# Review: Discrete Event Random Processes

Hongwei Zhang

http://www.cs.wayne.edu/~hzhang

## Outline

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network

## Outline

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network

## Markov chain

- Markov Chain
- Discrete-Time Markov Chains
- Calculating Stationary Distribution
- Global Balance Equations
- Generalized Markov Chains
- Continuous-Time Markov Chains

## Markov chain

- Markov Chain
- Discrete-Time Markov Chains
- Calculating Stationary Distribution
- Global Balance Equations
- Generalized Markov Chains
- Continuous-Time Markov Chains

## Markov Chain?

- Stochastic process that takes values in a *countable* set
  - Example: {0,1,2,…,m}, or {0,1,2,…}
  - Elements represent possible "states"
  - Chain transits from state to state
- *Memoryless (Markov) Property*: Given the present state, future transitions of the chain are independent of past history
- Markov Chains: discrete- or continuous- time

## Markov chain

- Markov Chain
- Discrete-Time Markov Chains
- Calculating Stationary Distribution
- Global Balance Equations
- Generalized Markov Chains
- Continuous-Time Markov Chains

## Discrete-Time Markov Chain (DTMC)

- Discrete-time stochastic process $\{X_n: n = 0,1,2,\ldots\}$
- Takes values in {0,1,2,…}
- Memoryless property:

Note: future and past are independent given the present; but they are not unconditionally independent.

$$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P\{X_{n+1} = j \mid X_n = i\}$$

$$P_{ij} = P\{X_{n+1} = j \mid X_n = i\}$$

Also written as $P_{i,j}$

- Transition probabilities $P_{ij}$

$$P_{ij} \geq 0, \quad \sum_{j=0}^{\infty} P_{ij} = 1$$

- Transition probability matrix $P=[P_{ij}]$

## Composition of DTMCs

- Given two *independent* DTMCs $X_n$, n≥0, on S and $Y_n$, n≥0, on T with transition probability matrices P and Q; then $Z_n = (X_n, Y_n)$ is a DTMC on S×T with

$$\Pr\left(Z_{n+1} = (s2, t2) \mid Z_n = (s1, t1)\right) = p_{s1,s2} q_{t1,t2}$$

- Multiple mutually independent DTMCs can be composed in a similar fashion

## Chapman-Kolmogorov Equations

- $n$ step transition probabilities

Also written as $P_{i,j}^{(n)}$

$$P_{ij}^{n} = P\{X_{n+m} = j \mid X_m = i\}, \quad n, m \geq 0,\ i, j \geq 0$$

- How to calculate?
  - Chapman-Kolmogorov equations

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^{n} P_{kj}^{m}, \quad n, m \geq 0,\ i, j \geq 0$$

    - $P_{ij}^{n}$ is element $(i, j)$ in matrix $P^n$
    - Recursive computation of *state* probabilities
  - Thus,

$$P^{(n)} = P^n$$

## State Probabilities – Stationary Distribution

- State probabilities (time-dependent)

$$\pi_j^n = P\{X_n = j\}, \qquad \pi^n = (\pi_0^n, \pi_1^n, \ldots)$$

$$P\{X_n = j\} = \sum_{i=0}^{\infty} P\{X_{n-1} = i\} P\{X_n = j \mid X_{n-1} = i\} \Rightarrow \pi_j^n = \sum_{i=0}^{\infty} \pi_i^{n-1} P_{ij}$$

In matrix form:

$$\pi^n = \pi^{n-1} P = \pi^{n-2} P^2 = \ldots = \pi^0 P^n$$

- If time-dependent distribution converges to a limit

$$\pi = \lim_{n \to \infty} \pi^n \qquad \pi = \pi P$$

π is called the *stationary distribution* (or *steady state distribution*)
  - existence depends on the *structure of Markov chain*

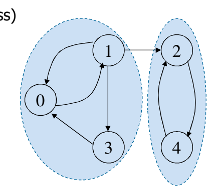
## Irreducibility of DTMC

- States i and j communicate:

$$\exists n, m:\ P_{ij}^{n} > 0,\ P_{ji}^{m} > 0$$

Denote as i↔j

- Binary relation ↔ is an *equivalence* (i.e., reflexive, symmetric, transitive); the equivalence classes induced by ↔ are called *communicating classes*
- Irreducible Markov chain: all states communicate (and thus form a single communicating class)



## First hit probabilities $f_{i,j}^{(n)}$

- Probability of *first hitting/visiting state j at time n*, when starting in state i at time 0

$$f_{i,j}^{(n)} = \Pr(X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j \mid X_0 = i)$$

$$f_{i,i}^{(0)} = 1, \text{ and for } j \neq i, f_{i,j}^{(0)} = 0$$

- $T_{ij}$: the first passage time from i to j
- Probability of visiting state j in *finite time* if starting in state i

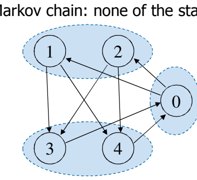
$$f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}$$

## Aperiodicity of DTMC

- Period $d_i$ of a state i:

$$d_i = \gcd\{n : f_{i,j}^{(n)} > 0\} = \gcd\{n : p_{i,j}^{(n)} > 0\}$$

  - Theorem: all the states in a communicating class of a DTMC have the same period.
- State i is *aperiodic* if $d_i = 1$
  - Special case: if $p_{j,j} > 0$, then j is aperiodic (why?)
- Aperiodic Markov chain: none of the states is periodic




## Limit Theorems

Theorem 0a: Irreducible aperiodic Markov chain

- For every state $j$, the following limit

$$\pi_j = \lim_{n\to\infty} P\{X_n = j \mid X_0 = i\}, \quad i = 0,1,2,\ldots$$

  exists and is independent of initial state $i$

- $N_j(k)$: number of visits to state $j$ up to time $k$

$$P\left\{\pi_j = \lim_{k\to\infty} \frac{N_j(k)}{k} \mid X_0 = i\right\} = 1$$

=>$\pi_j$: *frequency the process visits state j*

## Existence of Stationary Distribution (or steady state distribution)

Theorem 0b: Irreducible aperiodic Markov chain. There are two possibilities for scalars:

$$\pi_j = \lim_{n\to\infty} P\{X_n = j \mid X_0 = i\} = \lim_{n\to\infty} P_{ij}^n$$

1. $\pi_j = 0$, for all states $j$ ➡ No stationary distribution
2. $\pi_j > 0$, for all states $j$ ➡ $\pi$ is the *unique* stationary distribution

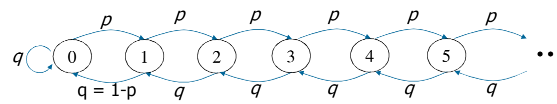
Remark: If the number of states is finite, case 2 is the only possibility

## Positivity

- A state j is *positive recurrent* if the process returns to state j "infinitely often"
- Formal definition:
  - A state j is *absorbing* if $p_{j,j} = 1$
  - A state j is *transient* if $f_{j,j} < 1$
  - A state j is *recurrent* (or *persistent*) if $f_{j,j} = 1$
    - A recurrent state j is *positive* if $\sum_{n=1}^{\infty} n f_{j,j}^{n} < \infty$ ; otherwise, it is *null*
  - Note: "positive recurrent => irreducible" always hold, but "irreducible => positive recurrent" is guaranteed to hold only for finite MC

---

- Example 0: a MC with countably infinite state space




  - All states are positive recurrent if p < ½, null recurrent if p = ½, and transient if p > ½

---

- Theorem D.2: for each communicating class of a DTMC $\{X_n\}$, exactly one of the following holds:
  - All the states in the class are transient
  - All the states in the class are null recurrent
  - All the states in the class are positive (recurrent)
- Thus, an irreducible DTMC is positive recurrent if any one of its state is positive

---

## Deciding positivity

- A communicating class C is *closed* if

$$\forall i, j : i \in C, j \notin C \Rightarrow p_{i,j} = 0$$

  Otherwise, the class is said to be *open*

- Theorem D.3: given a DTMC,
  - An open communicating class is transient
  - A *closed finite* communication class is positive recurrent

## What about *infinite* closed communicating classes?

- Theorem D.4: an irreducible DTMC on state space S is positive recurrent iff.

$$\exists \text{positive prob. distribution } \pi \text{ on S s.t. } \pi = \pi P.$$

where P is the state transition matrix.

  - Note: if such probability $\pi$ exists, it is unique and is called an *invariant* probability vector for the DTMC

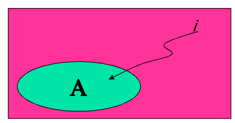
- If $\pi$ is invariant, and if $\Pr(X_0=i) = \pi_i$, then the DTMC so obtained is a *stationary* random process

---

- Alternative approach: drift analysis of a suitable *Lyapunov function* f(.)

- Theorem D.7: an irreducible DTMC $X_n$, $n\geq 0$, is *recurrent* if $\exists$ *nonnegative* function f(j), $j\in S$ (state space), s.t. $f(j)\to\infty$ as $j\to\infty$, and a *finite* set $A\subset S$ s.t.

$$\forall i \notin A : E\big(f(X_{n+1}) \mid X_n = i\big) \leq f(i)$$



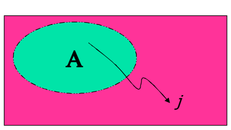

---

- Theorem D.8: an irreducible DTMC $X_n$, $n\geq 0$, is *transient* if $\exists$ nonnegative function f(j), $j\in S$, and a set $A\subset S$ s.t.

$$\forall i \notin A : E\big(f(X_{n+1}) \mid X_n = i\big) \leq f(i)$$

and $\exists j\notin A$ s.t.

$$\forall k \in A : f(j) < f(k)$$




---

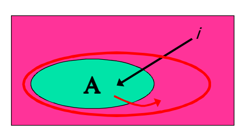
- Theorem D.9: an irreducible DTMC $X_n$, $n\geq 0$, is *positive recurrent* if $\exists$ nonnegative function f(j), $j\in S$, and a *finite* set $A\subset S$ s.t.

$$\forall i \notin A : E\big(f(X_{n+1}) \mid X_n = i\big) \leq f(i) - \varepsilon$$

for some $\varepsilon>0$, and

$$\forall k \in A : E\big(f(X_{n+1}) \mid X_n = k\big) \leq B$$

for some finite number B.



- Theorem D.10: an irreducible DTMC $X_n$, $n\geq 0$, on $i\in\{0,1,2,\ldots\}$ is *not positive recurrent* if $\exists$ finite values K>0 and B>0 s.t.

$$\forall i\geq 0:\ E(X_{n+1} \mid X_n = i) < \infty,\ \text{and}$$
$$\forall i\geq K:\ E(X_{n+1} \mid X_n = i) \geq i, \text{ and } E\left((X_n - X_{n+1})^+ \mid X_n = i\right) \leq B$$

Bounded downward drift

  - In the context of Theorems D.7-9, Theorem D.10 is for Lyapunov function f(j)=j
- This theorem is useful in establishing instability results
  - E.g., for a queue with finite # of servers where arrival rate is strictly greater than the overall service rate, $X_n$ = queue occupancy

- Exercise
  - Use Theorems D.7-9 to prove the results for Example 0 shown earlier

## Convergence of positive recurrent DTMC

- Given an *irreducible, positive* DTMC with period d and state space S,
  - $\forall j\in S, \lim_{n\to\infty} p^{nd}_{j,j} = d\pi_j$
  - If the DTMC is aperiodic (i.e., d=1),

$$\forall i, j\in S, \lim_{n\to\infty} p^{n}_{i,j} = \pi_j$$

## Ergodicity

- A state is *ergodic* if it is aperiodic and positive recurrent
- A MC is ergodic if every state is ergodic
- Ergodic chains have a unique stationary distribution

$\pi_j = 1/E(T_{jj})$, j = 0, 1, 2, …

where $T_{ij}$ is the first passage time from i to j

  - Note: Ergodicity ⇒ Time Averages = Stochastic Averages

## Markov chain

- Markov Chain
- Discrete-Time Markov Chains
- Calculating Stationary Distribution
- Global Balance Equations
- Generalized Markov Chains
- Continuous-Time Markov Chains

## Calculation of Stationary Distribution

A. Finite number of states

- Solve explicitly the system of equations

$$\pi_j = \sum_{i=0}^{m} \pi_i P_{ij}, \quad j = 0,1,\ldots,m$$
$$\sum_{i=0}^{m} \pi_i = 1$$

- Or, numerically from $P^n$ which converges to a matrix with rows equal to $\pi$
  - Suitable for a small number of states

B. Infinite number of states

- Cannot apply previous methods to problem of infinite dimension
- Guess a solution to recurrence:

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad j = 0,1,\ldots,$$
$$\sum_{i=0}^{\infty} \pi_i = 1$$

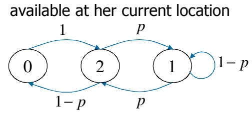
(detailed) balance equations can help the guess

## Example: Finite Markov Chain

- Absent-minded professor uses two umbrellas when commuting between home and office.
- If it rains and an umbrella is available at her location, she takes it. If it does not rain, she always forgets to take an umbrella.
- Let $p$ be the probability of rain each time she commutes.

Q: What is the probability that she gets wet on any given day?

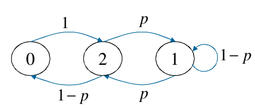
- Markov chain formulation
- $i$ is the number of umbrellas available at her current location




- Transition matrix

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1-p & p \\ 1-p & p & 0 \end{bmatrix}$$

## Example: Finite Markov Chain




$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1-p & p \\ 1-p & p & 0 \end{bmatrix}$$

$$\begin{cases} \pi = \pi P \\ \sum_i \pi_i = 1 \end{cases} \Leftrightarrow \begin{cases} \pi_0 = (1-p)\pi_2 \\ \pi_1 = (1-p)\pi_1 + p\pi_2 \\ \pi_2 = \pi_0 + p\pi_1 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases} \Leftrightarrow \pi_0 = \frac{1-p}{3-p}, \pi_1 = \frac{1}{3-p}, \pi_2 = \frac{1}{3-p}$$

$$P\{\text{gets wet}\} = \pi_0 p = p\frac{1-p}{3-p}$$

## Example: Finite Markov Chain

- Taking $p = 0.1$:

$$\pi = \left(\frac{1-p}{3-p}, \frac{1}{3-p}, \frac{1}{3-p}\right) = (0.310, 0.345, 0.345)$$

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0.9 & 0.1 \\ 0.9 & 0.1 & 0 \end{bmatrix}$$

- Numerically determine limit of $P^n$

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 0.310 & 0.345 & 0.345 \\ 0.310 & 0.345 & 0.345 \\ 0.310 & 0.345 & 0.345 \end{bmatrix} \quad (n \approx 150)$$

- Effectiveness depends on structure of $P$

## Markov chain

- Markov Chain
- Discrete-Time Markov Chains
- Calculating Stationary Distribution
- Global Balance Equations
- Generalized Markov Chains
- Continuous-Time Markov Chains

## Global Balance Equations

- Global Balance Equations (GBE)

$$\pi_j \sum_{i=0}^{\infty} P_{ji} = \sum_{i=0}^{\infty} \pi_i P_{ij} \Leftrightarrow \pi_j \sum_{i \neq j} P_{ji} = \sum_{i \neq j} \pi_i P_{ij}, \quad j \geq 0$$

- $\pi_j P_{ji}$ is the frequency of transitions from $j$ to $i$

$$\begin{pmatrix} \text{Frequency of} \\ \text{transitions out of } j \end{pmatrix} = \begin{pmatrix} \text{Frequency of} \\ \text{transitions into } j \end{pmatrix}$$

- Intuition: 1) $j$ visited infinitely often; 2) for each transition out of $j$ there must be a subsequent transition into $j$ with probability 1

## Global Balance Equations (contd.)

- Alternative Form of GBE

$$\sum_{j\in S} \pi_j \sum_{i\notin S} P_{ji} = \sum_{i\notin S} \pi_i \sum_{j\in S} P_{ij}, \quad S \subseteq \{0,1,2,...\}$$

- *If a probability distribution satisfies the GBE, then it is the unique stationary distribution of the Markov chain*
- Finding the stationary distribution:
  - Guess distribution from properties of the system
  - Verify that it satisfies the GBE
  - ☺ Special structure of the Markov chain simplifies task

## Global Balance Equations – Proof

First form:

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij} \text{ and } \sum_{i=0}^{\infty} P_{ji} = 1 \quad \Rightarrow$$

$$\pi_j \sum_{i=0}^{\infty} P_{ji} = \sum_{i=0}^{\infty} \pi_i P_{ij} \iff \pi_j \sum_{i \neq j} P_{ji} = \sum_{i \neq j} \pi_i P_{ij}$$

Second form:

$$\pi_j \sum_{i=0}^{\infty} P_{ji} = \sum_{i=0}^{\infty} \pi_i P_{ij} \Rightarrow \sum_{j \in S} \pi_j \sum_{i=0}^{\infty} P_{ji} = \sum_{j \in S} \sum_{i=0}^{\infty} \pi_i P_{ij} \Rightarrow$$

$$\sum_{j \in S} \pi_j \left( \sum_{i \in S} P_{ji} + \sum_{i \notin S} P_{ji} \right) = \sum_{j \in S} \left( \sum_{i \in S} \pi_i P_{ij} + \sum_{i \notin S} \pi_i P_{ij} \right) \Rightarrow$$

$$\sum_{j \in S} \pi_j \sum_{i \notin S} P_{ji} = \sum_{i \notin S} \pi_i \sum_{j \in S} P_{ij}$$

## Markov chain

- Markov Chain
- Discrete-Time Markov Chains
- Calculating Stationary Distribution
- Global Balance Equations
- Generalized Markov Chains
- Continuous-Time Markov Chains

## Generalized Markov Chains

- Markov chain on a set of states {0,1,…}, that whenever enters state $i$
  - The next state that will be entered is $j$ with probability $P_{ij}$
  - Given that the next state entered will be $j$, the time it spends at state $i$ until the transition occurs is a RV with distribution $F_{ij}$
- $\{Z(t)\colon t \geq 0\}$ describing the state of the chain at time $t$: *Generalized Markov chain*, or *Semi-Markov* process
  - Does GMC have the Markov property?
    - Future depends on 1) the present state, and 2) the length of time the process has spent in this state

## Generalized Markov Chains (contd.)

- $T_i$: time process spends at state $i$, before making a transition – *holding time*
- Probability distribution function of $T_i$

$$H_i(t) = P\{T_i \leq t\} = \sum_{j=0}^{\infty} P\{T_i \leq t \mid \text{next state } j\} P_{ij} = \sum_{j=0}^{\infty} F_{ij}(t) P_{ij}$$

$$E[T_i] = \int_0^{\infty} t \, dH_i(t)$$

- $T_{ii}$: time between successive transitions to $i$
- $X_n$ is the $n^{th}$ state visited. $\{X_n\colon n=0,1,\dots\}$
  - Is a Markov chain: embedded Markov chain
  - Has transition probabilities $P_{ij}$
- Semi-Markov process *irreducible*: if its embedded Markov chain is irreducible

## Limit Theorems

Given an irreducible semi-Markov process w/ $E[T_{ii}] < \infty$

- For any state $j$, the following limit
$$p_j = \lim_{t\to\infty} P\{Z(t) = j \mid Z(0) = i\}, \quad i = 0,1,2,...$$
exists and is independent of the initial state.
$$p_j = \frac{E[T_j]}{E[T_{jj}]}$$
- $T_j(t)$: time spent at state $j$ up to time $t$
$$P\left\{p_j = \lim_{t\to\infty} \frac{T_j(t)}{t} \mid Z(0) = i\right\} = 1$$
  - $p_j$ is equal to *the proportion of time spent at state j*

## Occupancy Distribution

Given an irreducible semi-Markov process where $E[T_{ii}] < \infty$, and the embedded Markov chain is ergodic w/ stationary distribution $\pi$

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij},\ j \geq 0; \quad \sum_{i=0}^{\infty} \pi_i = 1$$

then, with probability 1, the occupancy distribution of the semi-Markov process

$$p_j = \frac{\pi_j E[T_j]}{\sum_i \pi_i E[T_i]}, \quad j = 0,1,...$$

- $\pi_j$: proportion of transitions into state $j$
- $E[T_j]$: mean time spent at $j$
- Probability of being at $j$ is proportional to $\pi_j E[T_j]$

## Markov chain

- Markov Chain
- Discrete-Time Markov Chains
- Calculating Stationary Distribution
- Global Balance Equations
- Generalized Markov Chains
- Continuous-Time Markov Chains

## Continuous-Time Markov Chains (def.?)

Continuous-time process $\{X(t): t \geq 0\}$ taking values in $\{0,1,2,...\}$.

Whenever it enters state i

- Time it spends at state $i$ is *exponentially distributed* with parameter $\alpha_i$
- When it leaves state $i$, it enters state $j$ with probability $P_{ij}$, where $\Sigma_{j\neq i} P_{ij} = 1$
- Continuous-time Markov chain is a semi-Markov process with
$$F_{i,j}(t) = 1 - e^{-\alpha_i t}, \quad i, j = 0,1,2,...$$
- Exponential holding time => a continuous-time Markov chain has the Markov property

## CTMC: alternative definition

- {X(t)} on state space S is a *continuous time Markov chain* if

$$\forall t, s \ge 0, \forall j \in S,$$
$$\Pr\big(X(t+s) = j \mid X(u), u \le s\big) = \Pr\big(X(t+s) = j \mid X(s)\big)$$

- Assume time homogeneity, we write

$$p_{i,j}(t) := \Pr\big(X(t+s) = j \mid X(s) = i\big)$$

- For an arbitrary time t, time to next state transition W(t)

$$W(t) = \inf\{s > 0 : X(t+s) \ne X(t)\}$$

- Theorem D.11: for a CTMC {X(t)},

$$\forall i \in S, \forall t, u \ge 0 : \Pr\big(W(t) > u \mid X(t) = i\big) = e^{-\alpha_i u}$$
$$\text{for some constant } \alpha_i \ge 0.$$

  - Sojourn time at a state i is exponentially distributed with parameter $\alpha_i$ that only depend on i
  - A state $i \in S$ is called absorbing if $\alpha_i = 0$.

## Jump chain/embedded process

- Let $T_0=0$, $T_1$, $T_2$, … be the successive jump instants (i.e., instants when state changes) of a CTMC, and let $X_n=X(T_n)$
- Sequence $T_n$, $n \ge 0$, is called a sequence of *embedded instants*, and $X_n$, $n \ge 0$, is called an jump chain or an *embedded process*

- Theorem D.12: given a CTMC {X(t)} with jump instants $T_n$, $n \ge 0$, and jump chain $X_n$, $n \ge 0$, for $i_0, i_1, \ldots, i_{n-1}, i, j \in S$, $t_0, t_1, \ldots, t_n$, $u \ge 0$,

$$\Pr\left\{ X_{n+1} = j, T_{n+1} - T_n > u \;\middle|\; \begin{array}{c} X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i, \\ T_0 = t_0, \ldots, T_n = t_n \end{array} \right\} = p_{i,j} e^{-\alpha_i u}$$
$$\text{where } p_{i,j} \ge 0, \sum_{j \in S} p_{i,j} = 1, \text{ and if } \alpha_i > 0, \text{ then } p_{i,i} = 0$$

  - Sojourn time at a state i and the next state entered are independent, and only depend on state i
  - Thus, the embedded process is a DTMC with transition probability $p_{i,j}$

- A CTMC *is*
  - *Irreducible*: if its embedded Markov chain is irreducible, and
  - *Regular*: if number of transitions in a finite time interval is finite with probability 1
- Theorem D.13: $\{X(t)\}$ is a CTMC with embedded DTMC $\{X_n\}$ and sojourn time parameters $\alpha_i \in S$, then
  - If $\exists v$ s.t. $\alpha_i \le v$ for all i, then $\{X(t)\}$ is regular
  - If $\{X_n\}$ is recurrent, then $\{X(t)\}$ is regular

## CTMC: transience & recurrence

- Let $\tau_{j,j}$ = time until the process first returns to j after leaving it
- A state j in a CTMC is *recurrent* if $\Pr(\tau_{j,j}<\infty)=1$; otherwise, j is *transient*.

  A recurrent state j is *positive* if $E(\tau_{j,j}) < \infty$; otherwise, it is *null*.
- Same as in DTMC, the states of an irreducible CTMC are either all transient, all positive, or all null

- A state j is *recurrent* in CTMC *iff.* it is recurrent in the embedded DTMC;

  An irreducible CTMC is *recurrent iff.* the embedded DTMC is recurrent.
- Similar results doest NOT hold for positivity of CTMC states

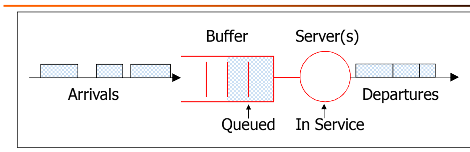
- Transition rate matrix Q
  - For $i,j \in S$, $i \neq j$, define $q_{i,j} = \alpha_i p_{i,j}$; can be interpreted as, conditional on being at state i, the rate of leaving i to enter j
  - for $i \in S$, $q_{i,i} = -\alpha_i$
    - Thus, the sum of each row of Q is 0
- Theorem D.14: an irreducible regular CTMC is *positive* iff. $\exists$positive prob. vector $\pi$ s.t. $\pi Q=0$ and $\sum_{i\in S}\pi_i = 1$. When such a $\pi$ exists, it is unique.
  - Note: the j-th equation is $\sum_{i\in S, i\neq j} \pi_i q_{i,j} = \pi_j \alpha_j$ , meaning the unconditional rate of entering j equals that of leaving j

- If the positive prob. vector $\pi$ exists,
  - it is also a stationary prob. vector; that is, if Pr(X(0)=i) = $\pi_i$, then Pr(X(t)=i) = $\pi_i$
  - $\pi_i = \frac{1/\alpha_i}{\tau_{i,i}}$
  - $\lim_{t\to\infty} p_{i,j}(t) = \pi_j$
    - No notion of periodicity for CTMC

## Example

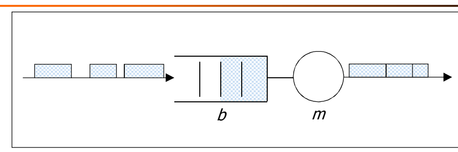
- M/M/1 queue

## Basic Queueing Model



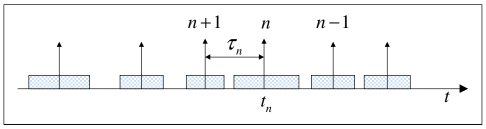

- A queue models any service station with:
  - One or multiple servers
  - A waiting area or buffer
- Customers arrive to receive service
- A customer that upon arrival does not find a free server waits in the buffer

## Characteristics of a Queue




- Number of servers $m$: one, multiple, infinite
- Buffer size $b$
- Service discipline (scheduling)
  - FCFS, LCFS, Processor Sharing (PS), etc
- Arrival process
- Service statistics

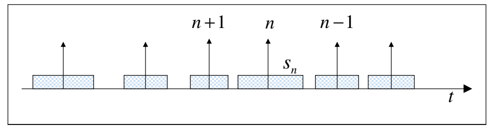
## Arrival Process




- $\tau_n$ : interarrival time between customers *n* and *n*+1
- $\tau_n$ is a random variable
- $\{\tau_n,\ n \geq 1\}$ is a stochastic process
  - Interarrival times are identically distributed and have a common mean

    $E[\tau_n] = E[\tau] = 1/\lambda$ , where $\lambda$ is called the *arrival rate*

## Service-Time Process




- $s_n$ : service time of customer $n$ at the server
- $\{s_n,\ n \geq 1\}$ is a stochastic process
  - Service times are identically distributed with common mean

    $E[s_n] = E[s] = \mu$ , where $\mu$ is called the service rate

*For packets, are the service times really random?*

## Queue Descriptors

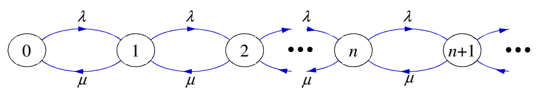
- Generic descriptor: A/S/m/k
  - *A* denotes the arrival process
    - For Poisson arrivals we use M (for Markovian)
  - *S* denotes the service-time distribution
    - M: exponential distribution
    - D: deterministic service times
    - G: general distribution
  - *m* is the number of servers
  - *k* is the max number of customers allowed in the system – either in the buffer or in service
    - k is omitted when the buffer size is infinite

## Queue Descriptors: Examples

- M/M/1: Poisson arrivals, exponentially distributed service times, one server, infinite buffer
- M/M/m: same as previous with m servers
- M/M/m/m: Poisson arrivals, exponentially distributed service times, m server, no buffering
- M/G/1: Poisson arrivals, identically distributed service times follows a general distribution, one server, infinite buffer
- */D/∞ : A constant delay system

## Example: M/M/1 Queue

- Arrival process: Poisson with rate λ
- Service times: iid, exponential with parameter μ
- Service times and interarrival times: independent
- Single server
- Infinite waiting room
- $X(t)$: Number of customers in system at time $t$ (state)




## Exponential Random Variables

- $X$: exponential RV with parameter λ
- $Y$: exponential RV with parameter μ
- $X$, $Y$: independent

Then:

1. $\min\{X, Y\}$: exponential RV with parameter λ+μ
2. $P\{X<Y\} = \lambda/(\lambda+\mu)$

<u>Proof:</u>

$$P\{\min\{X,Y\}>t\} = P\{X>t, Y>t\} = \\ = P\{X>t\}P\{Y>t\} = \\ = e^{-\lambda t}e^{-\mu t} = e^{-(\lambda+\mu)t} \Rightarrow$$
$$P\{\min\{X,Y\}\le t\} = 1-e^{-(\lambda+\mu)t}$$

$$P\{X<Y\} = \int_0^\infty \int_0^y f_{XY}(x,y)\,dx\,dy = \\ = \int_0^\infty \int_0^y \lambda e^{-\lambda x}\cdot \mu e^{-\mu y}\,dx\,dy = \\ = \int_0^\infty \mu e^{-\mu y}\int_0^y \lambda e^{-\lambda x}\,dx\,dy = \\ = \int_0^\infty \mu e^{-\mu y}(1-e^{-\lambda y})\,dy = \\ = \int_0^\infty \mu e^{-\mu y}dy - \frac{\mu}{\lambda+\mu}\int_0^\infty (\lambda+\mu)e^{-(\lambda+\mu)y}dy = \\ = 1-\frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}$$

## M/M/1 Queue: Markov Chain Formulation

- Jumps of $\{X(t): t\ge 0\}$ triggered by arrivals and departures
- $\{X(t): t\ge 0\}$ can jump only between neighboring states

Assume process at time $t$ is in state $i$: $N(t) = i \ge 1$

- $X_i$: time until the next arrival – exponential with parameter λ
- $Y_i$: time until the next departure – exponential with parameter μ
- $T_i = \min\{X_i, Y_i\}$: time process spends at state $i$
- $T_i$: exponential with parameter $\alpha_i = \lambda+\mu$

---

- $P_{i,i+1} = P\{X_i < Y_i\} = \lambda/(\lambda+\mu)$, $P_{i,i-1} = P\{Y_i < X_i\} = \mu/(\lambda+\mu)$
- $P_{01}=1$, and $T_0$ is exponential with parameter λ
- $\{N(t): t\ge 0\}$ is a CTMC with

$$q_{i,i+1} = \alpha_i p_{i,i+1} = \lambda,\ i\ge 0$$
$$q_{i,i-1} = \alpha_i p_{i,i-1} = \mu,\ i\ge 1$$
$$q_{0,0} = -\lambda$$
$$q_{i,j} = 0,\ |i-j|>1$$

- $\pi Q=0$ has a positive, summable (to 1) solution iff. $\lambda<\mu$
- If $\lambda<\mu$,
  - Prob{queue is non-empty} = $1-\rho$, where $\rho= \lambda/\mu$
  - $\pi_i = (1- \rho)\rho^i$, i = 0, 1, 2, …, is the stationary distribution

## Outline

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network

## Renewal process

- Given a sequence of mutually independent r.v.'s $X_k$, k=1,2,3,…, s.t. $X_k$, k>=2 are i.i.d., and $X_1$ can have a possibly different distribution, we define the *renewal instants*, $Z_k$, k>=1, as $Z_k = \sum_{i=1}^{k} X_i$
- The # of renewals in time (0, t] is called a ***renewal process*** M(t)
- Example: a CTMC B(t) with B(0) = i, and let's consider visits to state j
  - $X_1$: time to first visit j
  - $X_k$, k>=2: times between subsequent visits to j
  - M(t): # of visits to j up to time t

## Renewal reward process

- To associate a reward with each renewal interval
- Formally:
  - Given a renewal process with lifetimes $X_k$, $k\geq1$, associate $X_k$ with a reward $R_k$ s.t. $R_k$, $k\geq1$, are mutually independent; $R_k$ can depend on $X_k$
- Example: in the CTMC B(t), define $R_k$ as the time spent at a specific state i during the k-th renewal interval

- Let C(t) be the total reward accrued until time t, then the reward rate is $\lim_{t\to\infty}\frac{C(t)}{t}$
- [Renewal Reward Theorem D.15]: for $E(|R_k|)<\infty$ and $E(|X_k|)<\infty$, the following hold:
  - With probability 1, $\lim_{t\to\infty}\frac{C(t)}{t}=\frac{E(R_2)}{E(X_2)}$
  - $\lim_{t\to\infty}\frac{E(C(t))}{t}=\frac{E(R_2)}{E(X_2)}$
- Note: in general, $E(R_2)/E(X_2) \neq E(R_2/X_2)$

## Markov renewal process (MRP)

Generalized Markov Chain

- Let $X_n$, $n\geq 0$, be a random sequence with state space S, and let $T_0\leq T_1\leq T_2$ … be nondecreasing sequence of random times
- The random sequence $(X_n, T_n)$, $n\geq 0$, is a *Markov renewal process (MRP)* if for $i_0, i_1, \ldots, i_{n-1}, i, j\in S$, $t_0\leq t_1\leq \ldots \leq t_n$, $u\geq 0$

$$\Pr\left\{X_{n+1}=j, T_{n+1}-T_n\leq u \mid \begin{array}{c} X_0=i_0,\ldots,X_{n-1}=i_{n-1},X_n=i, \\ T_0=t_0,\ldots,T_n=t_n \end{array}\right\}=$$
$$\Pr\{X_{n+1}=j, T_{n+1}-T_n\leq u \mid X_n=i\}$$

  - MRP is a *generalization of CTMC*: 1) sojourn time may not be independent of the next state, and 2) sojourn time may not be exponentially distributed

- Let $p_{i,j}=\lim_{u\to\infty}\Pr\{X_{n+1}=j, T_{n+1}-T_n\leq u \mid X_n=i\}$ , assuming the limit does not depend on n;

  Then, $X_n$, $n\geq 0$, is a DTMC on S with transition prob. $p_{i,j}$, $i,j\in S$
- Distribution of sojourn time given the current and next states:

$$H_{i,j}(u)=\Pr\{T_{n+1}-T_n\leq u \mid X_n=i, X_{n+1}=j\}$$

- Theorem D.16:

$$\Pr\{T_1-T_0\leq u_1, T_2-T_1\leq u_2,\ldots,T_n-T_{n-1}\leq u_n\}=\prod_{i=1}^{n}H_{i-1,i}(u_i)$$

  - *Independent sojourn times* given the sequence of states at the end points

- Distribution of sojourn time:

$$H_i(u)=\sum_{j\in S}p_{i,j}H_{i,j}(u)$$

- Mean sojourn time at state i:

$$\sigma_i=\sum_{j\in S}p_{i,j}\sigma_{i,j}, \text{ where } \sigma_{i,j} \text{ is the mean of } H_{i,j}(u)$$

- Associate a reward $R_k$ with the interval $(T_{k-1}, T_k)$, for $k \geq 1$, s.t. $R_k$ is independent of anything else given $(X_{k-1}, X_k)$ and $(T_k - T_{k-1})$
- Let $r_j$ be the expected reward in an interval that begins in state j
- Suppose $X_k$, $k \geq 0$, is a positive recurrent DTMC on S with stationary prob. vector $\pi$. Then under the conditions that $\sum_{j \in S} \pi_j \sigma_j < \infty$,

$$\lim_{t \to \infty} \frac{C(t)}{t} = \frac{\sum_{j \in S} \pi_j r_j}{\sum_{j \in S} \pi_j \sigma_j}, \text{ with probability } 1$$

## Outline

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network

## Excess distribution, or excess-life/residual-life distribution

- Given a nonnegative r.v. X with distribution F(.) and finite mean EX= $\int_0^\infty (1 - F(u))du$ , the excess distribution is defined as

$$F_e(y) = \frac{\int_0^y (1 - F(u))du}{EX}$$

- Can be interpreted as the distribution function of the *residual life seen by a random observer* of a renewal process with i.i.d. lifetime X

## Interpretation of $F_e(y)$

- Consider the renewal process with i.i.d. lifetimes $X_k$, $k \geq 1$, with distribution F(.); define Y(t) as the *residual life* or *excess life* at a random time t, i.e., the time until the first renewal in $(t, \infty)$
- Consider
  - $\lim_{t \to \infty} \frac{1}{t} \int_0^t I_{\{Y(u) \leq y\}} du$ : long-run fraction of time that the excess life is ≤y
  - $\lim_{t \to \infty} \frac{1}{t} \int_0^t \Pr(Y(u) \leq y) du$ : time average prob. that the excess life is ≤y
- Then, by Theorem D.15, How?

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t I_{\{Y(u) \leq y\}} du \xrightarrow{w.p.1} F_e(y) \qquad \lim_{t \to \infty} \frac{1}{t} \int_0^t \Pr(Y(u) \leq y) du \to F_e(y)$$

- Proof: define reward function $R_k = \min\{X_k, y\}$, then

$$C(t) = \int_0^t I_{\{Y(u) \le y\}} du$$
$$E(R_k) = \int_0^\infty u f_{R_k}(u) du = \int_0^y u f_{X_k}(u) du + y(1 - F(y)) = \int_0^y (1 - F(u)) du$$
$$E(C(t)) = \int_0^t \Pr(Y(u) \le y) du$$
…

## Outline

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network

## Phase type distribution

- For a CTMC X(t) on state space {1,2, …, M, *a*} s.t. states {1,2, …, M} are all transient and α is absorbing, the *transition rate matrix* of X(t) is of the form $\begin{bmatrix} Q & q \\ 0 & 0 \end{bmatrix}$
  where Q is an M*M matrix, q is a column vector of size M; ∃probability vector α of size M (i.e., $0 \le \alpha_j \le 1, \sum_{j=1}^{M} \alpha_j = 1$ ) s.t., the CTMC starts in state j with prob. $\alpha_j$, and then evolves to absorption state *a*
- Then, the distribution of the *time until absorption* is said to be *phase type* with parameters (α, Q, q)
  - When the process is at state j, it is said to be at *phase j*

## Example

- For

$$\alpha = (1,0,0,0), \mathrm{Q} = \begin{bmatrix} -\mu & \mu & 0 & 0 \\ 0 & -\mu & \mu & 0 \\ 0 & 0 & -\mu & \mu \\ 0 & 0 & 0 & -\mu \end{bmatrix}, q = (0,0,0,\mu)^T$$

The phase type distribution is an Erlang distribution of order 4, with each stage being exponentially distributed with mean 1/μ

## Why phase-type distribution?

- Phase-type distribution can be used to approximate arbitrarily closely (in the sense of convergence in distribution) any distribution
  - This fact may not always be useful for numerical approximation, due to the large # of phases required for good approximation
  - But it is very useful for *theoretical purposes*:
    - We can often prove results using phase type distributions thanks to their simple structure; then
    - We can prove that the result holds for any distribution by considering a sequence of phase type distributions converging to the general distributions

## Overflow process of M/M/c/c system

- The sequence of times at which customers are denied service forms a renewal process, and the distribution of these times is phase type with

$$\alpha = (0,0,...,0,1),\ Q = \begin{bmatrix} -\lambda & \lambda & 0 & ... & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 & ... & 0 \\ 0 & \mu & ... & ... & 0 & ... \\ ... & ... & ... & ... & ... & 0 \\ 0 & 0 & 0 & \mu & -(\lambda+\mu) & \lambda \\ 0 & 0 & 0 & 0 & \mu & -(\lambda+\mu) \end{bmatrix}$$

$$q = (0,0,...,\lambda)^T$$

## Outline

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network

## Poisson arrivals see time averages (PASTA)

Observations of a process X(t) at *random* time points

vs.

Observations of a process X(t) *over all time*

## Motivating example

- Consider a stable D/D/1 queue where customers arrive periodically at intervals of length a and requires a service time b<a. Let X(t) be the number of customers in the system at time t
- Then
  - Average # of customers over all time is $\lim_{t\to\infty}\frac{1}{t}\int_0^t X(u)du = \frac{b}{a}$
  - Now, observe x(t) at $t_k$=ka, k≥0 (i.e., what arrivals see on average)

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} X(t_{k-}) = 0$$

Point observations *differ* from average behaviors ☹

## Formal characterization

- Let X(t), t≥0, be a random process, and B be a subset of the state space of X(t); A(t) be a Poisson arrival process with rate λ, and $t_k$, k ≥1, be the arrival points
- Then
  - $V^B(t) = \frac{1}{t}\int_0^t I_{\{X(u)\in B\}}du$
    is the fraction of time over (0, t] that the process X(.) is in B
  - $V_A^B(t) = \frac{1}{A(t)}\sum_{k=1}^{A(t)} I_{\{X(t_{k-})\in B\}}$
    is the fraction of arrivals over (0, t] that see the process X(.) in B

---

- *Lack of anticipation assumption*: for all t≥0, A(t+u)-A(t), u≥0, is independent of X(s), 0≤s≤t
  - i.e., for all t≥0, future arrivals are independent of the past of X(.)
  - Note: the assumption holds for independent Poisson arrival processes
- Theorem D.17: under the lack of anticipation assumption,

$$V^B(t) \xrightarrow{w.p.1} \overline{V}_B \quad \text{iff.} \quad V_A^B(t) \xrightarrow{w.p.1} \overline{V}_B$$

  - i.e., time average and arrival average are the same

## Bernoulli/Geometric arrivals see time averages (GASTA)

- For queueing processes that evolve at discrete times $t_k$=kT, k=0,1,2,…, let $X_k$ denote the discrete time queue embedded at instants $t_k$
- Consider a Bernoulli arrival process of rate p, i.e., at times $t_{k+}$ an arrival occurs with prob. p
  - Also called a Geometric process since inter-arrival times are geometrically distributed
- Due to lack of anticipation, results similar to PASTA holds for Bernoulli/geometric arrivals and can be called GASTA

## Outline

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network

## Level crossing analysis (LCA)

- When direct derivation of stationary prob. distributions (via $\pi=\pi P$ or other means such as balance equations) is difficult, LCA may help obtain ancillary equations that provide some information about stationary distribution
- Given r.p. X(t) on $[0, \infty)$ and a $x \geq 0$
  - Up-crossing rate $U_x(t)$: # of times that X(.) crosses the "level" x from below
  - Down-crossing rate $D_x(t)$: # of times that X(.) crosses the "level" x from above

---

- *Level crossing analysis* is based on the following facts
  - $|U_x(t) - D_x(t)| \leq 1$, and
    $\lim_{t\to\infty} \frac{1}{t} U_x(t) = \lim_{t\to\infty} \frac{1}{t} D_x(t)$, if either limit exists
  - The above limits can usually be written in terms of the stationary distribution of the r.p.

## Outline

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network

## Some important queueing models

- M/G/c/c queue
- Processor sharing queue
- Symmetric queues

## M/G/c/c queue

- Poisson arrivals with finite rate $\lambda$
- Service requirements are i.i.d. and generally distributed with distribution F(.) and finite mean $1/\mu$
  - Service requirement is also called the *holding time*, since a customer "holds" a dedicated server for the entire duration of its service
- Each arriving customer is assigned to a free server if one exists; otherwise, the arriving customer is denied admission and it goes away
- Given an example of M/G/c/c queue?

---

- X(t): # of customers in queue at time t
- Let $\rho = \lambda/\mu$
  - In M/G/c/c, $\rho$ equals to the average # of new arrivals during the holding time of a customer (by Little's Theorem)
- [Exercise D.3]: if F(.) is an exponential distribution function, then X(t) is a positive recurrent CTMC on state space {0,1,…,c}, with stationary distribution

$$\pi_n = \frac{\frac{\rho^n}{n!}}{\sum_{j=0}^{c} \frac{\rho^j}{j!}}$$

---

- When F(.) is not an exponential distribution function, X(t) is not Markovian
- But $(X(t), Y_1(t), \ldots, Y_{X(t)}(t))$ is a Markov process, where $Y_i(t)$ denotes the residual service requirement of the i-th customer in the system; and

$$\Pr(X(t)=n, Y_1 \le y_1, \ldots, Y_n \le y_n) = \pi_n \prod_{i=1}^{n} F_e(y_i),$$

where $\pi_n$ is as in the case of exponential hodling time,

$F_e(.)$ is the excess distribution of the holding time distribution F(.)

## Processor sharing queue: M/G/1 PS

- Poisson arrivals with finite rate λ
- Service requirements are i.i.d. and generally distributed with distribution F(.) and finite mean 1/μ
- Overall service rate: 1 unit per second

*(fair) processor sharing rule*: when there n customers in system, the unfinished work on the i-th customer decreases at rate 1/n

---

- Let $\rho = \frac{\lambda}{\mu}$
- X(t) denotes the # of customers at time t
- If F(.) is for a exponential distribution, then X(t) is a CTMC, and it is positive recurrent iff. <1, in which case the stationary distribution of X(t) is given by

$$\pi_n = (1-\rho)\rho^n$$

---

- When F(.) is not an exponential distribution function, X(t) is not Markovian
- But (X(t), $Y_1$(t),…, $Y_{X(t)}$(t)) is a Markov process, where $Y_i$(t) denotes the residual service requirement of the i-th customer in the system; and if ρ<1

$$\Pr(X(t)=n, Y_1 \le y_1, \ldots, Y_n \le y_n) = (1-\rho)\rho^n \prod_{i=1}^{n} F_e(y_i),$$
$$\text{where } F_e(.) \text{is the excess distribution of } F(.)$$

  - Note: the stationary distribution of X(t) in an M/G/1 PS queue is the same as that in an M/M/1 queue, and thus is insensitive to the distribution of F(.) (except through its mean)

---

## Sojourn times in M/G/1 PS

- Sojourn time W: amount of time that a customer stays in the system
- Since $\pi_n = (1-\rho)\rho^n$ , E(x) = ρ/(1- ρ); then, by Little's Theorem,

$$E(W) = \frac{E(S)}{1-\rho}, \text{where } E(S) = \frac{1}{\mu} \text{ and is the mean service requirement}$$

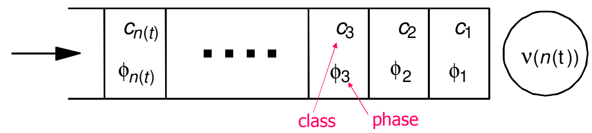
$$\text{Moreover, } E(W \mid S = s) = \frac{s}{1-\rho}$$

## Symmetric queue

- Consider the following queue:
  - Customers of class c, $c \in C$, arrive in independent Poisson processes of rate $\lambda_c$
  - Customers of class c have a phase type service requirement with parameter $(\alpha_c, Q_c)$ and mean $1/\mu_c$
  - An arriving customer finding (n-1) customers in the system joins in *position* l, $1 \le l \le n$, with prob. $\gamma(n, l)$
  - When there are n customers in the queue, the overall service rate applied is $\nu(n)$; and a *fraction* $\delta(n, l)$ of the service effort is applied to the customer at position l

---

- A system state:




- A aforementioned queueing system is said to be a *symmetric queue* if the functions $\delta(.,.)$ and $\gamma(.,.)$ are such that $\delta(n, l) = \gamma(n, l)$
  - Positioning implies priority

## Examples

- M/PH/1 queue with last-come-first-serve preemptive resume (LCFS-PR) discipline
  - $\nu(n)$ = constant $\nu$
  - $\gamma(n, 1) = 1$, and $\gamma(n, j) = 0$ for $j>1$
  - $\delta(n, 1) = 1$, and $\delta(n, j) = 0$ for $j>1$
- M/PH/1 processor sharing queue?
  - $\nu(n)$ = constant $\nu$
  - $\gamma(n, l) = \delta(n, l) = 1/n$, for $1 \le l \le n$

---

- M/PH/$\infty$ queue?
  - $\nu(n) = n\nu$
  - $\gamma(n, l) = \delta(n, l) = 1/n$, for $1 \le l \le n$

## Stationary distribution

- Let $\rho = \sum_{c \in C} \frac{\lambda_c}{\mu_c}$
- Theorem D.18: the stationary distribution of the # of customers in a symmetric queue is given by the prob. distribution of system state x

$$\pi(x) = G \frac{\rho^{|x|}}{\nu(1)\nu(2)...\nu(|x|)}$$

where $|x| = \#$ of customers at state x, and G is the normalization constant

  - Note: the distribution is insensitive to the service requirement distributions (except for their mean $1/\mu_c$, $c \in C$)

## Outline

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network

## Time Reversibility and Burke's Theorem

- Time-Reversal of Markov Chains
- Reversibility
- Truncating a Reversible Markov Chain
- Burke's Theorem
- Queues in Tandem


## Time Reversibility and Burke's Theorem

- Time-Reversal of Markov Chains
- Reversibility
- Truncating a Reversible Markov Chain
- Burke's Theorem
- Queues in Tandem

## Time-Reversed Markov Chains

- $\{X_n: n=0,1,\ldots\}$ irreducible aperiodic Markov chain with transition probabilities $P_{ij}$

$$\sum_{j=0}^{\infty} P_{ij} = 1, \quad i = 0,1,\ldots$$

- Unique stationary distribution ($\pi_j > 0$) iff. GBE holds, i.e.,

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad j = 0,1,\ldots$$

- Process in steady state:

$$\Pr\{X_n = j\} = \pi_j = \lim_{n\to\infty} \Pr\{X_n = j \mid X_0 = i\}$$

  - Starts at n=-∞, that is $\{X_n: n = \ldots,-1,0,1,\ldots\}$, or
  - Choose initial state according to the stationary distribution
- *How does $\{X_n\}$ look "reversed" in time?*

## Time-Reversed Markov Chains

- Define $Y_n=X_{\tau-n}$, for arbitrary $\tau>0$

  => $\{Y_n\}$ is the reversed process.

- Proposition 1:
  - $\{Y_n\}$ is a Markov chain with transition probabilities:

$$P^*_{ij} = \frac{\pi_j}{\pi_i} P_{ji}, \quad i, j = 0,1,2,\ldots$$

  - $\{Y_n\}$ has the same stationary distribution $\pi_j$ with the forward chain $\{X_n\}$
    - The reversed chain corresponds to the same process, looked at in the reversed-time direction

## Time-Reversed Markov Chains

Proof of Proposition 1:

$$\begin{aligned}
\Rightarrow P^*_{ij} &= P\{Y_m = j \mid Y_{m-1} = i, Y_{m-2} = i_2, \ldots, Y_{m-k} = i_k\} \\
&= P\{X_{\tau-m} = j \mid X_{\tau-m+1} = i, X_{\tau-m+2} = i_2, \ldots, X_{\tau-m+k} = i_k\} \\
&= P\{X_n = j \mid X_{n+1} = i, X_{n+2} = i_2, \ldots, X_{n+k} = i_k\} \\
&= \frac{P\{X_n = j, X_{n+1} = i, X_{n+2} = i_2, \ldots, X_{n+k} = i_k\}}{P\{X_{n+1} = i, X_{n+2} = i_2, \ldots, X_{n+k} = i_k\}} \\
&= \frac{P\{X_{n+2} = i_2, \ldots, X_{n+k} = i_k \mid X_n = j, X_{n+1} = i\} P\{X_n = j, X_{n+1} = i\}}{P\{X_{n+2} = i_2, \ldots, X_{n+k} = i_k \mid X_{n+1} = i\} P\{X_{n+1} = i\}} \\
&= \frac{P\{X_n = j, X_{n+1} = i\}}{P\{X_{n+1} = i\}} = P\{X_n = j \mid X_{n+1} = i\} = P\{Y_m = j \mid Y_{m-1} = i\} \\
&= \frac{P\{X_{n+1} = i \mid X_n = j\} P\{X_n = j\}}{P\{X_{n+1} = i\}} = \frac{P_{ji}\pi_j}{\pi_i}
\end{aligned}$$

$$\Rightarrow \sum_{i=0}^{\infty} \pi_i P^*_{ij} = \sum_{i=0}^{\infty} \pi_i \frac{\pi_j P_{ji}}{\pi_i} = \pi_j \sum_{i=0}^{\infty} P_{ji} = \pi_j$$

## Time Reversibility and Burke's Theorem

- Time-Reversal of Markov Chains
- Reversibility
- Truncating a Reversible Markov Chain
- Burke's Theorem
- Queues in Tandem

## Reversibility

- Stochastic process {X(t)} is called *reversible* if $(X(t_1), X(t_2),\ldots, X(t_n))$ and $(X(\tau-t_1), X(\tau-t_2),\ldots, X(\tau-t_n))$ have the same probability distribution, for all $\tau, t_1,\ldots, t_n$
- <u>Proposition D.1</u>: if $\{X(t), t\in R\}$ is stationary, then a time reversed process is also stationary
- <u>Proposition D.2</u>: a reversible process is stationary (and consequently any time reversal of a reversible process is stationary).

---

- Markov chain $\{X_n\}$ is *reversible* if and only if the transition probabilities of forward and reversed chains are equal, i.e.,

$$P_{ij} = P_{ij}^*$$

- Detailed Balance Equations ↔ Reversibility

$$\pi_i P_{ij} = \pi_j P_{ji}, \quad i, j = 0,1,\ldots$$

---

## Reversibility – Discrete-Time Chains

- <u>Theorem 1:</u> If there exists a set of positive numbers $\{\pi_j\}$, that sum up to 1 and satisfy:

$$\pi_i P_{ij} = \pi_j P_{ji}, \quad i, j = 0,1,\ldots$$

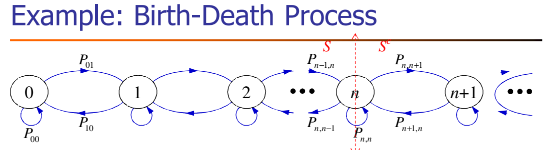
Then:

1. $\{\pi_j\}$ is the unique stationary distribution
2. The Markov chain is reversible

- <u>Example:</u> Discrete-time birth-death processes are reversible, since they satisfy the DBE

---

## Example: Birth-Death Process




- One-dimensional Markov chain with transitions only between neighboring states: $P_{ij}=0$, if $|i-j|>1$
- Detailed Balance Equations (DBE)

$$\pi_n P_{n,n+1} = \pi_{n+1} P_{n+1,n} \quad n = 0,1,\ldots$$

- Proof: GBE with $S=\{0,1,\ldots,n\}$ give:

$$\sum_{j=0}^{n}\sum_{i=n+1}^{\infty} \pi_j P_{ji} = \sum_{j=0}^{n}\sum_{i=n+1}^{\infty} \pi_i P_{ij} \Rightarrow \pi_n P_{n,n+1} = \pi_{n+1} P_{n+1,n}$$

## Time-Reversed Markov Chains (Revisited)

- Theorem 2: Irreducible Markov chain with transition probabilities $P_{ij}$. If there exist:
  - A set of transition probabilities $P_{ij}^*$, **with $\sum_j P_{ij}^*=1$**, $i \geq 0$, and
  - A set of positive numbers $\{\pi_j\}$, that sum up to 1, such that

$$\pi_i P_{ij}^* = \pi_j P_{ji}, \quad i, j \geq 0$$

  Then:
  - $P_{ij}^*$ are the transition probabilities of the reversed chain, and
  - $\{\pi_j\}$ is the stationary distribution of the forward and the reversed chains

- Remark: Used to find the stationary distribution, by guessing the transition probabilities of the reversed chain – even if the process is not reversible

## Continuous-Time Markov Chains

- $\{X(t): -\infty < t < \infty\}$ irreducible aperiodic Markov chain with transition rates $q_{ij}$, $i \neq j$
- Unique stationary distribution ($p_i > 0$) iff.

$$p_j \sum_{i \neq j} q_{ji} = \sum_{i \neq j} p_i q_{ij}, \quad j = 0,1,...$$

- Process in steady state – e.g., started at $t = -\infty$:

$$\Pr\{X(t) = j\} = p_j = \lim_{t \to \infty} \Pr\{X(t) = j \mid X(0) = i\}$$

- If $\{\pi_j\}$, is the stationary distribution of the embedded discrete-time chain:

$$p_j = \frac{\pi_j / \nu_j}{\sum_i \pi_i / \nu_i}, \quad \nu_j \equiv \sum_{i \neq j} q_{ji}, \quad j = 0,1,...$$

## Reversed Continuous-Time Markov Chains

- Reversed chain $\{Y(t)\}$, with $Y(t)=X(\tau-t)$, for arbitrary $\tau>0$
- Proposition 2:

1. $\{Y(t)\}$ is a continuous-time Markov chain with transition rates:

$$q_{ij}^* = \frac{p_j q_{ji}}{p_i}, \quad i, j = 0,1,..., i \neq j$$

2. $\{Y(t)\}$ has the same stationary distribution $\{p_j\}$ with the forward chain

- Remark: The transition rate out of state $i$ in the reversed chain is equal to the transition rate out of state $i$ in the forward chain

$$\sum_{j \neq i} q_{ij}^* = \frac{\sum_{j \neq i} p_j q_{ji}}{p_i} = \frac{p_i \sum_{j \neq i} q_{ij}}{p_i} = \sum_{j \neq i} q_{ij} = \nu_i, \quad i = 0,1,...$$

## Reversibility – Continuous-Time Chains

- Markov chain $\{X(t)\}$ is *reversible* iff. the transition rates of forward and reversed chains are equal $q_{ij} = q_{ij}^*$, or equivalently

$$p_i q_{ij} = p_j q_{ji}, \quad i, j = 0,1,..., i \neq j$$

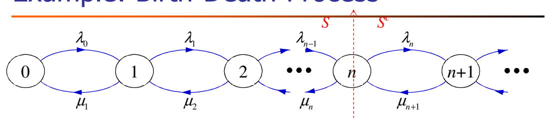
  i.e., Detailed Balance Equations ↔ Reversibility

- Theorem 3: If there exists a set of positive numbers $\{p_j\}$, that sum up to 1 and satisfy:

$$p_i q_{ij} = p_j q_{ji}, \quad i, j = 0,1,..., i \neq j$$

  Then:
  1. $\{p_j\}$ is the unique stationary distribution
  2. The Markov chain is reversible

## Example: Birth-Death Process




- Transitions only between neighboring states

  $q_{i,i+1}=\lambda_i,\quad q_{i,i-1}=\mu_i,\quad q_{ij}=0,\ |i-j|>1$

- Detailed Balance Equations

  $$\lambda_n p_n = \mu_{n+1} p_{n+1},\quad n=0,1,\ldots$$

- Proof: GBE with $S=\{0,1,\ldots,n\}$ give:

  $$\sum_{j=0}^{n}\sum_{i=n+1}^{\infty} p_j q_{ji} = \sum_{j=0}^{n}\sum_{i=n+1}^{\infty} p_i q_{ij} \Rightarrow \lambda_n p_n = \mu_{n+1} p_{n+1}$$

- M/M/1, M/M/c, M/M/∞

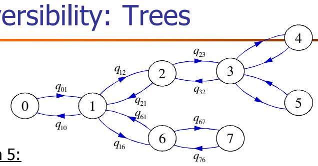
## Reversed Continuous-Time Markov Chains (Revisited)

- Theorem 4: Irreducible continuous-time Markov chain with transition rates $q_{ij}$. If there exist:
  - A set of transition rates $q_{ij}^*$, with $\sum_{j\neq i} q_{ij}^* = \sum_{j\neq i} q_{ij}$, $i \geq 0$, and
  - A set of positive numbers $\{p_j\}$, that sum up to 1, such that

  $$p_i q_{ij}^* = p_j q_{ji},\quad i,j\geq 0, i\neq j$$

  Then:
  - $q_{ij}^*$ are the transition rates of the reversed chain, and
  - $\{p_j\}$ is the stationary distribution of the forward and the reversed chains
- Remark: Used to find the stationary distribution, by guessing the transition probabilities of the reversed chain – even if the process is not reversible

## Reversibility: Trees




Theorem 5:

- Irreducible Markov chain, with transition rates that satisfy $q_{ij}>0 \leftrightarrow q_{ji}>0$
- Form a graph for the chain, where states are the nodes, and for each $q_{ij}>0$, there is a directed arc $i\to j$

Then, if graph is a tree – contains no loops – then Markov chain is reversible

Remarks:

- Sufficient condition for reversibility
- Generalization of one-dimensional birth-death process

## Kolmogorov's Criterion (Discrete Chain)

- Detailed balance equations determine whether a Markov chain is reversible or not, based on stationary distribution and *transition probabilities*
- Should be able to derive a reversibility criterion based only on the transition probabilities!
- Theorem D.20: A discrete-time Markov chain is reversible *iff.*

  $$P_{i_1 i_2} P_{i_2 i_3} \cdots P_{i_{n-1} i_n} P_{i_n i_1} = P_{i_1 i_n} P_{i_n i_{n-1}} \cdots P_{i_3 i_2} P_{i_2 i_1}$$

  for every finite sequence of states: $i_1, i_2, \ldots, i_n$, and any $n$
- Intuition: Probability of traversing any loop $i_1\to i_2\to\ldots\to i_n\to i_1$ is equal to the probability of traversing the same loop in the reverse direction $i_1\to i_n\to\ldots\to i_2\to i_1$

## Kolmogorov's Criterion (Continuous Chain)

- Detailed balance equations determine whether a Markov chain is reversible or not, based on stationary distribution and *transition rates*
- Should be able to derive a reversibility criterion based only on the transition rates!
- Theorem 7: A continuous-time Markov chain is reversible *if and only if*:

$$q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{n-1} i_n} q_{i_n i_1} = q_{i_1 i_n} q_{i_n i_{n-1}} \cdots q_{i_3 i_2} q_{i_2 i_1}$$

for any finite sequence of states: $i_1, i_2, \ldots, i_n$, and any $n$

- Intuition: Product of transition rates along any loop $i_1 \to i_2 \to \ldots \to i_n \to i_1$ is equal to the product of transition rates along the same loop traversed in the reverse direction $i_1 \to i_n \to \ldots \to i_2 \to i_1$

## Kolmogorov's Criterion (proof)

Proof of Theorem D.20:

- Necessary: If the chain is reversible the DBE hold

$$\left.\begin{aligned} \pi_1 P_{i_1 i_2} &= \pi_2 P_{i_2 i_1} \\ \pi_2 P_{i_2 i_3} &= \pi_3 P_{i_3 i_2} \\ &\vdots \\ \pi_{n-1} P_{i_{n-1} i_n} &= \pi_n P_{i_n i_{n-1}} \\ \pi_n P_{i_n i_1} &= \pi_1 P_{i_1 i_n} \end{aligned}\right\} \Rightarrow P_{i_1 i_2} P_{i_2 i_3} \cdots P_{i_{n-1} i_n} P_{i_n i_1} = P_{i_1 i_n} P_{i_n i_{n-1}} \cdots P_{i_3 i_2} P_{i_2 i_1}$$

- Sufficient: Fixing two states $i_1 = i$, and $i_n = j$ and summing over all states $i_2, \ldots, i_{n-1}$ we have

$$P_{i,i_2} P_{i_2 i_3} \cdots P_{i_{n-1},j} P_{ji} = P_{ij} P_{j,i_{n-1}} \cdots P_{i_3 i_2} P_{i_2,i} \Rightarrow P_{ij}^{n-1} P_{ji} = P_{ij} P_{ji}^{n-1}$$

Taking the limit $n \to \infty$

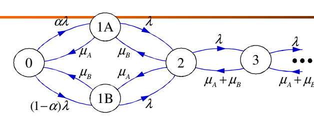
$$\lim_{n\to\infty} P_{ij}^{n-1} \cdot P_{ji} = P_{ij} \cdot \lim_{n\to\infty} P_{ji}^{n-1} \Rightarrow \pi_j P_{ji} = P_{ij} \pi_i$$

- Theorem D.21: A discrete-time Markov chain is reversible *iff*:

$$q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{n-1} i_n} q_{i_n i_1} = q_{i_1 i_n} q_{i_n i_{n-1}} \cdots q_{i_3 i_2} q_{i_2 i_1}$$

for every *minimal*, finite sequence of states: $i_1, i_2, \ldots, i_n$

## Example: M/M/2 Queue with Heterogeneous Servers




- M/M/2 queue. Servers A and B with service rates $\mu_A$ and $\mu_B$ respectively. When the system empty, arrivals go to A with probability α and to B with probability 1-α. Otherwise, the head of the queue takes the first free server.
- Need to keep track of which server is busy when there is 1 customer in the system. Denote the two possible states by: 1A and 1B.
- Reversibility: we only need to check the loop 0→1A→2→1B→0:

$$q_{0,1A} q_{1A,2} q_{2,1B} q_{1B,0} = \alpha\lambda \cdot \lambda \cdot \mu_A \cdot \mu_B \qquad q_{0,1B} q_{1B,2} q_{2,1A} q_{1A,0} = (1-\alpha)\lambda \cdot \lambda \cdot \mu_B \cdot \mu_A$$

  - Reversible if and only if α=1/2.

## Time Reversibility and Burke's Theorem

- Time-Reversal of Markov Chains
- Reversibility
- Truncating a Reversible Markov Chain
- Burke's Theorem
- Queues in Tandem

## Truncation of a Reversible Markov Chain

- Theorem D.22: {X(t)} reversible Markov process with state space *S*, and stationary distribution $\{p_j: j \in S\}$. Truncated to a set $E \subset S$, such that the resulting chain {Y(t)} is *irreducible*. Then, {Y(t)} is reversible and has stationary distribution:

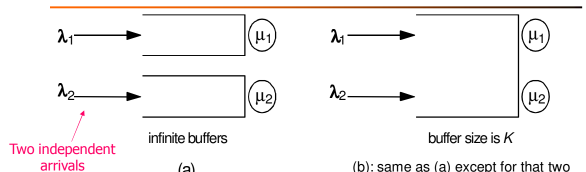
$$\tilde{p}_j = \frac{p_j}{\sum_{k \in E} p_k}, \quad j \in E$$

- Remark: This is the conditional probability that, in steady-state, the original process is at state *j*, given that it is somewhere in *E*
- Proof: Verify that:

$$\tilde{p}_j q_{ji} = \tilde{p}_i q_{ij} \Leftrightarrow \frac{p_j}{\sum_{k \in E} p_k} q_{ji} = \frac{p_i}{\sum_{k \in E} p_k} q_{ij} \Leftrightarrow p_j q_{ji} = p_i q_{ij}, \quad i, j \in S; i \neq j$$

$$\sum_{j \in E} \tilde{p}_j = \sum_{j \in E} \frac{p_j}{\sum_{k \in E} p_k} = 1$$

## Example




(a)

(b): same as (a) except for that two Arrivals share a capacity-limited buffer

- Joint process of queue length ($X_1(t)$, $X_2(t)$) is a CTMC
- For (a): $\pi^{(a)}_{n1,n2} = (1-\rho_1)\rho_1^{n1}(1-\rho_2)\rho_2^{n2}$
- (b) is a truncated version of (a) in the sense E = {(n1+n2)≥0: n1+n2≤K}, thus

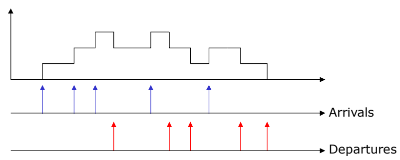
$$\pi^{(b)}_{n1,n2} = \frac{(1-\rho_1)\rho_1^{n1}(1-\rho_2)\rho_2^{n2}}{\sum_{(k1,k2)\in E}(1-\rho_1)\rho_1^{k1}(1-\rho_2)\rho_2^{k2}}$$

## Time Reversibility and Burke's Theorem

- Time-Reversal of Markov Chains
- Reversibility
- Truncating a Reversible Markov Chain
- Burke's Theorem
  - Birth-death processes: Poisson departures
- Queues in Tandem

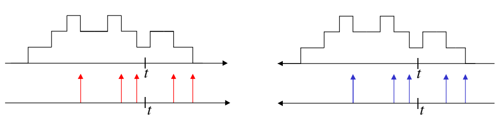
## Birth-death process

- {X(t)} birth-death process with stationary distribution $\{p_j\}$
- Arrival epochs: points of increase for {X(t)}
  Departure epoch: points of decrease for {X(t)}
- {X(t)} completely determines the corresponding arrival and departure processes




## Forward & reversed chains of birth-death processes

- Poisson arrival process: $\lambda_j=\lambda$, for all $j$
  - Birth-death process called a $(\lambda, \mu_j)$-process
  - Examples: M/M/1, M/M/c, M/M/∞ queues
- Poisson arrivals → LAA: for any time t, future arrivals are independent of {X(s): s≤t}
- $(\lambda, \mu_j)$-process at steady state is reversible: forward and reversed chains are stochastically identical
- => Arrival processes of the forward and reversed chains are stochastically identical
  - => Arrival process of the reversed chain is Poisson with rate λ
  - + "the arrival epochs of the reversed chain are the departure epochs of the forward chain" => **Departure process of the forward chain is Poisson with rate λ**




- Reversed chain: arrivals after time t are independent of the chain history up to time t (LAA)
- => Forward chain: departures prior to time t and future of the chain {X(s): s≥t} are independent

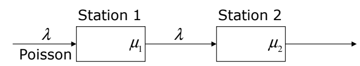
## Burke's Theorem

- <u>Theorem 10:</u> Consider a $(\lambda, \mu_j)$-process (e.g., those in M/M/1, M/M/c, or M/M/∞ systems). *Suppose that the system starts at steady-state.* Then:
  1. The departure process is Poisson with rate λ
  2. At each time *t*, the number of customers in the system is independent of the departure times prior to *t*
- Fundamental result for study of networks of M/M/* queues, where output process from one queue is the input process of another

## Time Reversibility and Burke's Theorem

- Time-Reversal of Markov Chains
- Reversibility
- Truncating a Reversible Markov Chain
- Burke's Theorem
- Queues in Tandem

## Single-Server Queues in Tandem

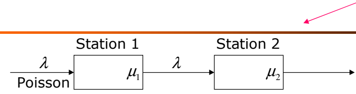



- Customers arrive at queue 1 according to Poisson process with rate λ.
- Service times exponential with mean $1/\mu_i$. *Assume service times of a customer in the two queues are independent.*
- Assume $\rho_i = \lambda/\mu_i < 1$
- What is the joint *stationary* distribution of $N_1$ and $N_2$ – number of customers in each queue?

$$p(n_1, n_2) = (1-\rho_1)\rho_1^{n_1} \cdot (1-\rho_2)\rho_2^{n_2} = p_1(n_1) \cdot p_2(n_2)$$

- Result: in *steady state* the queues are independent

Note: if N1(t) is not its stationary version, $N_1(t)$ and N2(t) are NOT independent. The asymptotic result, however, still holds.




- Q1 is a M/M/1 queue. At steady state its departure process is Poisson with rate λ. Thus Q2 is also M/M/1.
- Marginal stationary distributions:

$$p_1(n_1) = (1-\rho_1)\rho_1^{n_1}, \quad n_1 = 0,1,\ldots \qquad p_2(n_2) = (1-\rho_2)\rho_2^{n_2}, \quad n_2 = 0,1,\ldots$$

- To complete the proof: establish independence at steady state
- Q1 at steady state: at time t, $N_1(t)$ is independent of departures prior to t, which are arrivals at Q2 up to t. Thus $N_1(t)$ and $N_2(t)$ independent:

$$P\{N_1(t) = n_1, N_2(t) = n_2\} = P\{N_1(t) = n_1\}P\{N_2(t) = n_2\} = p_1(n_1) \cdot P\{N_2(t) = n_2\}$$

- Letting $t \to \infty$, the joint stationary distribution

$$p(n_1, n_2) = p_1(n_1) \cdot p_2(n_2) = (1-\rho_1)\rho_1^{n_1} \cdot (1-\rho_2)\rho_2^{n_2}$$

## Queues in Tandem

- Theorem: Network consisting of $K$ single-server queues in tandem. Service times at queue $i$ exponential with rate $\mu_i$, independent of service times at any queue $j \neq i$. Arrivals at the first queue are Poisson with rate λ. The stationary distribution of the network is:

$$p(n_1, \ldots, n_K) = \prod_{i=1}^{K} (1-\rho_i)\rho_i^{n_i}, \quad n_i = 0,1,\ldots; i = 1,\ldots,K$$

- At *steady state* the queues are independent; the distribution of queue $i$ is that of an isolated M/M/1 queue with arrival and service rates λ and $\mu_i$

$$p_i(n_i) = (1-\rho_i)\rho_i^{n_i}, \quad n_i = 0,1,\ldots$$

- Are the queues independent if not in steady state? Are stochastic processes $\{N_1(t)\}$ and $\{N_2(t)\}$ independent?

## Queues in Tandem: State-Dependent Service Rates

- Theorem 12: Network consisting of $K$ queues in tandem. Service times at queue $i$ exponential with rate $\mu_i(n_i)$ when there are $n_i$ customers in the queue – independent of service times at any queue $j \neq i$. Arrivals at the first queue are Poisson with rate $\lambda$. The stationary distribution of the network is:

$$p(n_1,\ldots,n_K) = \prod_{i=1}^{K} p_i(n_i), \quad n_i = 0,1,\ldots; i = 1,\ldots,K$$

where $\{p_i(n_i)\}$ is the stationary distribution of queue $i$ in isolation with Poisson arrivals with rate $\lambda$

- Examples: ./M/c and ./M/∞ queues
  - If queue $i$ is ./M/∞, then:

$$p_i(n_i) = \frac{(\lambda/\mu_i)^{n_i}}{n_i!} e^{-\lambda/\mu_i}, \quad n_i = 0,1,\ldots$$

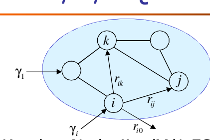
## Jackson Networks

- Open Jackson Networks
- Network Flows
- State-Dependent Service Rates
- Networks of Transmission Lines & Kleinrock's Assumption
- Closed Jackson Networks

## Jackson Networks

- Open Jackson Networks
- Network Flows
- State-Dependent Service Rates
- Networks of Transmission Lines & Kleinrock's Assumption
- Closed Jackson Networks

## Networks of ./M/1 Queues




- Network of $K$ nodes; Node $i$ is ./M/1-FCFS queue with service rate $\mu_i$
- External arrivals independent Poisson processes
  - $\gamma_i$: rate of external arrivals at node $i$
- *Markovian routing*: customer completing service at node $i$
  - is routed to node $j$ with probability $r_{ij}$ or
  - exits the network with probability $r_{i0}=1-\Sigma_j r_{ij}$
- Routing matrix $R=[r_{ij}]$ irreducible ⇒ *external arrivals eventually exit the system*

## Jackson Network

- Definition: A *Jackson network* is the CTMC $\{N(t)\}$, with $N(t)=(N_1(t),\dots, N_K(t))$ that describes the evolution of the previously defined network, where $N_i(t)$ = # of customers at node i
- Possible states: $n=(n_1, n_2,\dots, n_K)$, $n_i=1,2,\dots$, $i=1,2,..,K$
- For any state $n$, define the following operators:

$$A_i n = n + e_i \quad \text{arrival at } i$$
$$D_i n = n - e_i \quad \text{departure from } i$$
$$T_{ij} n = n - e_i + e_j \quad \text{transition from } i \text{ to } j$$

$e_i=(0,0,\dots,1,0,\dots,0)$ is unit vector with length K and the i-th position being 1

- Transition rates for the Jackson network:

$$q(n, A_i n) = \gamma_i$$
$$q(n, D_i n) = \mu_i r_{i0} \cdot \mathbf{1}\{n_i > 0\} \qquad i, j = 1,\dots, K$$
$$q(n, T_{ij} n) = \mu_i r_{ij} \cdot \mathbf{1}\{n_i > 0\}$$

while $q(n,m)=0$ for all other states $m$

## Jackson's Theorem for Open Networks

- $\lambda_i$: *total* arrival rate at node $i$

$$\lambda_i = \gamma_i + \sum_{j=1}^{K} \lambda_j r_{ji}, \quad i = 1,\dots, K$$

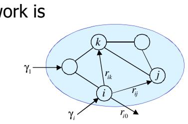
  - *Open network*: for some node $j$: $\gamma_j > 0$
    - Routing matrix is irreducible => Linear system has a unique solution $\lambda_1, \lambda_2,\dots, \lambda_K$

- Theorem 13: Consider a Jackson network, where $\rho_i=\lambda_i/\mu_i < 1$, for every node $i$. The stationary distribution of the network is

$$p(n) = \prod_{i=1}^{K} p_i(n_i), \quad n_1,\dots,n_K \geq 0$$

where for every node $i$ =1,2,…,$K$

$$p_i(n_i) = (1-\rho_i)\rho_i^{n_i}, \quad n_i \geq 0$$




## Jackson's Theorem (proof)

- Guess the reverse Markov chain and use Theorem 4
- Claim: The network reversed in time is a Jackson network with the same service rates, while the arrival rates and routing probabilities are

$$\gamma_i^* = \lambda_i r_{i0}, \quad r_{ij}^* = \frac{\lambda_j r_{ji}}{\lambda_i}, \quad r_{i0}^* = \frac{\gamma_i}{\lambda_i}$$

- Verify that for any states $n$ and $m \neq n$,

$$p(m) q^*(m,n) = p(n) q(n,m)$$

  Need to prove only for $m=A_i n$, $D_i n$, $T_{ij} n$. We show the proof for the first two cases – the third is similar

  - $q^*(A_i n, n) = q^*(A_i n, D_i A_i n) = \mu_i r_{i0}^* = \mu_i (\gamma_i / \lambda_i)$
    $p(A_i n) q^*(A_i n, n) = p(n) q(n, A_i n) \Leftrightarrow p(A_i n) \mu_i (\gamma_i / \lambda_i) = p(n)\gamma_i \Leftrightarrow p(A_i n) = \rho_i p(n)$
  - $q^*(D_i n, n) = q^*(D_i n, A_i D_i n) = \gamma_i^* = \lambda_i r_{i0}$
    $p(D_i n) q^*(D_i n, n) = p(n) q(n, D_i n) \Leftrightarrow p(D_i n)\lambda_i r_{i0} = p(n)\mu_i r_{i0} 1\{n_i > 0\}$
    $\Leftrightarrow \rho_i p(D_i n) = p(n) 1\{n_i > 0\}$

## Jackson's Theorem (proof cont.)

- Finally, verify that for any state $n$:

$$\sum_{m \neq n} q(n,m) = \sum_{m \neq n} q^*(n,m)$$

  - $\sum_{m \neq n} q(n,m) = \sum_i \gamma_i + \sum_{i,j} \mu_i r_{ij} 1\{n_i > 0\} + \sum_i \mu_i r_{i0} 1\{n_i > 0\}$
    $= \sum_i \gamma_i + \sum_i \mu_i [\sum_j r_{ij} + r_{i0}] \cdot 1\{n_i > 0\}$
    $= \sum_i \gamma_i + \sum_i \mu_i 1\{n_i > 0\}$
  - $\sum_{m \neq n} q^*(n,m) = \sum_i \gamma_i^* + \sum_i \mu_i 1\{n_i > 0\} = \sum_i \lambda_i r_{i0} + \sum_i \mu_i 1\{n_i > 0\}$

- Thus, we need to show that $\sum_i \gamma_i = \sum_i \lambda_i r_{i0}$

$$\sum_i \lambda_i r_{i0} = \sum_i \lambda_i - \sum_i \sum_j \lambda_i r_{ij} = \sum_i \lambda_i - \sum_j \sum_i \lambda_i r_{ij}$$
$$= \sum_i \lambda_i - \sum_j (\lambda_j - \gamma_j) = \sum_j \gamma_j$$

## Output Theorem for Jackson Networks

- Theorem 14: The reversed chain of a stationary open Jackson network is also a stationary open Jackson network with the same service rates, while the arrival rates and routing probabilities are

$$\gamma_i^* = \lambda_i r_{i0}, \quad r_{ij}^* = \frac{\lambda_j r_{ji}}{\lambda_i}, \quad r_{i0}^* = \frac{\gamma_i}{\lambda_i}$$

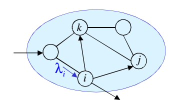
- Theorem 15: In a stationary open Jackson network, the departure process from the system at node *i* is Poisson with rate $\lambda_i r_{i0}$. The departure processes are independent of each other, and at any time t, their past up to t is independent of the state of the system N(t).
- *Remark*: 1) The total arrival process at a given node is not Poisson. The departure process from the node is not Poisson either. However, the process of the customers that *exit the network* at the node is Poisson. 2) In general, an open Jackson network need not be reversible

## Arrival Theorem in Open Jackson Networks

- The composite arrival process at node *i* in an open Jackson network has the "PASTA" property, although it need not be a Poisson process
- Theorem 16: In an open Jackson network at steady-state, the probability that a composite arrival at node *i* finds *n* customers at that node is equal to the (unconditional) probability of n customers at that node:

$$p_i(n) = (1-\rho_i)\rho_i^n, \quad n \geq 0, i = 1, \ldots, K$$

(Proof is omitted)



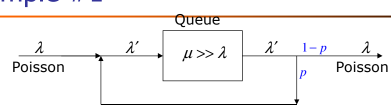

## Jackson Networks

- Open Jackson Networks
- Network Flows
- State-Dependent Service Rates
- Networks of Transmission Lines & Kleinrock's Assumption
- Closed Jackson Networks

## Non-Poisson Internal Flows

- Jackson's theorem: the numbers of customers in the queues are distributed as if each queue *i* is an isolated M/M/1 with arrival rate $\lambda_i$, independent of all others
- Total arrival process at a queue, however, need not be Poisson
  - "Loops" allow a customer to visit the same queue multiple times and introduce dependencies that violate the Poisson property
  - Internal flows are Poisson in *acyclic* networks
- Similarly. the departure process from a queue is not Poisson in general
  - The process of departures that *exit the network* at the node is Poisson according to the output theorem

## Example #1




- Example: Single queue with μ >> λ, where upon service completion a customer is fed back with probability $p \approx 1$, joining the end of the queue
- The total arrival process does not have independent interarrival times:
  - If an arrival occurs at time t, there is a very high probability that a feedback arrival will follow in (t, t+δ]
  - At arbitrary t, the probability of an arrival in (t, t+δ] is small since λ is small
- Arrival process consists of bursts, each burst triggered by a single customer arrival

## Jackson Networks

- Open Jackson Networks
- Network Flows
- State-Dependent Service Rates
- Networks of Transmission Lines & Kleinrock's Assumption
- Closed Jackson Networks

## State-Dependent Service Rates

- Service rate at node $i$ depends on the number of customers at that node: $\mu_i(n_i)$ when there are $n_i$ customers at node $i$
  - But service rate at i does not depend on the # of customers at other nodes
  - E.g, ./M/c and ./M/∞ queues
- Theorem 17: The stationary distribution of an open Jackson network where the nodes have state-dependent service rates is

$$p(n) = \prod_{i=1}^{K} p_i(n_i), \quad n_1, \ldots, n_K \geq 0$$

where for every node $i = 1,2,\ldots,K$

$$p_i(n_i) = \frac{1}{G_i} \frac{\lambda_i^{n_i}}{\mu_i(1)\cdots\mu_i(n_i)}, \quad n_i \geq 0$$

with normalization constant $G_i = \sum_{n_i=0}^{\infty} \frac{\lambda_i^{n_i}}{\mu_i(1)\cdots\mu_i(n_i)} < \infty$

(Proof follows identical steps with the proof of Theorem 13)

- Remark:
  - The stationary distribution has the product form; but if the network starts from some arbitrary initial state, the queues are not independent at any finite time
    - Similar to the example of two M/M/1 queues in tandem (as discussed at the end of Appendix D.3.1 of book R0)

## Jackson Networks

- Open Jackson Networks
- Network Flows
- State-Dependent Service Rates
- Networks of Transmission Lines & Kleinrock's Assumption
- Closed Jackson Networks

## Network of Transmission Lines

- Real Networks: Many transmission lines (queues) interact with each other
  - Output from one queue enters another queue,
  - Merging with other packet streams departing from the other queues
  - => 1) Interarrival times at various queues become strongly correlated with packet lengths; 2) Service times at various queues are not independent
    - Queueing models become analytically intractable
- Analytically Tractable Queueing Networks:
  - Independence of interarrival times and service times
  - Exponentially distributed service times
  - Network model: Jackson network
  - "Product-Form" stationary distribution

## Kleinrock Independence Assumption

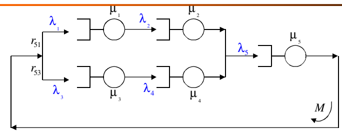
1. Interarrival times at various queues are independent
2. Service time of a given packet at the various queues are independent
   - Length of the packet is randomly selected each time it is transmitted over a network link
3. Service times and interarrival times: independent

- Assumption has been validated with experimental and simulation results – Steady-state distribution approximates the one described by Jackson's Theorems
- Good approximation when:
  - Poisson arrivals at entry points of the network
  - Packet transmission times "nearly" exponential
  - *Several packet streams merged on each link*
  - *Densely connected network*
  - *Moderate to heavy traffic load*

## Jackson Networks

- Open Jackson Networks
- Network Flows
- State-Dependent Service Rates
- Networks of Transmission Lines & Kleinrock's Assumption
- Closed Jackson Networks

## Closed Jackson Networks

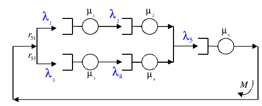



- Closed Network: $K$ nodes with exponential servers
  - No external arrivals ($\gamma_i$=0) , no departures ($r_{i0}$=0)
  - Fixed number $M$ of circulating customers
- Steady-state distribution is of "product-form" type (as shown later)

## Closed Jackson Network (contd.)

- Aggregate arrival rates

$$\lambda_i = \sum_{j=1}^{K} \lambda_j r_{ji}, \quad i = 1,\ldots,K$$



- The arrival rates are "relative" arrival rates – visit ratios between states
  - No unique solution, and can only be determined up to a multiplicative constant
  - Use an additional equation to obtain unique solution to the above system, e.g.
    - Set $\lambda_j$=1, for some node $j$
    - Set $\lambda_j$=$\mu_j$, for some node $j$
    - Set $\lambda_1$+ $\lambda_2$+…+ $\lambda_K$=1

## Closed Jackson Network (contd.)

- Let $x_i$ be the number of customers at station $i$, at steady state
  - Random variables $x_1$, $x_2$,…, $x_K$ are not independent – their sum must be equal to $M$
- The state $x$=($x_1$, $x_2$,…, $x_K$) of the closed network can take values $n$=($n_1$, $n_2$,…,$n_K$), with

$$n_i \ge 0 \;\text{ and }\; |n| = \sum_{i=1}^{K} n_i = M$$

  - Let $F(M)$ denote the set of all such states
- Jackson's theorem for closed networks gives the stationary distribution $p(n) = P\{x = n\} = P\{x_1 = n_1, \ldots, x_K = n_K\}$

## Jackson's Theorem for Closed Networks

- Theorem 1: The stationary distribution of a closed Jackson network is

$$p(n) = \frac{1}{G(M)} \prod_{i=1}^{K} \rho_i^{n_i}, \quad \text{for all } n \in \mathcal{F}(M) = \{n : n_i \ge 0, |n| = M\}$$

  where $\rho_i \equiv \lambda_i/\mu_i$ (note, this *is not* the actual utilization factor of station $i$), and the normalization constant $G(M)$ is a function of $M$
  - G(M) guarantees that $\{p(n)\}$ is a valid probability distribution

$$\sum_{n \in \mathcal{F}(M)} p(n) = 1 \Rightarrow G(M) = \sum_{n \in \mathcal{F}(M)} \prod_{i=1}^{K} \rho_i^{n_i}$$

- This stationary distribution is *said* to have a product-form
- However: at steady-state the queues *are not* independent
  - $\{p_i(n_i)\}$: marginal stationary distribution of queue i

$$p(n) \ne p_1(n_1) \cdots p_K(n_K)$$

## Jackson's Theorem for Closed Networks (proof)

- Theorem 2: The reversed chain of a stationary closed Jackson network is also a stationary closed Jackson network with the same service rates and routing probabilities: $r^*_{ij} = \lambda_j r_{ji} / \lambda_i$
- Proof of Theorems 1 & 2:
  - First, show that for the corresponding forward and reversed chains

$$p(m)q^*(m,n) = p(n)q(n,m), \quad n,m \in \mathcal{F}(M), n \neq m$$

Need to prove only for $m = T_{ij}n$

$$q(n,T_{ij}n) = \mu_i r_{ij} \mathbf{1}\{n_i > 0\}$$

$$q^*(T_{ij}n,n) = q^*(n - e_i + e_j, n) = \mu_j r^*_{ji} \mathbf{1}\{n_i + 1 > 0\} = \mu_j (\lambda_i r_{ij} / \lambda_j)$$

$$p(T_{ij}n)q^*(T_{ij}n,n) = p(n)q(n,T_{ij}n) \quad \Leftrightarrow \quad p(T_{ij}n)\mu_j(\lambda_i r_{ij}/\lambda_j) = p(n)\mu_i r_{ij} 1\{n_i > 0\}$$
$$\Leftrightarrow \quad p(T_{ij}n) = \rho_j \rho_i^{-1} p(n) 1\{n_i > 0\}$$

  - Verify, exactly as in the open-network case, that:

$$\sum_{m \neq n} q(n,m) = \sum_{m \neq n} q^*(n,m) = \sum_i \mu_i 1\{n_i > 0\}, \quad n \in \mathcal{F}(M)$$

## State-Dependent Service Rates

- Theorem: The stationary distribution of a closed Jackson network where the nodes have state-dependent service rates is

$$p(n) = \frac{1}{G(M)} \prod_{i=1}^{K} \frac{\lambda_i^{n_i}}{\mu_i(1)\cdots\mu_i(n_i)}, \quad \text{for all } n \in \mathcal{F}(K) = \{n : n_i \geq 0, |n| = K\}$$

where the normalization constant $G(M)$ is a function of $M$, the fixed number of customers in the network

  - Normalization constant:

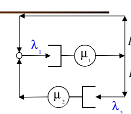
$$G(M) = \sum_{n \in \mathcal{F}(K)} \prod_{i=1}^{K} \frac{\lambda_i^{n_i}}{\mu_i(1)\cdots\mu_i(n_i)} \Rightarrow \sum_{n \in \mathcal{F}(K)} p(n) = 1$$

- Proof similar to the one for open networks

- Theorem D.27: a time reversal of a closed Jackson network is also a closed Jackson network.

## Example

Closed network model for CPU (rate $\mu_1$) and I/O (rate $\mu_2$) system. Upon service completion in 1, customer routed to 2 with probability $p_2 = 1 - p_1$, or back to 1 with probability $p_1$. $M$ =fixed number of customers




- $\lambda_1 = p_1\lambda_1 + \lambda_2, \lambda_2 = p_2\lambda_1$. Choose solution: $\lambda_1 = \mu_1$ and $\lambda_2 = p_2\mu_1$

$$\rho_1 = 1, \quad \rho_2 = \frac{p_2\mu_1}{\mu_2}$$

- Stationary distribution: $n$ customers in 2 and $M$-$n$ in 1

$$p(M-n,n) = \frac{1}{G(M)}\rho_1^{M-n}\rho_2^n = \frac{1}{G(M)}\rho_2^n, \quad n = 0,1,\ldots,M$$

  - Normalization constant

$$G(M) = \sum_{n=0}^{M} \rho_2^n = \frac{1-\rho_2^{M+1}}{1-\rho_2}$$

- Utilization factor and throughput of node 1:

$$U(M) = 1 - p(0,M) = 1 - \frac{\rho_2^M}{G(M)} = \frac{G(M-1)}{G(M)} \qquad \gamma_1(M) = U(M)\mu_1 = \mu_1 \frac{G(M-1)}{G(M)}$$

## Closed Networks: Normalization Constant

- Normalization constant as a function of $M$ and $K$:

$$G(M,K)=\sum_{n\in F(M)}\prod_{i=1}^{K}\rho_i^{n_i}=\sum_{\substack{n_1+\cdots+n_K=M\\ n_i\geq 0}}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_K^{n_K}$$

  - All performance measures of interest – throughput, average delay – can be obtained in terms of $G(M,K)$

- Computational complexity is exponential in $M$ and $K$:

$$\binom{M+K-1}{K-1}=\binom{M+K-1}{M}\ \text{terms in the summation}$$

- Recursive methods can be used to reduce complexity
- Iterative algorithm [due to Buzen]
  - Normalization constant will be treated as a function of both M and K and denoted $G(M,K)$ only in the context of the iterative algorithm

## Iterative Computation of $G(M)$

- For any $m$ and $k$ ($m$=0,…, $M$; $k$=1,…, $K$) define:

$$G(m,k)=\sum_{n\in F(m)}\prod_{i=1}^{k}\rho_i^{n_i}=\sum_{n_1+\cdots+n_k=m}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k}$$

- For a closed network of single-server queues $G(M,K)$ can be computed iteratively using the following recursive relation:

$$G(m,k)=G(m,k-1)+\rho_k G(m-1,k)$$

  with boundary conditions:

$$G(m,1)=\rho_1^m,\quad m=0,1,\dots,M$$
$$G(0,k)=1,\qquad k=1,2,\dots,K$$

## Iterative Algorithm (proof)

For $m>0$ and $k>1$ we split the sum into two sums over disjoint sets of states, corresponding to $n_k=0$, and $n_k>0$.

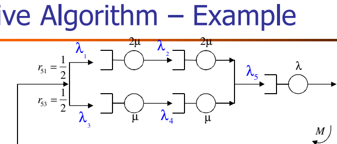
$$\begin{aligned}G(m,k)&=\sum_{n_1+\cdots+n_k=M}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k}\\ &=\sum_{\substack{n_1+\cdots+n_k=m\\ n_k=0}}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k}+\sum_{\substack{n_1+\cdots+n_k=m\\ n_k>0}}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k}\\ &=\sum_{n_1+\cdots+n_{k-1}=m}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_{k-1}^{n_{k-1}}+\sum_{\substack{n_1+\cdots+n_k=m\\ n_k>0}}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k}\end{aligned}$$

Note that the first sum is $G(m,k-1)$. For the second sum, observing that $n_k>0$, we define $n_k=n_k'+1$, where $n_k'\geq 0$. Then:

$$\begin{aligned}\sum_{\substack{n_1+\cdots+n_k=m\\ n_k>0}}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k}&=\sum_{\substack{n_1+\cdots+n_k'+1=m\\ n_k'\geq 0}}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k'+1}\\ &=\rho_k\sum_{\substack{n_1+\cdots+n_k'=m-1\\ n_k'\geq 0}}\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k'}=\rho_k G(m-1,k)\end{aligned}$$

Therefore: $G(m,k)=G(m,k-1)+\rho_k G(m-1,k)$

## Iterative Algorithm – Example




- $\lambda_1=\lambda_2,\ \lambda_3=\lambda_4$
  $r_{51}=r_{53}\Rightarrow\lambda_1=\lambda_3$
  $\lambda_5=\lambda_1+\lambda_3=2\lambda_1$

- $\rho_1=\rho_2=\lambda_1/2\mu,\ \ \rho_3=\rho_4=\lambda_3/\mu=2\rho_1$
  $\rho_5=\lambda_5/\lambda=2\lambda_1/\lambda=4\rho_1/\rho,\ \ \text{with}\ \rho\equiv\lambda/\mu$

- Visit ratios $\lambda_i$ determined up to a multiplicative constant
- Letting $\lambda_1=2\mu$, we have: $\rho_1=\rho_2=1,\ \rho_3=\rho_4=2,\ \rho_5=4/\rho$
- Calculation of $G(M,5)$ based on the iterative algorithm using these values

## Iterative Algorithm – Example

| $m$ \ $k$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 2 | 4 | 6 | 6+4/ρ |
| 2 | 1 | 3 | 11 | 23 | 23+(6+4/ρ)(4/ρ) |
| 3 | 1 | 4 | 26 | 72 | 72+[23+(6+4/ρ)(4/ρ)](4/ρ) |

- $\rho_1 = \rho_2 = 1,\ \rho_3 = \rho_4 = 2,\ \rho_5 = 4/\rho$
- Boundary conditions:

$$G(m,1) = \rho_1^m, \quad m = 0,1,\ldots,M$$
$$G(0,k) = 1, \quad k = 1,2,\ldots,K$$

- Iteration:

$$G(m,k) = G(m,k-1) + \rho_k G(m-1,k)$$

- Example:

$$G(1,2) = G(1,1) + \rho_2 G(0,2) = 1+1 = 2$$
$$G(1,3) = G(1,2) + \rho_3 G(0,3) = 2+2 = 4$$
$$G(2,2) = G(2,1) + \rho_2 G(1,2) = 1+2 = 3$$

## Summary

- Markov chains and some renewal theory
  - Markov chain
  - Renewal processes, renewal reward processes, Markov renewal processes
  - The excess distribution
  - Phase type distribution
  - PASTA
  - Level crossing analysis
- Some important queueing models
- Reversibility of Markov chains and Jackson Network