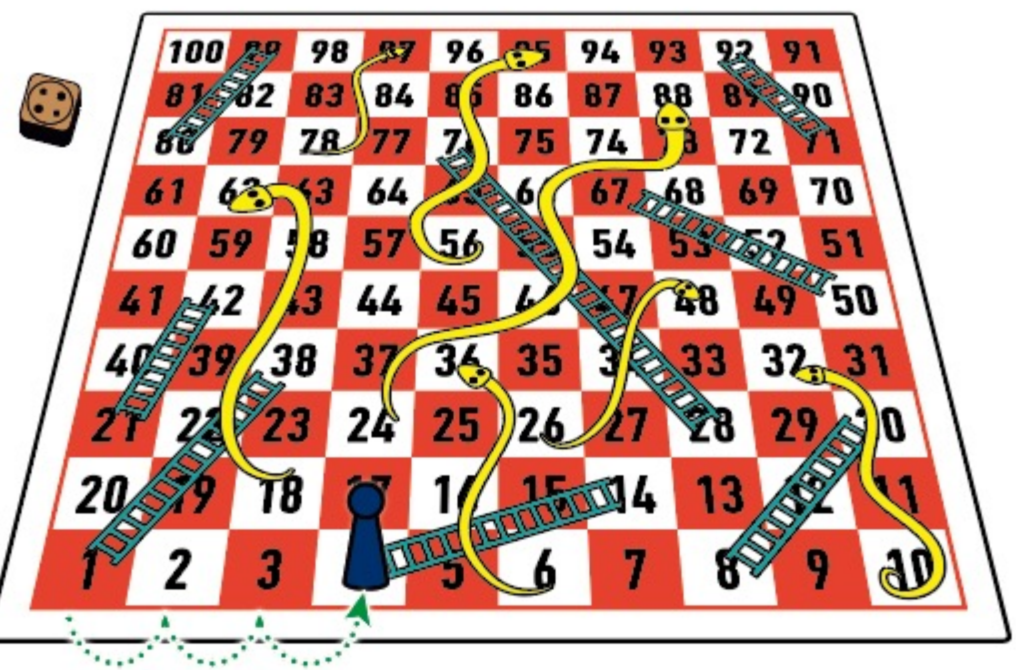
SHARIF
UNIVERSITY OF
TECHNOLOGY

Computer Simulation

Dr. Bardia Safaei

Chapter Four: Markov Chains

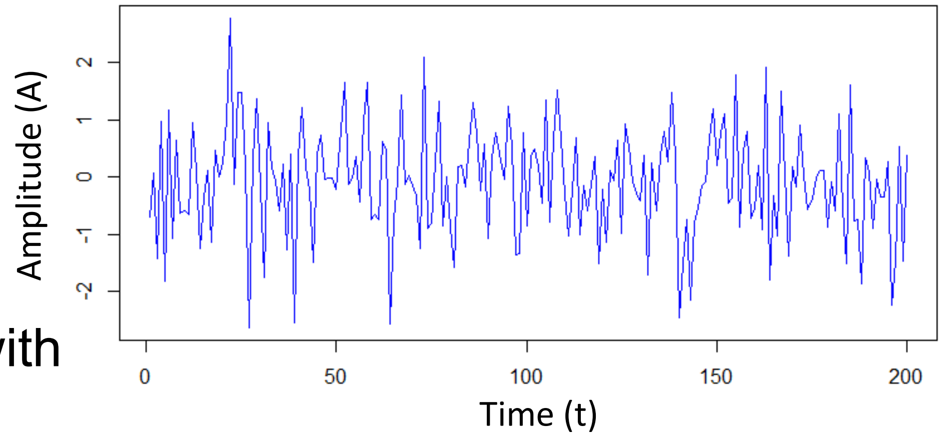


Outline

- Stochastic process
- Discrete Markov Chains (DMC)
- Continuous Markov Chains (CMC)

Stochastic Process (1)

- Consider a group of random variables X(t)
  - They are dependent on a real parameter (time)
    - They are indexed by t, indicating that they can change through the time
  - Defined on the same sample space S
- Example:
  - The **white noise** could be considered as a stochastic process
  - Noise does not have a deterministic pattern
    - Not for amplitude
    - Not for frequency
    - Not for phase
  - It is completely **random** with **unpredictable** behavior



Stochastic Process (2)

SHARIF UNIVERSITY OF TECHNOLOGY

- Why the noise wave is a stochastic process?
  - Now let's consider if I ask you to sketch your own white noise wave pattern
    - Everyone of you would probably draw a graph, which:
      - Is not 100% a noise, or not
        - It **may/could** be a white noise
      - Is not 100% identical to your friend's
    - In every graph, strength of the signal X could only get values in the amplitude space
    - Strength can change as time passes X(t)
  - Therefore, $\{X(t)|t \in T\}$ would be a stochastic process
- In the definition of stochastic process, T represents the time interval, where our process is under observation
  - Based on T, we have **discrete** and **continuous-**time random processes

Stochastic Process (3)

- Recall: The sample space S is the set of values acceptable for our random variable
- Since X(t) could range both in $t \in T$, and S, the variables in stochastic process could be represented as:

$$\{X(t,s)|s \in S, t \in T\}$$

  - Therefore, the family of variables could be denoted as a family of functions
- For a fixed $t = t_1, X(t_1, s) = X_{t_1}(s)$ is a **random variable** (denoted by X (t)) as s varies over the sample space S
- For a fixed sample points, $s_1 \in S$, the expression $X(t, s_1)$ $= X_{s_1}(t)$ is a single function of time t, called a **sample function** or a **realization** of the process

Stochastic Process (4)

- Based on T, and S, stochastic processes have four different variations:
  - Discrete-state/Discrete-time process
    - Example: A Bernoulli process for tossing a coin → $S \in \{0,1\}$, and T=experiment iteration number (as time)
  - Discrete-state/Continuous-time process
    - Example: Used in specifying connections or disconnections in telecommunications during a time interval
  - Continuous-state/Discrete-time process
    - Example: Measuring UV radiation every hour
  - Continuous-state/Continuous-time process
    - Example: Temperature alterations during a day

Stochastic Process (5)

- If the sample space of a stochastic process is discrete, then it is called a **discrete-state process**
  - Often referred to as a **chain**
  - Alternatively, if the state space is continuous, then we have a **continuous-state** process
- If the index set T is discrete, then we have a **discrete-time process**
  - It is denoted with $\{X_n | n \in T\}$, where observations are managed in discrete time steps
  - If the index set T is continuous, we have a **continuous time process**

Stochastic Process (6)

- For a fixed time $t = t_1$, the term $X(t_1)$ is a simple random variable that describes the state of the process at time t_1
  - In other words, $X(t_1)$ specifies the state of the systems @ t_1
- For a known fixed number x_1, the probability that $X(t_1) \leq x_1$ is given by the CDF of the random variable $X(t_1)$, denoted by:

$$F(x_1; t_1) = F_{X(t_1)}(x_1) = P[X(t_1) \leq x_1]$$

  - This gives us the probability @ t_1
- $F(x_1; t_1)$ is known as the first-order distribution of the process $\{X(t) | t \geq 0\}$

Stochastic Process (7)

- Given two time instants t_1, and t_2, $X(t_1)$, and $X(t_2)$ are two random variables on the same probability space
  - Recall: just consider the noise graph
    - $X(t_1)$ indicates the voltage of the noise at t_1, and $X(t_2)$ indicates the voltage at t_2
    - Both values are assigned from the amplitude sample space
- Their joint distribution is known as the **second-order distribution** of the process and is given by:

$$F_X(x_1, x_2; t_1, t_2) = P[X(t_1) \leq x_1, X(t_2) \leq x_2]$$

- The first order talks about 1 variable, while the second order takes 2 variables into account

Stochastic Process (8)

- In general, we define the nth-order joint distribution of the stochastic process $X(t), t \in T$:

$$F_X(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N) = P[X(t_1) \le x_1, X(t_2) \le x_2, \dots, X(t_N) \le x_N]$$

- The mentioned distributions are indicating the CDF
  - In order to obtain PDF, we use the derivation:

$$f_X(x_1; t_1) = \frac{\sigma}{\sigma_{x_1}} F_X(x_1; t_1)$$

1st order cumulation

$$f_X(x_1, x_2; t_1, t_2) = \frac{\sigma^2}{\sigma_{x_1}\sigma_{x_2}} F_X(x_1, x_2; t_1, t_2)$$

. . .

N^{th} order cumulation

$$f_X(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N) = \frac{\sigma^N}{\sigma_{x_1}\sigma_{x_2} \dots \sigma_{x_N}} F_X(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N)$$

Stationary Stochastic Process (1)

- A stochastic process $\{X(t)|t \in T\}$ is said to be **stationary** in the **strict sense**, if for $n \geq 1$, its nth-order joint CDF satisfies the following condition:

$$f_X(x_1; t_1) = f_X(x_1; t_1 + \tau)$$

  - Note: Do not confuse the concept of stationary stochastic process with the stationary concept in Poisson process

- Based on the above definition, if X(t) is a 1st order stochastic process, it could be proved that the CDF has also a similar equation:

$$F_X(x_1; t_1) = F_X(x_1; t_1 + \tau)$$

Stationary Stochastic Process (2)

- A stochastic process is called to be **Nth-order stationary** if the following is valid for $x \in \mathbb{R}^n$, t $\in T^n$:

$$F_X(x_1, \dots, x_N; t_1, \dots, t_N) = F_X(x_1, \dots, x_N; t_1 + \tau, \dots, t_N + \tau)$$

- What does it say?
  - Consider that we have sampled n instances in $t_1,\ t_1,\ \dots,\ t_N$
    - Probability of $X(t_1)\ \leq x_1$, **and** $X(t_2)\ \leq x_2$, **and** … $X(t_N)\ \leq x_N$ is P
    - If t is shifted with τ (+, or -), and the probability remains the same
- If this equation is valid for all $N \geq 1$, the stochastic process is said to be **strictly stationary**

Independent Processes

SHARIF UNIVERSITY OF TECHNOLOGY

- A stochastic process $\{X(t)|t \in T\}$ is said to be an **independent process** if its nth-order joint distribution satisfies the following condition:

$$F_X(x_1, \dots, x_N; t_1, \dots, t_N) = \prod_{i=1}^{n} F_X(x_i; t_i) = \prod_{i=1}^{n} P[X(t_i) \le x_i]$$

  - This indicates that the value of an independent nth-order joint $F_X(x; t)$ is equal to the multiplication of all the inequalities of each random variable, member of the process

- Note: do not confuse this with **independent processes**
  - Two stochastic processes are called to be independent if

$$F_{XY}(x_1, \dots, x_N, y_1, \dots, y_N; t_1, \dots, t_N, t'_1, \dots, t'_N) = F_X(x_1, \dots, x_N; t_1, \dots, t_N) \times F_Y(y_1, \dots, y_N; t'_1, \dots, t'_N)$$

Renewal Process

- A renewal process is defined as a **discrete-time independent process** denoted with $\{X_n | n = 1,2, \dots\}$, if $X_1, X_2, \dots$ are independent, and identically distributed (i.i.d), nonnegative random variables
- What does *i.i.d* mean?
  - Consider X_t $(t \in T)$ as an i.i.d random variable
  - If in two instances of time i, and j, $(i, j \in T)$ the value of X_i does not have any effect on the value X_j
  - And, probability of X_i is the same as probability of having X_j

Markov Dependency (1)

- Though the assumption of an independent process considerably simplifies analysis, such an assumption is often unwarranted
  - So, we are forced to consider some sort of dependency among these random variables
- The **simplest** and the **most important** type of dependency is the first-order dependency or Markov dependency
  - A stochastic process $\{X(t)|t \in T\}$ is called a Markov process, if for any instance of time $t_0 < t_1 < \cdots < t_{n-1} < t_n$, the probability distribution of random variable $X(t_n)$ only depends on $X(t_{n-1})$:

$$P[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}, X\,(t_{n-2}) = x_{n-2}, \dots, X(t_0) = x_0] \\ = \; P[X(t_n) = x_n \,| X(t_{n-1}) = x_{n-1}]$$

    - Probability of having $X(t_n)$ in t_n, only depends on **one instance earlier**, not the sequence of events happened before (during t_0 to t_{n-2})

Markov Dependency (2)

- Consider a system that we have determined its different states of operation
  - Assume every state as an event, which our random variable $X(t)$ could be assigned to, during time $\in [0, t]$
  - If the probability of being in state $X(t_n)$ at t (occurrence of $X(t_n)$) is only dependent on being in state $X(t_{n-1})$ (occurrence of $X(t_{n-1})$) at t_{n-1}
    - The dependency between the states (or our random variables) is called to be Markov
    - This dependency is first-order
- $X(t_n)$ could be considered as our **next state**, and $X(t_{n-1})$ could be considered as the **present state**
  - An important concept in reinforcement learning

Markov Dependency (3)

- The stochastic processes with a discrete sample space are called **chain**
  - Since Markov processes are categorized in such processes, they are called **Markov chains**
- Based on what we discussed earlier, in Markov chains we have:

$$P[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}, X\,(t_{n-2}) = x_{n-2}, \dots, X(t_0) = x_0]$$
$$= P[\underbrace{\boxed{X(t_n) = x_n}}_{\text{A}} \,|\, \underbrace{\boxed{X(t_{n-1}) = x_{n-1}}}_{\text{B}}]$$

- Assume that A indicates that in t_n, the system resides in state j, and B indicates that system resides in state i at t_{n-1}
  - Therefore, the equation could be represented as $P[X_n = j | X_{n-1} = \mathrm{i}]$

Homogeneous Markov Chains

- A Markov chain is said to be (time-) homogeneous if:

$$P[X_n = j|X_{n-1} = i] = P[X_1 = j|X_0 = i]$$

- For a homogeneous Markov chain, the **past history** of the process is completely summarized in the **current state**
  - Therefore, the distribution for the time Y, the process spends in a given state must be memory less
  - What does this really mean?
    - The probability of making a transition from one state to another does not depend on time
      - If at t_n, the transition probability from $X_{n-1} = i$ to $X_n = j$ is said to be $P_{i,j}(t_n)$, all transitions from i to j at any instances of time is also $P_{i,j}(t_n)$
        - Remove the time index → $P_{i,j}$ suffices for showing the transition probability

Discrete Time Markov Chain (DTMC) (1)

- In DTMC, we decide to observe the states of a system at a **discrete set of time-steps**
- Consider our successive observations from the system states are defined as random variables $X_1, X_2, \dots, X_n$ at time steps $0, 1, 2, \dots, n$, respectively
  - Recall: If $X_n = j$, then the state of the system at time-step n is s_j
  - s_0 is the initial state of the system, where the it has started operating
- A Markov property for DTMC can then be stated as:

$$P[X_n = s_n | X_{n-1} = s_{n-1}, X_{n-2} = s_{n-2}, \dots, X_0 = s_0] \\ = P[X_n = s_n | X_{n-1} = s_{n-1}]$$

  - The state of the system in the future (s_n) only depends on the state of the system right now (s_{n-1})

Discrete Time Markov Chain (DTMC) (2)

- The probability of being in state s_j at time-step n is denoted with $P\left(X_n = s_j\right) = P_j(n)$
  - This is the PMF of state s_j
- Since we are interested in homogeneous Markov chains, the time has no impact on the probability of transitions
  - The time-step index will be removed
  - The probability of having a transition from s_i to s_j with **n-step difference** is indicated with:

$$P_{i,j}(n) = P(X_{m+n} = s_j | X_m = s_i)$$

n-step transition probability

    - Where m is indicating an arbitrary time-step index
- One-step transition is of high importance, which is denoted with $P\left(X_n = S_j \middle| X_{n-1} = S_i\right) = P_{i,j}(1) = P_{i,j}$

Discrete Time Markov Chain (DTMC) (3)

- When the system starts operating, every state has a PMF, which indicates the probability of beginning from that state
  - These probabilities are indicated with a random variable X_0, often called the **initial probability vector**, and is specified as:

$$p(0) = [p_0(0), p_0(1), p_0(2), \dots, p_0(n)]$$

    - $P_0(i)$ denotes the probability of starting from s_i
    - n indicates the index of states in the systems model
  - Example: A system with 2 states s_0, and s_1: $p(0) = [p_0(0) = p, p_0(1) = 1 - p]$
- The one-step transition probabilities are compactly specified in the form of a **transition probability matrix**
  - It specifies the probability of having a transition from one state to another in only **one step move**

Transition Probability Matrix

SHARIF UNIVERSITY OF TECHNOLOGY

- The matrix for n-state system is as follows:

Example:

A 5-state Markov chain has 25-elemnt transition matrix

$$P = [P_{i,j}] = \begin{bmatrix} P_{0,0} & \cdots & P_{0,n} \\ P_{1,0} & & P_{1,n} \\ \vdots & \ddots & \vdots \\ P_{n,0} & \cdots & P_{n,n} \end{bmatrix}$$

- Recall: Since we are talking about chains (stochastic process with discrete-time and discrete-sample space), we are allowed to use **number of states**
- The entries of the matrix P satisfy the following two properties:
  - $0 \leq P_{i,j} \leq 1,\ i,j \in I$ (I indicates the size of the sample space)
  - $\sum_{j \in I} P_{i,j} = 1\ , i \in I$

State Transition Diagram

- An equivalent description of the one-step transition probabilities can be given by a **directed graph** called the state transition diagram of the Markov chain
  - State diagram for short
  - It depicts the states of the system and probabilities of transitions between them with a directed edge (branch)
- A node labeled *i* of the state diagram represents state *i* of the Markov chain
  - State 0 typically represents the initial state
- The branches labeled $P_{i,j}$, from node i to j implies that the conditional probability is: $P[X_n = j | X_{n-1} = i] = P_{i,j}$
  - Not forget that it is directional: $P_{i,j}$ is not necessarily $P_{j,i}$

Example (1)

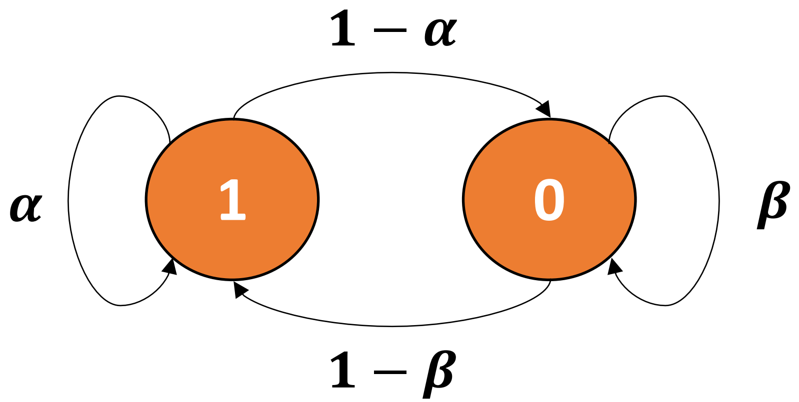
- A person could be in either 2 states: 1) Healthy, 2) Sick
- Consider a random variable X(n), where $n \in \{0,1,2,\ldots\}$
  - N indicates the day of observation
  - According to our system, X(n) could get two values, 0 or 1
    - States of the system

$$X(n) = \begin{cases} 1, if\ healthy \\ 0, \quad\quad if\ sick \end{cases}$$

  - Assume if the person is healthy, he stays healthy in the next step time-step with α: $P\{X_n = 1 | X_{n-1} = 1\} = P_{1,1} = \alpha$
  - Assume if the person is sick, he stays sick in the next step time-step with β: $P\{X_n = 0 | X_{n-1} = 0\} = P_{0,0} = \beta$

Example (2)

- State diagram of the system is illustrated as:



$$P = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix}$$

- What could we get from this figure?
  - If we reside in 1, with probability α we stay in state 1
  - If we reside in 1, with probability $1-\alpha$, we move to 0
  - If we reside in 0, with probabilityβ we stay in state 0
  - If we reside in 0, with probability $1-\beta$, we move to 1
- How did we obtain other values on the branches?

n-step Transition (1)

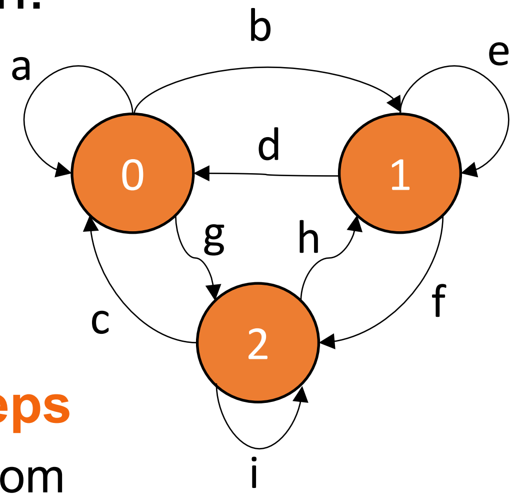
- In the previous example, we determined the transition probability matrix in one-step movement
- We are interested in obtaining an expression for evaluating the n-step transition probability based on the one-step probabilities
  - In other words, if system is residing in state i, with what probability, the system will reside in state j after n steps?
- Let P(n) be the matrix of n-step transition probabilities, whose (i, j) entry is $P_{i,j}(n)$, then we can write:

$$P(n) = P \times P(n-1) = P \times P \times P(n-2) = P^n$$

  - Why we could use this equation?

n-step Transition (2)

- Recall: For 2 independent events, their joint probability could be obtained by their individual multiplication
  - If there are more than one path towards a specific event, the multiplications of probabilities must be added for different paths
- Let's assume the following state diagram:
  - First, indicate the 1-step transition matrix



- Now, what is the probability of having a transition from state 0 to itself with 2 steps (based on the diagram)?
  - Specify all paths from s_0 to s_0 with only **2 steps**
    - Path refers to a sequence of states starting from s_i, ending at s_j
  - Calculate the probability in every path and add them together

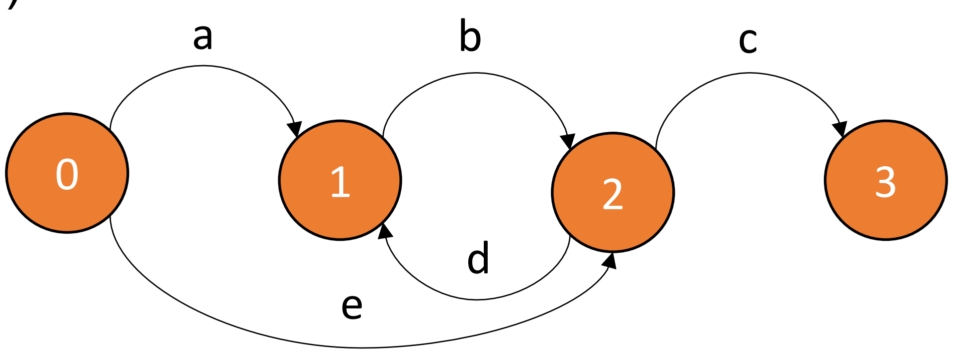
n-step Transition (3)

SHARIF UNIVERSITY OF TECHNOLOGY

- With utilizing the n-step transition matrix, and the initial probability vector, we can obtain the PMF for every state of the system after n steps:

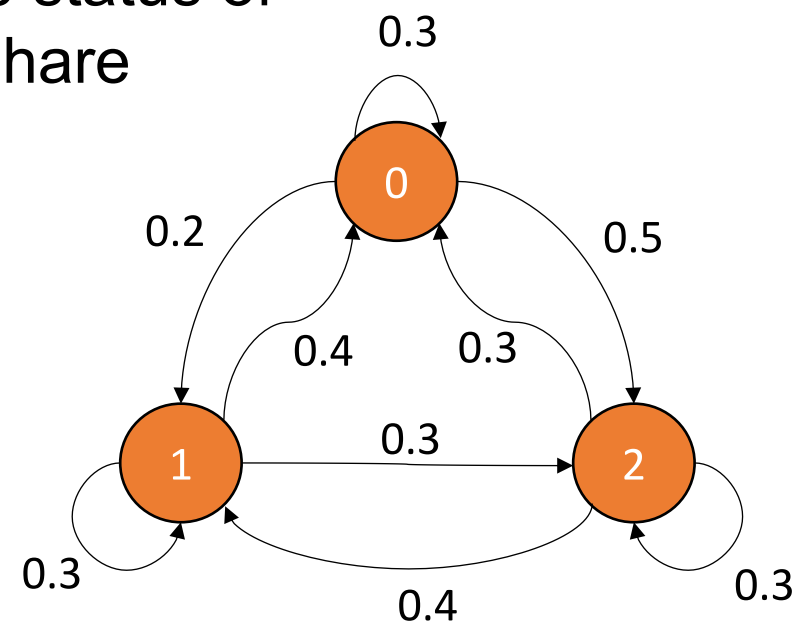
$$p(n) = p(0)P(n) = p(0)P^n$$

  - Elements of matrix p(n) gives the probability of being in every state of the system after n steps
  - Why we should multiply p(0)?
    - Because, the probability will be altered when the system starts from a different state
      - Calculate $p_3(2)$ starting from s_0, or s_1



Stock Exchange Example (1)

- Following diagram indicates the status of the market price for a specific share
  - State 0 → price rising
  - State 1 → price falling
  - State 2 → ranging
- What does it say?
- Should we always have a full graph?
  - No, we may have diagrams with few edges departing from a state to only a number of other states in the system
- Note: Sum of outsiding edges from every state equals to 1



Stock Exchange Example (2)

SHARIF UNIVERSITY OF TECHNOLOGY

- Determining the transition matrix
  - Recall: Indicate the title of the rows and columns for making it more convenient task to derive the matrix

$$
P = \begin{array}{c} \\ \\ \\ \\ \end{array}
\begin{array}{c}
\begin{array}{ccc} 0 & 1 & 2 \end{array} \\
\begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}
\end{array}
\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}
$$

- In this example, we assume that the system starts from **state 0** (rising)
  - The initial probability vector would be $p(0) = [1\ 0\ 0]$
  - Worthy to mention that we could have started from any other states, either falling or ranging
    - Only one state!

Stock Exchange Example (3)

- To calculate the n-step probability values:
  - After 1 step:

$$p(1) = p(0) \times P^1 = [1\ 0\ 0] \times \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} = [0.3\ 0.2\ 0.5]$$

  - If we start from the rising state:
    - With 30% chance, the price of the share will continue to rise after one unit of observation (day, month, …)
    - With 20% chance, the price of the share would tend to fall
    - With 50% chance, if we start from the rising state, the price of the share would tend to maintain its value
  - After 2 steps:

$$p(2) = p(0) \times P^2 = p(0) \times P^1 \times P^1 = p(1) \times P^1 =$$

$$[0.3\ 0.2\ 0.5] \times \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} = [0.32\ 0.32\ 0.36]$$

Stock Exchange Example (4)

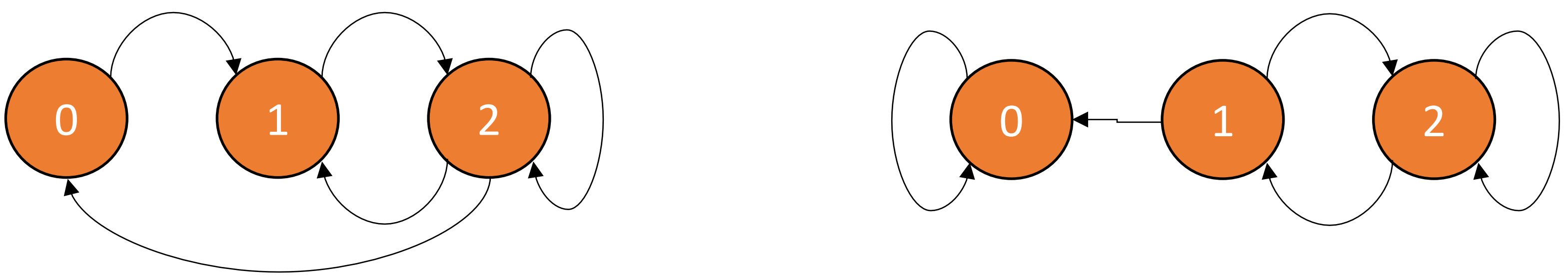
- This could be iterated for achieving any n-step probability vector:

$$p(3) = [0.332 \quad 0.304 \quad 0.364]$$
$$p(4) = [0.3304 \quad 0.303 \quad 0.3664]$$
$$\vdots$$
$$p(10) = [0.330357 \quad 0.303571 \quad 0.366072]$$

- Assume that we want to find the probabilities after $n \to \infty$ steps, what would be the probability vector for our states?
  - Calculate it for $p(0) = [0\ 1\ 0]$, and $p(0) = [0\ 0\ 1]$
  - Compare and analyze the outcomes for these three vectors

Transient and Recurrent States

- A state *i* is said to be **transient**, if there is a positive probability that the process will not return to this state if it is left by the process
  - The process cannot return even if it wants to
- A state *i* is said to be **recurrent** if starting from *i*, the process can eventually return to it with probability one
  - This return can happen in 1, or 2, or 3, …, n steps
    - The important thing is it could be returned independent from the time



a b

Period of States

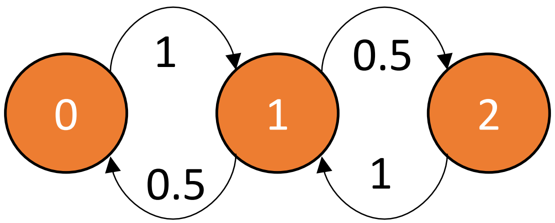
- For a recurrent state *i*, $p_{ii}(n) > 0, n \geq 1$, the period is denoted by d_{ii}, and is defined as the greatest common divisor (gcd) of the **set of positive integers n** such that $p_{ii}(n) > 0$
  - We may be able to return to i after different time-step (n) values
  - This depends on the complexity of the state diagram and the paths starting from state i, and ending to itself
- Formally, the period of a state is defined as:

$$d_{ii} = \gcd\{n|P(X_{m+n} = i|X_m = i) > 0\}$$

- If state *i* has period d_{ii}, any return to state i could occur in integer multiplications of d_{ii} time-steps

Aperiodic and Periodic States

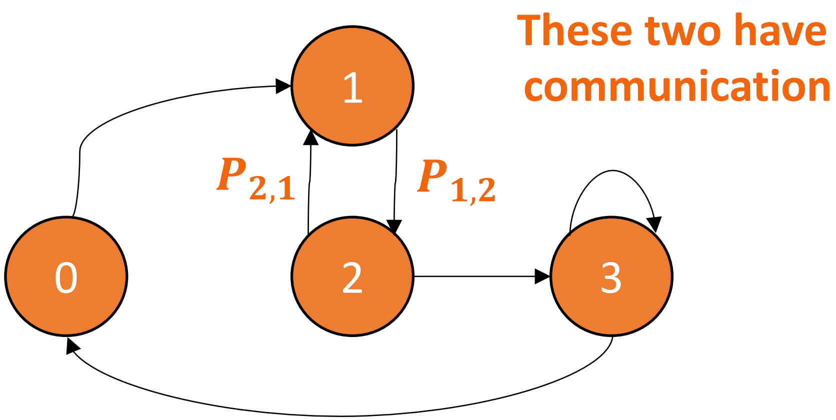
- A recurrent state i is said to be aperiodic if its period is $d_{ii} = 1$, and its called periodic if $d_{ii} > 1$
  - In the following state diagram, state 1 is a periodic state with $d_{ii} = 2$
    - Because, we can return to state 1 with 2 steps, or 2K steps



- How could we alter this Markov chain to make state 2 an aperiodic state?
  - Adding an edge from state 2 to itself
  - Update the transition probability values (to make sure $\sum = 1$)
- A state i is said to be an absorbing state if $P_{i,i} = 1$

More Definitions

- Two states *i* and *j* are in **communication** if **directed paths** from i to j and **vice-versa** exist in the state diagram
  - In other words, in the transition matrix, both $P_{i,j}$, and $P_{j,i}$ have values greater than 0
- A Markov chain is said to be **irreducible** if every recurrent state can be reached from every other state in a **finite number of steps**



  - In other words, for all $i, j \in I$, there is an integer $n \geq 1$ such that $P_{i,j}\,(n) > 0$
  - Simplest form of the diagram
  - In these chains, all of the states are recurrent

Continuous-Time Markov Chains (CTMC)

- Similar to DTMCs, in CTMC we confine our attention to **discrete-state** processes
  - This implies that, although the parameter t has a continuous range of values, the set of our random variables $X(t)$ (states) is discrete
- Recall: As we stated in the class, the main condition for a stochastic process to be Markov is the memoryless property
  - Probability of being in a future state only depends on the probability of being in the present state (not the history of other states)
  - Hence, in order to know where we are headed in the next move, we should know where we reside right now

$$P[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \ldots, X(t_0) = x_0] = P[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}]$$

CTMC Behavior Analysis (1)

- The behavior of the process is characterized by two elements:
  - The initial state probability given by the PMF of $X(t_0), P(X(t_0) = s_k), k \in \{0, 1, 2, \dots, n\}$
  - The transition probabilities: $P_{i,j}(v, t) = P(X(t) = s_j | X(v) = s_i)$
- For analyzing the CTMC, we have two options:
  - Transient (time-dependent) analysis
  - Steady-state analysis
- In Markov chains, in any instance of time, system resides in a specific state
  - Unlike the steady-state, in transient analysis, we freeze the system at time t, and take a snapshot
    - So, the system operation will be halted in a specific state

CTMC Behavior Analysis (2)

- Let denote the PMF of X(t) (or the state probabilities at time t) by:

$$\pi_j(t) = P\big(X(t) = s_j\big), j \in \{0,1,2, \dots, n\}, t \geq 0$$

  - This indicates the probability of being in s_j at time t

- Based on the principles, if we sum up all of the probabilities at time t, we have:

$$\sum_{j\in\{0,1,2,\dots,n\}} \pi_j = 1$$

✓ This is also applicable in DTMCs

✓ Independent from having time or time-steps, adding the probability values in any instance gives 1

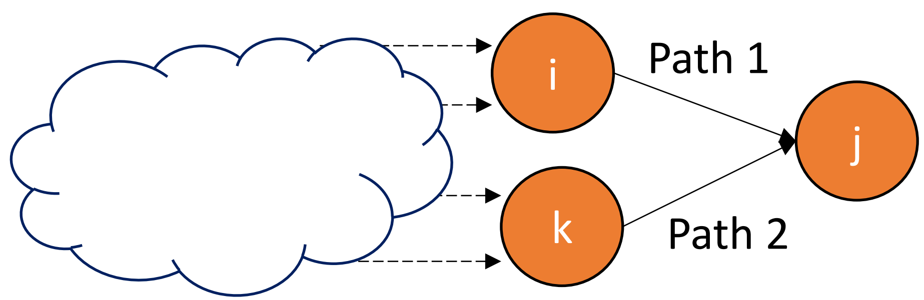
- What are we seeking in transient analysis?
  - Finding the probability values in a specific state at **time t**

CTMC Transient Behavior Analysis (1)

- A specific state could be reached via **different paths** in the state diagram
  - So, based on the theorem of total probability, for a given $t > v$, we can express the PMF of $X(t)$ in terms of the transition probabilities $P_{i,j}(v,t)$, and the PMF of $X(v)$

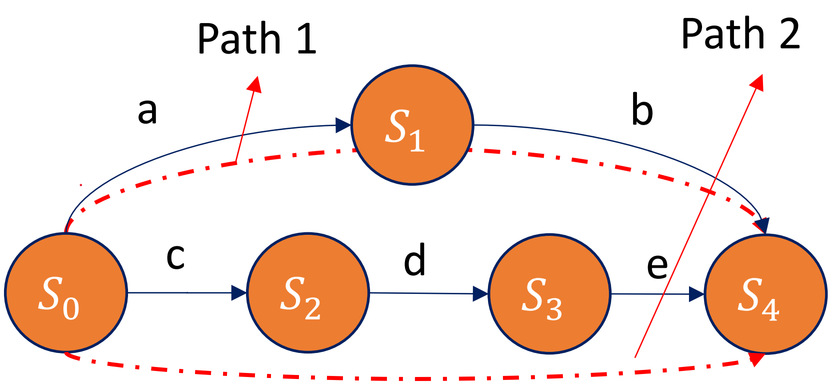
$$\pi_j(t) = P\left(X(t) = s_j\right) = ?$$

$$\pi_j(t) = \boxed{\sum_{i \in \{0,1,2,\ldots,n\}}} \boxed{P(X(t) = s_j | X(v) = s_i)} \times \boxed{P(X(v) = s_i)}$$



CTMC Transient Behavior Analysis (2)

- Now assume we **start** observing the system at $t = 0$, and move towards state j at time t
  - Then we have: $\pi_j(t) = \sum_{i\in\{0,1,2,\ldots,n\}} P_{i,j}(0,t)\pi_i(0)$
  - Where $P_{i,j}$ indicates the transition probabilities, and $\pi_i(0)$ denotes the probabilities at $t = 0$
- Example: To better understand the difference between $v = 0$, and $v \neq 0$ in our analysis, let's calculate $\pi_4(t)$
  - Since we are focusing on **1-step** transition
    - If $v = 0$
      - Only $\pi_i(0)$s, which are directly connected to s_4 are important
    - If $v \neq 0$
      - Roll back hop-by-hop until we reach the first state



CTMC Transient Behavior Analysis (3)

SHARIF UNIVERSITY OF TECHNOLOGY

- If we denote the probability of states in time t as a vector $\pi(t) = [\pi_0(t), \pi_1(t), \dots, \pi_n(t)]$, there exists a matrix, which establishes the following equation:

$$\frac{d\pi(t)}{dt} = \pi'(t) = \pi(t)Q$$

  - Where Q is the **Infinitesimal Generator matrix** composed of $q_{i,j}$
    - $q_{i,j}$ indicates the transition rate from state i to state j, where $i \neq j$, and $q_{i,j} \geq 0$
  - The diameter of Q is defined as follows:

$$q_{i,i} = -\sum_{j,j\neq i} q_{i,j} \;\; \rightarrow -\infty < q_{i,i} \leq 0$$

Sum of the transition rates departing state i must be equal to zero

Steady-State Analysis

- For a given CTMC, the steady-state probabilities are independent of time
  - Therefore, we could say $\lim_{t \to +\infty} \frac{d\pi(t)}{dt} = 0$
  - Based on what we discussed in the previous slide ($\pi'(t) = \pi(t)Q$), and the above equation, we have:

$$\sum_{i \in S} q_{i,j}\pi_i = 0, \qquad \forall j \in S$$

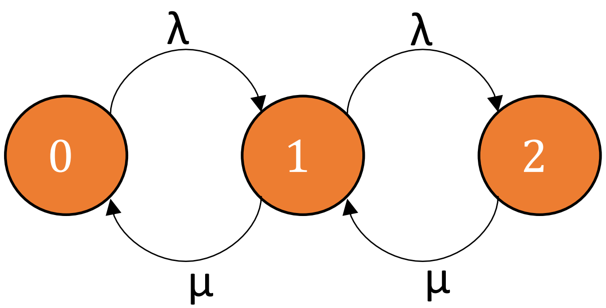
- Finally, in order to determine the steady-state unconditional probabilities, we can use the following matrix form:

$$\pi Q = 0$$

  - This gives us n-1 independent linear equations (n is the # states)
    - One more is required → Sum of probabilities is 1

Steady-State Analysis Example (1)

- Consider the following CTMC
  - We want to calculate the steady-state probabilities
- Some explanations about λ, and μ notations in Markov diagrams
  - According to the intended system, we may have **failure** or **repair** in the device



  - In reliability analysis, it is important to know the ability of having a transition to a repairable or unrepairable state
  - While failures are denoted with rate λ, repairs are denoted with μ
    - Multiplications of λ, and μ are also possible according to the structure of the system

$$\pi Q = 0 \rightarrow [\pi_1 \quad \pi_2 \quad \pi_3]\begin{bmatrix} -\lambda & \lambda & 0 \\ \mu & -(\mu+\lambda) & \lambda \\ 0 & \mu & -\mu \end{bmatrix} = [0\ 0\ 0] \qquad (1)$$

Steady-State Analysis Example (2)

- In any instance of time (including in steady-state $t \to \infty$), sum of probabilities is 1:

$$\pi_1 + \pi_2 + \pi_3 = 1 \quad \textbf{(2)}$$

  - This is the other equation needed
- Based on 1, and 2, it could be concluded that:

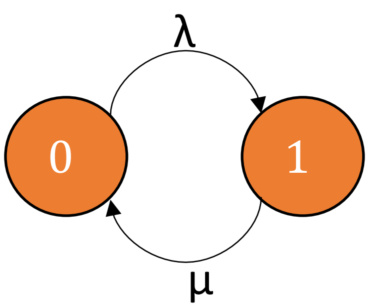
$$\pi_1 = \frac{1}{\rho^2 + \rho + 1}, \quad \pi_2 = \frac{\rho}{\rho^2 + \rho + 1}, \quad \pi_3 = \frac{\rho^2}{\rho^2 + \rho + 1}$$

  - For the simplicity of representing the π_i values, we have considered $\rho = {}^{\lambda}/_{\mu}$
- As you can see, the sum of probabilities are 1
  - This is a test for making sure you have calculated the right values

Step-by-Step Procedure for Transient Analysis

SHARIF UNIVERSITY OF TECHNOLOGY

- Transient analysis is more complicated than steady-state
- Let's discuss this on a simple 2-state example:
  - Obtain the transient probabilities $\pi_i(t)$ for the following CTMC
  - It has been assumed that the failure, and repair rates support exponential distribution
    - $f_1(t) = e^{-\lambda t}$, and $f_2(t) = e^{-\mu t}$
- In transient analysis, we use $\pi'(t) = \pi(t)Q$



$$[\pi_1'(t) \quad \pi_2'(t)] = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} [\pi_1(t) \quad \pi_2(t)]$$

$$\begin{cases} \pi_1'(t) = -\lambda\pi_1(t) + \mu\pi_2(t) \\ \pi_2'(t) = \lambda\pi_1(t) - \mu\pi_2(t) \end{cases}$$

How to solve this type of differential equations?

Laplace Transform (1)

SHARIF UNIVERSITY OF TECHNOLOGY

- Various approaches for solving differential equations
- One of the strongest methods is the **Laplace Transform**
  - It transforms a differential equation into a simple algebraic equation
  - The variation range is transformed from t to a new parameter S
- The Laplace transform for function $f(t)$ is denoted as:

$$L\{f(t)\} = F(S)$$

- To obtain the transformed version of $f(t)$ we use:

$$F(S) = \int_0^{\infty} e^{-St} f(t)dt$$

Laplace Transform (2)

- One of the useful transforms in the Laplace domain, is the transformation of nth-order derivation of $f(t)$
  - For instance, what is the Laplace transform of $\pi'(t)$?

We know: $\pi(S) = L\{\pi(t)\} = \int_0^\infty e^{-St}\pi(t)dt \rightarrow$

$$L\{\pi'(t)\} = \int_0^\infty \overset{v}{\boxed{e^{-St}}}\,\overset{du}{\boxed{\pi'(t)}}\,dt = \overset{v}{\boxed{e^{-St}}}\,\overset{u}{\boxed{\pi(t)}}[0,+\infty] - \int_0^\infty \overset{dv}{\boxed{(-S)e^{-St}}}\,\overset{u}{\boxed{\pi(t)}}\,dt$$

$$= 0 - \pi(0) - (-S)\int_0^\infty e^{-St}\pi(t)dt = S\pi(S) - \pi(0)$$

$$\Rightarrow \quad L\{\pi'(t)\} = \boldsymbol{S\pi(S) - \pi(0)}$$

Laplace transform for the 1st order deviation

Laplace Transform (3)

SHARIF UNIVERSITY OF TECHNOLOGY

- This process could be repeated to obtain higher order derivations
  - Example: What is the Laplace transform for the 2nd order derivation of function f(t) ($f''(t)$)?
    - Assume $g(t) = f'(t) \ \rightarrow L\{f''(t)\} = L\{g'(t)\}$
  - According to what we discussed in previous slide:

$$L\{f''(t)\} = SL\{g(t)\} - g(0) = SL\{f'(t)\} - f'(0) =$$

$$S[SF(S) - f(0)] - f'(0) = \boldsymbol{S^2F(S) - Sf(0) - f'(0)}$$

- Accordingly, the initial values ($f(0), f'(0), f''(0), \ldots$) must be provided

Simple 2-state Example (1)

- Now, we continue our previous example
- We must first transform the obtained differential equations into Laplace form:

$$\begin{cases} \pi_1'(t) = -\lambda\pi_1(t) + \mu\pi_2(t) \\ \pi_2'(t) = \lambda\pi_1(t) - \mu\pi_2(t) \end{cases} \Rightarrow \begin{cases} S\pi_1(S) - \pi_1(0) = -\lambda\pi_1(s) + \mu\pi_2(S) \\ S\pi_2(S) - \pi_2(0) = \lambda\pi_1(S) - \mu\pi_2(S) \end{cases}$$

- It could be derived that:

$$\pi_1(S) = \frac{\mu}{S+\lambda}\pi_2(S) + \frac{1}{S+\lambda}\pi_1(0) \quad \textbf{(1)} \qquad \pi_2(S) = \frac{\lambda}{S+\mu}\pi_1(S) + \frac{1}{S+\mu}\pi_2(0) \quad \textbf{(2)}$$

- We assume that the initial probability vector is $\pi(0) = [1\ 0]$
  - So, $\pi_1(0) = 1$, and $\pi_2(0) = 0$
  - Replace these values in the obtained equations

Simple 2-state Example (2)

- In this point, you achieve $\pi_1(S) = \frac{S+\mu}{S(S+\mu+\lambda)}$
- Now what?
  - One of the challenges in any Laplace transform procedure is the **inverse transform** to obtain t-based function based on the S-domain function
  - One approach is to use Mellin's inverse formula

$$f(t) = L^{-1}\{F(S)\}(t) = \frac{1}{2\pi i} \lim_{T\to\infty} \int_{\gamma - iT}^{\gamma + iT} e^{st} F(S) ds$$

  - This could be terrifying
- A table has been introduced, which denotes the inverse of well-known functions to simplify the transform from S to t

Laplace Inverse Transform

- It is recommended to make your $F(S)$ similar to the second column of this table
  - In order to simply detect $f(t)$
- Question: $\pi_1(S) = \dfrac{S+\mu}{S(S+\mu+\lambda)}$ is not similar to any of the indicated transforms
  - What should we do?
  - Use the **Separation of fractions**
    - To use:

$$f(t) = e^{at} \leftrightarrow F(S) = \frac{1}{S-a}$$

$f(t)$	$\mathcal{L}[f(t)] = F(s)$	
1	$\frac{1}{s}$	(1)
$e^{at} f(t)$	$F(s-a)$	(2)
$\mathcal{U}(t-a)$	$\frac{e^{-as}}{s}$	(3)
$f(t-a)\mathcal{U}(t-a)$	$e^{-as}F(s)$	(4)
$\delta(t)$	1	(5)
$\delta(t-t_0)$	e^{-st_0}	(6)
$t^n f(t)$	$(-1)^n \frac{d^n F(s)}{ds^n}$	(7)
$f'(t)$	$sF(s) - f(0)$	(8)
$f^n(t)$	$s^n F(s) - s^{(n-1)} f(0) - \cdots - f^{(n-1)}(0)$	(9)
$\int_0^t f(x) g(t-x) dx$	$F(s)G(s)$	(10)
$t^n \ (n = 0, 1, 2, \dots)$	$\frac{n!}{s^{n+1}}$	(11)
$t^x \ (x \geq -1 \in \mathbb{R})$	$\frac{\Gamma(x+1)}{s^{x+1}}$	(12)
$\sin kt$	$\frac{k}{s^2+k^2}$	(13)
$\cos kt$	$\frac{s}{s^2+k^2}$	(14)
e^{at}	$\frac{1}{s-a}$	(15)
$\sinh kt$	$\frac{k}{s^2-k^2}$	(16)
$\cosh kt$	$\frac{s}{s^2-k^2}$	(17)
$\frac{e^{at}-e^{bt}}{a-b}$	$\frac{1}{(s-a)(s-b)}$	(18)
$\frac{ae^{at}-be^{bt}}{a-b}$	$\frac{s}{(s-a)(s-b)}$	(19)
te^{at}	$\frac{1}{(s-a)^2}$	(20)
$t^n e^{at}$	$\frac{n!}{(s-a)^{n+1}}$	(21)
$e^{at} \sin kt$	$\frac{k}{(s-a)^2+k^2}$	(22)
$e^{at} \cos kt$	$\frac{s-a}{(s-a)^2+k^2}$	(23)
$e^{at} \sinh kt$	$\frac{k}{(s-a)^2-k^2}$	(24)
$e^{at} \cosh kt$	$\frac{s-a}{(s-a)^2-k^2}$	(25)
$t \sin kt$	$\frac{2ks}{(s^2+k^2)^2}$	(26)
$t \cos kt$	$\frac{s^2-k^2}{(s^2+k^2)^2}$	(27)
$t \sinh kt$	$\frac{2ks}{(s^2-k^2)^2}$	(28)
$t \cosh kt$	$\frac{s^2-k^2}{(s^2-k^2)^2}$	(29)
$\frac{\sin at}{t}$	$\arctan \frac{a}{s}$	(30)
$\frac{1}{\sqrt{\pi t}} e^{-a^2/4t}$	$\frac{e^{-a\sqrt{s}}}{\sqrt{s}}$	(31)
$\frac{a}{2\sqrt{\pi t^3}} e^{-a^2/4t}$	$e^{-a\sqrt{s}}$	(32)
$\operatorname{erfc}\left(\frac{a}{2\sqrt{t}}\right)$	$\frac{e^{-a\sqrt{s}}}{s}$	(33)

Separation of Fractions

SHARIF UNIVERSITY OF TECHNOLOGY

- In order to transform $\pi_1(S) = \frac{S+\mu}{S(S+\mu+\lambda)}$ to factors of $\frac{1}{S-a}$

$$\frac{S+\mu}{S(S+\mu+\lambda)} = \frac{1}{(S+\mu+\lambda)} + \frac{\boldsymbol{\mu}}{\boldsymbol{S(S+\mu+\lambda)}} = \frac{1}{(S+\mu+\lambda)} + \left[\frac{\boldsymbol{A_1}}{\boldsymbol{S}} + \frac{\boldsymbol{A_2}}{\boldsymbol{(S+\mu+\lambda)}}\right]$$

- Use simultaneous equations to find A_1, and A_2
- How to quickly find A_1, and A_2
  - Multiply the red sides by the denominator $(S+\mu+\lambda)$
  - Now consider obtained equation $\mu = A_1(S+\mu+\lambda) + A_2 S$
    - To find A1: Use root of the fist fraction denominator $(S=0)$
      - $\mu = A_1(\mu+\lambda) \to A_1 = \frac{\mu}{\mu+\lambda}$
    - To find A2: Use root of the second fraction denominator $(S=-\mu-\lambda)$
      - $\mu = A_2(-\mu-\lambda) \to A_2 = \frac{-\mu}{\mu+\lambda}$

Final Results of Our Simple 2-state Example

- Finally $\pi_1(S) = \frac{1}{(S+\mu+\lambda)} + \frac{\mu/_{\mu+\lambda}}{S} - \frac{\mu/_{\mu+\lambda}}{(S+\mu+\lambda)}$
- In slide 50, similar to what we did for $\pi_1(S)$, we could right down $\pi_2(S)$, based on $\pi_1(S)$
  - Accordingly, $\pi_2(S) = \frac{\lambda}{S+\lambda}\pi_1(S) = \frac{\lambda}{\cancel{S+\mu}} \times \frac{\cancel{S+\mu}}{S(S+\mu+\lambda)} = \frac{\lambda}{S(S+\mu+\lambda)}$
- By employing the separation of fractions technique, you will be able to obtain $\pi_2(S)$ as follows:

$$\pi_2(S) = \frac{\lambda/_{\mu+\lambda}}{S} - \frac{\lambda/_{\mu+\lambda}}{(S+\mu+\lambda)}$$

- Based on the table we have:

$$\pi_1(t) = \frac{\mu}{\mu+\lambda} + \frac{\mu}{\mu+\lambda}e^{-(\mu+\lambda)t} \qquad \pi_2(t) = \frac{\lambda}{\mu+\lambda} - \frac{\lambda}{\mu+\lambda}e^{-(\mu+\lambda)t}$$

Using Various Notations for Results (1)

- Consider the previous 2-state CTMC with following assumptions:
  - State 0 corresponds to the correct, and state 1 corresponds to the failed states
  - The device may fail with an exponential distribution with rate λ
    - This rate would not alter during the simulation
  - It could be repaired with an exponential distribution with rate μ
- Recall: The mean value for an exponentially distributed variable is obtained by reversing the rate value ($\frac{1}{\lambda}$)
  - The MTTF (or MTTR) for this device follows the same role

$$MTTF = 1/\lambda \qquad MTTR = 1/\mu$$

Using Various Notations for Results (2)

- If it is desired to obtain the steady-state probabilities for this example, what is the solution:
  - Using the $\pi Q = 0$, and doing the math from scratch
  - Use $\lim_{t\to\infty} \pi_i(t)$ for the intended states
- If we follow the second approach:

$$\pi_1 = \lim_{t\to\infty} \pi_1(t) = \lim_{t\to\infty} \frac{\mu}{\mu+\lambda} + \cancel{\frac{\mu}{\mu+\lambda} e^{-(\mu+\lambda)t}} = \frac{\frac{1}{MTTR}}{\frac{1}{MTTR}+\frac{1}{MTTF}} = \frac{MTTF}{MTTF+MTTR}$$

$$\pi_2 = \lim_{t\to\infty} \pi_2(t) = \lim_{t\to\infty} \frac{\lambda}{\mu+\lambda} - \cancel{\frac{\lambda}{\mu+\lambda} e^{-(\mu+\lambda)t}} = \frac{\frac{1}{MTTF}}{\frac{1}{MTTR}+\frac{1}{MTTF}} = \frac{MTTR}{MTTF+MTTR}$$

- We could use another notation: $\rho = \lambda/\mu$
  - $\pi_1 = \frac{1}{1+\rho}$, and $\pi_2 = \frac{\rho}{\rho+1}$

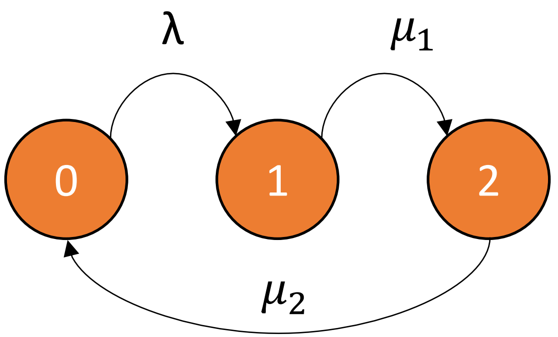
Repairing with Fault Detection and Correction (1)

- Let's modify our example a little bit
- Assume that the failure process is identical to what we had
- But, the repair is conducted in two phases:
  - First the system detects and locates the fault
  - Then it tries to repair the failed part
- The required time for both of these stages follow exponential distribution with rates μ_1, and μ_2
  - The mean time for theses phases: $1/\mu_1$, and $1/\mu_2$
- Depict the state diagram for this system?

Repairing with Fault Detection and Correction (2)

SHARIF UNIVERSITY OF TECHNOLOGY

- We have three states:
  - Correct state
  - Fault detected and located
  - Repair
- Specifying the rates on the edges
- Determining the infinitesimal generator matrix for future transient, and steady-state analysis



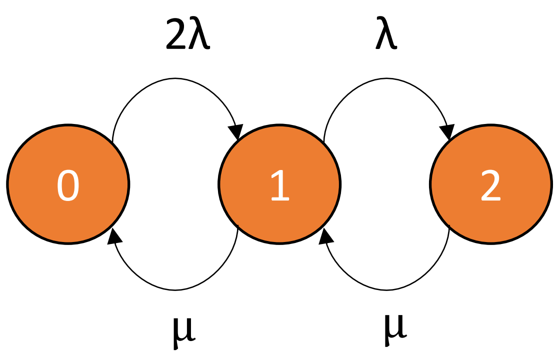
$$Q = \begin{bmatrix} -\lambda & \lambda & 0 \\ 0 & -\mu_1 & \mu_1 \\ \mu_2 & 0 & -\mu_2 \end{bmatrix}$$

If transient response is required $\pi Q = 0$

If steady-state response is required $\pi'(t) = \pi Q$

Multi-Component Systems

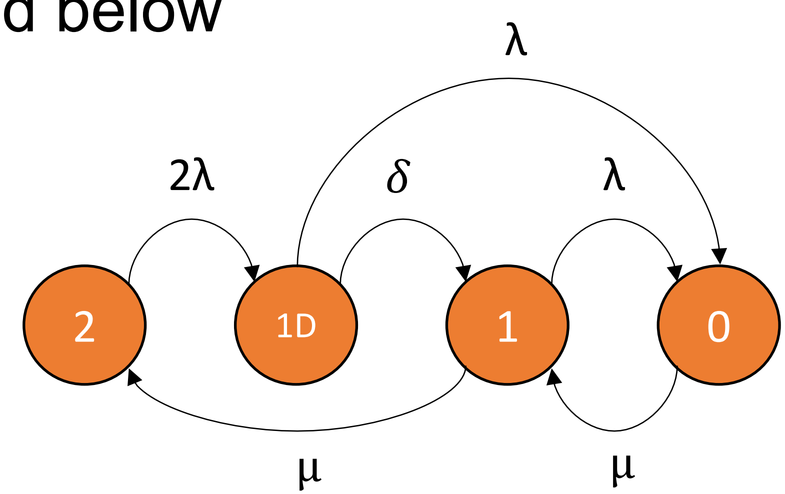
- Consider a two-component system, each component with failure rate λ
- Suppose there is **only a single repair facility** in the system, which services a failed component
- The system is unavailable to users if both components fail
  - Depict the state diagram for this system, and find the transition matrix



$$Q = \begin{bmatrix} -2\lambda & 2\lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & \mu & -\mu \end{bmatrix}$$

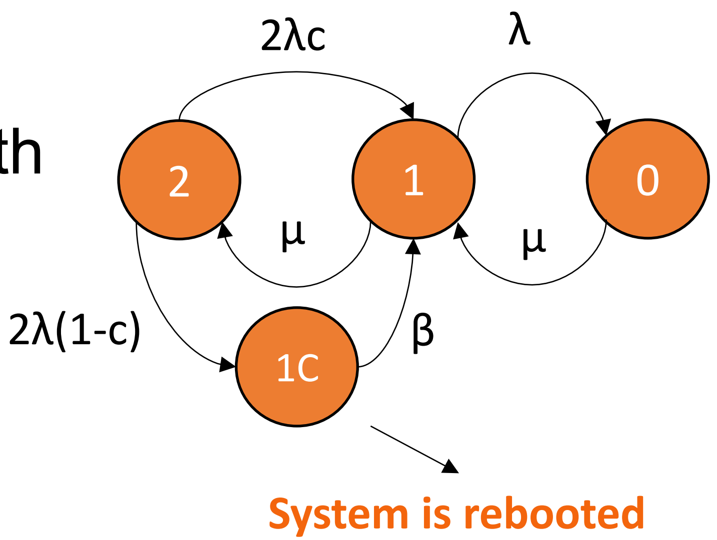
Fault Locating Delay

- We now introduce **detection delay** that is exponentially distributed with mean $1/\delta$
  - Suppose that it takes $1/\delta$ time units in average to detect a fault occurred in a component
  - The rate of fault detection is δ
- The state diagram for a system composed of fault detection mechanism is depicted below
- This systems has two devices
  - Before declaring a failure, system tries to find its source (faults and defects) in the devices
    - A state must be considered to show the **delay** for this effort



Failure Coverage

- The probability of some type of fault that can be detected during the test of any engineered system
  - High fault coverage is particularly valuable during manufacturing test
    - Techniques such as **Design For Test** and **automatic test pattern generation** are used to increase it
- Consider another variation of two-component system in which the failure is detected and handled with probability c and is not detected with probability $(1-c)$
  - If the system is not able to detect the failure, the whole system is rebooted



Class Workshop

- Find the **transient**, and **steady-state responses** for all of the existing states in the following CTMC
- Assumptions:
  - $\alpha = 0.04$
  - $\beta = 0.09$
  - $\lambda_1 = 0.3$
  - $\mu_1 = 2.5$
  - The initial probability vector is [0 1 0]