

COMPARISON OF CUTOFFS BETWEEN LAZY WALKS AND MARKOVIAN SEMIGROUPS

GUAN-YU CHEN,* *National Chiao Tung University*

LAURENT SALOFF-COSTE,** *Cornell University*

Abstract

We make a connection between the continuous time and lazy discrete time Markov chains through the comparison of cutoffs and mixing time in total variation distance. For illustration, we consider finite birth and death chains and provide a criterion on cutoffs using eigenvalues of the transition matrix.

Keywords: Markov chain; cutoff phenomenon

2010 Mathematics Subject Classification: Primary 60J10

Secondary 60J27

1. Introduction

Let Ω be a countable set and (Ω, K, π) be an irreducible Markov chain on Ω with transition matrix K and stationary distribution π . Let

$$H_t = e^{-t(I-K)} = \sum_{i=0}^{\infty} \frac{e^{-t} t^i K^i}{i!}$$

be the associated semigroup which describes the corresponding natural continuous time process on Ω . For $\delta \in (0, 1)$, set

$$K_\delta = \delta I + (1 - \delta)K, \tag{1.1}$$

where I is the identity matrix indexed by Ω . Clearly, K_δ is similar to K but with an additional holding probability depending of δ . We call K_δ the δ -lazy walk or δ -lazy chain of K . It is well known that if K is irreducible with stationary distribution π , then

$$\lim_{m \rightarrow \infty} K_\delta^m(x, y) = \lim_{t \rightarrow \infty} H_t(x, y) = \pi(y) \quad \text{for all } x, y \in \Omega, \delta \in (0, 1).$$

In this paper, we consider convergence in total variation. The total variation between two probabilities μ, ν on Ω is defined by $\|\mu - \nu\|_{\text{TV}} = \sup\{\mu(A) - \nu(A) \mid A \subset \Omega\}$. For any irreducible K with stationary distribution π , the (maximum) total variation distance is defined by

$$d_{\text{TV}}(m) = \sup_{x \in \Omega} \|K^m(x, \cdot) - \pi\|_{\text{TV}}, \tag{1.2}$$

and the corresponding mixing time is given by

$$T_{\text{TV}}(\varepsilon) = \inf\{m \geq 0 \mid d_{\text{TV}}(m) \leq \varepsilon\}.$$

Received 15 June 2012; revision received 12 December 2012.

* Postal address: Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan.

Email address: gychen@math.nctu.edu.tw

** Postal address: Malott Hall, Department of Mathematics, Cornell University, Ithaca, NY 14853-4201.

Email address: lsc@math.cornell.edu

We write the total variation distance and mixing time as $d_{TV}^{(c)}$ and $T_{TV}^{(c)}$ respectively for the continuous semigroup and as $d_{TV}^{(\delta)}$, $T_{TV}^{(\delta)}$ respectively for the δ -lazy walk.

A sharp transition phenomenon, known as cutoff, was introduced by Aldous and Diaconis in early 1980s. See, e.g. [5] and [8] for an introduction and a general review of cutoffs. In total variation, a family of irreducible Markov chains $(\Omega_n, K_n, \pi_n)_{n=1}^\infty$ is said to present a cutoff if

$$\lim_{n \rightarrow \infty} \frac{T_{n,TV}(\varepsilon)}{T_{n,TV}(\eta)} = 1 \quad \text{for all } 0 < \varepsilon < \eta < 1. \tag{1.3}$$

The family is said to present a (t_n, b_n) cutoff if $b_n = o(t_n)$ and

$$|T_{n,TV}(\varepsilon) - t_n| = O(b_n) \quad \text{for all } 0 < \varepsilon < 1.$$

The cutoff for the associated continuous semigroups is defined in a similar way. This paper contains the following general result.

Theorem 1.1. *Consider a family of irreducible and positive recurrent Markov chains $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) \mid n = 1, 2, \dots\}$. For $\delta \in (0, 1)$, let \mathcal{F}_δ be the family of associated δ -lazy walks and let \mathcal{F}_c be the family of associated continuous semigroups. Suppose $T_{n,TV}^{(c)}(\varepsilon_0) \rightarrow \infty$ for some $\varepsilon_0 \in (0, 1)$. Then, the following are equivalent.*

- (a) \mathcal{F}_δ has a cutoff in total variation.
- (b) \mathcal{F}_c has a cutoff in total variation.

Furthermore, if \mathcal{F}_c has a cutoff, then

$$\lim_{n \rightarrow \infty} \frac{T_{n,TV}^{(c)}(\varepsilon)}{T_{n,TV}^{(\delta)}(\varepsilon)} = 1 - \delta \quad \text{for all } \varepsilon \in (0, 1).$$

Theorem 1.2. *Let \mathcal{F} be the family in Theorem 1.1. Assume that $t_n \rightarrow \infty$. Then, the following are equivalent.*

- (a) \mathcal{F}_c has a (t_n, b_n) cutoff.
- (b) For $\delta \in (0, 1)$, \mathcal{F}_δ has a $(t_n/(1 - \delta), b_n)$ cutoff.

We refer the reader to Theorems 3.1, 3.2, 3.3, and 3.4 for more detailed discussions. In a private communication with Y. Peres and D. Levin, we understand that a similar approach to the proof of Theorems 1.1 and 1.2 is also used in [12] to study the same comparison issue. See Proposition 3.1 and [12, Theorem 20.3] for details.

For an illustration, we consider finite birth and death chains. For $n \geq 1$, let $\Omega_n = \{0, 1, \dots, n\}$ and K_n be the transition kernel of a birth and death chain on Ω_n with birth rate $p_{n,i}$, death rate $q_{n,i}$, and holding rate $r_{n,i}$, where $p_{n,n} = q_{n,0} = 0$ and $p_{n,i} + q_{n,i} + r_{n,i} = 1$. Suppose that K_n is irreducible with stationary distribution π_n . For the family $\{(\Omega_n, K_n, \pi_n) \mid n = 1, 2, \dots\}$, Ding *et al.* [10] showed that, in the discrete time case, if $\inf_{i,n} r_{n,i} > 0$, then the cutoff in total variation exists if and only if the product of the total variation mixing time and the spectral gap, which is defined to be the smallest nonzero eigenvalue of $I - K$, tends to infinity. There is also a similar version for the continuous time case. The next theorem, which is a summary of Theorem 4.8, is an application of the above result and Theorem 1.1.

Theorem 1.3. Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) \mid n = 1, 2, \dots\}$ be a family of irreducible birth and death chains as above. For $n \geq 1$, let $0, \lambda_{n,1}, \dots, \lambda_{n,n}$ be eigenvalues of $I - K_n$ and set

$$\lambda_n = \min_{1 \leq i \leq n} \lambda_{n,i}, \quad s_n = \sum_{i=1}^n \lambda_{n,i}^{-1}.$$

Then, the following are equivalent.

- (a) \mathcal{F}_c has a total variation cutoff.
- (b) For $\delta \in (0, 1)$, \mathcal{F}_δ has a total variation cutoff.
- (c) $s_n \lambda_n \rightarrow \infty$.

The remainder of this article is organized as follows. In Section 2, the concepts of cutoffs and mixing times are introduced and fundamental results are reviewed. In Section 3, a detailed comparison of the cutoff time and window size is made between the continuous time and lazy discrete time cases, where the state space is allowed to be infinite. In Section 4, we focus on finite birth and death chains and provide a criterion on total variation cutoffs using the eigenvalues of the transition matrices.

2. Cutoffs in total variation

Throughout this paper, for any two sequences s_n and t_n of positive numbers, we write $s_n = O(t_n)$ if there exists $C > 0$ and $N > 0$ satisfying $|s_n| \leq C|t_n|$ for all $n \geq N$. If $s_n = O(t_n)$ and $t_n = O(s_n)$, we write $s_n \asymp t_n$. If $t_n/s_n \rightarrow 1$ as $n \rightarrow \infty$, we write $t_n \sim s_n$.

Consider the following definitions.

Definition 2.1. Referring to the notation in (1.2), a family $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) \mid n = 1, 2, \dots\}$ is said to present a total variation

- (a) precutoff if there is a sequence t_n and $B > A > 0$ satisfying

$$\lim_{n \rightarrow \infty} d_{n,\text{TV}}(\lceil Bt_n \rceil) = 0, \quad \liminf_{n \rightarrow \infty} d_{n,\text{TV}}(\lfloor At_n \rfloor) > 0;$$

- (b) cutoff if there is a sequence t_n satisfying, for all $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} d_{n,\text{TV}}(\lceil (1 + \varepsilon)t_n \rceil) = 0, \quad \lim_{n \rightarrow \infty} d_{n,\text{TV}}(\lfloor (1 - \varepsilon)t_n \rfloor) = 1;$$

- (c) (t_n, b_n) cutoff if $b_n = o(t_n)$ and

$$\lim_{c \rightarrow \infty} \overline{F}(c) = 0, \quad \lim_{c \rightarrow -\infty} \underline{F}(c) = 1,$$

where

$$\overline{F}(c) = \limsup_{n \rightarrow \infty} d_{n,\text{TV}}(\lceil t_n + cb_n \rceil), \quad \underline{F}(c) = \liminf_{n \rightarrow \infty} d_{n,\text{TV}}(\lfloor t_n + cb_n \rfloor).$$

In definition 2.1, t_n is called a cutoff time and b_n is called a window for t_n . The cutoffs for continuous semigroups are the same except for the deletion of $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$.

Remark 2.1. In Definition 2.1, if $t_n \rightarrow \infty$ (or equivalently $T_{n,\text{TV}}(\varepsilon) \rightarrow \infty$ for some $\varepsilon \in (0, 1)$), then the cutoff is consistent with (1.3). This is also true for cutoffs in continuous semigroups without the assumption $t_n \rightarrow \infty$.

The following lemma characterizes the total variation convergence using specific subsequences of indices and events, which is useful in proving and disproving cutoffs.

Lemma 2.1. *Consider a family of irreducible and positive recurrent Markov chains $\{(\Omega_n, K_n, \pi_n) \mid n = 1, 2, \dots\}$. Let t_n be a sequence of nonnegative integers. Then, the following are equivalent.*

- (a) $d_{n,\text{TV}}(t_n) \rightarrow 0$.
- (b) *For any increasing sequence of positive integers n_k , any $A_{n_k} \subset \Omega_{n_k}$, and any $x_{n_k} \in \Omega_{n_k}$, there is a subsequence m_k for which*

$$\lim_{k \rightarrow \infty} |K_{m_k}^{t_{m_k}}(x_{m_k}, A_{m_k}) - \pi_{m_k}(A_{m_k})| = 0.$$

Proof of Lemma 2.1. (a) implies (b) is obvious. For (b) implies (a), choose $A_n \subset \Omega_n$ and $x_n \in \Omega_n$ satisfying $d_{n,\text{TV}}(t_n) \leq 2|K_n^{t_n}(x_n, A_n) - \pi_n(A_n)|$. Let n_k be an increasing sequence of positive integers and choose a subsequence m_k for which

$$\lim_{k \rightarrow \infty} |K_{m_k}^{t_{m_k}}(x_{m_k}, A_{m_k}) - \pi_{m_k}(A_{m_k})| = 0.$$

This implies $d_{m_k,\text{TV}}(t_{m_k}) \rightarrow 0$, as desired.

Remark 2.2. Lemma 2.1 also holds in continuous time by allowing t_n to take values in the positive real numbers. See [4] and [5] for further discussions on cutoffs.

3. Comparisons of cutoffs

In this section, we establish the relationship between the cutoffs of lazy walks and continuous semigroups. Let Ω be a countable set and K be a transition matrix indexed by Ω . In the notation of (1.1), the δ -lazy walk evolves in accordance with

$$(K_\delta)^t = \sum_{i=0}^t \binom{t}{i} \delta^{t-i} (1-\delta)^i K^i \quad \text{for all } \delta \in (0, 1), t \geq 0,$$

whereas the continuous time chain follows

$$H_t = e^{-t(I-K)} = \sum_{i=0}^{\infty} \left(e^{-t} \frac{t^i}{i!} \right) K^i.$$

Observe that $I - K = (I - K_\delta)/(1 - \delta)$. This implies that

$$d_{\text{TV}}^{(c)}(t) \leq e^{-t/(1-\delta)} \sum_{i=0}^m \frac{[t/(1-\delta)]^i}{i!} + d_{\text{TV}}^{(\delta)}(m).$$

Concerning the cutoff times and windows, we discuss each of them in detail.

3.1. Cutoff times

Theorem 3.1. *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) \mid n = 1, 2, \dots\}$ be a family of irreducible Markov chains on countable state spaces with stationary distributions. For $\delta \in (0, 1)$, let $\mathcal{F}_\delta = \{(\Omega_n, K_{n,\delta}, \pi_n) \mid n = 1, 2, \dots\}$ and $\mathcal{F}_c = \{(\Omega_n, H_{n,t}, \pi_n) \mid n = 1, 2, \dots\}$. Suppose there is $\varepsilon_0 > 0$ for which $T_{n,\text{TV}}^{(\delta)}(\varepsilon_0) \rightarrow \infty$ or $T_{n,\text{TV}}^{(c)}(\varepsilon_0) \rightarrow \infty$. Then, the following are equivalent.*

- (a) \mathcal{F}_δ has a cutoff (respectively precutoff) in total variation.
- (b) \mathcal{F}_c has a cutoff (respectively precutoff) in total variation.

Furthermore, if \mathcal{F}_c has a cutoff, then

$$\lim_{n \rightarrow \infty} \frac{T_{n,TV}^{(c)}(\varepsilon)}{T_{n,TV}^{(\delta)}(\varepsilon)} = 1 - \delta \quad \text{for all } \varepsilon \in (0, 1).$$

The above theorem is in fact a simple corollary of the following proposition.

Proposition 3.1. *Let $\mathcal{F}_\delta, \mathcal{F}_c$ be the families in Theorem 3.1 and t_n, r_n be sequences tending to infinity. Fix $\delta \in (0, 1)$.*

- (a) *If $d_{n,TV}^{(\delta)}(\lceil t_n \rceil) \rightarrow 0$, then*

$$\lim_{n \rightarrow \infty} d_{n,TV}^{(c)}((1 - \delta)t_n + cb_n) = 0$$

for all $c > 0$ and for any sequence b_n satisfying $\sqrt{t_n} = o(b_n)$.

- (b) *If $d_{n,TV}^{(c)}(r_n) \rightarrow 0$, then*

$$\lim_{n \rightarrow \infty} d_{n,TV}^{(\delta)}\left(\left\lceil \frac{r_n}{1 - \delta} + cb_n \right\rceil\right) = 0$$

for all $c > 0$ and for any sequence b_n satisfying $\sqrt{r_n} = o(b_n)$.

- (c) *If $d_{n,TV}^{(c)}(r_n) \rightarrow 1$, then*

$$\lim_{n \rightarrow \infty} d_{n,TV}^{(\delta)}\left(\left\lfloor \frac{r_n}{1 - \delta} \right\rfloor\right) = 1.$$

- (d) *If $d_{n,TV}^{(\delta)}(\lfloor t_n \rfloor) \rightarrow 1$, then*

$$\lim_{n \rightarrow \infty} d_{n,TV}^{(c)}((1 - \delta)t_n) = 1.$$

Proof. We prove (a); the proof of (b) is similar and is therefore omitted. Suppose that $d_{n,TV}^{(\delta)}(\lceil t_n \rceil) \rightarrow 0$. Since $\sqrt{t_n} = o(b_n)$, it is clear that

$$\lim_{n \rightarrow \infty} d_{n,TV}^{(\delta)}(\lceil t_n + cb_n + c'\sqrt{t_n} \rceil) = 0 \quad \text{for all } c > 0, c' \in \mathbb{R}. \tag{3.1}$$

Fix $c > 0$ and let $x_n \in \Omega_n$ and $A_n \subset \Omega_n$. Given any increasing sequence n_l , we may choose, according to Lemma 3.1, a subsequence m_l for which $\pi_{m_l}(A_{m_l}) \rightarrow \alpha \in [0, 1]$ and, for all $c' \in \mathbb{R}$,

$$\lim_{l \rightarrow \infty} K_{m_l, \delta}^{\lceil t_{m_l} + cb_{m_l} + c'\sqrt{t_{m_l}} \rceil}(x_{m_l}, A_{m_l}) = \frac{1}{\sqrt{2\pi\delta}} \int_{-\infty}^{\infty} e^{-(x - \sqrt{1 - \delta}c')^2 / (2\delta)} f(x) \, dx$$

and

$$\lim_{l \rightarrow \infty} H_{m_l, (1 - \delta)(t_{m_l} + cb_{m_l})}(x_{m_l}, A_{m_l}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} f(x) \, dx,$$

where f is nonnegative and bounded by 1. Note that $f(x)$ is not exactly the function in Lemma 3.2 and, in fact, should be replaced by $f(\sqrt{1 - \delta}x)$. As the dilation $f(\sqrt{1 - \delta}x)$ of

$f(x)$ does not matter in this proof, we simply write $f(x)$ for it. By (3.1) and Lemma 3.2, f is equal to α almost everywhere and, by Lemma 2.1, this implies that $d_{n,\text{TV}}^{(c)}((1 - \delta)t_n + cb_n) \rightarrow 0$ as $n \rightarrow \infty$ for all $c > 0$.

The proofs for (c) and (d) are similar and we only give the details for (4). First, we choose sequences $x_n \in \Omega_n$ and $A_n \subset \Omega_n$ satisfying

$$\lim_{n \rightarrow \infty} \pi_n(A_n) = 1, \quad \lim_{n \rightarrow \infty} K_{n,\delta}^{\lfloor t_n \rfloor}(x_n, A_n) = 0.$$

Let n_l be a sequence tending to ∞ . Applying Lemma 3.1 with $c = 0$ and $a_{n,m} = K_n^m(x_n, A_n)$, we may choose a subsequence, say m_l , for which

$$\lim_{l \rightarrow \infty} H_{m_l, (1-\delta)t_{m_l}}(x_{m_l}, A_{m_l}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} g(x) dx$$

and

$$\lim_{l \rightarrow \infty} K_{m_l, \delta}^{\lfloor t_{m_l} \rfloor}(x_{m_l}, A_{m_l}) = \frac{1}{\sqrt{2\pi} \delta} \int_{-\infty}^{\infty} e^{-x^2/(2\delta)} g(x) dx,$$

where g is a nonnegative measurable function bounded by 1. This leads to $g = 0$ almost everywhere and

$$\lim_{l \rightarrow \infty} d_{m_l, \text{TV}}^{(c)}((1 - \delta)t_{m_l}) = 1.$$

The following is a simple corollary of Proposition 3.1 (a) and (b).

Corollary 3.1. *Let \mathcal{F}_δ and \mathcal{F}_c be the families in Theorem 3.1 and t_n, r_n be sequences tending to infinity. Fix $\delta \in (0, 1)$.*

(a) *If $d_{n,\text{TV}}^{(\delta)}(\lceil t_n \rceil) \rightarrow 0$, then*

$$\lim_{n \rightarrow \infty} d_{n,\text{TV}}^{(c)}((1 + \varepsilon)(1 - \delta)t_n) = 0 \quad \text{for all } \varepsilon > 0.$$

(b) *If $d_{n,\text{TV}}^{(c)}(r_n) \rightarrow 0$, then*

$$\lim_{n \rightarrow \infty} d_{n,\text{TV}}^{(\delta)}\left(\left\lceil \frac{(1 + \varepsilon)r_n}{1 - \delta} \right\rceil\right) = 0 \quad \text{for all } \varepsilon > 0.$$

Proof of Theorem 3.1. Set $r_n = T_{n,\text{TV}}^{(\delta)}(\varepsilon_0)$ and $s_n = T_{n,\text{TV}}^{(c)}(\varepsilon_0)$. Suppose that $r_n \rightarrow \infty$. By Corollary 3.1 (b), if

$$\liminf_{n \rightarrow \infty} d_{n,\text{TV}}^{(c)}\left(\frac{(1 - \delta)r_n}{2}\right) = 0,$$

then

$$\liminf_{n \rightarrow \infty} d_{n,\text{TV}}^{(\delta)}\left(\left\lceil \frac{(1 + \varepsilon)r_n}{2} \right\rceil\right) = 0 \quad \text{for all } \varepsilon > 0.$$

But, taking $\varepsilon = \frac{1}{2}$ implies that, for large enough n ,

$$d_{n,\text{TV}}^{(\delta)}\left(\left\lceil \frac{(1 + \varepsilon)r_n}{2} \right\rceil\right) \geq d_{n,\text{TV}}^{(\delta)}(r_n - 1) > \varepsilon_0 > 0.$$

This makes a contradiction and, hence, if $r_n \rightarrow \infty$, then

$$\liminf_{n \rightarrow \infty} d_{n,\text{TV}}^{(c)}\left(\frac{(1 - \delta)r_n}{2}\right) > 0.$$

Similarly, if $s_n \rightarrow \infty$, then Corollary 3.1 (a) implies that

$$\liminf_{n \rightarrow \infty} d_{n,TV}^{(\delta)}(\lceil s_n \rceil) > 0.$$

This proves the equivalence

$$T_{n,TV}^{(\delta)}(\varepsilon_0) \rightarrow \infty \text{ for some } \varepsilon_0 > 0 \iff T_{n,TV}^{(c)}(\varepsilon_0) \rightarrow \infty \text{ for some } \varepsilon_0 > 0.$$

For the equivalence of (a) and (b), the proof, for precutoffs, is given by Corollary 3.1 (a) and (b), while the proof for cutoffs also uses Proposition 3.1 (c) and (d).

3.2. Cutoff windows

This section is devoted to the comparison of cutoff windows introduced in Definition 2.1.

Theorem 3.2. *Let \mathcal{F} be a family of irreducible positive recurrent Markov chains and $\mathcal{F}_\delta, \mathcal{F}_c$ be the associated families of lazy walks and continuous semigroups. Let t_n and b_n be sequences of positive reals and assume that $t_n \rightarrow \infty$. If \mathcal{F}_δ (respectively \mathcal{F}_c) presents a (t_n, b_n) cutoff in total variation, then $\sqrt{t_n} = O(b_n)$.*

Remark 3.1. There are examples with cutoffs but the order of any window size must be bigger than $\sqrt{t_n}$. Consider the Ehrenfest chain on $\{0, \dots, n\}$, which is a birth and death chain with rates $p_{n,i} = 1 - i/n, q_{n,i} = i/n$, and $r_{n,i} = 0$. It is obvious K_n is irreducible and periodic with stationary distribution $\pi_n(i) = 2^{-n} \binom{n}{i}$. An application of the representation theory shows that, for $0 \leq i \leq n, 2i/n$ is an eigenvalue of $I - K_n$. Let $\lambda_n = 2/n$ and $s_n = \sum_{i=1}^n n/2i = \frac{1}{2}n \log n + O(n)$. By Theorem 4.1, since $\lambda_n s_n$ tends to infinity, both \mathcal{F}_c and \mathcal{F}_δ have a total variation cutoff. For a detailed computation on the total variation and the L^2 -distance; see, e.g. [7]. It is well known that \mathcal{F}_c has a $(\frac{1}{4}n \log n, n)$ total variation cutoff. By Theorem 3.3, \mathcal{F}_δ has a $(n \log(n)/4(1 - \delta), n)$ total variation cutoff for $\delta \in (0, 1)$, which is nontrivial. For the continuous time Ehrenfest chains, Theorem 3.2 says that the window size is at least $\sqrt{n \log n}$, while n is the correct order.

Proof of Theorem 3.2. We prove the continuous time case. The lazy discrete time case can be treated similarly. Assume the converse that the sequence $\sqrt{t_n}/b_n$ is not bounded. By considering the subsequence of $\sqrt{t_n}/b_n$ which tends to ∞ , we lose no generality by assuming that $b_n = o(\sqrt{t_n})$. According to the definition of cutoffs, we may choose $C > 0, x_n \in \Omega_n$, and $A_n \subset \Omega_n$ satisfying

$$\liminf_{n \rightarrow \infty} |H_{n,t_n+Cb_n}(x_n, A_n) - \pi_n(A_n)| > 0.$$

By Lemma 3.1, we may choose a sequence n_l tending to ∞ for which $\pi_{n_l}(A_{n_l})$ converges to $\alpha \in [0, 1]$ and

$$\lim_{l \rightarrow \infty} H_{n_l,t_{n_l}+Cb_{n_l}}(x_{n_l}, A_{n_l}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} f(x) dx \neq \alpha,$$

where f is positive and bounded by 1. Let $c \in \mathbb{R}$. For any $\varepsilon > 0$, choose $N > 0$ satisfying, for $n \geq N$,

$$\left| H_{n,t_n+cb_n}(x_n, A_n) - \sum_{i: |i-t_n| \leq N\sqrt{t_n}} \left(e^{-(t_n+cb_n)} \frac{(t_n+cb_n)^i}{i!} \right) K_n^i(x_n, A_n) \right| < \varepsilon.$$

Note that

$$e^{-(t_n+cb_n)} \frac{(t_n + cb_n)^i}{i!} = e^{-(t_n+Cb_n)} \frac{(t_n + Cb_n)^i}{i!} (1 + o(1)) \quad \text{as } n \rightarrow \infty,$$

where $o(1)$ is uniform for $|i - t_n| \leq N\sqrt{t_n}$. This implies that

$$\lim_{l \rightarrow \infty} H_{n_l, t_{n_l} + cb_{n_l}}(x_{n_l}, A_{n_l}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} f(x) dx \quad \text{for all } c \in \mathbb{R}.$$

Since \mathcal{F}_c presents a (t_n, b_n) cutoff, the right-hand side integral is equal to α , a contradiction.

Theorem 3.3. *Let $\mathcal{F}_\delta, \mathcal{F}_c$ be the families in Theorem 3.2 and $t_n \rightarrow \infty$. Then, the following are equivalent.*

- (a) \mathcal{F}_δ has a (t_n, b_n) cutoff.
- (b) \mathcal{F}_c has a $((1 - \delta)t_n, b_n)$ cutoff.

To prove this theorem, we need the following proposition.

Proposition 3.2. *Let $\mathcal{F}_\delta, \mathcal{F}_c$ be as in Theorem 3.3 and t_n, r_n be sequences tending to ∞ .*

- (a) *If \mathcal{F}_δ has a (t_n, b_n) cutoff then \mathcal{F}_c has a $((1 - \delta)t_n, d_n)$ cutoff for any sequence satisfying $d_n = o(t_n)$ and $b_n = o(d_n)$.*
- (b) *If \mathcal{F}_c has a (r_n, b_n) cutoff then \mathcal{F}_δ has a $(r_n/(1 - \delta), d_n)$ cutoff for any sequence satisfying $d_n = o(r_n)$ and $b_n = o(d_n)$.*

Proof. Immediately from Theorem 3.2 and Proposition 3.1.

Proof of Theorem 3.3. We prove (a) implies (b), while the reasoning for (b) implies (a) is similar. Suppose that \mathcal{F}_δ has a (t_n, b_n) cutoff with $t_n \rightarrow \infty$. Fix $\varepsilon \in (0, 1)$ and set $c_n = |T_{n, \text{TV}}^{(c)}(\varepsilon) - (1 - \delta)t_n|$. By [5, Proposition 2.3], it remains to show that $c_n = O(b_n)$. Assume the converse, that is, there exists a subsequence $\xi = \{n_l \mid l = 1, 2, \dots\}$ for which $c_{n_l}/b_{n_l} \rightarrow \infty$ as $l \rightarrow \infty$. Respectively let $\mathcal{F}_\delta(\xi), \mathcal{F}_c(\xi)$ be families of $\mathcal{F}_\delta, \mathcal{F}_c$ restricted to ξ . This implies that $\mathcal{F}_\delta(\xi)$ has a (t_{n_l}, b_{n_l}) cutoff, but $\mathcal{F}_c(\xi)$ has no $((1 - \delta)t_{n_l}, \sqrt{b_{n_l}c_{n_l}})$ cutoff, a contradiction with Proposition 3.2.

3.3. Chains with specified initial states

For any probability μ on a countable set Ω , we write (μ, Ω, K, π) for an irreducible Markov chain on Ω with transition matrix K , stationary distribution π , and initial distribution μ . The total variation distances for the associated δ -lazy walk and continuous time chain are defined by

$$d_{\text{TV}}^{(\delta)}(\mu, n) = \|\mu K_\delta^n - \pi\|_{\text{TV}}, \quad d_{\text{TV}}^{(c)}(\mu, t) = \|\mu H_t - \pi\|_{\text{TV}}. \tag{3.2}$$

Denote by $T_{\text{TV}}^{(\delta)}(\mu, \varepsilon)$ and $T_{\text{TV}}^{(c)}(\mu, \varepsilon)$ the corresponding mixing times, and the concept of cutoffs can be defined similarly to how they appear in Definition 2.1 according to (3.2). It is an easy exercise to achieve a similar version of Lemma 2.1 for cutoffs with specified initial distributions. The proofs for Propositions 3.1 and 3.2 and Corollary 3.1 can be adapted to the case where the initial distribution is prescribed. This gives the following theorems.

Theorem 3.4. *Let $\mathcal{F} = \{(\mu_n, \Omega_n, K_n, \pi_n) \mid n = 1, 2, \dots\}$ be a family of irreducible Markov chains and $\mathcal{F}_\delta, \mathcal{F}_c$ be the families of associated δ -lazy walks and continuous time chains.*

- (a) \mathcal{F}_δ has a cutoff (precutoff) if and only if \mathcal{F}_c has a cutoff (respectively precutoff).
- (b) If \mathcal{F}_δ has a cutoff then $T_{n,\text{TV}}^{(c)}(\mu_n, \varepsilon) \sim (1 - \delta)T_{n,\text{TV}}^{(\delta)}(\mu_n, \varepsilon)$ as n tends to ∞ for all $\varepsilon \in (0, 1)$.

Let $t_n \rightarrow \infty$ and $b_n > 0$.

- (c) \mathcal{F}_δ has a (t_n, b_n) cutoff if and only if \mathcal{F}_c has a $((1 - \delta)t_n, b_n)$ cutoff.
- (d) If \mathcal{F}_δ has a (t_n, b_n) cutoff then $\sqrt{t_n} = O(b_n)$.

3.4. Proofs

This subsection collects required techniques for the proof of theorems in Sections 3.1 and 3.2.

Lemma 3.1. *Let $a_{n,m} \in [0, 1]$, $t_n > 0$ and $c \in \mathbb{R}$. Suppose that $t_n \rightarrow \infty$. Then, there exists a subsequence n_k of positive integers and a nonnegative measurable function f bounded by 1 for which*

$$\lim_{k \rightarrow \infty} \sum_{m=0}^{\infty} \left(e^{-t_{n_k} - c\sqrt{t_{n_k}}} \frac{(t_{n_k} + c\sqrt{t_{n_k}})^m}{m!} \right) a_{n_k,m} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-c)^2/2} f(x) dx$$

for all $c \in \mathbb{R}$, and

$$\begin{aligned} \lim_{k \rightarrow \infty} \sum_{m \geq 0} \binom{\lfloor (t_{n_k} + c\sqrt{t_{n_k}})/(1 - \delta) \rfloor}{m} (1 - \delta)^m \delta^{\lfloor (t_{n_k} + c\sqrt{t_{n_k}})/(1 - \delta) \rfloor - m} a_{n_k,m} \\ = \frac{1}{\sqrt{2\pi\delta}} \int_{-\infty}^{\infty} e^{-(x-c)^2/(2\delta)} f(x) dx \end{aligned}$$

for all $c \in \mathbb{R}$, $\delta \in (0, 1)$, where $\lceil z \rceil$ is any of $\lceil z \rceil$, $\lfloor z \rfloor$.

Proof. For $n \geq 1$ and any Borel set $A \subset \mathbb{R}$, set

$$\mu_n(A) = \frac{1}{\sqrt{t_n}} \sum_{\{m: m - t_n/\sqrt{t_n} \in A\}} a_{n,m}.$$

Let n_k be a subsequence of \mathbb{N} for which

$$\lim_{k \rightarrow \infty} \mu_{n_k}((a, b]) = \mu((a, b]) \quad \text{for all } a, b \in \mathbb{Q}, a < b. \tag{3.3}$$

Clearly, $\mu((a, b]) \leq b - a$ for $a < b$ with $a, b \in \mathbb{Q}$. This implies that the convergence in (3.3) holds for all $a < b$ and $\mu((a, b]) \leq b - a$. As a consequence of the Carathéodory extension theorem, μ can be extended to a measure on \mathbb{R} . It is obvious that μ is absolutely continuous with respect to the Lebesgue measure and we write f as the Radon–Nykodym derivative.

Let $\varepsilon > 0$ and choose $M > 0$ satisfying, for $n \geq M$,

$$\sum_{\{m: |m - t_n|/\sqrt{t_n} \notin (-M, M)\}} e^{-t_n - c\sqrt{t_n}} \frac{(t_n + c\sqrt{t_n})^m}{m!} < \varepsilon.$$

For any integer $N > 1$, set $x_i = iM/N$ and $A_{n,i} = \{m \geq 0 \mid |m - t_n|/\sqrt{t_n} \in (x_i, x_{i+1}]\}$. By Stirling’s formula, it is easy to see

$$e^{-t_n - c\sqrt{t_n}} \frac{(t_n + c\sqrt{t_n})^m}{m!} = \frac{1 + o(1)}{\sqrt{2\pi t_n}} \exp \left\{ -\frac{1}{2} \left(\frac{m - t_n}{\sqrt{t_n}} - c \right)^2 \right\} \quad \text{as } n \rightarrow \infty,$$

where $o(1)$ is uniform for $m \in A_{n,i}$ with $-N \leq i < N$. This implies that

$$\sum_{m \in A_{n,i}} \left(e^{-t_n - c\sqrt{t_n}} \frac{(t_n + c\sqrt{t_n})^m}{m!} \right) a_{n,m} \begin{cases} \leq \frac{M_i \mu_n(A_{n,i})}{\sqrt{2\pi}} + o(1), \\ \geq \frac{m_i \mu_n(A_{n,i})}{\sqrt{2\pi}} + o(1), \end{cases}$$

where $M_i = \sup\{e^{-(x-c)^2/2} \mid x \in (x_i, x_{i+1}]\}$ and $m_i = \inf\{e^{-(x-c)^2/2} \mid x \in (x_i, x_{i+1}]\}$. Summing over the index i and replacing n with n_k yields

$$\limsup_{k \rightarrow \infty} \sum_{m=0}^{\infty} \left(e^{-t_{n_k} - c\sqrt{t_{n_k}}} \frac{(t_{n_k} + c\sqrt{t_{n_k}})^m}{m!} \right) a_{n_k,m} \leq \frac{1}{\sqrt{2\pi}} \sum_{i=-N}^{N-1} M_i \mu((x_i, x_{i+1}]) + \varepsilon$$

and

$$\liminf_{k \rightarrow \infty} \sum_{m=0}^{\infty} \left(e^{-t_{n_k} - c\sqrt{t_{n_k}}} \frac{(t_{n_k} + c\sqrt{t_{n_k}})^m}{m!} \right) a_{n_k,m} \geq \frac{1}{\sqrt{2\pi}} \sum_{i=-N}^{N-1} m_i \mu((x_i, x_{i+1}]) - \varepsilon.$$

Letting $N \rightarrow \infty$ and then $\varepsilon \rightarrow 0$ gives the desired limit. The proof of the second limit is similar and is therefore omitted.

Lemma 3.2. *Let f be a bounded nonnegative measurable function and let F be the function $F(t) = \int_{-\infty}^{\infty} e^{-(x-t)^2} f(x) dx$. If F is constant then f is constant almost everywhere.*

Proof. Set $A = F(t)$, $B^{-1} = \int_{-\infty}^{\infty} e^{-x^2/2} f(x) dx$ and write

$$e^{-(x-t)^2/2} f(x) = B^{-1} \sqrt{2\pi} e^{t^2/4} \left(\frac{1}{\sqrt{2\pi}} e^{-(t-x)^2/2} \right) (B e^{-x^2/2} f(x)).$$

Note that $AB/(\sqrt{2\pi}e^{t^2/4})$ is the density of $X+Y$, where X has the standard normal distribution, Y is continuous with density function $B e^{-x^2/2} f(x)$, and X, Y are independent. This implies that $AB = 1/\sqrt{2}$ and

$$e^{-u^2} = \mathbb{E}(e^{iu(X+Y)}) = e^{-u^2/2} \mathbb{E}(e^{iuY}) \quad \text{for all } u \in \mathbb{R}.$$

Clearly, Y has the standard normal distribution and, thus, f is a constant almost everywhere.

3.5. A remark on the spectral gap and mixing time

In this subsection, we make a comparison between the spectral gaps of continuous time chains and δ -lazy discrete time chains. Let (Ω, K, π) be an irreducible and reversible finite Markov chain with spectral gap λ , the smallest nonzero eigenvalue of $I - K$. First, we consider the continuous time case. Since (K, π) is reversible, there is a function f defined on $\{0, 1, \dots, n\}$ for which $Kf = (1 - \lambda)f$. This implies that

$$d_{TV}^{(c)}(t) = \frac{1}{2} \|H_t - \pi\|_{\infty \rightarrow \infty} \geq \frac{\|(H_t - \pi)f\|_{\infty}}{2\|f\|_{\infty}} = \frac{e^{-\lambda t}}{2},$$

where $\|A\|_{\infty \rightarrow \infty} := \sup\{\|Ag\|_{\infty} : \|g\|_{\infty} = 1\}$. Consequently, we obtain

$$T_{TV}^{(c)}(\varepsilon) \geq \frac{-\log(2\varepsilon)}{\lambda}.$$

For the lazy discrete time case, a similar discussion yields

$$d_{\text{TV}}^{(\delta)}(t) \geq \frac{\beta_\delta^t}{2}, \quad T_{\text{TV}}^{(\delta)}(\varepsilon) \geq \left\lceil \frac{\log(2\varepsilon)}{\log \beta_\delta} \right\rceil,$$

where β_δ is the second largest absolute value of all nontrivial eigenvalues of K_δ . By setting $\delta_0 = \inf\{\delta \in (0, 1) \mid \beta_\delta = 1 - (1 - \delta)\lambda\}$, it is easy to see that $\delta_0 \leq \frac{1}{2}$ and, for $\delta \in [\delta_0, 1)$, $\beta_\delta = 1 - (1 - \delta)\lambda$. As a function of δ , β_δ is decreasing on $(0, \delta_0)$ and increasing on $(\delta_0, 1)$. Note that $|1 - (1 - \delta)\lambda| \leq \beta_\delta \leq \max\{1 - 2\delta, 1 - (1 - \delta)\lambda\}$. The first inequality implies $1 - \beta_\delta \leq (1 - \delta)\lambda$. Using the second inequality, if $\beta_\delta > 1 - 2\delta$ then $1 - \beta_\delta = (1 - \delta)\lambda$. If $\beta_\delta \leq 1 - 2\delta$ then $1 - \beta_\delta \geq 2\delta \geq \delta\lambda$, where the last inequality uses the fact $\lambda \leq 2$. We summarize the discussion in the following lemma.

Lemma 3.3. *Let K be an irreducible transition matrix on a finite set Ω with stationary distribution π . For $\delta \in (0, 1)$, let K_δ be the δ -lazy walk given by (1.1). Suppose that (π, K) is reversible, that is $\pi(x)K(x, y) = \pi(y)K(y, x)$ for all $x, y \in \Omega$, and let λ be the smallest nonzero eigenvalue of $I - K$ and β_δ be the largest absolute value of all nontrivial eigenvalues of K_δ . Then, it holds true that*

$$\min\{1 - \delta, \delta\}\lambda \leq 1 - \beta_\delta \leq 1 - |1 - (1 - \delta)\lambda| \leq (1 - \delta)\lambda \quad \text{for all } \delta \in (0, 1).$$

Furthermore, for $\varepsilon \in (0, \frac{1}{2})$,

$$T_{\text{TV}}^{(c)}(\varepsilon) \geq \frac{-\log(2\varepsilon)}{\lambda}, \quad T_{\text{TV}}^{(\delta)}(\varepsilon) \geq \left\lceil \frac{\log(2\varepsilon)}{\log \beta_\delta} \right\rceil \geq \left\lceil \frac{-\log(2\varepsilon)}{2 \max\{1 - \delta, \log(2/\delta)\}\lambda} \right\rceil,$$

where the last inequality assumes that $|\Omega| \geq 2/\delta$.

Proof. It remains to prove the second inequality in the lower bound of the mixing time for the δ -lazy chain. Note that if $\lambda \leq \frac{1}{2}$, then

$$-\log \beta_\delta \leq -\log(1 - (1 - \delta)\lambda) \leq 2(1 - \delta)\lambda,$$

where the last inequality uses the fact $\log(1 - x) \geq -2x$ for $x \in (0, \frac{1}{2})$. For $\lambda \geq \frac{1}{2}$, let $\theta_1(\delta), \dots, \theta_{|\Omega|}(\delta)$ be eigenvalues of K_δ . Then, $\theta_i(\delta) = \delta + (1 - \delta)\theta_i(0)$ and $\sum_{i=1}^{|\Omega|} \theta_i(0) \geq 0$. See [13] for a reference on the second inequality. This implies that

$$1 + (|\Omega| - 1)\beta_\delta \geq \sum_{i=1}^{|\Omega|} \theta_i(\delta) \geq |\Omega|\delta.$$

Assuming $|\Omega| \geq 2/\delta$, the above inequality yields

$$\beta_\delta \geq \frac{|\Omega|\delta - 1}{|\Omega| - 1} \geq \frac{\delta}{2}, \quad -\log \beta_\delta \leq \left(2 \log \frac{2}{\delta}\right)\lambda.$$

4. Finite birth and death chains

In this section, we consider the total variation cutoff for birth and death chains. A birth and death chain on $\{0, 1, \dots, n\}$, with birth rate p_i , death rate q_i , and holding rate r_i , is a Markov chain with transition matrix K given by

$$K(i, i + 1) = p_i, \quad K(i, i - 1) = q_i, \quad K(i, i) = r_i \quad \text{for all } 0 \leq i \leq n,$$

where $p_i + q_i + r_i = 1$ and $p_n = q_0 = 0$. It is obvious that K is irreducible if and only if $p_i q_{i+1} > 0$ for $0 \leq i < n$. Under the assumption of irreducibility, the unique stationary distribution π of K is given by $\pi(i) = c(p_0 \cdots p_{i-1}) / (q_1 \cdots q_i)$, where c is a positive constant satisfying $\sum_{i=0}^n \pi(i) = 1$.

In the next two subsections, we recall some results developed in [9] and [10], and make an improvement on them using the result in Section 3. In the third subsection, we go back to the issue of cutoffs and make a comparison of total variation and separation cutoffs.

4.1. The total variation cutoff

Throughout this subsection, we let

$$\mathcal{F} = \{(\Omega_n, K_n, \pi_n) \mid n = 1, 2, \dots\} \tag{4.1}$$

denote a family of irreducible birth and death chains with $\Omega_n = \{0, 1, \dots, n\}$ and transition matrix

$$K_n(i, i + 1) = p_{n,i}, \quad K_n(i, i - 1) = q_{n,i}, \quad K_n(i, i) = r_{n,i} \quad \text{for all } 0 \leq i \leq n,$$

where $p_{n,i} + q_{n,i} + r_{n,i} = 1$ and $p_{n,n} = q_{n,0} = 0$. Write $\lambda_n = \lambda(K_n)$ as the spectral gap of K_n . As before, \mathcal{F}_c denotes the family of associated continuous semigroups and, for $\delta \in (0, 1)$, \mathcal{F}_δ denotes the family of δ -lazy chains. Recall one of the main results in [10] as follows.

Theorem 4.1. (Theorems 3 and 3.1 of [10].) *Consider the family in (4.1). For $n \geq 1$, let λ_n be the smallest nonzero eigenvalue of $I - K_n$ and let $\beta_{n,\delta}$ be the second largest absolute value of all nontrivial eigenvalues of $K_{n,\delta}$. Then, \mathcal{F}_c (respectively \mathcal{F}_δ with $\delta \in (0, 1)$) has a total variation cutoff if and only if $T_{n,\text{TV}}^{(c)}(\frac{1}{4})\lambda_n \rightarrow \infty$ (respectively $T_{n,\text{TV}}^{(\delta)}(\frac{1}{4})(1 - \beta_{n,\delta}) \rightarrow \infty$). Moreover, if \mathcal{F}_c (respectively \mathcal{F}_δ) has a cutoff, then the window has size at most $\sqrt{T_{n,\text{TV}}^{(c)}(\frac{1}{4})/\lambda_n}$ (respectively $\sqrt{T_{n,\text{TV}}^{(\delta)}(\frac{1}{4})/(1 - \beta_{n,\delta})}$).*

Remark 4.1. By Lemma 3.3, the total variation cutoff in the discrete time case is equivalent to $T_{n,\text{TV}}^{(\delta)}(\frac{1}{4})\lambda_n \rightarrow \infty$. By Theorems 3.1 and 4.1, and Lemma 3.3, if \mathcal{F}_c or \mathcal{F}_δ has a cutoff, then the window size is at most $\sqrt{T_{n,\text{TV}}^{(c)}(\frac{1}{4})/\lambda_n}$ or $\sqrt{T_{n,\text{TV}}^{(\delta)}(\frac{1}{4})/\lambda_n}$.

Remark 4.2. There are examples with cutoffs, but the order of the optimal window size is less than $\sqrt{T_{n,\text{TV}}^{(c)}(\frac{1}{4})/\lambda_n}$. See Remark 3.1.

The combination of the above theorem and Theorem 3.1 yields the following.

Theorem 4.2. *Referring to Theorem 4.1, the following are equivalent.*

- (a) \mathcal{F}_c has a total variation cutoff.
- (b) \mathcal{F}_δ has a total variation cutoff.
- (c) \mathcal{F}_c has a total variation precutoff.
- (d) \mathcal{F}_δ has a total variation precutoff.
- (e) $T_{n,\text{TV}}^{(c)}(\varepsilon)\lambda_n \rightarrow \infty$ for some $\varepsilon \in (0, 1)$.
- (f) $T_{n,\text{TV}}^{(\delta)}(\varepsilon)\lambda_n \rightarrow \infty$ for some $\varepsilon \in (0, 1)$.

Proof of Theorem 4.2. It remains to show (c) implies (e) and this is given by the inequality $d_{n,\text{TV}}^{(c)}(t) \geq e^{-\lambda_n t} / 2$.

Theorem 4.3. Consider the family in (4.1). It holds true that $T_{n,\text{TV}}^{(c)}(\varepsilon/2) \asymp T_{n,\text{TV}}^{(\delta)}(\eta/2)$ for all $\varepsilon, \eta, \delta \in (0, 1)$. Furthermore, if there is $\varepsilon_0 \in (0, 1)$ for which $T_{n,\text{TV}}^{(c)}(\varepsilon_0/2)\lambda_n$ or $T_{n,\text{TV}}^{(\delta)}(\varepsilon_0/2)\lambda_n$ is bounded, then $T_{n,\text{TV}}^{(c)}(\varepsilon/2) \asymp 1/\lambda_n$ and $T_{n,\text{TV}}^{(\delta)}(\varepsilon/2) \asymp 1/\lambda_n$ for all $\varepsilon, \delta \in (0, 1)$.

Proof of Theorem 4.3. Assume that there is a subsequence n_k and $\varepsilon, \eta \in (0, \frac{1}{2})$ for which either $T_{n_k,\text{TV}}^{(c)}(\varepsilon)/T_{n_k,\text{TV}}^{(\delta)}(\eta) \rightarrow \infty$ or $T_{n_k,\text{TV}}^{(\delta)}(\eta)/T_{n_k,\text{TV}}^{(c)}(\varepsilon) \rightarrow \infty$. By Lemma 3.3, we have $T_{n_k,\text{TV}}^{(c)}(\varepsilon)\lambda_{n_k} \rightarrow \infty$ or $T_{n_k,\text{TV}}^{(\delta)}(\eta)\lambda_{n_k} \rightarrow \infty$. In either case, Theorems 3.1 and 4.1 imply that the subfamily indexed by $(n_k)_{k=1}^\infty$ has a cutoff in both continuous time and δ -lazy discrete time cases. As a consequence of Theorem 3.1, we obtain $T_{n_k,\text{TV}}^{(c)}(\varepsilon) \sim (1 - \delta)T_{n_k,\text{TV}}^{(\delta)}(\eta)$, which contradicts the assumption.

Concerning the window size, a combination of Theorem 3.2 and Theorem 4.1 yields the following Theorem.

Theorem 4.4. Let \mathcal{F}, λ_n be as in Theorem 4.1. Suppose that \mathcal{F}_c or \mathcal{F}_δ has a total variation cutoff and $\lambda_n \asymp 1$. Then, for any $\varepsilon, \eta \in (0, 1)$ with $\varepsilon \neq \eta$,

$$|T_{n,\text{TV}}^{(c)}(\varepsilon) - T_{n,\text{TV}}^{(c)}(\eta)| \asymp \sqrt{T_{n,\text{TV}}^{(c)}(\varepsilon)} \asymp |T_{n,\text{TV}}^{(\delta)}(\varepsilon) - T_{n,\text{TV}}^{(\delta)}(\eta)|.$$

4.2. The separation cutoff

In this subsection, we apply the results obtained in the previous subsection to the separation cutoff. First, we give a definition of the separation. Given an irreducible finite Markov chain K on Ω with initial distribution μ and stationary distribution π , the separation distance at time m is defined by

$$d_{\text{sep}}(\mu, m) := \max_{x \in \Omega} \left\{ 1 - \frac{\mu K^m(x)}{\pi(x)} \right\}.$$

Aldous and Diaconis [1] introduce the concept of the strong stationary time to identify the separation distance. Set $d_{\text{sep}}(m) = \max_i d_{\text{sep}}(i, m)$. A well known bound on the separation is achieved by Aldous and Fill in Lemma 7 of [2, Chapter 4], which states that

$$\bar{d}(m) \leq d_{\text{sep}}(m), \quad d_{\text{sep}}(2m) \leq 1 - (1 - \bar{d}(m))^2,$$

where $\bar{d}(m) := \max_{i,j} \|K^m(i, \cdot) - K^m(j, \cdot)\|_{\text{TV}}$. It is clear from the definitions that $d_{\text{TV}}(m) \leq \bar{d}(m) \leq 2d_{\text{TV}}(m)$. Let $T_{\text{sep}}(\varepsilon)$ be the separation mixing time. The above inequalities imply

$$T_{\text{TV}}(\varepsilon) \leq T_{\text{sep}}(\varepsilon) \leq 2T_{\text{TV}}\left(\frac{\varepsilon}{4}\right) \quad \text{for all } \varepsilon \in (0, 1). \tag{4.2}$$

Note that the above discussions are also valid for the continuous time case. As the separation distance is between (0, 1), the separation cutoff is similar to the total variation cutoff as given in Definition 2.1. By (4.2), we obtain the following lemma.

Lemma 4.1. Let \mathcal{F} be a family of finite Markov chains in either the discrete or continuous time case. Assume that $T_{n,\text{TV}}(\varepsilon) \rightarrow \infty$ or $T_{n,\text{sep}}(\varepsilon) \rightarrow \infty$ for some $\varepsilon \in (0, 1)$ in the discrete time case. Then, \mathcal{F} has a total variation precutoff if and only if \mathcal{F} has a separation precutoff.

For birth and death chains, the application of (4.2) to Theorem 4.3 leads to the following theorem.

Theorem 4.5. *Theorem 4.3 also holds in separation. Furthermore, for $\varepsilon, \eta \in (0, \frac{1}{2})$, we have $T_{n,TV}^{(c)}(\varepsilon) \asymp T_{n,sep}^{(c)}(\eta)$.*

Let K be an irreducible birth and death chain on $\{0, 1, \dots, n\}$ with stationary distribution π . The authors of [10] obtain the following fact:

$$d_{sep}^{(c)}(t) = 1 - \frac{H_t(0, n)}{\pi(n)}, \quad d_{sep}^{(\delta)}(m) = 1 - \frac{K_\delta^m(0, n)}{\pi(n)} \quad \text{for all } \delta \in [\frac{1}{2}, 1). \tag{4.3}$$

The authors of [9] provide a criterion on the separation cutoff for continuous time chains and monotone discrete time chains. The result says that a separation cutoff exists if and only if the product of the spectral gap and the separation mixing time tends to infinity. The next theorem, which is also obtained in [10], is a consequence of this fact and Theorems 4.2 and 4.5.

Theorem 4.6. *Let \mathcal{F} be a family of birth and death chains given by (4.1). The following are equivalent.*

- (a) \mathcal{F}_c has a cutoff in total variation.
- (b) For $\delta \in (0, 1)$, \mathcal{F}_δ has a cutoff in total variation.
- (c) \mathcal{F}_c has a cutoff in separation.
- (d) For $\delta \in [\frac{1}{2}, 1)$, \mathcal{F}_δ has a cutoff in separation.

The next theorem is a simple corollary of Theorems 4.2 and 4.6, and Lemma 4.1.

Theorem 4.7. *Theorem 4.2 also holds in separation distance with $\delta \in [\frac{1}{2}, 1)$.*

4.3. The cutoff time in total variation and separation

In this subsection, we introduce a spectral representation of the total variation mixing time. Let K be the transition kernel of an irreducible birth and death chain on $\{0, 1, \dots, n\}$. Suppose that K is irreducible with stationary distribution π and let $0 < \lambda_1 < \dots < \lambda_n$ be the eigenvalues of $I - K$. Consider the continuous time case. Using [9, Theorem 4.1] and [10, Corollary 4.5], we have

$$d_{sep}^{(c)}(t) = 1 - \frac{H_t(0, n)}{\pi(n)} = 1 - \frac{H_t(n, 0)}{\pi(0)} = \mathbb{P}(S > t),$$

where S is a sum of n independent exponential random variables with parameters $\lambda_1, \dots, \lambda_n$. By the one-sided Chebyshev inequality, we have

$$\mathbb{E}(S) - \sqrt{\frac{\text{var}(S)}{1/\varepsilon - 1}} \leq T_{sep}^{(c)}(\varepsilon) \leq \mathbb{E}(S) + \sqrt{\left(\frac{1}{\varepsilon} - 1\right) \text{var}(S)} \quad \text{for all } \varepsilon \in (0, 1).$$

Note that

$$\mathbb{E}(S) = \sum_{i=1}^n \frac{1}{\lambda_i}, \quad \text{var}(S) = \sum_{i=1}^n \frac{1}{\lambda_i^2} \leq (\mathbb{E}(S))^2.$$

Clearly, this implies

$$\frac{(\sqrt{1 - \varepsilon} - \sqrt{\varepsilon})\mathbb{E}(S)}{\sqrt{1 - \varepsilon}} \leq T_{sep}^{(c)}(\varepsilon) \leq \frac{(\sqrt{\varepsilon} + \sqrt{1 - \varepsilon})\mathbb{E}(S)}{\sqrt{\varepsilon}} \quad \text{for all } \varepsilon \in (0, 1). \tag{4.4}$$

The above equation states that, given $\varepsilon \in (0, \frac{1}{2})$, the separation mixing time is bounded by $\sum_{i=1}^n \lambda_i^{-1}$ up to universal constants. The above discussion is also valid for the discrete time case with the assumption that $K(i, i + 1) + K(i + 1, i) \leq 1$ for $0 \leq i < n$. See [9] for the details. The next proposition is an application of (4.2) and (4.4).

Proposition 4.1. *Let K be an irreducible birth and death chain on $\{0, 1, \dots, n\}$. Let $0, \lambda_1, \dots, \lambda_n$ be eigenvalues of K and set $s = \sum_{i=1}^n \lambda_i^{-1}$. Then,*

$$\left(\frac{\sqrt{1-\varepsilon} - \sqrt{\varepsilon}}{\sqrt{1-\varepsilon}}\right)^s \leq T_{\text{sep}}^{(c)}(\varepsilon) \leq \left(\frac{\sqrt{\varepsilon} + \sqrt{1-\varepsilon}}{\sqrt{\varepsilon}}\right)^s \text{ for all } \varepsilon \in (0, \frac{1}{2}),$$

and

$$\frac{1}{2} \left(\frac{\sqrt{1-4\varepsilon} - \sqrt{4\varepsilon}}{\sqrt{1-4\varepsilon}}\right)^s \leq T_{\text{TV}}^{(c)}(\varepsilon) \leq \left(\frac{\sqrt{\varepsilon} + \sqrt{1-\varepsilon}}{\sqrt{\varepsilon}}\right)^s \text{ for all } \varepsilon \in (0, \frac{1}{8}).$$

The above also holds in the discrete time case with the assumption that $K(i, i + 1) + K(i + 1, i) \leq 1$ for $0 \leq i < n$.

Applying Proposition 4.1 to Theorems 4.2 and 4.6 yields the following theorem, where the result in separation is included in [9] and the result in total variation is implicitly obtained in [10].

Theorem 4.8. (Cutoffs from the spectrum.) *Let \mathcal{F} be the family in (4.1). For $n \geq 1$, let $\lambda_{n,1}, \dots, \lambda_{n,n}$ be nonzero eigenvalues of $I - K_n$ and set*

$$\lambda_n = \min_{1 \leq i \leq n} \lambda_{n,i}, \quad s_n = \frac{1}{\lambda_{n,1}} + \dots + \frac{1}{\lambda_{n,n}}.$$

Then, the following are equivalent.

- (a) \mathcal{F}_c has a total variation cutoff.
- (b) For $\delta \in (0, 1)$, \mathcal{F}_δ has a total variation cutoff.
- (c) \mathcal{F}_c has a total variation precutoff.
- (d) For $\delta \in (0, 1)$, \mathcal{F}_δ has a total variation precutoff.
- (e) $s_n \lambda_n \rightarrow \infty$.

The above also holds in separation with $\delta \in [\frac{1}{2}, 1)$. In particular, if (e) holds, then, for $\varepsilon \in (0, 1)$,

$$\frac{1}{2} \leq \liminf_{n \rightarrow \infty} \frac{T_{n,\text{TV}}^{(c)}(\varepsilon)}{s_n} \leq \limsup_{n \rightarrow \infty} \frac{T_{n,\text{TV}}^{(c)}(\varepsilon)}{s_n} \leq 1.$$

The last result establishes a relation between the mixing time and birth and death rates. Consider an irreducible birth and death chain $(X_m)_{m=0}^\infty$ on $\{0, 1, \dots, n\}$ with transition matrix K and stationary distribution π . Let N_t be a Poisson process of parameter 1 that is independent of X_m and set, for $0 \leq i \leq n$,

$$\tau_i := \inf\{t \geq 0 \mid X_{N_t} = i\}.$$

Brown and Shao discuss the distribution of τ_i in [3] and obtain the following result:

$$\mathbb{P}_0(\tau_n > t) = \sum_{j=1}^n \left(\prod_{k \neq j} \frac{\theta_k}{\theta_k - \theta_j} \right) e^{-\theta_j t},$$

where \mathbb{P}_i is the conditional probability given $X_0 = i$ and $\theta_1, \dots, \theta_n$ are eigenvalues of the submatrix of $I - K$ indexed by $\{0, 1, \dots, n - 1\}$. Let \mathbb{E}_i be the conditional expectation given $X_0 = i$. Clearly, this implies $\mathbb{E}_0(\tau_n) = \sum_{j=1}^n 1/\theta_j$. Note that $\mathbb{E}_0\tau_n$ can be formulated by the birth and death rates using the strong Markov property. This leads to

$$\mathbb{E}_0(\tau_n) = \sum_{j=1}^n \frac{1}{\theta_j} = \sum_{k=0}^{n-1} \frac{\pi([0, k])}{\pi(k)p_k}, \tag{4.5}$$

where $\pi(A) := \sum_{i \in A} \pi(i)$. Now fix $0 \leq i_0 \leq n$. By (4.5), we have

$$\mathbb{E}_0(\tau_{i_0}) = \sum_{i=1}^{i_0} \frac{1}{\lambda'_i}, \quad \mathbb{E}_n(\tau_{i_0}) = \sum_{i=1}^{n-i_0} \frac{1}{\lambda''_i},$$

where $\lambda'_1, \dots, \lambda'_{i_0}$ and $\lambda''_1, \dots, \lambda''_{n-i_0}$ are eigenvalues of the submatrices of $I - K$ indexed by $\{0, \dots, i_0 - 1\}$ and $\{i_0 + 1, \dots, n\}$ respectively. Let $\bar{\lambda}_1 \leq \dots \leq \bar{\lambda}_n$ be a rearrangement of $\lambda'_1, \dots, \lambda'_{i_0}, \lambda''_1, \dots, \lambda''_{n-i_0}$. Clearly, $\bar{\lambda}_1, \dots, \bar{\lambda}_n$ are eigenvalues of the submatrix obtained by removing the i_0 th row and the i_0 th column of $I - K$. Let $\lambda_1 < \dots < \lambda_n$ be nonzero eigenvalues of $I - K$. By Theorem 4.3.8 in [11], we have $\bar{\lambda}_i \leq \lambda_i \leq \bar{\lambda}_{i+1}$ and this leads to

$$\sum_{i=2}^n \frac{1}{\bar{\lambda}_i} \leq \sum_{i=1}^n \frac{1}{\lambda_i} \leq \sum_{i=1}^n \frac{1}{\bar{\lambda}_i} = \sum_{k=0}^{i_0-1} \frac{\pi([0, k])}{\pi(k)p_k} + \sum_{k=i_0+1}^n \frac{\pi([k, n])}{\pi(k)q_k},$$

where the first equality uses (4.5). By Proposition 4.1, we obtain, for $\varepsilon \in (0, 1)$,

$$T_{TV}^{(c)}(\varepsilon) \leq T_{sep}^{(c)}(\varepsilon) \leq \left(\frac{\sqrt{\varepsilon} + \sqrt{1-\varepsilon}}{\sqrt{\varepsilon}} \right) \min_{0 \leq i \leq n} \left\{ \sum_{k=0}^{i-1} \frac{\pi([0, k])}{\pi(k)p_k} + \sum_{k=i+1}^n \frac{\pi([k, n])}{\pi(k)q_k} \right\}.$$

The above discussion also holds in the discrete time case with the assumption that $p_i + q_{i+1} \leq 1$ for all $0 \leq i < n$. This includes the δ -lazy chain for $\delta \in [\frac{1}{2}, 1)$ and we apply it to get the following corollary.

Corollary 4.1. *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) \mid n = 1, 2, \dots\}$ be a family of irreducible birth and death chains in (4.1) with birth, death, and holding rates $p_{n,i}, q_{n,i}$, and $r_{n,i}$. For $n \geq 1$, set*

$$t_n = \min_{0 \leq i \leq n} \left\{ \sum_{k=0}^{i-1} \frac{\pi_n([0, k])}{\pi_n(k)p_{n,k}} + \sum_{k=i+1}^n \frac{\pi_n([k, n])}{\pi_n(k)q_{n,k}} \right\}.$$

If \mathcal{F}_c or \mathcal{F}_δ has a total variation cutoff then, for $\varepsilon \in (0, 1)$ and $\delta \in [\frac{1}{2}, 1)$,

$$\limsup_{n \rightarrow \infty} \frac{T_{n,sep}^{(c)}(\varepsilon)}{t_n} \leq 1, \quad \limsup_{n \rightarrow \infty} \frac{T_{n,sep}^{(\delta)}(\varepsilon)}{t_n} \leq \frac{1}{1-\delta}$$

and, for $\varepsilon \in (0, 1)$,

$$\limsup_{n \rightarrow \infty} \frac{T_{n,TV}^{(c)}(\varepsilon)}{t_n} \leq 1 \quad \limsup_{n \rightarrow \infty} \frac{T_{n,TV}^{(\delta)}(\varepsilon)}{t_n} \leq \frac{1}{1-\delta}.$$

Remark 4.3. In [6], the constant t_n in Corollary 4.1 is proved to be of the same order as the constant s_n in Theorem 4.8 and the following term

$$\sum_{k=0}^{i_n-1} \frac{\pi_n([0, k])}{\pi_n(k)p_{n,k}} + \sum_{k=i_n+1}^n \frac{\pi_n([k, n])}{\pi_n(k)q_{n,k}},$$

where i_n satisfies $\pi_n([0, i_n]) \geq \frac{1}{2}$ and $\pi_n([i_n, n]) \geq \frac{1}{2}$.

Remark 4.4. The bound in Corollary 4.1 is also obtained implicitly in [10] using a coupling argument.

Acknowledgements

We thank the anonymous referee for her/his very careful reading of the manuscript. We also thank Y. Peres and D. Levin for pointing out the similarity between some of the techniques used in this paper and techniques used in [12]. The first author is partially supported by NSC grant NSC100-2918-I-009-010 and the second author is partially supported by NSF grant DMS-1004771.

References

- [1] ALDOUS, D. AND DIACONIS, P. (1987). Strong uniform times and finite random walks. *Adv. Appl. Math.* **8**, 69–97.
- [2] ALDOUS, D. AND FILL, J. *Reversible Markov Chains and Random Walks on Graphs*. Available at <http://www.stat.berkeley.edu/users/aldous/RWG/book.html>.
- [3] BROWN, M. AND SHAO, Y.-S. (1987). Identifying coefficients in the spectral representation for first passage time distributions. *Prob. Eng. Inf. Sci.* **1**, 69–74.
- [4] CHEN, G.-Y. (2006). The cutoff phenomenon for finite Markov chains. Doctoral Thesis, Cornell University.
- [5] CHEN, G.-Y. AND SALOFF-COSTE, L. (2008). The cutoff phenomenon for ergodic Markov processes. *Electron. J. Prob.* **13**, 26–78.
- [6] CHEN, G.-Y. AND SALOFF-COSTE, L. (2013). On the mixing time and spectral gap for birth and death chains. *ALEA Lat. Amer. J. Prob. Math. Statist.* **10**, 293–321.
- [7] DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, CA.
- [8] DIACONIS, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. USA.* **93**, 1659–1664.
- [9] DIACONIS, P. AND SALOFF-COSTE, L. (2006). Separation cut-offs for birth and death chains. *Ann. Appl. Prob.* **16**, 2098–2122.
- [10] DING, J., LUBETZKY, E. AND PERES, Y. (2010). Total variation cutoff in birth-and-death chains. *Prob. Theory Relat. Fields* **146**, 61–85.
- [11] HORN, R. A. AND JOHNSON, C. R. (1990). *Matrix Analysis*. Cambridge University Press.
- [12] LEVIN, D. A., PERES, Y. AND WILMER, E. L. (2009). *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI.
- [13] SALOFF-COSTE, L. (1997). Lectures on finite Markov chains. In *Lectures on Probability Theory and Statistics* (St-Flour, 1996; Lecture Notes Math. **1665**), Springer, Berlin, pp. 301–413.