Lecture 11. Continuous Time Markov Chain

Sim, Min Kyu, Ph.D.
mksim@seoultech.ac.kr

서울과학기술대학교 데이터사이언스학과

BK21 FOUR 데이터사이언스와 비즈니스포텐셜

Motivation - You kinda already know this!

- Suppose there are total $n \geq 1$ customers in the 1-server post office. (1 in svc, $n-1$ in waiting)
- $T_A \sim exp(2)$ (arrival), $T_S \sim exp(3)$ (svc).

  interarrival time   service time

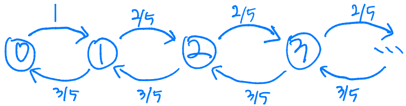
  service completion   customer arrival

- (# of total customer) $=n$ will become either [$n-1$], or [$n+1$]. And we call this event "transition" here.
  - *Q)* Difference between "transition" [here] and [in DTMC]?
  - *A)* No self-transition ($n \to n$)
    Time to transition is not constant

  $X_1 \sim exp(\lambda_1)$, $X_2 \sim exp(\lambda_2)$ X_1 & X_2 independent.
  $\Rightarrow \mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$

- You already know this!
  - What is the math. expression for event transition to $n+1$? $T_A > T_S$
  - What is the prob. of transition to $n+1$? $\mathbb{P}[T_A < T_S] = \frac{2}{2+3} = \frac{2}{5}$
  - What is the prob. of transition to $n-1$? $1-\frac{2}{5}=\frac{3}{5}$
  - What is the dist. of 'time to transition'? $\min(T_A, T_S) \sim exp(2+3)$

- Draw DTMC type transition diagram.



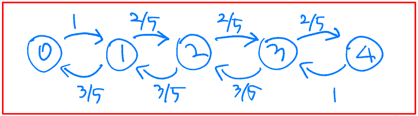
- Transition in DTMC and here
  - In DTMC, $n = 1, 2, ..$ is when transition occurs.
  - In here, transition occurs when an event occurs.
    $\rightarrow$ We need description for 'time to transition'.
- What is CTMC?

DTMC + description for time to transition for all states = CTMC

time to transition follows an exponential distribution

transition occurs in Poisson Process fashion

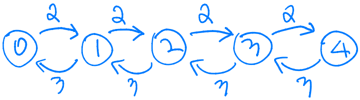
- Suppose $S = \{0, 1, 2, 3, 4\}$ in post office. (max. 3 waiting) M/M/1/3
  1. Draw DTMC type transition diagram; DTMC type transition matrix



$$\begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & & 1 & & & \\ 1 & 3/5 & & 2/5 & & \\ 2 & & 3/5 & & 2/5 & \\ 3 & & & 3/5 & & 2/5 \\ 4 & & & & 1 & \end{array}$$

  2. Add time to transition information
  3. Can you combine 1 and 2?

from state 0 : exp(2)
1 : exp(5)
2 : exp(5)
3 : exp(5)
4 : exp(3)



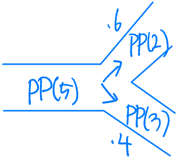
(transition prob. + time to transition) = CTMC's rate diagram

  4. Matrix representation

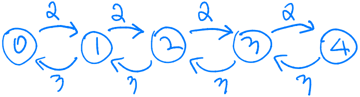
CTMC's rate matrix

$$G = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & -2 & 2 & & & \\ 1 & 3 & -5 & 2 & & \\ 2 & & 3 & -5 & 2 & \\ 3 & & & 3 & -5 & 2 \\ 4 & & & & 3 & -3 \end{array}$$

Σ row $= 0$.



Stationary Distribution - by flow balance equation



	inflow	outflow
State 0	$3\pi_1$	$2\pi_0$
State 1	$2\pi_0+3\pi_2$	$3\pi_1+2\pi_1$
State 2	$2\pi_1+3\pi_3$	$3\pi_2+2\pi_2$
State 3	$2\pi_2+3\pi_4$	$3\pi_3+2\pi_3$
State 4	$2\pi_3$	$3\pi_4$

$$\sum \pi_i = 1, \quad \pi = \left[\tfrac{81}{211} \quad \tfrac{54}{211} \quad \tfrac{36}{211} \quad \tfrac{24}{211} \quad \tfrac{16}{211} \right]$$

Stationary Distribution - by cutting method

$$\begin{tikzcd} 0 \arrow[r, bend left, "2"] & 1 \arrow[l, bend left, "3"] \arrow[r, bend left, "2"] & 2 \arrow[l, bend left, "3"] \arrow[r, bend left, "2"] & 3 \arrow[l, bend left, "3"] \arrow[r, bend left, "2"] & 4 \arrow[l, bend left, "3"] \end{tikzcd}$$

	$L \to R$		$R \to L$
$\{0\}$ vs. $\{1,2,3,4\}$	$2\pi_0$	$=$	$3\pi_1$
$\{0,1\}$ vs. $\{2,3,4\}$	$2\pi_1$	$=$	$3\pi_2$
$\{0,1,2\}$ vs. $\{3,4\}$	$2\pi_2$	$=$	$3\pi_3$
$\{0,1,2,3\}$ vs $\{4\}$	$2\pi_3$	$=$	$3\pi_4$

$\sum \pi_i = 1$,

Stationary Distribution - by $\pi G = 0$ rate matrix. G.

↳ equivalent to flow balance equation

$$G = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & -2 & 2 & & & \\ 1 & 3 & -5 & 2 & & \\ 2 & & 3 & -5 & 2 & \\ 3 & & & 3 & -5 & 2 \\ 4 & & & & 3 & -3 \end{array}$$

$$(\pi_0 \ \pi_1 \ \pi_2 \ \pi_3 \ \pi_4) \begin{bmatrix} -2 & 2 & & & \\ 3 & -5 & 2 & & \\ & 3 & -5 & 2 & \\ & & 3 & -5 & 2 \\ & & & 3 & -3 \end{bmatrix} = (0 \ 0 \ 0 \ 0 \ 0)$$

$2\pi_0 - 5\pi_1 + 3\pi_2 = 0.$

$\Rightarrow 2\pi_0 + 3\pi_2 = 5\pi_1$

Stationary distribution, π

$\begin{cases} \text{DTMC}: \ \pi P = \pi \\ \text{CTMC}: \ \pi G = 0 \end{cases}$

CTMC-Definition

Definition - CTMC (Continuous Time Markov Chain)

- Consider a continuous time stochastic process $\{X(t), t \geq 0\}$ in state space S. It is called a **Continuous Time Markov Chain (CTMC)** if for all $s, t \geq 0$ and $i, j \in S$ and $X(u) \in S$ for $0 \leq u < s$, outdated info
$\mathbb{P}[X(t+s) = j | X(s) = i, \cancel{X(u) = x(u), 0 \leq u < s}] = \mathbb{P}[X(t+s) = j | X(s) = i]$

- Continuous time version of Markov property.
- Key observations
  - Continuous time stochastic process that depends on the "most recent information".
  - History does not matter.
  - def. doesn't explicitly mention exp. dist.
  - Time to transition is exp. dist. ∵ Markov property is equivalent to exponentially distributed "time to transition"

blank

Def. Stationary Dist. in CTMC.

In CTMC, π is said to be a stationary distribution if for all $t \geq 0$

$\sum \pi_i = 1$, $\pi P(t) = \pi$, $P(t)$ = CTMC transition matrix.

cf) $\pi G = 0$ gives stationary dist.

in DTMC

$\pi P = \pi$

$\pi P^2 = \pi$

$\vdots$

$\pi P^n = \pi$ for all $n \in \mathbb{N}$

blank

CTMC transition matrix

- **Question**: Consider the first class example of CTMC. Suppose there are two customers now, what is the probability that there will be three customers in 6.5 minutes later? $\mathbb{P}(X(6.5)=3 \mid X(0)=2) = [P(6.5)]_{2,3}$
- **Motivation**: We need "CTMC type of transition matrix" to answer this question.
- As opposed to DTMC transition matrix **P** that contains "one step transition probability", we need CTMC transition matrix $\mathbf{P}(t)$
- Specifically, CTMC transition matrix should be given as $[\mathbf{P}(t)]_{i,j} = \mathbb{P}[X(t) = j | X(0) = i]$
- The original question asks $[\mathbf{P}(6.5)]_{2,3} = \mathbb{P}[X(6.5) = 3 | X(0) = 2]$

How to get $\mathbf{P}(t)$*?*

Claim, $\mathbf{P}(t) = e^{tG}$ $P(6.5) = e^{6.5G}$

pf)(very rough proof)

Remind Taylor's 1st order approximation of $f(t) \approx f(0) + t \cdot f'(0)$.

Similarly, $\mathbf{P}(t) \approx \mathbf{P}(0) + t \cdot \mathbf{P}'(0) = \mathbf{I} + t\mathbf{P}'(0)$.

For the differential equation $\mathbf{P}(t) = \mathbf{I} + t\mathbf{P}'(0)$, try $\mathbf{P}(t) = e^{tG}$ and get $e^{tG} = \mathbf{I} + tG$.

Taylor

The last equation above holds in Taylor's 1st order approximation's sense.

($\because e^x \approx 1 + x$).

Therefore, $\mathbf{P}(t) = e^{tG}$ is verified.

Example

- So, what is $[\mathbf{P}(6.5)]_{2,3} = \mathbb{P}[X(6.5) = 3 | X(0) = 2]$ for the class example?

$$P(6.5) = e^{6.5G} = \begin{bmatrix} .38 & .26 & .17 & .11 & .08 \\ \_\_\_ & & '' & & \_\_\_ \\ \_\_\_ & & '' & & \_\_\_ \\ \_\_\_ & & '' & & \_\_\_ \\ \_\_\_ & & '' & & \_\_\_ \end{bmatrix}$$

$$P(6.5)_{2,3} = 0.11$$

CTMC version of Chapman-Kolmogrov

- Remind in DTMC,
  - $\mathbf{P}^{n+m} = \mathbf{P}^n \mathbf{P}^m$.
  - $\mathbb{P}(X_{n+m} = j | X_0 = i) = \sum_{k \in S} \mathbb{P}(X_n = k | X_0 = i) \mathbb{P}(X_m = j | X_0 = k)$
- Now in CTMC,
  - $\mathbf{P}(s + t) = \mathbf{P}(s)\mathbf{P}(t)$
  - Even more, $\mathbf{P}(as + bt) = [\mathbf{P}(s)]^a [\mathbf{P}(t)]^b$

Stationary distribution

$\pi P(t) = \pi$

- Why is it possible to do $\pi G = 0$ to get stationary distribution?
- flow of proof)
  - Begin with $\pi = \pi \mathbf{P}(t)$ just like DTMC $\pi = \pi P$
  - Take derivative both side by t and evaluate at $t = 0$ to get...
  - $0 = \pi \mathbf{P}'(0) \Rightarrow 0 = \pi G$
- proof) claim, $\pi = \pi P(t)$ for all $t \geq 0$. $\Rightarrow \pi G = 0$.

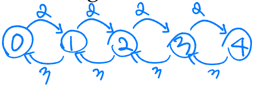
$\pi = \pi P(t) = \pi e^{tG}$

$\Rightarrow \frac{d}{dt}\pi = \frac{d}{dt}\pi e^{tG} \Rightarrow 0 = \pi e^{tG} \cdot G$

$t=0$, $0 = \pi G$.

Revisit the first example (with $S = \{0, 1, 2, 3, 4\}$)

- Rate diagram & Rate matrix



$$\begin{bmatrix} -2 & 2 & & & \\ 3 & -5 & 2 & & \\ & 3 & -5 & 2 & \\ & & 3 & -5 & 2 \\ & & & 3 & -3 \end{bmatrix}$$

- Stationary distribution

$\pi = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4) = (\frac{81}{211}, \frac{54}{211}, \frac{36}{211}, \frac{24}{211}, \frac{16}{211})$. Limiting prob.

- Questions of interest

  1. What is the long run fraction of time the system is empty? $\pi_0 = 81/211$
  2. What is the long run fraction of time the server is busy? $1 - \pi_0 = 130/211$
  3. What is the probability that a customer is not accepted to system? π_4
  4. What is the expected # of customer in the system?

     $L_{sys} = 0 \cdot \pi_0 + 1 \cdot \pi_1 + 2 \cdot \pi_2 + 3\pi_3 + 4\pi_4 = 1.24$

  5. What is the expected # of customer in the queue?

$L_{queue} = 0 \cdot \pi_0 + 0 \cdot \pi_1 + 1\pi_2 + 2\pi_3 + 3\pi_4 = 0.63.$

More questions M/M/1/3

- Revisit Little's Law: $L = \lambda W$ (LN 5 , p 16)
- More questions of interest
  6. What is expected total time spent in the system for a customer? (Waiting + Service time)

     $W_{sys} = \frac{L_{sys}}{\lambda} = \frac{1.24}{2} = 0.62$ min ?

     $W_{sys} = \frac{L_{sys}}{\lambda_{eff}} = \frac{1.24}{2(1-\pi_4)} = 0.67$ min

  7. What is expected waiting time for a customer?

     $W_q = \frac{L_q}{\lambda} = \frac{0.63}{2} = 0.315$ min ?

     $W_q = \frac{L_q}{\lambda_{eff}} = \frac{0.63}{2(1-\pi_4)} = 0.34$ min

  8. What is expected service time for a customer?

     $W_{svc} = W_{sys} - W_q = 0.295$ min ?

     $W_{svc} = W_{sys} - W_q = 0.33$ min

  9. What is expected service time for a customer?

     $W_{svc} = E T_s = \frac{1}{3} = 0.333$ min

  10. What is TH(throughput)?

      $TH = \lambda_{eff} = 2(1-\pi_4)$

      $\because$ stable system

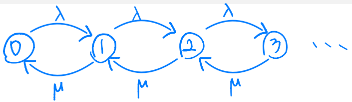
CTMC and Queuing Theory

- Revisit Kendall's notation: (LN 5 , p 7)
- M/M/·/· queuing model (Both inter-arrival and service times are exponentially distributed) can be analysed by CTMC, where the state is defined as the number of customers in the system!
- Examples
  - M/M/1/3
  - M/M/1/∞
  - M/M/k/∞
  - M/M/∞/∞

$M/M/1/\infty$

- Arrival is PP(λ /minute) and service completion is PP(μ /minute).
- $T_A \sim exp(\lambda)$
  $T_S \sim exp(\mu)$
  $\xrightarrow{\lambda} \boxed{\mu}$
- Assume $\lambda < \mu$ for stability. ($\rho := \lambda/\mu < 1$)
- Let $X(t)$ be the number of customers in the system at time t. Then, $\{X(t), t \geq 0\}$ is CTMC with state space $S = \{0, 1, 2, ...\}$.
- Rate diagram

Stationary distribution (by cutting)



	L to R		R to L	
$\{0\}$ vs $\{1,2,3,\cdots\}$	$\lambda\pi_0$	$=$	$\mu\pi_1$	$\Rightarrow \pi_1 = \left(\frac{\lambda}{\mu}\right)\pi_0$
$\{0,1\}$ vs $\{2,3,4,\cdots\}$	$\lambda\pi_1$	$=$	$\mu\pi_2$	$\Rightarrow \pi_2 = \left(\frac{\lambda}{\mu}\right)\pi_1$
$\{0,1,2\}$ vs $\{3,4,5,\cdots\}$	$\lambda\pi_2$	$=$	$\mu\pi_3$	$\Rightarrow \pi_3 = \left(\frac{\lambda}{\mu}\right)\pi_2$

$$\sum_0^\infty \pi_i = \quad ?$$

$$\pi_i = \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right) \text{ for } i = 0,1,2,\cdots$$

- Stationary distribution traffic intensity ρ.
  - $\pi_0 = ?$ $\quad 1-\frac{\lambda}{\mu} = 1-\rho$

$$\pi_i = \left(\frac{\lambda}{\mu}\right)^i \left(1-\frac{\lambda}{\mu}\right) \text{ for } i=0,1,2,\cdots$$
$$= \rho^i(1-\rho)$$

1. The long run average # of customers in the system (L_{sys})?

$$L_{sys} = 0\cdot\pi_0 + 1\cdot\pi_1 + \cdots$$
$$= 0\cdot(1-\rho) + 1\cdot\rho(1-\rho) + 2\cdot\rho^2(1-\rho) + \cdots$$
$$= (1-\rho)\{1\cdot\rho + 2\cdot\rho^2 + 3\rho^3 + \cdots\}$$
$$= (1-\rho)\frac{\rho}{(1-\rho)^2} = \frac{\rho}{(1-\rho)} = \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}} = \frac{\lambda}{\mu-\lambda}$$

cf) $S = \rho + 2\rho^2 + 3\rho^3 + \cdots$
$-\rho S = \rho^2 + 2\rho^3 + \cdots$

$(1-\rho)S = \rho + \rho^2 + \rho^3 + \cdots$
$= \frac{\rho}{1-\rho}$ $\quad \therefore S = \frac{\rho}{(1-\rho)^2}$

2. The expected time spent in the system by a customer (W_{sys})?

$$W_{sys} = \frac{L_{sys}}{\lambda} = \frac{1}{\lambda}\left(\frac{\rho}{1-\rho}\right) = \frac{1}{\lambda}\cdot\frac{\lambda}{\mu}\cdot\frac{1}{1-\rho} = \frac{1}{\mu}\left(\frac{1}{1-\rho}\right) = \frac{E[S]}{1-\rho}$$

$\lambda=2, \mu=3 \Rightarrow \rho = \frac{2}{3} \Rightarrow W_{sys} = \frac{E[S]}{1-\frac{2}{3}} = 3E[S]$

3. The long run average # of customers in the queue (L_q)?

$$L_q = 0\cdot\pi_0 + 0\cdot\pi_1 + 1\cdot\pi_2 + 2\pi_3 + \cdots$$
$$= ?$$

4. The expected time spent in the queue by a customer (W_q)?

i) $W_q = \dfrac{L_q}{\lambda} = ?$

ii) $W_q = W_{sys} - W_{svs} = \dfrac{\mathbb{E}T_s}{1-\rho} - \mathbb{E}T_s = \mathbb{E}T_s\left(\dfrac{1-(1-\rho)}{1-\rho}\right) = \mathbb{E}T_s\left(\dfrac{\rho}{1-\rho}\right)\left(\dfrac{1^2+1^2}{2}\right)$

cf) Kingman's $\Rightarrow W_q = \mathbb{E}V\cdot \dfrac{\rho}{1-\rho}\left(\dfrac{C_a^2+C_s^2}{2}\right)$

- Discussion on W_q

blank

Homework

factorial

- Consider M/M/∞/∞, M/M/k/∞. For each queuing model, respectively, do the followings:
  - Define the CTMC with arrival of PP(λ) and service of PP(μ).
  - Define state space
  - Draw rate diagram
  - Find the stationary distribution. While doing so, make necessary assumption so that the stationary distribution exists.
  - Find $L_{sys}, W_{sys}, L_q, W_q$.

?

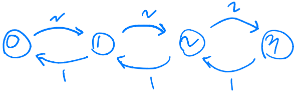
blank

blank

blank

Discussion on stability and stationary distribution

- Example. Rate Diagram



- Is this system stable?

Yes. because of finite state space, despite $\lambda > \mu$.

- What is being "stable"?

queue is not increasing to inf.

- Does this sound related to existence of stationary distribution?

stable dist. $\Rightarrow \exists$ stationary dist / unstable $\Rightarrow \nexists$ stationary dist.

- i.e. can unstable system have a stationary distribution?

- Review on random walk.

Stationary distribution in CTMC

- Remind that, *finite*, *irreducible*, *aperiodic* DTMC has a unique stationary distribution.

- Periodicity is not a concern in CTMC.

  ∵ continuous time for transition occuring

- Irreducibility is not a big concern because most of CTMC has only one class. Why?

  CTMC, in general, concerns with ongoing continuing stochastic process.

- Thus, a finite state CTMC has a unique stationary distribution.

- How about the *infinite* state CTMC?

  ∃ stationary dist. as long as stable.

Conclusion: Stationary distribution in CTMC

- As long as irreducible,
  - Finite CTMC has a unique stationary distribution
  - Infinite CTMC has a unique stationary distribution if and only if it is stable.
- If you can calculate unique stationary distribution, then the CTMC is stable. (regardless of finite or infinite state)

$M/M/2/1$

\# server

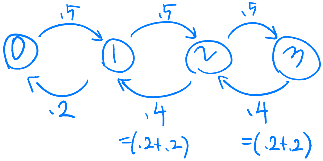
\# waiting space

-
- Suppose arrival is PP(0.5/minutes) and each server's service time follows exponential distribution with mean 5 minutes. How would you define CTMC?

  $T_A \sim \exp(\frac{1}{2})$

  $T_S \sim \exp(\frac{1}{5})$
- State space

  $S = \{0, 1, 2, 3\}$
- Rate diagram



$X_A \sim \exp(\lambda_A)$

$X_B \sim \exp(\lambda_B)$

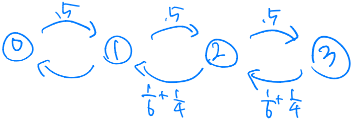
X_A & X_B are independent.

$\Rightarrow \min(X_A, X_B) \sim \exp(\lambda_A + \lambda_B)$

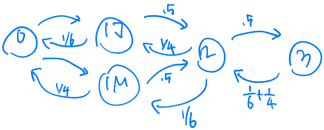
M/M/2/1 with different service rate

- Let arrival process be PP(0.5/minutes), same as the previous one.
  - Suppose the two servers (John and Mary) have service time exponentially distributed with mean 6 and 4 minutes, respectively.
- State space and rate diagram

$S = \{0, 1, 2, 3\}$



$S = \{0, 1J, 1M, 2, 3\}$



Priority rule (Assignment rule)

ex) flip the coin.

$$\begin{tikzcd} & 1J \\ 0 \arrow[ur, ".25"] \arrow[dr, ".25"'] & \\ & 1M \end{tikzcd}$$

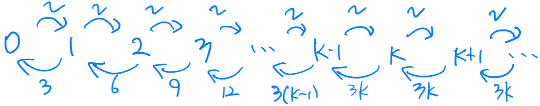
ex) always John

$$\begin{tikzcd} & 1J \\ 0 \arrow[ur, ".5"] \arrow[dr, "0"'] & \\ & 1M \end{tikzcd}$$

blank

$M/M/k/\infty$

- Suppose $\lambda = 2/min$ and $\mu = 3/min$
- Stationary distribution



$$\pi_i = \frac{1}{\sum_{0}^{k} \frac{2^i}{3^i\, i!} + \sum_{k+1}^{\infty} \frac{2^i}{3^i\, k^{i-k}\, k!}}$$

blank

blank

$M/M/2/\infty$

blank

$M/M/\infty/\infty$

blank

blank

```
"Learn from yesterday, live for today, hope for tomorrow. The important thing is
not to stop questioning. - Albert Einstein"
```