Continuous Time Markov Chains (CTMC)
Lecture #6

Contents

1 Markov Process (Continuous Time Markov Chain)

- The main difference from DTMC is that transitions from one state to another can occur at any instant of time.
- In order to satisfy the Markov property, the time the system spends in any given state should be memoryless $\Rightarrow$ the state sojourn time is exponentially distributed.

1.1 Mathematical Representation

- A Markov process X_t is completely determined by the so called ***generator matrix*** or ***transition rate matrix*** $Q = [q_{ij}]$

$$q_{ij} = \lim_{\Delta t \to 0} \frac{P\{X_{t+\Delta t} = j | X_t = i\}}{\Delta t} \quad i \neq j$$

  where q_{ij} transition rate or transition intensity and represents the probability per time unit that the system makes a transition from state i to state j.

  - The total transition rate out of state i, denoted as, can be expressed as $q_i = \sum_{i \neq j} q_{ij}$
  - This is the rate at which the probability of state i decreases. Define $q_{ii} = -q_i$

- Hence, one can express the generator matrix as

$$Q = \begin{pmatrix} q_{00} & q_{01} & . & . & . \\ q_{10} & q_{11} & . & . & . \\ . & & . & & . \\ . & & . & & . \\ . & & . & & . \end{pmatrix} = \begin{pmatrix} -q_0 & q_{01} & . & . & . \\ q_{10} & -q_1 & . & . & . \\ . & & . & & . \\ . & & . & & . \\ . & & . & & . \end{pmatrix}$$

Note that the sum of each row equals ***zero*** *indicating that the probability mass flowing out of state i will go to some other states (is conserved)*

1.2 Transient State Probabilities

- State probability vector $\pi(t)$ is now a function of time evolving as follows

$$\frac{d}{dt}\pi(t) = \pi(t)Q \tag{1}$$

- Generally, studying the behavior of CTMC is not a simple task. Even in homogeneous chains, the study of such chains is not generally tractable
- The transient solution for the state probabilities $\pi(t)$ can be expressed as

$$\pi(t) = \pi_0 e^{Qt}.$$

A closed-form expression for the transient behavior is not easy to obtain even for simple chains.

1.3 Steady State Analysis

- For the steady state analysis, *an irreducible CTMC has a limiting distribution that is independent of the initial state distribution.*
- For irreducible Markov Chains at steady state

$$\pi = \lim_{t\to\infty} \pi(t) \quad \pi Q = 0 \tag{2}$$

  - The solution is unique up to a constant factor.
  - The solution is uniquely determined by the normalization condition ($\sum_i \pi_i = 1$).
  - π is the (left) eigenvector belonging to the eigenvalue 0.

- Hence, Global balance condition which expresses the balance of probability flows

$$\pi_j q_j = \sum_{i\neq j} \pi_i q_{ij}$$

$$\pi_j \sum_{i\neq j} q_{ji} = \sum_{i\neq j} \pi_i q_{ij}$$

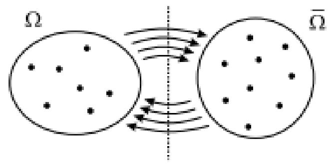
- Solving Balance Equations. Similar to the discrete case, we have

$$\pi Q = 0 \;\; and \;\; \pi E = e$$

Hence, the steady state solution will be

$$\pi = e(Q + E)^{-1}$$

- Note that the global balance equation can be applied for a set of states.



Proof

$$\pi_j \sum_{i \neq j} q_{ji} = \sum_{i \neq j} \pi_i q_{ij}$$

Now let us add a set of these equation corresponding to a subset S of the model states and split the above summations into two summations over states $\in S$ and states $\notin S$

$$\sum_{j \in S} \pi_j \sum_{i,\, i \neq j} q_{ji} = \sum_{j \in S} \sum_{i,\, i \neq j} \pi_i q_{ij}$$

$$\sum_{j \in S} \pi_j \left[\sum_{i \notin S} q_{ji} + \sum_{i \in S} q_{ji} \right] = \sum_{j \in S} \left[\sum_{i \in S} \pi_i q_{ij} + \sum_{i \notin S} \pi_i q_{ij} \right]$$

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

Example

Consider a birth death chain in which births occur with a rate of λ and death occurs at a rate of μ. The chain represent the system population.

Solution:

1.4 Embedded Markov Chain

- Also commonly known as jump chain.
  - Focus is on the transitions of X_t (when they occur), i.e. on the sequence of (different) states visited by X_t.
  - Let the state transitions of X_t occur at instants $t_0, t_1, ...$ Define $X_n^{(e)}$ to be the value of X_t immediately after the transition at time t_n (at the instant t_n^+) or the value of X_t in (t_n, t_{n+1}).
- Since X_t is a Markov process, the embedded chain $X_n^{(e)}$ constitutes a Markov chain.
- The transition probabilities of the embedded chain

$$p_{ij} = \begin{cases} \frac{q_{ij}}{\sum_j q_{ij}} & , i \neq j \\ 0 & , i = j \end{cases}$$

- Let π be the steady state probability of the Markov process and $\pi^{(e)}$ be the steady state probability of the embedded Markov chain.

$$\pi_i = \frac{\pi_i^{(e)}/q_i}{\sum_j \pi_j^{(e)}/q_j} \iff \pi_i^{(e)} = \frac{\pi_i q_i}{\sum_j \pi_j q_j}$$

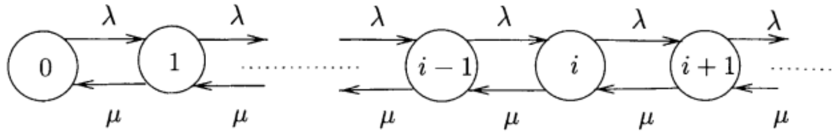
- Note that
  - π_i= proportion of time that the X_t spends in state i
  - $\pi_i^{(e)}$= relative frequency with which state i occurs in the jump chain $X_n^{(e)}$

2 Queuing Systems

2.1 M/M/1 Queuing System

- jobs arrive with a negative exponential interarrival time distribution with rate X.
- The job service requirements are also negative exponentially distributed with mean $E[S] = 1/\mu$.

2.1.1 Average Performance Metric Derivations



- This is similar to the birth death example discussed above

$$\pi_i = \frac{\lambda}{\mu}\pi_{i-1} \quad \sum_i \pi_i = 1$$

and hence we can express the steady state distribution as

$$\pi_i = \rho^i \pi_0 = \rho^i(1-\rho)$$

where $\rho = \lambda/\mu = \lambda E[S]$.

- ρ is commonly known as the system utilization ($\pi_0 = 1-\rho$)
- Hence, the expected number of users can be expressed as

$$N = \sum_i i\pi_i = (1-\rho)\sum_i i\rho^i$$

Using the computer science cheat sheet (prove that $\sum_i i\rho^i = \rho/(1-\rho)^2$

$$N = (1-\rho)\rho/(1-\rho)^2 = \frac{\rho}{1-\rho}$$

2.1.2 System Time Distribution

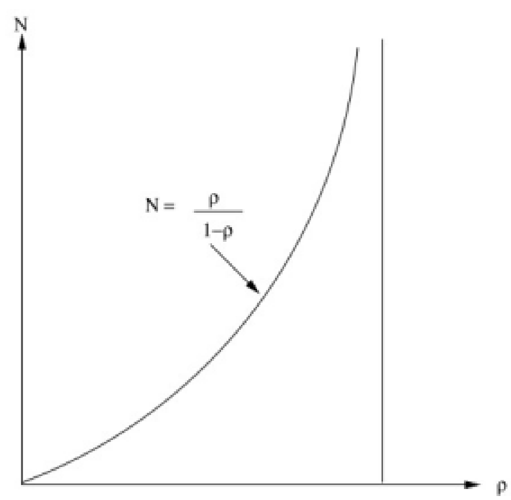
The conditional distribution assuming that the user find n users in the system would be Erlang-(n+1) (why?)

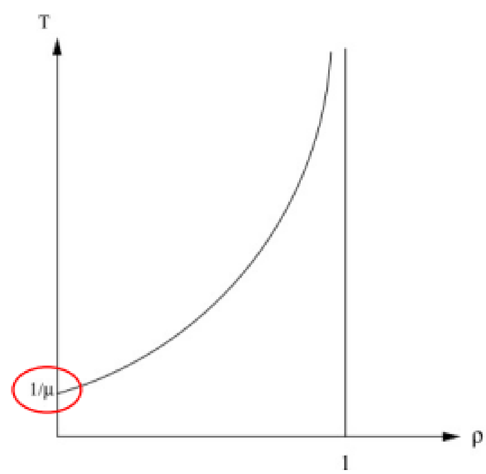
$$f_{T|N}(t|N=n) = \frac{\mu^{n+1}t^n}{n!}e^{-\mu t}$$

Hence, the unconditional waiting time distribution can be derived as follows

$$\begin{aligned} f_T(t) &= \sum_n f_{T|N}(t|N=n)p_n \\ &= \sum_n \frac{\mu^{n+1}t^n}{n!}e^{-\mu t}\rho^n(1-\rho) \\ &= e^{-\mu t}\mu(1-\rho)\ e^{\rho\mu t} \\ &= [\mu(1-\rho)]e^{-[\mu(1-\rho)]t} \end{aligned}$$

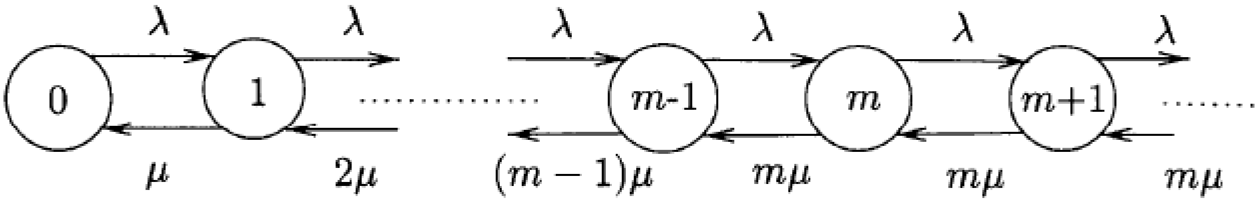
HEnce, the mean system time is $\frac{1}{\mu-\lambda}$





2.2 M/M/m Systems

- jobs arrive with a negative exponential interarrival time distribution with rate λ.
- The job service requirements are also negative exponentially distributed with mean $E[S] = 1/\mu$.
- The system has m servers



- Note the death rate of this chain?
- Using Global Balance condition, one can express the state probabilities as

$$\pi_i = \begin{cases} \frac{m^i}{i!}\rho^i \pi_o & \forall i = 0, ..., m-1 \\ \frac{m^m}{m!}\rho^i \pi_o & \forall i \geq m \end{cases}$$

where $\rho = \lambda/m\mu$. For stability, $\rho < 1$.

- Using the normalization equation, one can express π_o as

$$\pi_o = \left[\sum_{j=0}^{m-1} \frac{(m\rho)^j}{j!} + \frac{(m\rho)^m}{(1-\rho)\, m!} \right]^{-1}$$

and the expected number of the users in the system can be calculated as

$$N = \sum_{i=1}^{\infty} i p_i = m\rho + \rho \frac{(m\rho)^m}{m!} \frac{\pi_o}{(1-\rho)^2}$$

- The probability of queuing (also commonly known as Erlang-C blocking probability) can be estimated as

$$p_Q = \sum_{i=m}^{\infty} p_i = \frac{(m\rho)^m}{m!(1-\rho)} \pi_o$$

and represent the blocking probability for systems that enqueue access request when all the servers are busy.

- Similarly, the average number of jobs in queue can be estimated as

$$N_Q = \sum_{n=0}^{\infty} n p_{n+m} = P_Q \left[\frac{\rho}{1-\rho} \right] \tag{3}$$

- Eq. (3) suggests that M/M/m system behaves identically to an M/M/1 system with a service rate $m\mu$ once all servers are busy.

3 Homework

- what if we scale the Arrival rate of M/M/1?
- what if we scale the service rate of M/M/1?
- what if we scale both service and arrival rates of M/M/1?
- Evaluate the performance of M/M/1/K, M/M/m, M/M/m/K, M/M/∞ queuing systems