol{T}) = \frac{k}{R^2}$$

Modeling the Arrival Process

The Erlang Distribution



If we model interarrival times as an Erlang distribution with shape parameter k, we are really saying that the interarrival process is equivalent to a customer going through k **phases** (each of which has the no-memory property) before arriving.

Modeling the Service Process

- We assume that the service times of different customers are independent random variables and that each customers service time is governed by a random variable $\boldsymbol{S}$ having a density function $s(t)$.
- We let $\frac{1}{\mu} = \int_0^\infty ts(t)dt$ be then **mean service time** for a customer.
- The variable $\frac{1}{\mu}$ will have units of *hours per customer*, so μ has units of *customers per hour*. For this reason, we call μ **the service rate**.

Modeling the Service Process

- **For example**, $\mu = 5$ means that if customers were always present, the server could serve an average of 5 customers per hour, and the average service time of each customer would be $\frac{1}{5}$ hour.
- Unfortunately, actual service times may not be consistent with the no-memory property.
- For this reason, we often assume that $s(t)$ is an Erlang distribution with shape parameters k and rate parameter $k\mu$.

Modeling the Service Process

- In certain situations, interarrival or service times may be modeled as having *zero variance*; in this case, interarrival or service times are considered to be **deterministic**.
- If interarrival times are deterministic, then each interarrival time will be exactly $\frac{1}{\lambda}$, and if service times are deterministic, each customers service time is exactly $\frac{1}{\mu}$.

The Kendall–Lee Notation for Queuing Systems

- Standard notation used to describe many queuing systems.
- The notation is used to describe a queuing system in which all arrivals wait in a single line until one of s identical parallel servers is free. Then the first customer in line enters service, and so on.
- To describe such a queuing system, Kendall devised a notation in which each queuing system is described by six characters: 1/2/3/4/5/6

The Kendall–Lee Notation for Queuing Systems

1 / **2** / **3** / **4** / **5** / **6**

- Specifies the **nature of the arrival process**.
- The following standard abbreviations are used:

M = Interarrival times are independent, identically exponential distributed (iid)

D = Interarrival times are iid and deterministic

E_k = Interarrival times are iid Erlangs with shape parameter k.

GI = Interarrival times are iid and governed by some general distribution

The Kendall–Lee Notation for Queuing Systems

1 / 2 / 3 / 4 / 5 / 6

- Specifies the **nature of the service process**.
- The following standard abbreviations are used:

M = Service times are iid and exponentially distributed

D = Service times are iid and deterministic

E_k = Service times are iid Erlangs with shape parameter k.

GI = Service times are iid and governed by some general distribution

The Kendall–Lee Notation for Queuing Systems

1 / 2 / 3 / 4 / 5 / 6

- Specifies the **number of parallel servers.**

The Kendall–Lee Notation for Queuing Systems

1 / 2 / 3 / 4 / 5 / 6

- Specifies the **queue discipline**.
- The following standard abbreviations are used:

$$FCFS = \text{First come, first served}$$

$$LCFS = \text{Last come, first served}$$

$$SIRO = \text{Service in random order.}$$

$$GD = \text{General queue discipline}$$

The Kendall–Lee Notation for Queuing Systems

1 / 2 / 3 / 4 / 5 / 6

- Specifies the **maximum allowable number of customers in the system**.

- Specifies the **size of the population from which customers are drawn**.

The Kendall–Lee Notation for Queuing Systems

- In many important models $4/5/6$ is $GD/1/1$. If this is the case, then $4/5/6$ is often omitted.
- **Example:** $M/E_2/8/FCFS/10/\infty$ might represent a *health clinic* with 8 doctors, exponential interarrival times, two-phase Erlang service times, an FCFS queue discipline, and a total capacity of 10 patients.

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [20]
Queuing Theory

Section: [20.3]
Birth–Death Processes

Introduction

- We use **birth-death** processes to answer questions about several different types of queuing systems.
- We define the number of people present in any queuing system at time t to be the **state of the queuing systems at time** t.
- We call π_j the **steady state (equilibrium) probability**, of state j.
- For $t = 0$, the state of the system will equal the number of people initially present in the system.
- The quantity $P_{ij}(t)$ is defined as the probability that j people will be present in the queuing system at time t, given that at time 0, i people are present.

Introduction

- Note that $P_{ij}(t)$ is **analogous (مُماثل)** to the $n-$step transition probability $P_{ij}(t)$ (the probability that after n transitions, a Markov chain will be in state j, given that the chain began in state i).
- The behavior of $P_{ij}(t)$ before the steady state is reached is called **the transient behavior of the queuing system**.
- The question of how large t must be before the steady state is approximately reached is difficult to answer. For now, when we analyze the behavior of a queuing system, we assume that the steady state has been reached.

Introduction

- For a birth–death process, it is easy to determine the steady-state probabilities (if they exist).
- A **birth-death process** is a continuous-time stochastic process for which the system's state at any time is a nonnegative integer.
- If a birth–death process is in state j at time t, then the motion of the process is governed by the following laws.

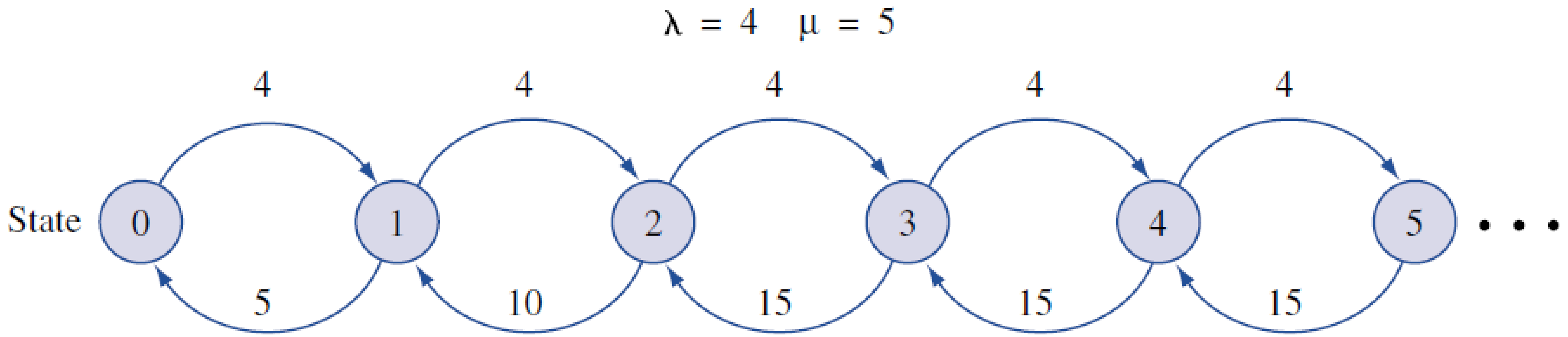
Laws of Motion for Birth–Death Processes

- A **birth** increases the system state by 1, to $j + 1$.
  - The variable j is called the birth rate in state j.
  - In most queuing systems, a birth is simply an *arrival*.
- A **death** decreases the system state by 1, to $j - 1$.
  - The variable j is the death rate in state j.
  - In most queuing systems, a death is a *service completion*.
  - Note that $\mu_0 = 0$ must hold, or a negative state could occur.
- Births and deaths are **independent** of each other.

Laws of Motion for Birth–Death Processes

Example: Consider $M/M/3/FCFS/\infty/\infty$ queuing system in which interarrival times are exponential with $\lambda = 4$ and service times are exponential with $\mu = 5$. To model this system as a birth-death process, we would use the following parameters

$$\lambda_j = 4 \text{ for } j = 0,1,2,\cdots$$
$$\mu_0 = 0$$
$$\mu_1 = 5$$
$$\mu_2 = 10$$
$$\mu_j = 15 \text{ for } j = 3,4,5,\cdots$$



Computing Steady-State Probabilities for Birth-Death Processes

In steady-state, the following balance equation must hold for every state j:

Rate IN = Rate OUT

- **At state 0:**

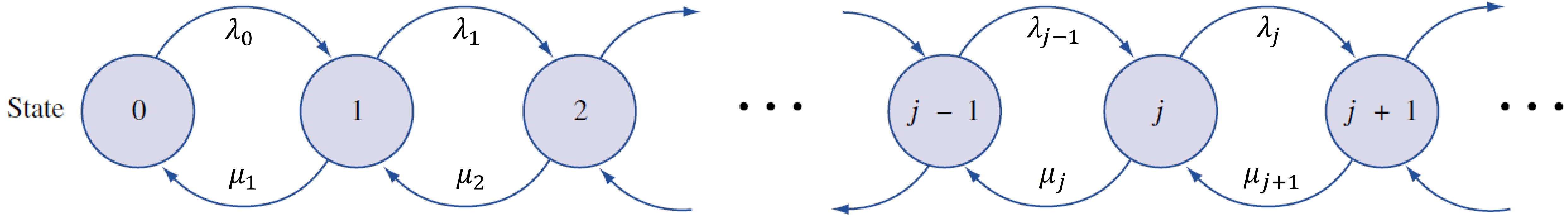
$$\lambda_0 \pi_0 = \mu_1 \pi_1$$

- **At state j:**

$$(\lambda_j + \mu_j)\pi_j = \lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1}$$

for $j = 1,2,3,\cdots$

- **Remember:** $\sum \pi_j = 1$



Computing Steady-State Probabilities for Birth-Death Processes

We obtain the flow balance equations for a birth–death process:

$$
\begin{aligned}
\lambda_0\pi_0 &= \mu_1\pi_1 && (j = 0)\\
(\lambda_1 + \mu_1)\pi_1 &= \lambda_0\pi_0 + \mu_2\pi_2 && (j = 1)\\
(\lambda_2 + \mu_2)\pi_2 &= \lambda_1\pi_1 + \mu_3\pi_3 && (j = 2)\\
&\ \vdots && \quad\vdots\\
(\lambda_j + \mu_j)\pi_j &= \lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1} && (\text{any } j)
\end{aligned}
$$

Solution of Birth–Death Flow Balance Equations

- To solve the previous set of equations, we begin by expressing all the π_j's in terms of π_0.
- From the equation $\lambda_0\pi_0 = \mu_1\pi_1$, we obtain $\pi_1 = \frac{\lambda_0\pi_0}{\mu_1}$.
- Substituting this result into the equation $(\lambda_1 + \mu_1)\pi_1 = \lambda_0\pi_0 + \mu_2\pi_2$ yields:

$$(\lambda_1 + \mu_1)\frac{\lambda_0\pi_0}{\mu_1} = \lambda_0\pi_0 + \mu_2\pi_2$$

$$\pi_2 = \frac{\lambda_0\lambda_1\pi_0}{\mu_1\mu_2}$$

- In general,

$$\pi_j = \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j}\pi_0$$

Solution of Birth–Death Flow Balance Equations

- If we let $c_j = \frac{\lambda_0 \lambda_1 \cdots \lambda_{j-1}}{\mu_1 \mu_2 \cdots \mu_j}$, we obtain $\pi_j = c_j \pi_0$ for $j = 1,2,3,\cdots$.
- Since at any given time, we must be in some state, the steady-state probabilities must sum to 1:

$$\begin{aligned} \pi_0 + \pi_1 + \pi_2 + \cdots &= 1 \\ \pi_0 + c_1\pi_0 + c_2\pi_0 + c_3\pi_0 + \cdots &= 1 \\ \pi_0(1 + c_1 + c_2 + c_3 + \cdots) &= 1 \\ \pi_0 &= \frac{1}{1 + \sum_{j=1}^{\infty} c_j} \end{aligned}$$

- If $\sum_{j=1}^{\infty} c_j$ is finite (converges), then we can obtain π_0.
- It can be shown that if $\sum_{j=1}^{\infty} c_j$ is infinite (diverges), then no steady-state distribution exists.

Solution of Birth–Death Flow Balance Equations

Example: A grocery operates with three check-out counters. The manager uses the following schedule to determine the number of counters in operation, depending on the number of customers in store:

Number of customers in store	**Number of counters in operation**
1, 2, 3	1
4, 5, 6	2
More that 6	3

Customers arrive in the counters area according to a Poisson distribution with mean rate 10 customers per hour. The average check-out time per customer is exponential with mean 12 minutes. Determine the steady-state probability π_n of n customers in check-out area.

$$\lambda_n = 10 \text{ for } n = 0,1,2,\cdots$$

$$\mu_1 = \mu_2 = \mu_3 = \frac{60}{12} = 5$$

$$\mu_4 = \mu_5 = \mu_6 = 2 \times 5 = 10$$

$$\mu_j = 3 \times 5 = 15 \text{ for } j = 7,8,9,\cdots$$

Solution of Birth–Death Flow Balance Equations

Example: (*Continue*)

$$\lambda_n = 10 \text{ for } n = 0,1,2,\cdots$$

$$\mu_1 = \mu_2 = \mu_3 = \frac{60}{12} = 5$$

$$\mu_4 = \mu_5 = \mu_6 = 2 \times 5 = 10$$

$$\mu_j = 3 \times 5 = 15 \text{ for } j = 7,8,9,\cdots$$

$$c_1 = \frac{\lambda_0}{\mu_1} = \frac{10}{5} = 2$$

$$c_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} = \frac{\lambda_1}{\mu_2} c_1 = \frac{10 \cdot 2}{5} = 4$$

$$c_3 = \frac{\lambda_2}{\mu_3} c_2 = \frac{10 \cdot 4}{5} = 8$$

$$c_4 = \frac{\lambda_3}{\mu_4} c_3 = \frac{10 \cdot 8}{10} = 8$$

$$c_5 = \frac{\lambda_4}{\mu_5} c_4 = \frac{10 \cdot 8}{10} = 8$$

$$c_6 = \frac{\lambda_5}{\mu_6} c_5 = \frac{10 \cdot 8}{10} = 8$$

$$c_7 = \frac{\lambda_6}{\mu_7} c_6 = \frac{10 \cdot 8}{15} = 8 \cdot \left(\frac{2}{3}\right)$$

$$c_8 = \frac{\lambda_7}{\mu_8} c_7 = \frac{10 \cdot 8 \cdot \left(\frac{2}{3}\right)}{15} = 8 \cdot \left(\frac{2}{3}\right)^2$$

$$\vdots$$

$$c_n = 8 \cdot \left(\frac{2}{3}\right)^{n-6} \quad \text{for } n = 7,8,9,\cdots$$

Solution of Birth–Death Flow Balance Equations

Example: (*Continue*)

$$c_1 = 2\ ,\ c_2 = 4\ ,\ c_3 = 8\ ,\ c_4 = 8\ ,\ c_5 = 8\ ,\ c_6 = 8\ ,\ \ c_n = 8 \cdot \left(\frac{2}{3}\right)^{n-6} \quad \text{for } n = 7,8,9,\cdots$$

To evaluate π_0 we need to find $\sum_{j=1}^{\infty} c_j$:

$$\sum_{j=1}^{\infty} c_j = c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + \underbrace{c_7 + c_8 + \cdots}_{\sum_{j=7}^{\infty} 8 \cdot \left(\frac{2}{3}\right)^{n-6} = \sum_{j=0}^{\infty} 8 \cdot \left(\frac{2}{3}\right)^{n+1}}$$

$$\sum_{j=1}^{\infty} c_j = 2 + 4 + 8 + 8 + 8 + 8 + \frac{8 \cdot \left(\frac{2}{3}\right)}{1 - \frac{2}{3}} = 38 + 16 = 54$$

Solution of Birth–Death Flow Balance Equations

Example: (*Continue*)

$$c_1 = 2\ ,\ c_2 = 4\ ,\ c_3 = 8\ ,\ c_4 = 8\ ,\ c_5 = 8\ ,\ c_6 = 8\ ,\ \ c_n = 8 \cdot \left(\frac{2}{3}\right)^{n-6} \quad \text{for } n = 7,8,9,\cdots$$

$$\therefore\ \pi_0 = \frac{1}{1 + \sum_{j=1}^{\infty} c_j} = \frac{1}{1+54} = \frac{1}{55} = 0.0\overline{18}$$

$$\pi_1 = c_1\pi_0 = \frac{2}{55}$$

$$\pi_2 = c_2\pi_0 = \frac{4}{55}$$

$$\pi_n = c_n\pi_0 = \frac{8}{55} \cdot \left(\frac{2}{3}\right)^{n-6} \quad \text{for } n = 7,8,\cdots$$

$$\pi_3 = c_3\pi_0 = \frac{8}{55} = \pi_4 = \pi_5 = \pi_6$$

Solution of Birth–Death Flow Balance Equations

Example: (*Continue using* Excel)

STATE J	LAMBDA	MU	CJ		PROB.	
0	10	0	1.00000		0.0182	=1/SUM(D2:D22)
1	10	5	2.00000	=B2/C3*D2	0.0364	=D3*F2
2	10	5	4.00000	=B3/C4*D3	0.0728	=D4*F2
3	10	5	8.00000	=B4/C5*D4	0.1456	=D5*F2
4	10	10	8.00000	=B5/C6*D5	0.1456	=D6*F2
5	10	10	8.00000	=B6/C7*D6	0.1456	=D7*F2
6	10	10	8.00000	=B7/C8*D7	0.1456	=D8*F2
7	10	15	5.33333	=B8/C9*D8	0.097066	=D9*F2
8	10	15	3.55556	=B9/C10*D9	0.064711	=D10*F2
9	10	15	2.37037	=B10/C11*D10	0.043141	=D11*F2
10	10	15	1.58025	=B11/C12*D11	0.02876	=D12*F2
11	10	15	1.05350	=B12/C13*D12	0.019174	=D13*F2
12	10	15	0.70233	=B13/C14*D13	0.012782	=D14*F2
13	10	15	0.46822	=B14/C15*D14	0.008522	=D15*F2
14	10	15	0.31215	=B15/C16*D15	0.005681	=D16*F2
15	10	15	0.20810	=B16/C17*D16	0.003787	=D17*F2
16	10	15	0.13873	=B17/C18*D17	0.002525	=D18*F2
17	10	15	0.09249	=B18/C19*D18	0.001683	=D19*F2
18	10	15	0.06166	=B19/C20*D19	0.001122	=D20*F2
19	10	15	0.04111	=B20/C21*D20	0.000748	=D21*F2
20	10	15	0.02740	=B21/C22*D21	0.000499	=D22*F2

Solution of Birth–Death Flow Balance Equations

Example: Indiana Bell customer service representatives receive an average of 1,700 calls per hour. The time between calls follows an exponential distribution. A customer service representative can handle an average of 30 calls per hour. The time required to handle a call is also exponentially distributed. Indiana Bell can put up to 25 people on hold. If 25 people are on hold, a call is lost to the system. Indiana Bell has 75 service representatives.

1. What fraction of the time are all operators busy?
2. What fraction of all calls are lost to the system?

$$\lambda_j = 1700 \text{ for } j = 0,1,\cdots,99 \qquad \mu_j = 30j \text{ for } j = 1,2,\cdots,75$$

$$\lambda_{100} = 0 \qquad \mu_j = 30 \times 75 = 2250 \text{ for } j > 75$$

Solution of Birth–Death Flow Balance Equations

Example: Using Excel:

1. What fraction of the time are all operators busy?

   ```
   =SUMIF(A2:A102,">75",H2:H102)
   ```

   ```
   0.009638
   ```

2. What fraction of all calls are lost to the system?

   ```
   =H102
   ```

   ```
   2.82465E-06
   ```

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [20]

Queuing Theory

Section: [20.4]

The M/M/1/GD/∞/∞ Queuing System and the Queuing Formula $L = \lambda W$

Derivation of Steady-State Probabilities

- An $M/M/1/GD/\infty/\infty$ queuing system may be modeled as a birth-death process with the following parameters:

$$\lambda_j = \lambda \quad (j = 0,1,2,\cdots)$$

$$\mu_0 = 0$$

$$\mu_j = \mu \quad (j = 1,2,3,\cdots)$$

- Then, the steady-state probabilities are:

$$\pi_1 = \frac{\lambda}{\mu}\pi_0 \quad , \quad \pi_2 = \frac{\lambda^2}{\mu^2}\pi_0 \quad , \quad \pi_3 = \frac{\lambda^3}{\mu^3}\pi_0 \quad , \quad \cdots \quad , \quad \pi_j = \frac{\lambda^j}{\mu^j}\pi_0$$

- Define $\rho = \frac{\lambda}{\mu}$. Then we obtain

$$\pi_1 = \rho\pi_0 \quad , \quad \pi_2 = \rho^2\pi_0 \quad , \quad \pi_3 = \rho^3\pi_0 \quad , \quad \cdots \quad , \quad \pi_j = \rho^j\pi_0$$

Derivation of Steady-State Probabilities

- Since $\pi_0 + \pi_1 + \pi_2 + \cdots = 1$, then:

$$\pi_0(1 + \rho + \rho^2 + \cdots) = 1$$
$$\pi_0\left(\frac{1}{1-\rho}\right) = 1 \ \text{ if } |\rho| < 1$$
$$\pi_0 = 1 - \rho \ \text{ if } \rho < 1 \text{ or } 0 \leq \lambda < \mu$$

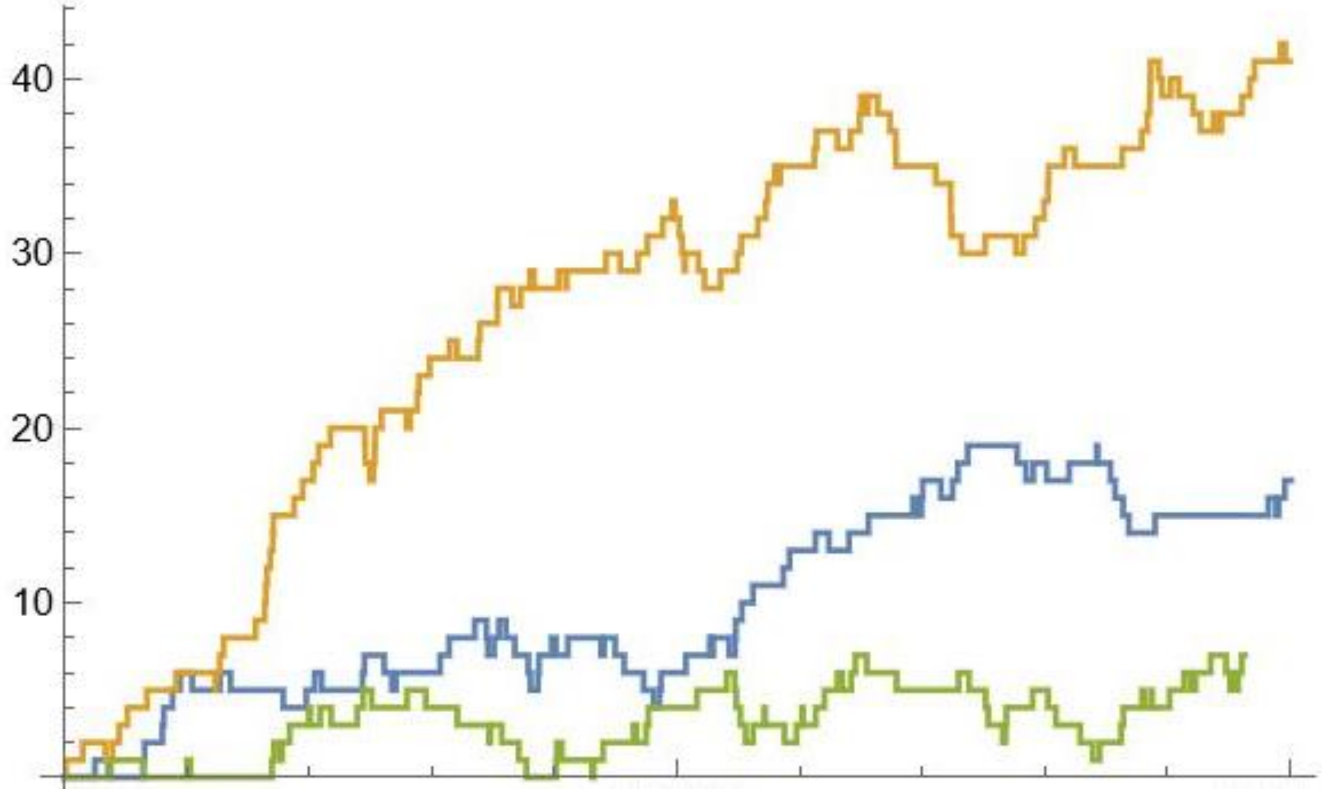
- So that, $\pi_j = \rho^j(1-\rho)$ where $0 \leq \rho < 1$.
- If $\rho \geq 1$, **no** steady-state distribution exists

Derivation of Steady-State Probabilities

Example: if $\lambda = 6$ and $\mu = 4$, then $\rho = 1.5 \geq 1$. Even if the server were working all the time, she could only serve an average of 4 people per hour. Thus, the average number of customers in the system would grow by at least $6 - 4 = 2$ customers per hour. This means that after a long time, the number of customers present would "blow up," and no steady-state distribution could exist.

Derivation of Steady-State Probabilities

```
model = QueueingProcess[6, 4];
data = RandomFunction[model, {0, 10}, 3];
ListStepPlot[data]
```



Derivation of L

- Throughout the rest of this section, we assume that $\rho < 1$, ensuring that a steady-state probability distribution does exist.
- We now use the steady-state probability distribution to determine several quantities of interest.
- **For example,** assuming that the steady state has been reached, **the average number of customers present in the queuing system** (call it L) is given by

$$L = \sum_{j=0}^{\infty} j\pi_j = \sum_{j=0}^{\infty} j\rho^j(1-\rho) = (1-\rho)\sum_{j=0}^{\infty} j\rho^j$$

Derivation of L

- Define $S = \sum_{j=0}^{\infty} j\rho^j = \rho + 2\rho^2 + 3\rho^3 + \cdots$
- Multiply S by ρ to obtain $\rho S = \rho^2 + 2\rho^3 + 3\rho^4 + \cdots$
- So,

$$
\begin{array}{rl}
S = & \rho + 2\rho^2 + 3\rho^3 + \cdots \\
-\quad \rho S = & \rho^2 + 2\rho^3 + 3\rho^4 + \cdots \\
\hline
S - \rho S = & \rho + \rho^2 + \rho^3 + \cdots = \dfrac{\rho}{1-\rho}
\end{array}
$$

$$\boxed{S = \frac{\rho}{(1-\rho)^2}}$$

$$L = (1-\rho)\sum_{j=0}^{\infty} j\rho^j = (1-\rho)\frac{\rho}{(1-\rho)^2} = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$$

Derivation of L_q

- In some situations, we are interested in **the expected number of people waiting in the queue**, and we denote this number by L_q.
- Note that if 0 or 1 customer is present in the system, then nobody is waiting in line, but if j people are present $(j \geq 1)$, there will be $j-1$ people waiting in line.
- Thus, if we are in the steady state,

$$L_q = \sum_{j=1}^{\infty}(j-1)\pi_j = \sum_{j=1}^{\infty} j\pi_j - \sum_{j=1}^{\infty}\pi_j$$

$$= \frac{\rho}{1-\rho} - (1-\pi_0) = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

Derivation of L_s

- Also of interest is L_s, **the expected number of customers in service**.
- For *any* queuing system,

$$L = L_s + L_q$$

$$\frac{\rho}{1-\rho} = L_s + \frac{\rho^2}{1-\rho}$$

$$L_s = \frac{\rho}{1-\rho} - \frac{\rho^2}{1-\rho} = \rho = \frac{\lambda}{\mu}$$

The Queuing Formula $L = \lambda W$

- Often we are interested in **the amount of time that a typical customer spends in a queuing system**.
- We define W as the expected time a customer spends in the queuing system, *including* time in line plus time in service, and W_q as the expected time a customer spends waiting in line.
- Both W and W_q are computed under the assumption that the steady state has been reached.
- By using a powerful result known as **Little's queuing formula**, W and W_q may be easily computed from L and L_q.

The Queuing Formula $L = \lambda W$

We first define (for any queuing system or any subset of a queuing system) the following quantities:

- $\lambda =$ average number of arrivals entering the system per unit time.
- $L =$ average number of customers present in the queuing system.
- $L_q =$ average number of customers waiting in line.
- $L_s =$ average number of customers in service.
- $W =$ average time a customer spends in the system.
- $W_q =$ average time a customer spends in line.
- $W_s =$ average time a customer spends in service.

The Queuing Formula $L = \lambda W$

Theorem: For any queuing system in which a steady-state distribution exists, the following relations hold:

$$L = \lambda W$$

$$L_q = \lambda W_q$$

$$L_s = \lambda W_s$$

- For an $M/M/1/GD/\infty/\infty$ queuing system, we have

$$W = \frac{1}{\mu - \lambda} \quad , \quad W_q = \frac{\lambda}{\mu(\mu - \lambda)} \quad , \quad W_s = \frac{1}{\mu}$$

The Queuing Formula $L = \lambda W$

Example [1]: An average of 10 cars per hour arrive at a single-server drive-in teller. Assume that the average service time for each customer is 4 minutes, and both interarrival times and service times are exponential. Answer the following questions:

$$\lambda = 10$$
$$\mu = \frac{60}{4} = 15$$
$$\rho = \frac{\lambda}{\mu} = \frac{10}{15} = \frac{2}{3}$$

1. What is the probability that the teller is idle?

   Solution: $\pi_0 = 1 - \rho = 1 - \frac{2}{3} = \frac{1}{3}$

2. What is the average number of cars waiting in line for the teller? (A car that is being served is not considered to be waiting in line.)

   Solution: $L_q = \frac{\rho^2}{1-\rho} = \frac{\left(\frac{2}{3}\right)^2}{1-\frac{2}{3}} = \frac{4}{3}$ customers

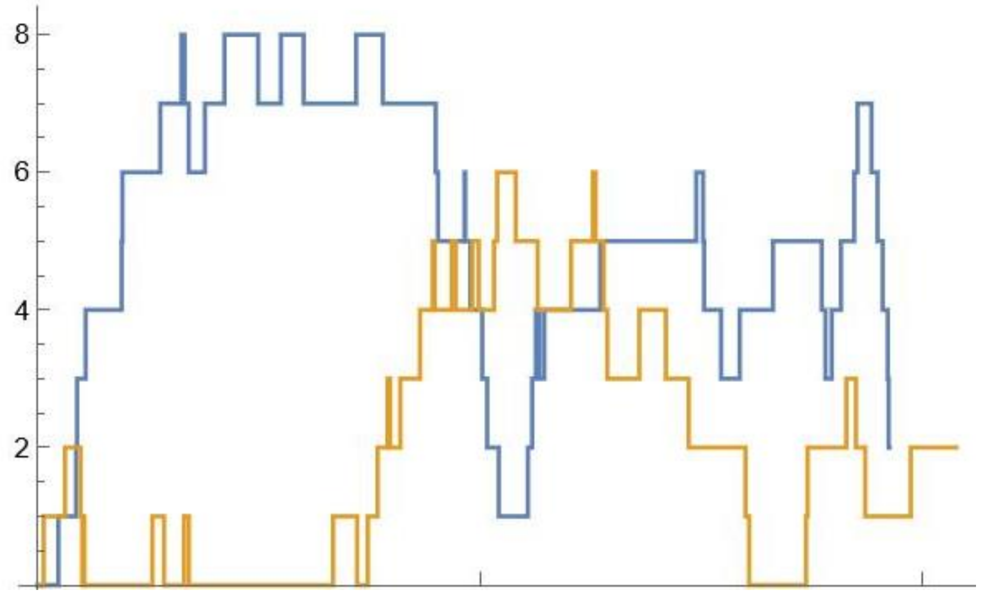
The Queuing Formula $L = \lambda W$

Example [1]: An average of 10 cars per hour arrive at a single-server drive-in teller. Assume that the average service time for each customer is 4 minutes, and both interarrival times and service times are exponential. Answer the following questions:

3. What is the average amount of time a drive-in customer spends in the bank parking lot (including time in service)?

Solution: $W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda} = \frac{1}{5}$ Hours $= 12$ minutes

```
model = QueueingProcess[10, 15];
data = RandomFunction[model, {0, 2}, 2];
ListStepPlot[data]
```



The Queuing Formula $L = \lambda W$

Example [2]: Our local MacDonald's uses an average of 10,000 pounds of potatoes per week. The average number of pounds of potatoes on hand is 5,000. On the average, how long do potatoes stay in the restaurant before being used? Find μ.

Solution:

$$W = \frac{L}{\lambda} = \frac{5000}{10000} = \frac{1}{2} \text{ Week}$$

$$L = \frac{\lambda}{\mu - \lambda} \Rightarrow 5000 = \frac{10000}{\mu - 10000} \Rightarrow \mu = 10002$$

The Queuing Formula $L = \lambda W$

Example [3]: Suppose that all car owners fill up when their tanks are exactly half full. At the present time, an average of 7.5 customers per hour arrive at a single-pump gas station. It takes an average of 4 minutes to service a car. Assume that interarrival times and service times are both exponential.

$$\lambda = 7.5$$
$$\mu = \frac{60}{4} = 15$$
$$\rho = \frac{\lambda}{\mu} = \frac{7.5}{15} = \frac{1}{2}$$

1. For the present situation, compute L and W.
   Solution:

$$L = \frac{\rho}{1 - \rho} = 1$$

$$W = \frac{L}{\lambda} = \frac{2}{15} \approx 0.1\overline{3} \text{ hours} = 8 \text{ minutes}$$

The Queuing Formula $L = \lambda W$

Example [3]:

2. Suppose that a gas shortage occurs and panic buying takes place. To model this phenomenon, suppose that all car owners now purchase gas when their tanks are exactly three-quarters full. Since each car owner is now putting less gas into the tank during each visit to the station, we assume that the average service time has been reduced to $10/3$ minutes. How has panic buying affected L and W?

Solution:

$$\lambda = 2 \times 7.5 = 15$$

$$\mu = \frac{60}{10/3} = 18$$

$$\rho = \frac{\lambda}{\mu} = \frac{15}{18} = \frac{5}{6}$$

$$L = \frac{\rho}{1-\rho} = 5$$

$$W = \frac{L}{\lambda} = \frac{5}{15} \approx 0.\bar{3} \text{ hours} = 20 \text{ minutes}$$

Thus, panic buying has caused long lines.

The Queuing Formula $L = \lambda W$

M/M/1 QUEUE			Service Station Example	
The Results				
LAMBDA	MU	RHO	STATE j	PI_j
7.5	15	1/2	0	0.500000000000
			1	0.250000000000
L	1.0000		2	0.125000000000
Lq	0.5000		3	0.062500000000
Ls	0.5000		4	0.031250000000
W	0.1333		5	0.015625000000
Wq	0.0667		6	0.007812500000
Ws	0.0667		7	0.003906250000
			8	0.001953125000
			9	0.000976562500
			10	0.000488281250
			11	0.000244140625
			12	0.000122070313
			13	0.000061035156
			14	0.000030517578
			15	0.000015258789

M/M/1 QUEUE			Service Station Example	
The Results				
LAMBDA	MU	RHO	STATE j	PI_j
15	18	5/6	0	0.166666666667
			1	0.138888888889
L	5.0000		2	0.115740740741
Lq	4.1667		3	0.096450617284
Ls	0.8333		4	0.080375514403
W	0.3333		5	0.066979595336
Wq	0.2778		6	0.055816329447
Ws	0.0556		7	0.046513607872
			8	0.038761339894
			9	0.032301116578
			10	0.026917597148
			11	0.022431330957
			12	0.018692775797
			13	0.015577313165
			14	0.012981094304
			15	0.010817578586

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [20]

Queuing Theory

Section: [20.5]

The $M/M/1/GD/c/\infty$ Queuing System

The Steady-State Probabilities

- The $M/M/1/GD/c/\infty$ system is identical to the $M/M/1/GD/\infty/\infty$ system except for the fact that when c customers are present, all arrivals are turned away and are forever lost to the system.
- An $M/M/1/GD/c/\infty$ queuing system may be modeled as a birth-death process with the following parameters:

$$\lambda_j = \lambda \quad (j = 0,1,2,\cdots,c-1)$$

$$\lambda_c = 0$$

$$\mu_0 = 0$$

$$\mu_j = \mu \quad (j = 1,2,3,\cdots,c)$$

The Steady-State Probabilities

- The steady-state probabilities are:

$$\pi_1 = \frac{\lambda}{\mu}\pi_0 \quad , \quad \pi_2 = \frac{\lambda^2}{\mu^2}\pi_0 \quad , \quad \pi_3 = \frac{\lambda^3}{\mu^3}\pi_0 \quad , \quad \cdots \quad , \quad \pi_c = \frac{\lambda^c}{\mu^c}\pi_0$$

- Define $\rho = \frac{\lambda}{\mu}$. Then we obtain

$$\pi_1 = \rho\pi_0 \quad , \quad \pi_2 = \rho^2\pi_0 \quad , \quad \pi_3 = \rho^3\pi_0 \quad , \quad \cdots \quad , \quad \pi_c = \rho^c\pi_0$$

- Since $\pi_0 + \pi_1 + \pi_2 + \cdots + \pi_c = 1$, then:

$$\pi_0(1 + \rho + \rho^2 + \cdots + \rho^c) = 1$$

$$\pi_0\left(\frac{1 - \rho^{c+1}}{1 - \rho}\right) = 1 \;\text{ if } \rho \neq 1$$

$$\pi_0 = \frac{1 - \rho}{1 - \rho^{c+1}} \;\text{ if } \rho \neq 1 \text{ or } \lambda \neq \mu$$

The Steady-State Probabilities

- So if $\lambda \neq \mu$, the steady-state probabilities for the $M/M/1/GD/c/\infty$ model are given by

$$\pi_0 = \frac{1-\rho}{1-\rho^{c+1}}$$

$$\pi_j = \frac{\rho^j(1-\rho)}{1-\rho^{c+1}} \text{ for } j = 1,2,\cdots,c$$

$$\pi_j = 0 \text{ for } j = c+1, c+2, \cdots$$

- Also, we can show that when $\lambda \neq \mu$,

$$L = \frac{\rho[1-(c+1)\rho^c + c\rho^{c+1}]}{(1-\rho^{c+1})(1-\rho)}$$

The Steady-State Probabilities

- If $\lambda = \mu$, then all the c_j's equal 1, and all the π_j's must be equal, and the steady-state probabilities for the $M/M/1/GD/c/\infty$ system are

$$\pi_j = \frac{1}{c+1} \text{ for } j = 0,1,2,\cdots,c$$

$$L = \frac{c}{2}$$

- For the other formulas, we have:

$$L_s = 1 - \pi_0 \quad , \quad L_q = L - L_s \quad ,$$

$$W = \frac{L}{\lambda(1-\pi_c)} \quad , \quad W_q = \frac{L_q}{\lambda(1-\pi_c)} \quad , \quad W_s = W - W_q$$

- For an $M/M/1/GD/c/\infty$ system, a steady state will exist even if $\lambda \geq \mu$. This is because the finite capacity of the system prevents the number of people in the system from "blowing up."

The Steady-State Probabilities

Important Note:

- Recall that, λ represents the average number of customers per unit time who actually enter the system.
- In our finite capacity model, an average of λ arrivals per unit time arrive, but $\lambda\pi_c$ of these arrivals find the system filled to capacity and leave.
- Thus, an average of $\lambda - \lambda\pi_c = \lambda(1 - \pi_c)$ arrivals per unit time will actually enter the system.

Example

Barber Shop: A one-man barber shop has a total of 10 seats. Interarrival times are exponentially distributed, and an average of 20 prospective customers arrive each hour at the shop. Those customers who find the shop full do not enter. The barber takes an average of 12 minutes to cut each customer's hair. Haircut times are exponentially distributed.

1. On the average, how many haircuts per hour will the barber complete?

   Solution: $\pi_0 = \frac{1-\rho}{1-\rho^{c+1}} = \frac{1-4}{1-4^{11}} = \frac{1}{1398101}$.

   Also, $\pi_{10} = \frac{\rho^j(1-\rho)}{1-\rho^{c+1}} = \frac{4^{10}(1-4)}{1-4^{11}} = 0.75$. Thus, an average of $20(1-0.75) = 5$ customers per hour will receive haircuts.

$c = 10$

$\lambda = 20$

$\mu = \frac{60}{12} = 5$

$\rho = \frac{\lambda}{\mu} = \frac{20}{5} = 4$

Example

Barber Shop: A one-man barber shop has a total of 10 seats. Interarrival times are exponentially distributed, and an average of 20 prospective customers arrive each hour at the shop. Those customers who find the shop full do not enter. The barber takes an average of 12 minutes to cut each customer's hair. Haircut times are exponentially distributed.

2. On the average, how much time will be spent in the shop by a customer who enters?

   Solution: To determine W, we compute

$$L = \frac{4[1-(11)4^{10}+10\cdot 4^{11}]}{(1-4^{11})(1-4)} = 9.67 \text{ customers}$$

So, $W = \frac{L}{\lambda(1-\pi_c)} = \frac{9.67}{20(1-0.75)} = 1.93 \text{ hours}$

$c = 10$

$\lambda = 20$

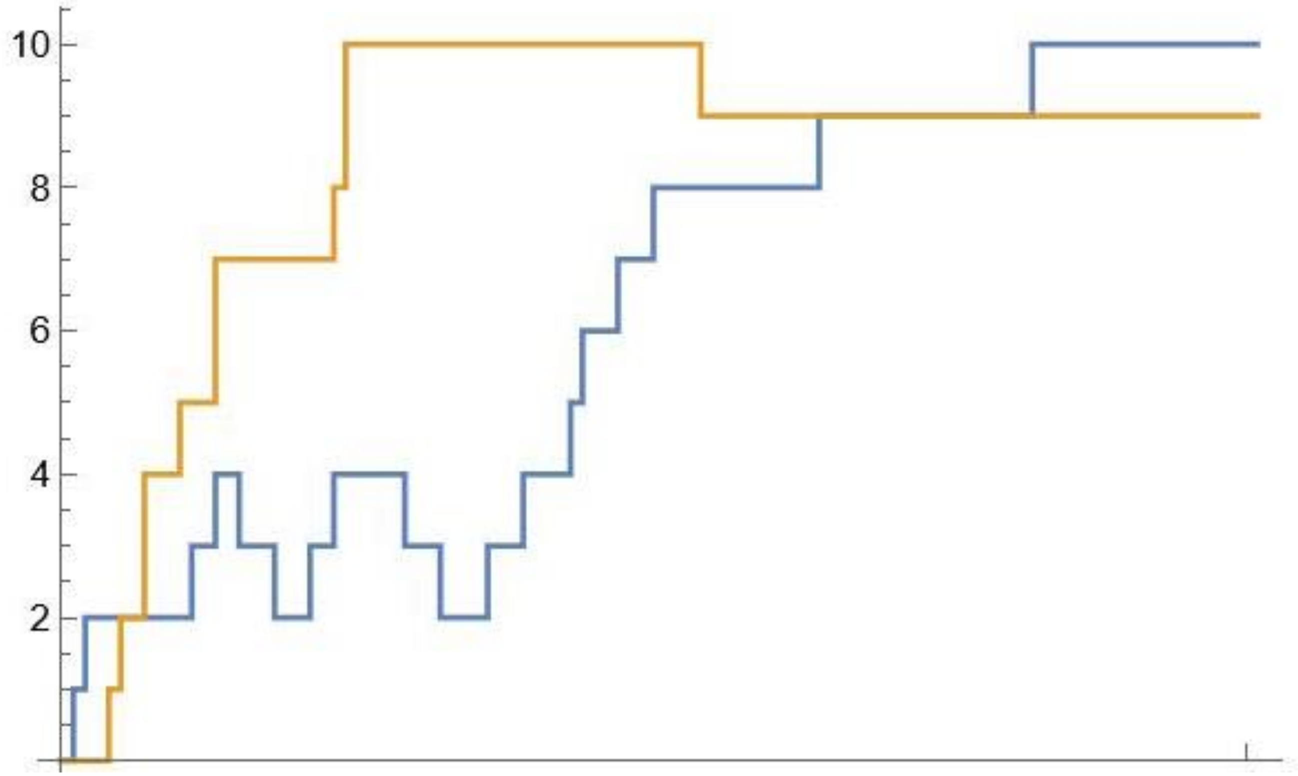
$\mu = \frac{60}{12} = 5$

$\rho = \frac{\lambda}{\mu} = \frac{20}{5} = 4$

Example

Barber Shop:

```
model = QueueingProcess[20, 5, 1, 10, 0];
data = RandomFunction[model, {0, 1, 0.01}, 2];
ListStepPlot[data]
```



Course: Applied Probability MW [2:15 – 3:30]

Chapter: [20]

Queuing Theory

Section: [20.6]

The M/M/s/GD/∞/∞ Queuing System

Assumptions and Notes

- We assume that interarrival times are exponential (with rate λ), service times are exponential (with rate μ), and there is a single line of customers waiting to be served at one of s parallel servers.
- If $j \leq s$ customers are present, then all j customers are in service
- If $j > s$ customers are present, then all s servers are occupied, and $j - s$ customers are waiting in line.
- Any arrival who finds an idle server enters service immediately, but an arrival who does not find an idle server joins the queue of customers awaiting service.

Model Parameters

- The $M/M/s/GD/\infty/\infty$ system can be modeled as a birth–death process with parameters

$$\lambda_j = \lambda \text{ for } j = 0,1,2,\cdots$$

$$\mu_j = j\mu \text{ for } j = 0,1,\cdots,s$$

$$\mu_j = s\mu \text{ for } j = s+1, s+2,\cdots$$

- Define $\rho = \frac{\lambda}{s\mu}$.

Steady-State Probabilities

- For $\rho < 1$, the steady-state probabilities are:

$$\pi_0 = \frac{1}{\frac{(s\rho)^s}{s!\,(1-\rho)} + \sum_{i=0}^{s-1} \frac{(s\rho)^i}{i!}}$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{j!} \text{ for } j = 1,2,\cdots,s$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{s! \cdot s^{j-s}} \text{ for } j = s, s+1, s+2, \cdots$$

- If $\rho \geq 1$, no steady state exists.
- It can be shown that the steady-state probability that all servers are busy is given by

$$P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!\,(1-\rho)}$$

Steady-State Probabilities

$P(j \geq s)$ for the $M/M/s/GD/\infty/\infty$ Queuing System

ρ	$s = 2$	$s = 3$	$s = 4$	$s = 5$	$s = 6$	$s = 7$
.10	.02	.00	.00	.00	.00	.00
.20	.07	.02	.00	.00	.00	.00
.30	.14	.07	.04	.02	.01	.00
.40	.23	.14	.09	.06	.04	.03
.50	.33	.24	.17	.13	.10	.08
.55	.39	.29	.23	.18	.14	.11
.60	.45	.35	.29	.24	.20	.17
.65	.51	.42	.35	.30	.26	.21
.70	.57	.51	.43	.38	.34	.30
.75	.64	.57	.51	.46	.42	.39
.80	.71	.65	.60	.55	.52	.49
.85	.78	.73	.69	.65	.62	.60
.90	.85	.83	.79	.76	.74	.72
.95	.92	.91	.89	.88	.87	.85

The Queuing Formulas

- It can also be shown that:

$$L_q = \frac{P(j \geq s)\rho}{1 - \rho}$$

$$L_s = \frac{\lambda}{\mu} = s\rho$$

$$L = L_s + L_q$$

$$W_q = \frac{L_q}{\lambda} = \frac{P(j \geq s)}{s\mu - \lambda}$$

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu}$$

$$W = W_s + W_q$$

Example

Consider a bank with two tellers. An average of 80 customers per hour arrive at the bank and wait in a single line for an idle teller. The average time it takes to serve a customer is 1.2 minutes. Assume that interarrival times and service times are exponential. Determine

1. The expected number of customers present in the bank.
   Solution: From the previous table, $P(j \geq 2) = 0.71$. So,
$$L_s = \frac{\lambda}{\mu} = \frac{80}{50} = 1.6$$
$$L_q = \frac{P(j \geq 2)\rho}{1-\rho} = \frac{(0.71)(0.8)}{0.2} = 2.84$$
$$\therefore L = L_s + L_q = 1.6 + 2.84 = 4.44 \text{ customers}$$

$$s = 2$$
$$\lambda = 80$$
$$\mu = \frac{60}{1.2} = 50$$
$$\rho = \frac{\lambda}{s\mu} = \frac{80}{100} = 0.8$$

Example

Consider a bank with two tellers. An average of 80 customers per hour arrive at the bank and wait in a single line for an idle teller. The average time it takes to serve a customer is 1.2 minutes. Assume that interarrival times and service times are exponential. Determine

$$s = 2$$
$$\lambda = 80$$
$$\mu = \frac{60}{1.2} = 50$$
$$\rho = \frac{\lambda}{s\mu} = \frac{80}{100} = 0.8$$

2. The expected length of time a customer spends in the bank.

   Solution:

$$\therefore W = \frac{L}{\lambda} = \frac{4.44}{80} = 0.055 \text{ hours} = 3.3 \text{ minutes}$$

Example

Consider a bank with two tellers. An average of 80 customers per hour arrive at the bank and wait in a single line for an idle teller. The average time it takes to serve a customer is 1.2 minutes. Assume that interarrival times and service times are exponential. Determine

$$s = 2$$
$$\lambda = 80$$
$$\mu = \frac{60}{1.2} = 50$$
$$\rho = \frac{\lambda}{s\mu} = \frac{80}{100} = 0.8$$

3. The fraction of time that a particular teller is idle.
   Solution: We need to evaluate the value $\pi_0 + 0.5\pi_1$.

$$P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!\,(1-\rho)} \Rightarrow 0.71 = \frac{(2 \times 0.8)^2 \pi_0}{2! \times 0.2}$$

$$\Rightarrow \pi_0 = 0.11$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{j!} \Rightarrow \pi_1 = \frac{2 \times 0.8 \times 0.11}{1!} = 0.176$$

$$\therefore \pi_0 + 0.5\pi_1 = 0.198$$

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [21]

Simulation

Section: [21.1]

Basic Terminology

Introduction

- **Simulation (المحاكاة)** is a very powerful and widely used management science technique for the analysis and study of complex systems.
- **Simulation** may be *defined as* a technique that imitates (يُقلّد) the operation of a real-world system as it evolves over time.
- The major *advantage* of simulation is that simulation methods are easier to apply than analytical methods.
- Once a model is built, it can be used repeatedly to analyze different policies, parameters, or designs.
- It must be emphasized that simulation is not an optimizing technique. It is most often used to analyze "**what if**" types of questions.

Definitions

- A **system** is a collection of entities (وحدات) that act and interact toward the accomplishment (تحقيق، إنجاز) of some logical end (نهاية منطقية).
- The **state** of a system is the collection of variables necessary to describe the status of the system at any given time.
- In a system, an object of interest is called an **entity**, and any properties of an entity are called **attributes**.
- Systems may be classified as **discrete** or **continuous**.
- A **discrete system** is one in which the state variables change only at discrete or countable points in time.

Definitions

- A **continuous system** is one in which the state variables change continuously over time.
- There are two types of simulation models: **static (ثابت)** and **dynamic (مرن)**.
- A **static simulation** model is a representation of a system at a particular point in time.
- A **dynamic simulation** is a representation of a system as it evolves over time.

Course: Applied Probability MW [2:15 – 3:30]

Chapter: [21]

Simulation

Section: [21.3]

Random Numbers and Monte Carlo Simulation

Monte Carlo Process

- A large proportion of the applications of simulations are for **probabilistic models**.
- The **Monte Carlo technique** is defined as a technique for selecting numbers randomly from a probability distribution for use in a trial (computer run) of a simulation model.
- The basic principle behind the process is the same as in the operation of gambling devices in casinos (such as those in Monte Carlo, Monaco).

The Use of Random Numbers (Example)

The manager of ComputerWorld, a store that sells computers and related equipment (مُعدّات), is attempting to determine how many laptop PCs the store should order each week.

- A primary consideration in this decision is the average number of laptop computers that the store will sell each week and the average weekly revenue generated from the sale of laptop PCs. A laptop sells for $4,300.
- The number of laptops demanded each week is a random variable (which we will define as x) that ranges from 0 to 4.
- From past sales records, the manager has determined the frequency of demand for laptop PCs for the past 100 weeks.

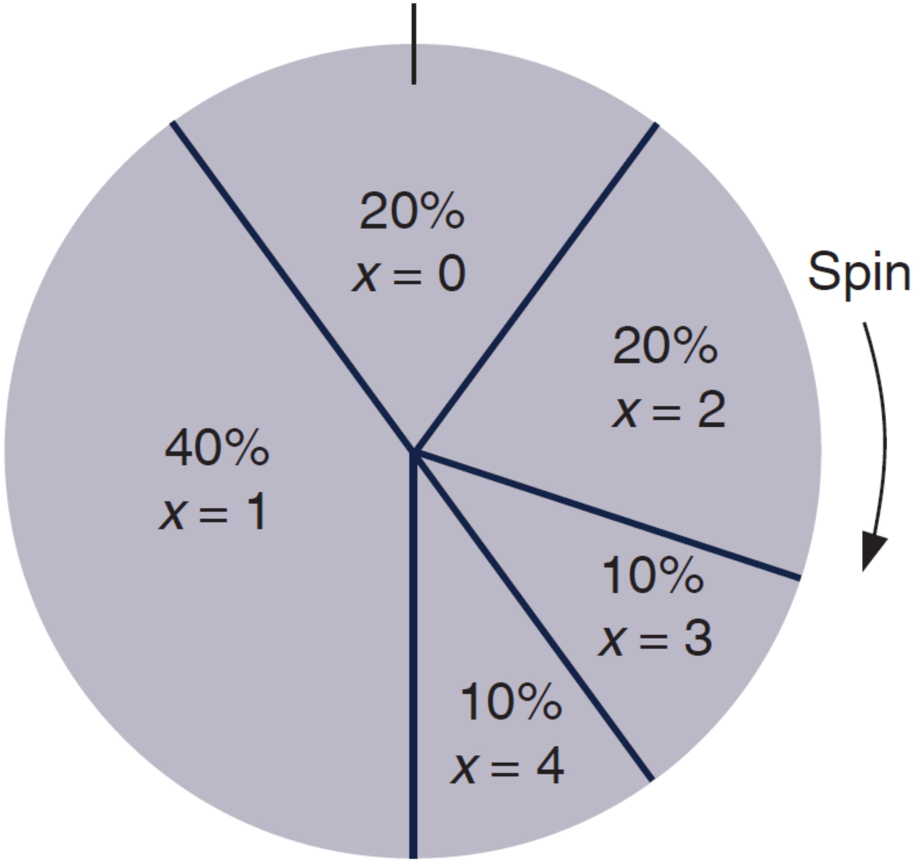
The Use of Random Numbers (Example)

The purpose of the Monte Carlo process is to **generate the random variable**, *demand*, by sampling from the probability distribution $P(x)$.

PCs Demanded per Week	Frequency of Demand	Probability of Demand, *P*(*x*)
0	20	.20
1	40	.40
2	20	.20
3	10	.10
4	10	.10
	100	1.00

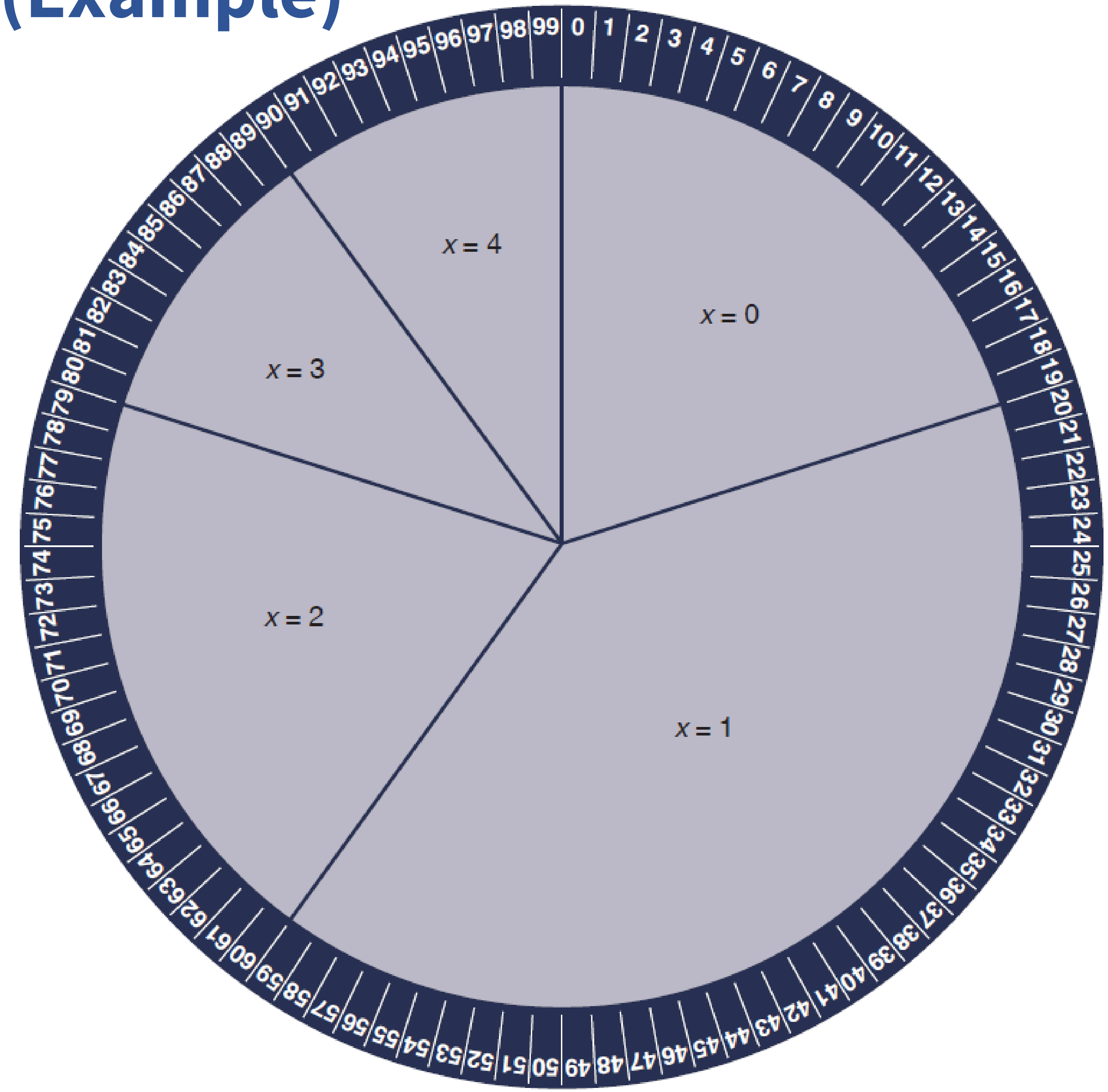
The Use of Random Numbers (Example)

- The partitioned roulette wheel replicates the probability distribution for demand if the values of demand occur in a random manner.
- The segment at which the wheel stops indicates demand for one week.



The Use of Random Numbers (Example)

- In addition to partitioning the wheel into segments corresponding to the probability of demand, we will put numbers along the outer rim, as on a real roulette wheel.
- When the manager spins this new wheel, the actual demand for PCs will be determined by a number.



The Use of Random Numbers (Example)

- For example, if the number 71 comes up on a spin, the demand is 2 PCs per week.
- Obviously, it is not generally practical to generate weekly demand for PCs by spinning a wheel. Alternatively, the process of spinning a wheel can be replicated by using random numbers alone.
- First, we will transfer the ranges of random numbers for each demand value from the roulette wheel to a table

The Use of Random Numbers (Example)

- Next, instead of spinning the wheel to get a random number, we will select a random number using computer, *Excel* for example.

Demand, x	Ranges of Random Numbers, r	
0	0–19	
1	20–59	$r = 39$
2	60–79	
3	80–89	
4	90–99	

- In Excel, by entering the command `=RANDBETWEEN(0,99)` in a cell, we generate a random number between 0 and 99.
- Random numbers between 0 and 1 can be generated in Excel by entering the formula `=RAND()` in a cell.

The Use of Random Numbers (Example)

- If you attempt to replicate this spreadsheet, you will generate **different** random numbers from those you have got before. In fact, any time you recalculate anything on your spreadsheet, the random numbers *will change*.
- The **more periods simulated**, the more **accurate** the results.
- Simulation results will **not equal** *analytical results* unless *enough trials* have been conducted to reach steady state.
- It is often difficult to validate results of simulation (that true steady state has been reached and that simulation model truly replicates reality).
- When analytical analysis is not possible, there is no analytical standard of comparison, thus making validation even more difficult.