CS 3.307
Performance Modeling for Computer Systems

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Markov Process

- There are two versions of Markov chains- Discrete time and Continuous time.
- A stochastic process $\{X_n, n \in \mathbb{Z}_+\}$ is a discrete time Markov chain if for any $n_1 < n_2 < \ldots < n_k < n$,

$$P(X_n = j | X_{n_1} = x_1, .., X_{n_k} = i) = P(X_n = j | X_{n_k} = i)$$

- $\{X(t), t \geq 0\}$ is a Markov process (ctmc) if for $t_1 < t_2 < \ldots t_n < t$,

$$P(X(t) = j | X(t_1) = x_1, .., X(t_n) = i) = P(X(t) = j | X(t_n) = i)$$

- This is known as the Markov property.
- State space in both cases can be integers or general ($\mathbb{R}^d$)
- We will stick with integer or finite state space

Example: Coin with memory!

- In a Markovian coin with memory, the outcome of the next toss depends on the current toss.
- $X_n = 1$ for heads and $X_n = -1$ otherwise. $\mathcal{S} = \{+1, -1\}$.
- Sticky coin : $P(X_{n+1} = 1 | X_n = 1) = 0.9$ and $P(X_{n+1} = -1 | X_n = -1) = 0.8$ for all n.
- Flippy Coin: $P(X_{n+1} = 1 | X_n = 1) = 0.1$ while $P(X_{n+1} = -1 | X_n = -1) = 0.3$ for all n.
- This can be represented by a transition diagram (see board)
- The transition probability matrix P for the two cases is $P_s = \begin{bmatrix} 0.9 & .1 \\ 0.2 & 0.8 \end{bmatrix}$ and $P_f = \begin{bmatrix} 0.1 & 0.9 \\ 0.7 & 0.3 \end{bmatrix}$
- The row corresponds to present state and the column corresponds to next state.

Running example: Dice with memory!

- In a markovian dice with memory, the outcome of the next roll depends on the current roll.
- $X_n = i$ for $i \in \mathcal{S}$ where $\mathcal{S} = \{1, \ldots, 6\}$.
- Example transition probability matrix

$$P = \begin{bmatrix} 0.9 & .1 & 0 & 0 & 0 & 0 \\ 0 & .9 & .1 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.9 & 0.1 \\ 0.1 & 0 & 0 & 0 & 0 & 0.9 \end{bmatrix}$$

- In the ctmc counterpart for these examples, imagine the coin tosses itself/ dice rolls itself after waiting in the state for a random time that is exponentially distributed. (more later)

Time-homogenous Markov Process

- A DTMC is said to be time homogeneous if the one step transition probabilities are same at all time.
- $P(X_{n+1} = j | X_n = i) = P(X_{n+1+s} = j | X_{n+s} = i) := p_{ij}$
- One step transition probability matrix $P = [[p_{ij}]]$
- $i, j \in \mathcal{S}$ which is countable and $|\mathcal{S}| \leq \infty$

For a CTMC ...

- For a time homogeneous CTMC, we have

$$\begin{aligned} P(X(t) = j | X(t_n) = i) &= P(X(t+s) = j | X(t_n + s) = i) \\ &= P(X(t - t_n) = j | X(0) = i). \end{aligned}$$

- We have a transition probability matrix with entries $p_{ij}(t)$, i.e., $P(t) = [[p_{ij}(t)]]$.

DTMC – Time spent in a state

- For a time homogeneous DTMC, we have a transition probability matrix with entries p_{ij}, i.e., $P = [[p_{ij}]]$.
- Let $Y_n = inf\{s > 0 : X_{n+s} \neq X_n\}$
- Y_n is the remaining time that the process spends in whichever state it is in, at time n.
- Consider a Markov coin, its state transition matrix and diagram
- Y_n is geometric random variable.
- What would be the time spent in a state for a continuous time Markov chain ?

CTMC – Time spent in a state

- For a time homogeneous CTMC, we have a transition probability matrix with entries $p_{ij}(t)$, i.e., $P(t) = [[p_{ij}(t)]]$.
- Let $Y_t = inf\{s > 0 : X(t+s) \neq X(t)\}$
- Y_t is the remaining time that the process spends in whichever state it is in, at time t.
- Intuitively, the time spent in a state should depend only on what state it is in, and not on the previous state.

Theorem

$$P(Y_t > u | X(t) = i) := \bar{G}_i(u) = e^{-a_i u}$$

for all $i \in \mathcal{S}$ *and* $t \geq 0, u \geq 0$ *and for some real number* $a_i \in [0, \infty]$.

Proof 1

- $\bar{G}_i(u+v) = P(X(s) = i, s \in [t, t+u+v] | X(t) = i)$
- $\bar{G}_i(u+v) = P(X(s) = i, s \in [t+u, t+u+v]; X(p) = i, p \in [t, t+u] | X(t) = i)$
- $P(AB|C) = P(A|BC)P(B|C)$
- Due to Markov property we have $P(AB|C) = P(A|B)P(B|C)$
- $P(X(s) = i, s \in [t+u, t+u+v] | X(p) = i, p \in [t, t+u]) =$
- $P(X(s) = i, s \in [t+u, t+u+v] | X(t+u) = i) = \bar{G}_i(v)$
- $P(X(p) = i, p \in [t, t+u] | X(t = i)) = \bar{G}_i(u)$
- $\bar{G}_i(u+v) = \bar{G}_i(u)\bar{G}_i(v)$
- Only CCDF function which satisfies this equation is the exponential distribution. This requires a proof. We will skip this part.

Simpler Proof

- Let τ_i denote the time the CTMC spends in state i before moving out. Suppose the CTMC is in state i at time 0.
- What is $P(\tau_i > s + t | \tau_i > s)$?
- Note that $X(s) = i$ and therefore from the Markov property,

$$
\begin{aligned}
P(\tau_i > s + t | \tau_i > s) &= P(X(u) = i, u \in [s, s+t] | X(t) = i, t \in [0, s]) \\
&= P(X(u) = i, u \in [s, s+t] | X(s) = i) \\
&= P((X(u) = i, u \in [0, t] | X(0) = i) \\
&= P(\tau_i > t).
\end{aligned}
$$

- Since $P(\tau_i > s + t | \tau_i > s) = P(\tau_i > t)$, this implies the distribution has memoriless property and must be exponential.

Finite dimensional distributions

- Consider a DTMC $\{X_n, n \geq 0\}$ with tpm denoted by P.
- We assume M states and X_0 denotes the initial state.
- You can start in any starting state or may pick your starting state randomly.
- Let $\bar{\mu} = (\mu_1, \dots, \mu_M)$ denote the initial distribution.
- How does one obtain the finite dimensional distribution $P(X_0 = x_0, X_1 = x_1, \dots X_k = x_k)$?

Finite dimensional distributions

- Consider a CTMC $\{X_t, t \geq 0\}$ with t-time pm given by $P(t)$.
- We assume M states and X_0 denotes the initial state.
- Let $\bar{\mu} = (\mu_1, \ldots, \mu_M)$ denote the initial distribution.
- How does one obtain the finite dimensional distribution $P(X_0 = x_0, X_{t_1} = x_1, \ldots X_{t_k} = x_k)$?

Chapman Kolmogorov Equations for DTMC

- $P = [[p_{ij}]]$ denotes the one step transition probability matrix.
- Let $P^{(n)}$ denote the n-step transition probability matrix.
- CK equation tells us that $P^{(n+l)} = P^{(n)}P^{(l)}$.
- $p_{ij}^{(n+l)} = P(X_{n+l} = j | X_0 = i) = \sum_k P(X_{n+l} = j, X_n = k | X_0 = i)$
- $p_{ij}^{(n+l)} = \sum_k P(X_{n+l} = j | X_n = k, X_0 = i) P(X_n = k | X_0 = i)$
- $p_{ij}^{(n+l)} = \sum_k P(X_{n+l} = j | X_n = k) P(X_n = k | X_0 = i)$
- $p_{ij}^{(n+l)} = \sum_k p_{ik}^{(n)} p_{kj}^{(l)} = [P^{(n)}P^{(l)}]_{ij}$
- At which step did we use time homogeneity and the Markov property?

n step transition probabilities

- $P = [[p_{ij}]]$ denotes the one step transition probability matrix.
- Let $P^{(n)}$ denote the n-step transition probability matrix.
- From the CK equation we know that $P^{(n+l)} = P^{(n)}P^{(l)}$.
- It is easy to see that $P^{(n)} = P^{(n-1)}P$.
- For an M state DTMC, $p_{ij}^{(2)} = \sum_{k=1}^{M} p_{ik}p_{kj}$.
- This implies that that the n-step transition probability matrix can be obtained as $P^{(n)} = P^n$
- Given X_0 and P, you can generate n-step probabilities or $P_{X_0}(X_n)$

Chapman Kolmogorov Equations for CTMC

- Let $P(t)$ denote the t-time transition probability matrix.
- CK equation for a CTMC is $P(t+l) = P(t)P(l)$.
- $p_{ij}(t+l) = P(X(t+l) = j | X(0) = i)$
- $= \sum_k P(X(t+l) = j, X(t) = k | X(0) = i)$
- $= \sum_k P(X(t+l) = j | X(t) = k, X(0) = i) P(X(t) = k | X(0) = i)$
- $= \sum_k P(X(t+l) = j | X(t) = k) P(X(n) = k | X(0) = i)$
- $p_{ij}(t+l) = \sum_k p_{ik}(t) p_{kj}(l) = [P(t)P(l)]_{ij}$

What generates a CTMC ?

- $P(t+l) = P(t)P(l)$.
- In DTMC, we could use P to generate the chain on Matlab.
- What about CTMC ? Can we use $P(t)$?
- What is $\lim_{h \to 0} P(h)$?
- What is $\frac{dP(h)}{dh}$ evaluated at $h = 0$?

What generates a CTMC ?

- Lets look at $\frac{dP(h)}{dh}|_{h=0} = \lim_{h\to 0} \frac{P(h)-P(0)}{h} = \lim_{h\to 0} \frac{P(h)-I}{h}$.
- Define $Q := \lim_{h\to 0} \frac{P(h)-I}{h}$
- **Does it always exist ?** Yes! (Proposition 2.2 and 2.4 (Anderson))
- Q has terms of the form q_{ii} and q_{ij} for $i, j \in \{1, 2, \ldots, M\}$.
- $q_{ii} = \frac{dp_{ii}(h)}{dh}|_{h=0}$. Similarly $q_{ij} = \frac{dp_{ij}(h)}{dh}|_{h=0}$

What generates a CTMC ?

Theorem

Let $P(t)$ be a transition function. Then the generator matrix $Q = \lim_{h \to 0} \frac{P(h) - I}{h}$ exists.

Theorem

$P(Y_t > u | X(t) = i) = e^{-a_i u}$ *where* $a_i > 0$.

Theorem

(Proposition 2.8 Anderson)
$P(Y_t > u | X(t) = i) := e^{q_{ii} u}$, *i.e.,* $q_{ii} = -a_i$.
$P\left(X(t + Y_t) = j | X(t) = i\right) = \frac{q_{ij}}{|q_{ii}|}$.

Q generates the CTMC

- Cannot generate CTMC directly from $P(t)$.
- From $P(t)$, obtain Q using $Q = \frac{dP(h)}{dh}|_{h=0}$
- Consider Y_t when $X(t) = i$.
- Now use the following theorem for generating the CTMC on a computer

Theorem

(Proposition 2.8 Anderson: we won't see proof)
$P(Y_t > u | X(t) = i) := e^{q_{ii} u}$, i.e., $q_{ii} = -a_i$.
$P\left(X(t + Y_t) = j | X(t) = i\right) = \frac{q_{ij}}{|q_{ii}|}$.

Properties of a conservative Q

Theorem

$P(Y_i > u | X(t) = i) := e^{q_{ii} u}$, *i.e.*, $q_{ii} = -a_i$.
$P\left(X(t + Y_t) = j | X(t) = i\right) = \frac{q_{ij}}{|q_{ii}|}$ *where* $q_{ij} \geq 0$.

- Suppose Q is conservative.
- Recall that q_{ii} is negative. A conservative Q implies $q_{ii} = -\sum_{j \neq i} q_{ij}$.
- $|q_{ii}|$ is the exponential rate at which you leave state i.
- q_{ij} is the exponential rate at which you leave state i to go to state j.
- minimum of exponentials is exponential with aggregated rate.
- This justifies the rate of leaving state i to be $\sum_{j \neq i} q_{ij}$.

Equivalent definition of a CTMC using Q

- Suppose Q is conservative.
- Then in the CTMC, you stay in state i for a random duration that has exponential($|q_{ii}|$) distribution.
- From i, you will move to state j with probability $\frac{q_{ij}}{|q_{ii}|}$.
- Equivalently, in state i, you have $M-1$ exponential(q_{ij}) clocks for $j = 1, 2, \ldots, i-1, i+1, \ldots M$.
- You move to that state whose clock rings first!

RECAP

CK Equations: $P(t+l) = P(t)P(l)$

Theorem

Let $P(t)$ be a transition function. Then the generator matrix $Q = \lim_{h\to 0} \frac{P(h)-I}{h}$ exists.

Theorem

$P(Y_t > u | X(t) = i) = e^{-a_i u}$ *where* $a_i > 0$.

Theorem

For a CTMC with Q matrix, we have
$P(Y_t > u | X(t) = i) := e^{q_{ii} u}$, *i.e.*, $q_{ii} = -a_i$.
$P\left(X(t+Y_t) = j | X(t) = i\right) = \frac{q_{ij}}{|q_{ii}|}$.

Kolmogorov forward/backward equations CTMC

- $\frac{dP(t)}{dt} = \lim_{s\to\infty} \frac{P(t+s)-P(t)}{s}$

- $\frac{dP(t)}{dt} = P(t) \lim_{s\to\infty} \frac{P(s)-I}{s}$

- $\frac{dP(t)}{dt} = P(t)Q.$

- $P(t) = e^{tQ}$ satisfies the above. (Calculus of Matrix exponentials)

- $P(t) = e^{tQ} := I + tQ + \ldots + \frac{(tQ)^n}{n!} \ldots$

Example: Poisson process $N(t)$ as a CTMC

- States $\mathcal{S} = Z_{\geq 0}$.
- Why is it a Markov process / Markov property satisfied?
- $P(N(t) = k|N(t_1) = k_1, \ldots, N(t_m) = k_m) = P(N(t) = k|N(t_m) = k_m)$?
- $P(N(t) = k|N(t_1) = k_1, \ldots, N(t_m) = k_m) = P(N(t - t_k) = k - k_m)$. Therefore the above is true.
- $p_{ij}(t) = P(N(t) = j|N(0) = i)$. $max(j - i, 0)$ arrivals in time t.
- We know that this has Possion distribution.
- How does $P(t)$ look for a Poisson process ?

Example: Poisson process $N(t)$ as a CTMC

- How does $P(t) = [[P(N(t) = j | N(0) = i)]]$ look ?
- Entries below the diagonal are zero.
- Diagonal entries have the value $e^{-\lambda t}$
- ijth entry above the diagonal has the value $e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}$

Example: Poisson process $N(t)$ as a CTMC

- How does $Q = \frac{dP(h)}{dh}|_{h=0}$ look ?
- ijth entry above the diagonal $p_{ij}(t) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}$
- what is $\frac{d}{dt}(e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!})|_{t=0}$?
- If $j - i = 1$, then $\frac{d}{dt}(e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}) = \lambda$.
- If $j - i > 1$, then $\frac{d}{dt}(e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}) = 0$.
- How does Q for Poisson process look like ?
- $P(t) = e^{tQ} = I + tQ + \ldots + \frac{(tQ)^n}{n!} + \ldots$

Example 3: Binomial process as a DTMC

- DO IT YOURSELF!

Embedded DTMC in a CTMC

- Consider a CTMC over state space $\mathcal{S}$.
- Let $Y_n, n \geq 0$ denote the sequence of times spent in successive states of the CTMC
- Define T_n to be the jump times of the CTMC, i.e., the times of successive state transitions.
- Then $T_n = \sum_{k=1}^{n} Y_k$.
- Define $X_n = X(T_n)$ for $n \geq 0$. X_n is the DTMC embedded in the CTMC.
- The corresponding TPM has $p_{ij} = \frac{q_{ij}}{|q_{ii}|}$.
- $\{X_n, \}$ is such that there are no one step transitions from a state to itself, i.e., $p_{ii} = 0$.

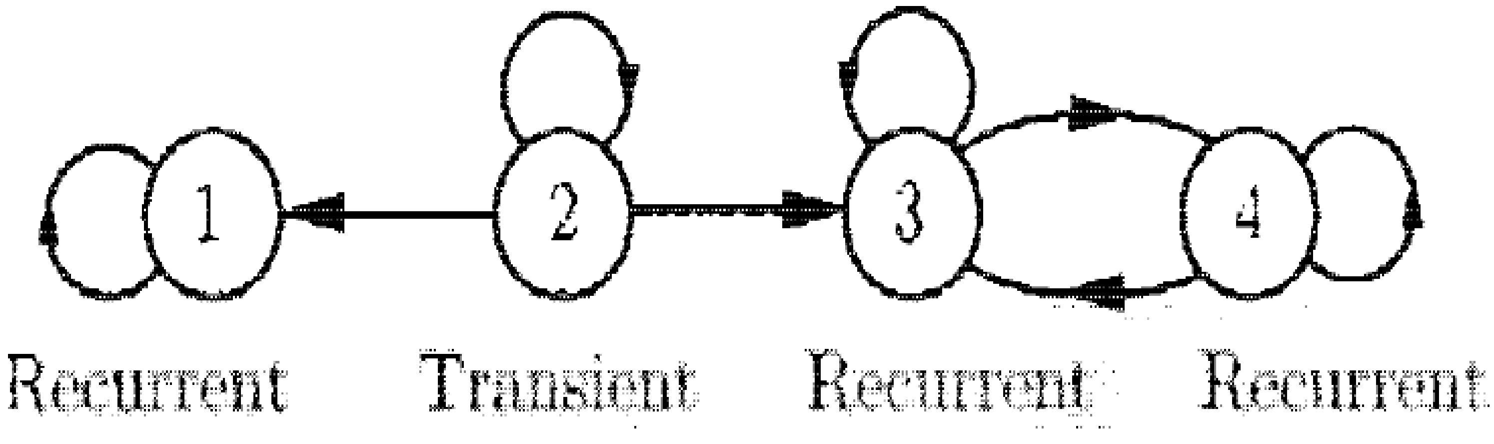
Classification of states

- Consider a Markov process with state space $\mathcal{S}$
- We say that j is accessible from i if $p_{ij}^n > 0$ for some n.
- This is denoted by $i \rightarrow j$.
- if $i \rightarrow j$ and $j \rightarrow i$ then we say that i and j communicate. This is denoted by $i \leftrightarrow j$.

> A chain is said to be irreducible if $i \leftrightarrow j$ for all $i, j \in \mathcal{S}$.

Recurrent and Transient states

- We say that a state i is recurrent if $F_{ii} = P($ ever returning to i having started in i $) = 1$.
- F_{ii} is not easy to calculate. (We will se ethis after Quiz)
- If a state is not recurrent, it is transient.
- For a transient state i, $F_{ii} < 1$.
- If $i \leftrightarrow j$ and i is recurrent, then j is recurrent.



Limiting probabilities

- $P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0.6 & 0.4 \\ 0.6 & 0.4 & 0 \end{bmatrix}$ $P^5 = \begin{bmatrix} .06 & .3 & .64 \\ .18 & .38 & .44 \\ .38 & .44 & .18 \end{bmatrix}$ $P^{30} = \begin{bmatrix} .23 & .385 & .385 \\ .23 & .385 & .385 \\ .23 & .385 & .385 \end{bmatrix}$

- $P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$ $\lim_{n\to\infty} P^n = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$

- What is the interpretation of $\lim_{n\to\infty} p_{ij}^{(n)} = [\lim_{n\to\infty} P^n]_{ij}$?

- $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$ denotes the probability of being in state j at time n when starting in state i.

- For an M state DTMC, $\bar{\pi} = (\pi_1, \ldots, \pi_M)$ denotes the limiting distribution.

Limiting probabilities

- Do the limiting probabilities always exist ?
- $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ In this case the limiting probabilities do not exist.
- Now suppose $\bar{\mu} = [.5, .5]$. Then $P(X_1 = 1) = 0.5$. But this is true for every X_n, i.e., $P(X_n = 1) = 0.5$. (already in steady state)
- What is happening ?
- Now suppose $\bar{\mu} = [.1, .9]$. Then $P(X_1 = 1) = 0.9$. But this is **not** true for every X_n, i.e.,
  $P(X_2 = 1) = 0.9, P(X_3 = 1) = 0.1$. (Never in steady-state)

Stationary distribution

- $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is a periodic chain for which the the limiting probabilities do not exist.
- What however always exists is known as stationary distribution (not necessarily unique)
- A **stationary distribution** is a probability (row) vector on $\mathcal{S}$ that satisfies $\pi = \pi P$ in case of DTMC.
- For a CTMC, we know that $\frac{dP(t)}{dt} = P(t)Q$. When $\lim_{t\to\infty} P(t) = \Pi$, this means that at stationarity $\frac{dP(t)}{dt} = 0$. Therefore we have $\pi Q = 0$ in case of CTMC.
- If the limiting distribution exists, it is equal to its stationary distirbution.