Notes for Continuous Time Markov Chains

Xiaoyu Chen

1 Basic Idea

For a Markov chain P, we may think of it as a function $f^P : \mathbb{Z}_{\geq 0} \to \mathbb{R}^{n \times n}$ defined as

$$f^P(t) = P^t$$

. Now, to make P a continuous Markov chain, what we want to achieve is to find a new function $\tilde{f^P} : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n \times n}$ while make sure that $\tilde{f^P}(t) = f^P(t)$ when $t \in \mathbb{Z}_{\geq 0}$.

Note that if

$$\begin{aligned} e^Q &= I + Q + \frac{Q^2}{2!} + \frac{Q^3}{3!} + \cdots \\ &= \sum_{n=0}^{\infty} \frac{Q^n}{n!} \end{aligned}$$

, and if $e^Q = P$, then we could define the function $\tilde{f^P}$ as

$$\tilde{f^P}(t) = e^{tQ}$$

Fact 1.1. $e^{Q_1 + Q_2} = e^{Q1} e^{Q2}$

Proof.

$$\begin{aligned} e^{Q_1+Q_2} &= \sum_{n=0}^{\infty} \frac{(Q_1 + Q_2)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} Q_1^k Q_2^{n-k} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{(n-k)!k!} Q_1^k Q_2^{n-k} \\ &= \sum_{k=0}^{\infty} \frac{Q_1^k}{k!} \sum_{n=k}^{\infty} \frac{Q_2^{n-k}}{(n-k)!} \\ &= e^{Q_1} e^{Q_2} \end{aligned}$$

□

Fact 1.2. *$e^I = e \cdot I$ and thus $e^{t(Q+I)} = e^t e^{tQ}$.*

Fact 1.3. $\frac{\mathrm{d}}{\mathrm{d}t} e^{tQ} = Qe^{tQ} = e^{tQ}Q$

Proof.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} e^{tQ} &= \frac{\mathrm{d}}{\mathrm{d}t} \sum_{n=0}^{\infty} \frac{t^n Q^n}{n!} \\
&= \sum_{n=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t} \frac{t^n Q^n}{n!} \\
&= \sum_{n=1}^{\infty} \frac{n t^{n-1} Q^n}{n!} \\
&= \sum_{n=1}^{\infty} \frac{t^{n-1} Q^{n-1}}{(n-1)!} Q \\
&= e^{tQ} Q = Q e^{tQ}
\end{aligned}$$

□

2 Exponential Distribution

Definition 2.1 (Exponential Distribution)**.** *A random variable $T : \Omega \to [0, \infty)$ has exponential distribution of parameter $\lambda (0 \le \lambda < \infty)$ if*

$$\Pr[T > t] = e^{-\lambda t} \quad \textit{for all } t \ge 0$$

. λ is somethimes called "rate" of the exponential distribution. We write $T \sim E(\lambda)$ for short. If $\lambda > 0$, then T has density function

$$f_T(t) = \lambda e^{-\lambda t}$$

. The mean of T is given by

$$\mathbb{E}(T) = \int_0^{\infty} \Pr[T > t] \mathrm{d}t = 1/\lambda$$

Theorem 2.1 (Memoryless)**.** *A random variable $T : \Omega \to [0, \infty)$ has an exponential distribution iff it has the following memoryless property:*

$$\Pr[T > s + t | T > s] = \Pr[T > t]$$

Proof. $\Rightarrow$: if $T \sim E(\lambda)$, then

$$\Pr[T > s + t | T > s] = \frac{\Pr[T > s + t]}{\Pr[T > s]} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \Pr[T > t]$$

$\Leftarrow$: if T has memoryless property, then let $g(t) = \Pr[T > t]$. First, note that $g(1)$ is a constant, so there is some constant λ such that $g(1) = e^{-\lambda}$. Then, for any rational number p/q, we have

$$g(p/q) = g(1/q)^p = g(1)^{p/q}$$

Finally, note that g is decreasing, so for any real number t, we have $r \leq t \leq s$ (r, s are rational numbers) and $g(r) \geq g(t) \geq g(s)$. Since r and s could be arbitrarily close to t, so we have $g(t) = e^{-\lambda t}$. $\square$

3 Possion Process

Definition 3.1 (Possion Distribution)**.** *A random variable $X : \Omega \to \mathbb{Z}_{\geq 0}$ has Possion distribution of parameter $\lambda (0 \leq \lambda < \infty)$ if*

$$\Pr[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

We write $X \sim P(\lambda)$ for short.

Fact 3.1. *If $X \sim P(\lambda)$, then we have $\lambda = \mathbb{E}(X) = \mathrm{Var}(X)$.*

Proof.

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \cdot k \\ &= \sum_{k=1}^{\infty} \lambda \cdot \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} \\ &= \lambda \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \end{aligned}$$

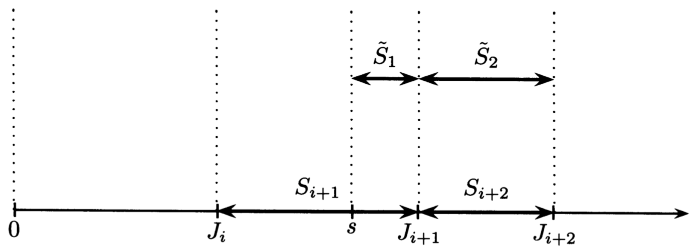
$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \cdot k^2 \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} \cdot k \\ &= \lambda e^{-\lambda} \left[(k-1) \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} \right] \\ &= \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda}) \\ &= \lambda^2 + \lambda \end{aligned}$$

$\square$

Now we can define Possion process as follow.

Definition 3.2 (Possion Process)**.** *Let $T_1, T_2, \cdots$ be i.i.d. exponential random varibles of rate r (see 2.1). Let $S_k := \sum_{i=1}^{k} T_i$ for $k \geq 1$. Let $X_t := k$ for $S_k \leq t < S_{k+1}$. Then, we say $(X_t)_{t \geq 0}$ is a Possion Process of rate r.*

Theorem 3.1 (Memoryless (Markov Property))**.** *If X_t is a Possion process of rate r, then for some constant $s > 0$, $(X_{s+t} - X_s)_{t \geq 0}$ is also a possion process of rate r.*

Proof. 

□

Lemma 3.1. *If X_t is a Possion process of reate r, and let $S_k = \sum_{i=1}^{k} T_i$, then S_k has a gamma distribution with shape parameter k and rate parameter r, i.e. its density function is*

$$f_k(s) = \frac{r^k s^{k-1} e^{-rs}}{(k-1)!}$$

.

Proof. We prove this by induction. As a base case, we have

$$\begin{aligned}
f_2(s) &= \int_{-\infty}^{\infty} f_{T_1}(t) f_{T_2}(s-t)\mathrm{d}t \\
&= \int_0^s re^{-rt} re^{-r(s-t)}\mathrm{d}t \\
&= \int_0^s r^2 e^{-rs}\mathrm{d}t \\
&= r^2 e^{-rs} s
\end{aligned}$$

For the inductive case, we have:

$$\begin{aligned}
f_{k+1}(s) &= \int_0^s f_k(t) f_{T_{k+1}}(s-t)\mathrm{d}t \\
&= \int_0^s \frac{r^k t^{k-1} e^{-rt}}{(k-1)!} re^{-r(s-t)}\mathrm{d}t \\
&= \int_0^s \frac{r^{k+1} t^{k-1} e^{-rs}}{(k-1)!}\mathrm{d}t \\
&= \int_0^s \frac{r^{k+1} e^{-rs}}{k!}\mathrm{d}t^k \\
&= \frac{r^{k+1} s^k e^{-rs}}{k!}
\end{aligned}$$

□

Theorem 3.2 ([LP17], Exercise 20.3)**.** *If $(X_t)_{t\geq 0}$ is a Possion process of rate r, then for any $t, s \geq 0$, $X_{s+t} - X_s$ forms a Possion distribution with parameter rt*

Proof. Due to Theorem 3.1 we only need to show that for a fixed t, $X_t - X_0 = X_t$ forms a Possion distribution. Assume that $S_k = \sum_{i=1}^{k} T_i$.

We prove this by induction.

For the base case, we have $\Pr[X_t = 0] = \Pr[S_1 > t] = \Pr[T_1 > t] = e^{-rt}$. For the inductive case, we have

$$\begin{aligned}
\Pr[X_t = k] = \Pr[S_k \le t < S_{k+1}] &= \Pr[S_k \le t < S_k + T_{k+1}] \\
&= \int_0^t \Pr[S_k = s] \cdot \Pr[T_{k+1} > t - s] \\
&= \int_0^t f_k(s)\mathrm{d}s \cdot \Pr[T_{k+1} > t - s] \\
&= \int_0^t \frac{r^k s^{k-1} e^{-rs}}{(k-1)!}\mathrm{d}s \cdot e^{-r(t-s)} \\
&= \int_0^t \frac{r^k s^{k-1} e^{-rt}}{(k-1)!}\mathrm{d}s \\
&= \frac{(rt)^k e^{-rt}}{k!}
\end{aligned}$$

□

4 Construct CTMC

There are mainly two different but equivalent way to construct a countinuous time Markov chain (CTMC) by using Possion clock. We give details about them one by one.

Definition 4.1 (Possion Clock)**.** *A Possion clock of rate r is a clock that rings at time T, where T is an exponential random variable with rate r.*

4.1 CTMC with a global Possion Clock

Definition 4.2 (CTMC in [LP17])**.** *For a discrete Markov chain P and a Possion clock with rate r, we could make each transition of P happens only in the moment where the clock rings. And thus get a CTMC $(X_t)_{t\ge 0}$.*

Let N_t denotes the number of ringings made by the clock. By Theorem 3.2, we know that once t is fixed to some constant, N_t is a random variable that has Possion distribution with rate rt. And note that

$$\Pr[X_t = y | N_t = k \wedge X_0 = x] = P^k(x, y)$$

. So we have

$$\begin{aligned}
\tilde{P}^t(x,y) &= \sum_{k=0}^{\infty} \Pr[X_t = y | N_t = k \wedge X_0 = x] \cdot \Pr[N_t = k] \\
&= \sum_{k=0}^{\infty} P^k(x,y) \cdot \frac{(rt)^k e^{-rt}}{k!} \\
&= e^{-rt} \sum_{k=0}^{\infty} \frac{(rt)^k P^k(x,y)}{k!} \\
&= e^{-rt} e^{rtP} \\
&= e^{rt(P-I)}
\end{aligned}$$

In [LP17], they also denote $\tilde{P}^t$ as H_t (called ***heat kernel***).

4.2 CTMC with Possion Clocks on Each States

Definition 4.3 (CTMC in [Nor98])**.** *For a discrete Markov chain P, i.e. $(Y_t)_{t\geq 0}$, before its t-th transition the current states, we put a Possion clock on each states of P with different rate. Usually, we have a rate matrix $Q = P - I$. Then, Y_t is determined by the first ringing.*

The following theorem implies that the two definitions we given above are equivalent (some factors may not equal).

Theorem 4.1 ([Nor98], Theorem 2.3.3.)**.** *Let I be a countable set and let T_k, $k \in I$, be independet random variables with $T_k \sim E(q_k)$ and $0 < q := \sum_{k\in I} q_k < \infty$. Set $T = \inf_k T_k$. Then this infimum is attained at a unique random value K of k, with probability 1. Moreover, T and K are independent, with $T \sim E(q)$ and $\Pr[K = k] = q_k/q$.*

Proof. Set $K = k$ if $T_k < T_j$ for all $j \neq k$, otherwise let K be undefined. Then

$$\begin{aligned}
\Pr[K = k \wedge T \geq t] &= \Pr[T_k \geq t \wedge T_j > T_k, \forall j \neq k] \\
&= \int_t^{\infty} q_k e^{-q_k s} \Pr[T_j > s, \forall j \neq k] \mathrm{d}s \\
&= \int_t^{\infty} q_k e^{-q_k s} \prod_{j\neq k} e^{-q_j s} \mathrm{d}s \\
&= \int_t^{\infty} q_k e^{-qs} \mathrm{d}s = \frac{q_k}{q} e^{-qt}
\end{aligned}$$

Hence $\Pr[K = k \text{ for some } k] = 1$ and T and K have the claimed joint distribution. ∎

5 Different Mixing Times

Definition 5.1 (ℓ^p distance)**.** *For any function $f : \Omega \to \mathbb{R}$, we have:*

$$\| f \|_{p,\pi} := \left(\sum_{x \in \Omega} \pi(x) |f(x)|^p \right)^{1/p}$$

Definition 5.2 (An Entropy-like Measure)**.**

$$\mathrm{Ent}_\pi(f) := \mathbb{E}_\pi[f \log f] - (\mathbb{E}_\pi f) \log \mathbb{E}_\pi f$$

Note that, if $\mathbb{E}_\pi f = 1$, we have

$$\mathrm{Ent}_\pi(f) = \mathbb{E}_\pi[f \log f]$$

There are many ways to measure the distance between $P^t(x, \cdot)$ and π.

Definition 5.3. *For a discrete Markov chain P, let $k_t^x(y) := P^t(x, y)/\pi(y)$.*

Fact 5.1.

$$\| P^t(x, \cdot) - \pi \|_{TV} = \frac{1}{2} \| k_t^x - 1 \|_{1,\pi}$$

Fact 5.2.

$$\mathrm{Var}_\pi(k_t^x) = \| k_t^x - 1 \|_{2,\pi}$$

Fact 5.3.

$$D(P^t(x, \cdot) \| \pi) = \sum_{y \in \Omega} \pi(y) \frac{P^t(x, y)}{\pi(y)} \log \frac{P^t(x, y)}{\pi(y)} = \mathrm{Ent}_\pi(k_t^x)$$

Definition 5.4 (Some Mixing Times)**.**

$$\begin{aligned} \tau(\varepsilon) &= \min\{n : \forall x \in \Omega, \| p^n(x, \cdot) - \pi \|_{TV} \leq \varepsilon\} \\ \tau_D(\varepsilon) &= \min\{n : \forall x \in \Omega, D(p^n(x, \cdot) \| \pi) \leq \varepsilon\} \\ \tau_2(\varepsilon) &= \min\{n : \forall x \in \Omega, \| p^n(x, \cdot) - \pi \|_{2,\pi} \leq \varepsilon\} \end{aligned}$$

Fact 5.4 ([LP17], Exercise 4.5)**.**

$$\| P^t(x, \cdot) - \pi \|_{TV} \leq \mathrm{Var}_\pi(k_t^x)$$

, and thus $\tau_2(\varepsilon) \geq \tau(\varepsilon)$.

Fact 5.5 ([JC05], Section 9.7; Pinsker's Ineq)**.**

$$2 \| P^t(x, \cdot) - \pi \|_{TV}^2 \leq D(P^t(x, \cdot) \| \pi) = \mathrm{Ent}_\pi(k_t^x)$$

6 Bounds in Continuous Time

See Section 1.2 of [MT06] for more details.

References

[JC05] Mark Jerrum and Sampling Counting. Integrating: Algorithms and complexity, lectures in mathematics, eth zürich. chapter 9. Birkhauser Verlag, Basel, 2005. http://www.maths.qmul.ac.uk/~mj/ETHbook/chapter9.pdf.

[LP17] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[MT06] Ravi R Montenegro and Prasad Tetali. *Mathematical aspects of mixing times in Markov chains*. Now Publishers Inc, 2006.

[Nor98] James R Norris. *Markov chains*. Number 2. Cambridge university press, 1998.