Markov chain mixing time		The (separation) cutoff phenomenon
	Lecture 3-4	
Lecturer: Michael Choi		*Scribe: Michael Choi*

1 Goal of this lecture

In our first lecture, we introduce the notion of total variation mixing time. One popular alternative in the literature, apart from the total variation distance, is known as the separation distance. We shall study basic properties of the separation distance, and discuss a phenomenon known as cutoff phenomenon that appears in studying the rate of convergence of a sequence of Markov chains. Most of the material is taken from Diaconis and Saloff-Coste (2006); Levin et al. (2009).

2 Separation distance and strong stationary time

Let (X_t) be an irreducible Markov chain with stationary distribution π, where the Markov chain run in discrete time. Suppose that $(\mathcal{F}_t)$ is a filtration, and (X_t) is adapted to $(\mathcal{F}_t)$. A **stationary time** τ for (X_t) is a $(\mathcal{F}_t)$-stopping time, possibly depending on the starting position x, such that the distribution of X_τ is π:

$$\mathbb{P}_x(X_\tau = y) = \pi(y), \quad \text{for all } y.$$

Example 1 *Let ξ be a $\mathcal{X}$-valued random variable with distribution π, and define*

$$\tau := \min\{t \geq 0;\ X_t = \xi\}.$$

Let $\mathcal{F}_t = \sigma(\xi, (X_s)_{0 \leq s \leq t})$. The time τ is a $(\mathcal{F}_t)$-stopping time, and because $X_\tau = \xi$, τ is a stationary time.

Suppose that the chain starts at x_0. Then $\tau = 0$ implies $X_\tau = x_0$; therefore, τ and X_τ are not independent.

A **strong stationary time** for (X_t) and starting position x is an $(\mathcal{F}_t)$-stopping time τ, such that for all times t and all y,

$$\mathbb{P}_x(\tau = t, X_\tau = y) = \mathbb{P}_x(\tau = t)\pi(y).$$

In words, X_τ has distribution π and is independent of τ.

Remark 1 *If τ is a strong stationary time starting from x, then*

$$\begin{aligned}
\mathbb{P}_x(\tau \leq t, X_t = y) &= \sum_{s \leq t} \sum_z \mathbb{P}_x(\tau = s, X_s = z, X_t = y) \\
&= \sum_{s \leq t} \sum_z P^{t-s}(z, y)\mathbb{P}_x(\tau = s)\pi(z) \\
&= \mathbb{P}_x(\tau \leq t)\pi(y).
\end{aligned}$$

We can now introduce the separation distance. The **separation distance** is defined to be, for $x \in \mathcal{X}$,

$$\begin{aligned} sep(\mu, \nu) &:= \max_{y \in \mathcal{X}} \left[1 - \frac{\mu(y)}{\nu(y)} \right], \\ s_x(t) &:= sep(P^t(x, \cdot), \pi) = \max_{y \in \mathcal{X}} \left[1 - \frac{P^t(x,y)}{\pi(y)} \right], \\ s(t) &:= \max_{x \in \mathcal{X}} s_x(t). \end{aligned}$$

Note that separation distance is not symmetric and is not a distance between probability measures.

Lemma 1 (Relationship between strong stationary time and separation distance) *If τ is a strong stationary time for starting state x, then*

$$s_x(t) \le \mathbb{P}_x(\tau > t).$$

Proof: Fix $x \in \mathcal{X}$. For every $y \in \mathcal{X}$,

$$\begin{aligned} 1 - \frac{P^t(x,y)}{\pi(y)} &= 1 - \frac{\mathbb{P}_x\left(X_t = y\right)}{\pi(y)} \\ &\le 1 - \frac{\mathbb{P}_x\left(X_t = y, \tau \le t\right)}{\pi(y)} \\ &= 1 - \frac{\pi(y)\mathbb{P}_x(\tau \le t)}{\pi(y)} \\ &= \mathbb{P}_x(\tau > t), \end{aligned}$$

where the second equality follows from Remark 1. □

Lemma 2 (Relationship between total variation distance and separation distance) *The separation distance $s_x(t)$ satisfies*

$$\left\| P^t(x, \cdot) - \pi \right\|_{\mathrm{TV}} \le s_x(t),$$

and therefore $d(t) \le s(t)$.

Proof: By Proposition 1 item 2 in lecture 1, we have

$$\begin{aligned} \left\| P^t(x, \cdot) - \pi \right\|_{\mathrm{TV}} = \sum_{P^t(x,y) < \pi(y)} \left[\pi(y) - P^t(x,y) \right] &= \sum_{y \in \mathcal{X}} \pi(y) \left[1 - \frac{P^t(x,y)}{\pi(y)} \right] \\ &\le \max_{y \in \mathcal{X}} \left[1 - \frac{P^t(x,y)}{\pi(y)} \right] = s_x(t). \end{aligned}$$

□

Combining Lemma 1 and Lemma 2, we have

Corollary 3

$$\left\|P^t(x,\cdot)-\pi\right\|_{\mathrm{TV}} \le \mathbb{P}_x(\tau > t).$$

Given starting state x, a state y is a **halting state** for a stopping time τ if $X_t = y$ implies $\tau \le t$.

Proposition 4 (Fastest strong stationary time) *If there exists a halting state for starting state x associated with a strong stationary time τ, then τ is the fastest strong stationary time, i.e.*

$$s_x(t) = \mathbb{P}_x(\tau > t).$$

In words, any other strong stationary time stochastically dominates τ under $\mathbb{P}_x$.

Proof: We repeat the proof in Lemma 1, except that we have equality throughout the proof.

$$\begin{aligned}
1-\frac{P^t(x,y)}{\pi(y)} &= 1-\frac{\mathbb{P}_x\left(X_t = y\right)}{\pi(y)} \\
&= 1-\frac{\mathbb{P}_x\left(X_t = y, \tau \le t\right)}{\pi(y)} \\
&= 1-\frac{\pi(y)\mathbb{P}_x(\tau \le t)}{\pi(y)} \\
&= \mathbb{P}_x(\tau > t),
\end{aligned}$$

□

3 Fastest strong stationary time of the birth-death process

In the previous section, we have seen that the the total variation distance and separation distance are closely related to the fastest strong stationary time. It is therefore useful to obtain the distribution (and hence the tail probability) of the fastest strong stationary time in order to bound the mixing time. For a class of Markov chains known as birth-death processes, we shall see that the fastest strong stationary time is tractable and distributed as a convolution of exponential distributions.

A Markov chain on $\mathcal{X} = \{0,1,\ldots,m\}$ is said to be a **birth-death** chain if $P(x,y) = 0$ unless $|x-y| \le 1$. It can be checked easily that birth-death chains are reversible. In the following, it is perhaps more convenient for us to work in cotinuous-time, and all our previous results can be generalized to continuous-time in a similar fashion. Let $G := P - I$ be the generator, and $P^t := e^{Gt}$ be the continuized Markov chain for $t \ge 0$. For ergodic P, the eigenvalues of $-G$ are arranged in ascending order as $\lambda_0(-G) = 0 < \lambda_1(-G) \le \lambda_2(-G) \le \ldots \le \lambda_m(-G) \le 2$. We state without proof that the fastest strong stationary time of birth-death process starting at 0 is a convolution of exponential distributions:

Theorem 5 (Fastest strong stationary time of birth-death process starting at 0**)** *Let G be the infinitesimal generator of a continuous-time ergodic birth-death process on $\mathcal{X} = \{0,1,\ldots,m\}$. Then*

$$s_0(t) = \mathbb{P}_0(\tau > t),$$

where $\tau = \sum_{i=1}^{m} S_i$, and each S_i is independent exponential random variable with parameter $\lambda_i(-G)$. In particular,

$$\mathbb{E}_0(\tau) = \sum_{i=1}^{m} \frac{1}{\lambda_i(-G)}, \quad \mathrm{Var}_0(\tau) = \sum_{i=1}^{m} \frac{1}{\lambda_i(-G)^2}.$$

4 The separation cutoff phenomenon

Suppose that we have a sequence of Markov chains, indexed by n, on state space $\mathcal{X}_n$, transition semigroup $(P_n^t)_{t\geq 0}$ and stationary distribution π_n. These Markov chains may exhibit abrupt convergence to stationarity, known as the cutoff phenomenon.

Definition 1 (Separation cutoff) *A family of Markov chains $(\mathcal{X}_n, \pi_n, (\mu_n^t)_{t\geq 0})_{n=1,2,\ldots}$, where $\mu_n^t = P_n^t(x_n, \cdot)$, exhibit separation cutoff if there exists a sequence (t_n) of positive reals such that for any $\epsilon \in (0,1)$,*

$$\lim_{n\to\infty} sep(\mu_n^{(1+\epsilon)t_n}, \pi_n) = 0,$$
$$\lim_{n\to\infty} sep(\mu_n^{(1-\epsilon)t_n}, \pi_n) = 1.$$

Remark 2 *Clearly this definition can be generalized to other notions of cutoff, where we replace the separation distance by appropriate distance such as total variation or L^2 distance.*

We now study the separation cutoff phenomenon for birth-death processes. We have a family of continuous-time birth-death processes with generators $G_n = P_n - I_n$ on $\mathcal{X}_n = \{0, 1, \ldots, m_n\}$. Let τ_n be the fastest strong stationary time of the n-th chain starting at 0, and define t_n and σ_n^2 to be the mean and variance of τ_n under $\mathbb{P}_0$:

$$t_n := \mathbb{E}_0(\tau_n) = \sum_{i=1}^{m_n} \frac{1}{\lambda_i(-G_n)}, \quad \sigma_n^2 := \mathrm{Var}_0(\tau_n) = \sum_{i=1}^{m_n} \frac{1}{\lambda_i(-G_n)^2}.$$

Theorem 6 *Suppose that we have a family of continuous-time birth-death processes with generators $G_n = P_n - I_n$ on $\mathcal{X}_n = \{0, 1, \ldots, m_n\}$. Let τ_n be the fastest strong stationary time of the n-th chain starting at 0, and define t_n and σ_n^2 to be the mean and variance of τ_n under $\mathbb{P}_0$. Define $\lambda_{1,n} := \lambda_1(-G_n)$ be the spectral gap and $N_n := \lambda_{1,n} t_n$, then if $N_n \to \infty$,*

$$\lim_{n\to\infty} sep(P_n^{(1+\epsilon)t_n}(0,\cdot), \pi_n) = 0,$$
$$\lim_{n\to\infty} sep(P_n^{(1-\epsilon)t_n}(0,\cdot), \pi_n) = 1.$$

Remark 3 *Indeed $N_n \to \infty$ is necessary and sufficient condition for separation cutoff of birth-death processes, see Diaconis and Saloff-Coste (2006).*

Before we begin the proof, we first state a version of Chebyshev's inequality applied on τ_n:

Lemma 7 (Chebyshev's inequality) *Suppose that we are in the setting of Theorem 6. Then for $a > 0$,*

$$\mathbb{P}_0\left(\tau_n > t_n + a\sigma_n\right) \le \frac{1}{1+a^2}, \quad \mathbb{P}_0\left(\tau_n < t_n - a\sigma_n\right) \le \frac{1}{1+a^2}.$$

Proof of Theorem 6: First, we note that

$$\begin{aligned}\sigma_n^2 = \sum_{i=1}^{m_n} \frac{1}{\lambda_i(-G_n)^2} &= \lambda_{1,n}^{-2} \sum_{i=1}^{m_n} \frac{\lambda_{1,n}^2}{\lambda_i(-G_n)^2} \\ &\le \lambda_{1,n}^{-2} \sum_{i=1}^{m_n} \frac{\lambda_{1,n}}{\lambda_i(-G_n)} = \lambda_{1,n}^{-1} t_n,\end{aligned}$$

where we use $\lambda_{1,n}/\lambda_i(-G_n) \le 1$ in the inequality. As a result, $\sigma_n \le N_n^{-1/2} t_n$. Now, by Lemma 7 and Theorem 5,

$$\begin{aligned}sep(P_n^{(1+\epsilon)t_n}(0,\cdot),\pi_n) &= \mathbb{P}_0(\tau_n > (1+\epsilon)t_n) \\ &= \mathbb{P}_0(\tau_n > t_n + \epsilon N_n^{1/2} N_n^{-1/2} t_n) \\ &\le \mathbb{P}_0(\tau_n > t_n + \epsilon N_n^{1/2}\sigma_n) \\ &\le \frac{1}{1+\epsilon^2 N_n} \to 0.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{P}_0(\tau_n < (1-\epsilon)t_n) &= \mathbb{P}_0(\tau_n < t_n - \epsilon N_n^{1/2} N_n^{-1/2} t_n) \\ &\le \mathbb{P}_0(\tau_n < t_n - \epsilon N_n^{1/2}\sigma_n) \\ &\le \frac{1}{1+\epsilon^2 N_n} \to 0.\end{aligned}$$

This yields $sep(P_n^{(1-\epsilon)t_n}(0,\cdot),\pi_n) \to 1$ as $n \to \infty$. □

References

P. Diaconis and L. Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006.

D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009.