Front. Math. China 2014, 9(4): 881–898
DOI 10.1007/s11464-014-0380-8

RESEARCH ARTICLE

Explicit criteria on separation cutoff for birth and death chains

Yonghua MAO, Yuhui ZHANG

School of Mathematical Sciences, Beijing Normal University, Laboratory of Mathematics and Complex Systems, Ministry of Education, Beijing 100875, China

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2014

Abstract The criteria on separation cutoff for birth and death chains were obtained by Diaconis and Saloff-Coste in 2006. These criteria are involving all eigenvalues. In this paper, we obtain the explicit criterion, which depends only on the birth and death rates. Furthermore, we present two ways to estimate moments of the fastest strong stationary time and then give another but equivalent criterion explicitly.

Keywords Separation cutoff, birth and death chain, hitting time, fastest strong stationary time (FSST), eigenvalue, stochastic monotonicity, duality, boundary theory
MSC 60B10, 60J05, 60J27

1 Introduction and main results

In [6], Diaconis wrote: "At present writing, proof of a cutoff is a difficult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors ...." Diaconis and Saloff-Coste [9] gave a necessary and sufficient condition for separation cutoff of a sequence of continuous or discrete time birth and death chains. This condition is involving all eigenvalues of the generator or transition matrix of the birth and death chains. In this paper, we give an explicit criterion, via three different ways. First of all, based on their criterion, we will give the explicit criterion by using the celebrated estimates of the spectral gap for the birth and death processes [5], and an eigentime identity for the birth and death processes [16]. Second, cutoff in separation is closely related to the *fastest strong stationary time* (FSST) [1,7]. A use of Chebyshev inequality involving the mean and variance of the FSST will give the criterion of cutoff. See [9]. We will recall briefly this argument in Section 2. So our second

Received March 3, 2014; accepted April 3, 2014
Corresponding author: Yonghua MAO, E-mail: maoyh@bnu.edu.cn

way for the criterion is to calculate explicitly the mean and variance of FSST for the birth and death chains. Third, from the definition of the separation between two probability measures, a halting state is crucial, especially for a stochastically monotone Markov chain. See [13] and Section 3. The third way we use is to give the moments of FSST via the moments of the hitting times. Actually, all three criteria are nevertheless equivalent, but appear in disguises.

Let X_t, $t \geqslant 0$, be a continuous time birth and death process on the state space $E = \{0, 1, \ldots, m\}$. Its generator is the matrix $Q = (q_{ij})$:

$$q_{ij} = \begin{cases} b_i, & 0 \leqslant i \leqslant m-1,\ j = i+1; \\ a_i, & 1 \leqslant i \leqslant m,\ j = i-1; \\ -(a_i + b_i), & 0 \leqslant i = j \leqslant m. \end{cases}$$

Here, $a_i > 0$ for $1 \leqslant i \leqslant m$, $b_i > 0$ for $0 \leqslant i \leqslant m-1$, and $a_0 = b_m = 0$. This process is ergodic (positive recurrent and irreducible). Let $\pi_0 = 1/Z$, and

$$\pi_i = \frac{1}{Z}\frac{b_0 \cdots b_{i-1}}{a_1 \cdots a_i}, \quad 1 \leqslant i \leqslant m, \tag{1.1}$$

where Z is such that $\pi = (\pi_i,\, 0 \leqslant i \leqslant m)$ is a probability measure. Then π is the stationary distribution.

The separation between probability measures μ, ν on finite space E is

$$\operatorname{sep}(\mu, \nu) = \max_{i \in E}\Big\{1 - \frac{\mu_i}{\nu_i}\Big\}. \tag{1.2}$$

Applying the separation to the birth and death chains as above, we have

$$\lim_{t \to \infty} \operatorname{sep}(\mu(t), \pi) = 0, \tag{1.3}$$

where $\mu(t)$ is the distribution of the chain at time t.

Separation cutoff is concerning a family of birth and death processes. For $n = 0, 1, 2, \ldots$, let $X_t^{(n)}$ be a sequence of birth and death process on $E_n = \{0, 1, \ldots, m_n\}$ with generator matrix $Q^{(n)}$. Let $\pi^{(n)}$ be the corresponding stationary distribution. Let

$$0 = \lambda_0^{(n)} < \lambda_1^{(n)} < \cdots < \lambda_{m_n}^{(n)}$$

be the eigenvalues of $-Q^{(n)}$ and set

$$T^{(n)} = \sum_{i=1}^{m_n} \frac{1}{\lambda_i^{(n)}}.$$

Let $\mu^{(n)}(t)$ be the distribution $X_t^{(n)}$ starting from state 0:

$$\mu_i^{(n)}(t) = \mathbb{P}[X_t^{(n)} = i \mid X_0^{(n)} = 0], \quad i \in E_n.$$

It is proved in [9] that whenever $\lim_{n\to\infty}\lambda_1^{(n)}T^{(n)}=\infty$,

$$\lim_{n\to\infty}\operatorname{sep}(\mu^{(n)}(cT^{(n)}),\pi^{(n)})=\begin{cases}0, & c>1;\\ 1, & c<1.\end{cases}\tag{1.4}$$

The expression in (1.4) is referred to *separation cutoff* for birth and death processes.

A similar separation cutoff was also proved in [9] for discrete-time birth and death chains, under the additional assumption that the processes are stochastically monotone. In the discrete time case, we keep the notations as in the continuous time as same as possible. Let $P=(p_{ij})$ be transition probability matrix on $E=\{0,1,\dots,m\}$:

$$p_{ij}=\begin{cases}b_i, & 0\leqslant i\leqslant m-1,\ j=i+1;\\ a_i, & 1\leqslant i\leqslant m,\ j=i-1;\\ c_i, & 0\leqslant i=j\leqslant m.\end{cases}\tag{1.5}$$

Here, all a's and b's are positive, and c's are nonnegative. Assume

$$a_i+b_i+c_i=1\ (1\leqslant i\leqslant m-1),\quad b_0+c_0=1,\quad a_m+c_m=1.$$

The chain P is stochastically monotone whenever $b_i+a_{i+1}\leqslant 1$ for $0\leqslant i<m$. Let π be defined in the same way as in (1.1) and set $0=\lambda_0<\lambda_1<\cdots<\lambda_m$ to be the eigenvalues of $I-P$, where I is the identity matrix. Then it was proved in [9] that a family of ergodic and stochastically monotone discrete time birth and death chains has the separation cutoff if and only if

$$\lim_{n\to\infty}\lambda_1^{(n)}T^{(n)}=\infty.$$

In this paper, for the sake of simplicity, we focus on the continuous time birth and death processes. For the discrete time birth and death chains, we only point out the difference.

The rest of the paper is organized as follows. In Section 2, we will give a general sufficient and necessary condition for separation cutoff, and revisit the criterion in [9]. In Section 3, we present three explicit criteria. Finally, in Section 4, we apply the criteria to the restricted chains and the Metropolis chains.

2 Fastest strong stationary time

For the continuous time birth and death process $X(t)$ on $\{0,1,\dots,m\}$ starting at 0, let $\mu(t)$ be the distribution of X_t. It is known from [7] that there exists a FSST τ such that

$$\operatorname{sep}(\mu(t),\pi)=\mathbb{P}[\tau>t].\tag{2.1}$$

And the distribution of τ is given in Laplace transformation:

$$\mathbb{E}\mathrm{e}^{-\lambda\tau} = \prod_{\nu=1}^{m} \frac{\lambda_\nu}{\lambda+\lambda_\nu}, \quad \lambda \geqslant 0, \tag{2.2}$$

where $\lambda_1 < \cdots < \lambda_m$ are positive eigenvalues of the matrix $-Q$.

For the discrete time birth and death chains on $\{0, 1, \ldots, m\}$ starting at 0, assume that the chain is stochastically monotone. It was proved in [11] that (2.1) holds and the distribution of τ is given by

$$\mathbb{E}s^{\tau} = \prod_{\nu=1}^{m} \frac{\lambda_\nu}{\lambda_\nu + s}, \quad s \in [0, 1). \tag{2.3}$$

Diaconis and Saloff-Coste [9] used this formula and a Chebyshev inequality to give the cutoff for the birth and death process. The sufficient part in the following proposition is essentially due to them, here we give a direct proof for the necessary part.

Proposition 2.1 *For the family of birth and death chains $X_t^{(n)}$ on the state space $\{0, 1, \ldots, m_n\}$ starting at 0, let $\tau^{(n)}$ be the FSST. Then there is the separation cutoff* (1.4) *if and only if*

$$\frac{(\mathbb{E}\tau^{(n)})^2}{\mathrm{Var}(\tau^{(n)})} \to \infty, \quad \textit{or, equivalently} \quad \frac{(\mathbb{E}\tau^{(n)})^2}{\mathbb{E}(\tau^{(n)})^2} \to 1. \tag{2.4}$$

Proof Assume first (2.4). The following Chebyshev type inequality can be found on [10; p. 152]). For a non-negative ξ with mean $\mathbb{E}\xi$ and variance $\mathrm{Var}(\xi)$, we have, for any $c > 0$,

$$\mathbb{P}[\xi \geqslant \mathbb{E}\xi + c\sqrt{\mathrm{Var}(\xi)}\,] \leqslant \frac{1}{1+c^2}, \quad \mathbb{P}[\xi \leqslant \mathbb{E}\xi - c\sqrt{\mathrm{Var}(\xi)}\,] \leqslant \frac{1}{1+c^2}.$$

Apply this Chebyshev inequality to the sequence of fastest strong stationary times $\tau^{(n)}$ to get that under (2.4),

$$\begin{aligned}\mathbb{P}[\tau^{(n)} \geqslant (1+c)\mathbb{E}\tau^{(n)}] &= \mathbb{P}\Big[\tau^{(n)} \geqslant \mathbb{E}\tau^{(n)} + \frac{c\mathbb{E}\tau^{(n)}}{\sqrt{\mathrm{Var}(\tau^{(n)})}}\sqrt{\mathrm{Var}(\tau^{(n)})}\,\Big] \\ &\leqslant \frac{1}{1 + c^2\frac{[\mathbb{E}\tau^{(n)}]^2}{\mathrm{Var}(\tau^{(n)})}} \\ &\to 0,\end{aligned}$$

and similarly,

$$\mathbb{P}[\tau^{(n)} \leqslant (1-c)\mathbb{E}\tau^{(n)}] \to 0.$$

Conversely, assume (1.4). Let $\xi^{(n)} = \tau^{(n)}/T^{(n)}$. Then (1.4) implies that $\xi^{(n)}$ converges to 1 in probability. Since it follows from (2.2) that

$$\mathbb{E}(\tau^{(n)})^3 \leqslant 6(T^{(n)})^3, \quad \text{or} \quad \mathbb{E}(\xi^{(n)})^3 \leqslant 6,$$

we have (cf. [3])

$$\lim_{n\to\infty} \mathbb{E}|\xi^{(n)} - 1|^2 = 0.$$

In particular, the limit

$$\lim_{n\to\infty} \mathbb{E}\xi^{(n)} = \lim_{n\to\infty} \mathbb{E}(\xi^{(n)})^2 = 1.$$

Therefore,

$$\lim_{n\to\infty} \frac{(\mathbb{E}\tau^{(n)})^2}{\mathbb{E}(\tau^{(n)})^2} = \lim_{n\to\infty} \frac{(\mathbb{E}\xi^{(n)})^2}{\mathbb{E}(\xi^{(n)})^2} = 1.$$

The assertion is proved. □

We remark that in general, we can derive from (2.2) that

$$\mathbb{E}(\tau^{(n)})^k \leqslant k!\,(T^{(n)})^k, \quad k = 1, 2, \ldots.$$

For the FSST $\tau^{(n)}$, it follows from (2.2) that

$$\mathbb{E}\tau^{(n)} = T^{(n)} = \sum_{\nu=1}^{n} \frac{1}{\lambda_\nu^{(n)}}, \quad \mathrm{Var}(\tau^{(n)}) = \sum_{\nu=1}^{n} \Big(\frac{1}{\lambda_\nu^{(n)}}\Big)^2.$$

Since

$$\Big(\frac{1}{\lambda_1^{(n)}}\Big)^2 \leqslant \mathrm{Var}(\tau^{(n)}) \leqslant \frac{1}{\lambda_1^{(n)}}\, T^{(n)},$$

we get

$$\lambda_1^{(n)} T^{(n)} \leqslant \frac{(\mathbb{E}\tau^{(n)})^2}{\mathrm{Var}(\tau^{(n)})} \leqslant [\lambda_1^{(n)} T^{(n)}]^2.$$

From this and (2.4), we obtain (1.4) whenever $\lim_{n\to\infty} \lambda_1^{(n)} T^{(n)} = \infty$, as did in [9].

3 Explicit criteria

In this section, we will present three explicit criteria for separation cutoff of the birth and death chains.

3.1 Eigenvalues

Our first explicit criterion is rather a direct use of the celebrated estimates for $\lambda_1^{(n)}$ in [5; Corollary 6.6], and the explicit formula for $T^{(n)}$ in [16].

The formula for $T^{(n)}$ comes from the so-called eigentime identity, appearing first in [2; Chapter 4]. Let $\lambda_\nu^{(n)}$, $\nu \geqslant 1$, be the non-zero eigenvalues of $-Q^{(n)}$. Then it holds that

$$\sum_{\nu=1}^{m_n} \frac{1}{\lambda_\nu^{(n)}} = \sum_{i,j=0}^{m_n} \pi_i^{(n)} \pi_j^{(n)} \mathbb{E}_i \tau_j^{(n)} = \sum_{j=0}^{m_n} \pi_j^{(n)} \mathbb{E}_i \tau_j^{(n)} \quad \text{for all } i.$$

According to [16], for birth and death chains, we have

$$T^{(n)} = \sum_{\nu=1}^{m_n} \frac{1}{\lambda_\nu^{(n)}} = \sum_{i=0}^{m_n-1} \frac{1}{\pi_i^{(n)} b_i^{(n)}} \sum_{j=i+1}^{m_n} \pi_j^{(n)} \sum_{j=0}^{i} \pi_j^{(n)}. \tag{3.1}$$

Next, we introduce an elegant estimate of spectral gap for ergodic birth and death chains derived recently in [5]. Let

$$\kappa^{(n)} = \min_{0\leqslant \ell < m \leqslant m_n} \left[\left(\sum_{i=0}^{\ell} \pi_i^{(n)}\right)^{-1} + \left(\sum_{i=m}^{m_n} \pi_i^{(n)}\right)^{-1}\right]\left(\sum_{i=\ell}^{m-1} \frac{1}{\pi_i^{(n)} b_i^{(n)}}\right)^{-1}. \tag{3.2}$$

Then

$$\frac{1}{4}\kappa^{(n)} \leqslant \lambda_1^{(n)} \leqslant \kappa^{(n)}. \tag{3.3}$$

This estimate is a generalization of the classical Hardy inequality. The obvious advantage is the universal constant 1/4 appears in (3.3). See [5] for a comprehensive study on this topic.

The following is the first explicit criterion for the separation cutoff.

Theorem 3.1 *Separation cutoff occurs if and only if*

$$\lim_{n\to\infty} \kappa^{(n)} T^{(n)} = \infty.$$

The only seeming obstacle is taking minimum in two variables ℓ, m in (3.2).

For the discrete time case, the eigentime identity and the formula in (3.1) remain true. Now, $\lambda_1^{(n)}$ is eigenvalue for $I-P$, and $P-I$ can be seen as the generator of a continuous time birth and death chain, thus the estimates in (3.2) still work. Therefore, Theorem 3.1 holds for stochastically monotone discrete time birth and death chains.

3.2 Duality

From the above routine of the proof for the separation cutoff, we see that one has only to get the first two moments of the FSST. In the sequel, we will focus on this for a birth and death process, instead of a sequence of birth and death processes.

For discrete time birth and death process P given in (1.5), Diaconis-Fill [7] constructed a dual absorbing birth-death process, whose absorption time τ^* has the same distribution as τ the FSST, after then Fill [11] obtained the same conclusion for the continuous time birth and death process. Set for $0 \leqslant i \leqslant m$, $H_i = \sum_{j\leqslant i} \pi_j$ and for $i \leqslant m-1$,

$$a_i^* = \frac{H_{i-1}}{H_i} b_i, \quad b_i^* = \frac{H_{i+1}}{H_i} a_{i+1}, \quad a_m^* = b_m^* = 0.$$

Let

$$\pi_i^* = \frac{b_0 H_i^2}{\pi_i b_i}.$$

Let a_i^* $(1 \leqslant i \leqslant m)$ and b_i^* $(0 \leqslant i \leqslant m)$ be the birth and death rates for a dual process $X^*(t)$, respectively. So m is an absorbing state for $X^*(t)$. Assume that both $X(t)$ and $X^*(t)$ start at 0. Then the absorption time (hitting time to m) of $X^*(t)$ has the same distribution as the FSST τ of $X(t)$. From these and [15; Theorem 4.1], we have the first and second order moments of τ:

$$\mathbb{E}\tau = \mathbb{E}\tau^* = \sum_{i=0}^{m-1} \frac{1}{\pi_i^* b_i^*} \sum_{j=0}^{i} \pi_j^*,$$

$$\begin{aligned}\mathbb{E}\tau^2 &= \mathbb{E}(\tau^*)^2 \\ &= 2\sum_{i=0}^{m-1} \frac{1}{\pi_i^* b_i^*} \sum_{j=0}^{i} \pi_j^* \sum_{k=j}^{m-1} \frac{1}{\pi_k^* b_k^*} \sum_{l=0}^{j} \pi_l^* \\ &= 2(\mathbb{E}\tau)^2 - 2\sum_{i=0}^{m-1} \frac{1}{\pi_i^* b_i^*} \sum_{j=0}^{i} \pi_j^* \sum_{k=0}^{j-1} \frac{1}{\pi_k^* b_k^*} \sum_{l=0}^{j} \pi_l^*.\end{aligned}$$

Denote

$$S = (\mathbb{E}\tau)^2 - \frac{1}{2}\,\mathbb{E}\tau^2 = \sum_{i=0}^{m-1} \frac{1}{\pi_i^* b_i^*} \sum_{j=0}^{i} \pi_j^* \sum_{k=0}^{j-1} \frac{1}{\pi_k^* b_k^*} \sum_{l=0}^{j} \pi_l^*.$$

To get the expression of these moments by a_i, b_i, we need the following lemma.

Lemma 3.2 *Let α_i be a positive sequence. For $k \leqslant m$, we have*

$$\sum_{i=0}^{k-1} \frac{1}{\pi_i^* b_i^*} \sum_{j=0}^{i} \pi_j^* \alpha_j = \frac{1}{H_k} \sum_{j=0}^{k-1} \frac{H_j(H_k - H_j)}{\pi_j b_j}\, \alpha_j. \tag{3.4}$$

Proof Note that

$$\pi_i^* = \frac{b_0 H_i^2}{\pi_i b_i}, \qquad \frac{1}{\pi_i^* b_i^*} = \frac{\pi_{i+1}}{b_0 H_{i+1} H_i} = b_0\Big[\frac{1}{H_i} - \frac{1}{H_{i+1}}\Big]. \tag{3.5}$$

We have

$$\begin{aligned}\sum_{i=0}^{k-1} \frac{1}{\pi_i^* b_i^*} \sum_{j=0}^{i} \pi_j^* \alpha_j &= \sum_{j=0}^{k-1} \pi_j^* \alpha_j \sum_{i=j}^{k-1} \frac{1}{\pi_i^* b_i^*} \\ &= \sum_{j=0}^{k-1} \frac{b_0 \alpha_j H_j^2}{\pi_j b_j} \sum_{i=j}^{k-1} \frac{\pi_{i+1}}{b_0 H_{i+1} H_i} \\ &= \sum_{j=0}^{k-1} \frac{\alpha_j H_j^2}{\pi_j b_j} \sum_{i=j}^{k-1} \Big[\frac{1}{H_i} - \frac{1}{H_{i+1}}\Big] \\ &= \frac{1}{H_k} \sum_{j=0}^{k-1} \frac{H_j(H_k - H_j)}{\pi_j b_j}\, \alpha_j.\end{aligned}$$

The proof is finished. □

By letting $k=m$, $\alpha_j=1$ in (3.4) and noting that $H_m=1$, we get

$$\mathbb{E}\tau=\sum_{j=0}^{m-1}\frac{1}{\pi_j b_j}\,H_j(1-H_j)=T.$$

Again in (3.4), letting $k=m$ and

$$\alpha_j=\sum_{i=0}^{j-1}\frac{1}{\pi_i^* b_i^*}\sum_{k=0}^{i}\pi_k^*=\frac{1}{H_j}\sum_{k=0}^{j-1}\frac{1}{\pi_k b_k}\,H_k(H_j-H_k),$$

we have

$$S=\sum_{j=0}^{m-1}\frac{1-H_j}{\pi_j b_j}\sum_{k=0}^{j-1}\frac{H_k(H_j-H_k)}{\pi_k b_k}. \tag{3.6}$$

Since

$$\frac{\mathbb{E}\tau^2}{(\mathbb{E}\tau)^2}=2-2\,\frac{S}{(\mathbb{E}\tau)^2},$$

Proposition 2.1 yields the following result.

Theorem 3.3 *Separation cut-off occurs if and only if*

$$\lim_{n\to\infty}\frac{S^{(n)}}{[T^{(n)}]^2}=\frac{1}{2}, \tag{3.7}$$

where $T^{(n)}$ *is given in* (3.1) *and*

$$S^{(n)}=\sum_{j=0}^{m_n-1}\frac{1-H_j^{(n)}}{\pi_j^{(n)}b_j^{(n)}}\sum_{k=0}^{j-1}\frac{H_k^{(n)}(H_j^{(n)}-H_k^{(n)})}{\pi_k^{(n)}b_k^{(n)}} \tag{3.8}$$

with $H_j^{(n)}=\sum_{k=0}^{j}\pi_k^{(n)}$.

For the discrete time case, the only difference is that

$$\mathbb{E}\tau^2=\mathbb{E}\tau+2(\mathbb{E}\tau)^2-S.$$

But for discrete time birth and death chains,

$$\mathbb{E}\tau=\mathbb{E}\tau^*\geqslant\sum_{i=0}^{m-1}\frac{1}{b_i^*}\to\infty,\quad m\to\infty,\ \text{as } b_i^*\leqslant 1.$$

Therefore, Theorem 3.3 also holds for stochastically monotone discrete time birth and death chains.

3.3 Stochastic monotonicity

In this section, we will obtain the distribution of the FSST for the birth and death process via the halting state. We will deduce Theorem 3.3 once again.

However, this method can be used to deal with general stochastically monotone Markov processes.

A stochastic matrix $P = (p_{ij}, 0 \leqslant i, j \leqslant m)$ is said to be stochastically monotone if for all i, j with $i \leqslant j$,

$$\sum_{\ell \geqslant k} p_{ik} \leqslant \sum_{\ell \geqslant k} p_{jk}$$

for each k. A Markov chain is monotone if its transition probability matrices are monotone. It is known that the continuous time birth and death process is monotone, and the discrete time birth and death chain is stochastically monotone whenever $b_i + a_{i+1} \leqslant 1$.

Following [13], $i \in E$ is called a halting state for the FSST τ if for all $t \geqslant 0$,

$$\mathbb{P}[\tau > t] = 1 - \frac{\mu_i(t)}{\pi_i} = \mathrm{sep}(\mu(t), \pi). \tag{3.9}$$

Since

$$\sup_j \left(1 - \frac{\mu_j(t)}{\pi_j}\right) = \mathrm{sep}(\mu(t), \pi), \tag{3.10}$$

i is the halting state if and only if for $t \geqslant 0$,

$$\frac{\mu_i(t)}{\pi_i} = \min_j \frac{\mu_j(t)}{\pi_j}. \tag{3.11}$$

We remark that the definition of "halting state" appeared in a draft version of [13], but disappeared in the final version published by AMS. Whatever, we learned this definition from them.

Lemma 3.4 *For the continuous time or stochastically monotone discrete time birth and death process chain on state space $E = \{0, 1, \ldots, m\}$, starting at 0, state m is a halting state for the FSST τ.*

Proof Let

$$p_{ij}(t) = \mathbb{P}[X_t = j \mid X_0 = i].$$

Then $\pi_i p_{ij}(t) = \pi_j p_{ji}(t)$ by the symmetry, so that

$$\frac{\mu_m(t)}{\pi_m} = \frac{p_{0m}(t)}{\pi_m} = \frac{p_{m0}(t)}{\pi_0}.$$

But by the stochastic monotonicity,

$$p_{m0}(t) = \min_i p_{i0}(t).$$

It follows that

$$\frac{\mu_m(t)}{\pi_m} = \min_i \frac{p_{i0}(t)}{\pi_0} = \min_i \frac{p_{0i}(t)}{\pi_i} = \min_i \frac{\mu_i(t)}{\pi_i}.$$

The assertion is proved. □

Next, we will use Lemma 3.4 to give the distribution of the FSST in Laplace transformation form. And then, we use the results in [14,15] to calculate the first two order moments of the FSST by that of hitting times.

Theorem 3.5 *Let τ be the FSST for a discrete time or continuous time birth and death process on $\{0, 1, \dots, m\}$, starting at 0.*

(1) *For the discrete-time case, assume further that the chain is stochastically monotone, that is, $b_i + a_{i+1} \leqslant 1$. Let*

$$\psi_m(s) = \sum_{n=0}^{\infty} \mathbb{P}[X_n = m] s^n.$$

Then

$$\mathbb{E}s^\tau = (1-s)\frac{\psi_m(s)}{\pi_m}, \quad s \in [0,1). \tag{3.12}$$

(2) *For the continuous-time case, by letting*

$$\psi_m(\lambda) = \int_0^\infty \mathrm{e}^{-\lambda t} \mathbb{P}[X_t = m] \mathrm{d}t,$$

we have

$$\mathbb{E}\mathrm{e}^{-\lambda\tau} = \frac{\lambda \psi_m(\lambda)}{\pi_m}, \quad \lambda \geqslant 0. \tag{3.13}$$

Proof (1) Since $\tau \geqslant 1$, for $s \in (0,1)$, we have

$$\mathbb{E}s^\tau = \sum_{k\geqslant 1} \mathbb{P}[\tau = k] s^k = \sum_{k \geqslant 1} \mathbb{P}[\tau = k](1-s)\sum_{l \geqslant k} s^l = (1-s)\sum_{l\geqslant 1} s^l \mathbb{P}[\tau \leqslant l].$$

Hence, by Lemma 3.4 and (3.9), we get

$$\mathbb{E}s^\tau = (1-s)\sum_{n\geqslant 1} s^n \frac{\mu_m(s)}{\pi_m} = (1-s)\frac{\psi_m(s)}{\pi_m}.$$

(2) For $\lambda \geqslant 0$, the integral by parts gives

$$\mathbb{E}\mathrm{e}^{-\lambda\tau} = \int_0^\infty \mathrm{e}^{-\lambda t} \mathrm{d}(\mathbb{P}[\tau \leqslant t]) = \lambda \int_0^\infty \mathrm{e}^{-\lambda t} \mathbb{P}[\tau \leqslant t] \mathrm{d}t.$$

Then by Lemma 3.4 and (3.9), we get

$$\mathbb{E}\mathrm{e}^{-\lambda\tau} = \lambda \int_0^\infty \mathrm{e}^{-\lambda t} \frac{\mu_m(t)}{\pi_m} \mathrm{d}t = \frac{\lambda \psi_m(\lambda)}{\pi_m}.$$

The proof is finished. □

To compute moments of τ of birth and death chains, we will use the construction theory for birth and death chains on boundary. Let τ_k be the hitting time of state k. The following result was proved in [12; Theorem 7.4].

Lemma 3.6 *For the continuous time birth death process X_t, let*

$$\phi_{im}(\lambda) = 0, \quad \phi_{ij}(\lambda) = \int_0^\infty e^{-\lambda t}\mathbb{P}[X_t = j,\, t < \tau_m \mid X_0 = i]\mathrm{d}t, \quad 0 \leqslant i, j < m,$$

and for $0 \leqslant i, j \leqslant m$,

$$\psi_{ij}(\lambda) = \int_0^\infty e^{-\lambda t}\mathbb{P}[X_t = j \mid X_0 = i]\mathrm{d}t.$$

It holds that for $0 \leqslant i, j \leqslant m$,

$$\psi_{ij}(\lambda) = \phi_{ij}(\lambda) + \frac{\xi_i(\lambda)\pi_j\xi_j(\lambda)}{\lambda\sum_{i=0}^m \pi_i\xi_i(\lambda)}, \tag{3.14}$$

where

$$\xi_i(\lambda) = \mathbb{E}[e^{-\lambda\tau_m} \mid X_0 = i].$$

From Lemma 3.6 and Theorem 3.5, we can deduce the relationship between the FSST τ and τ_k.

Theorem 3.7 *For the continuous-time birth and death process starting at* 0,

$$\mathbb{E}e^{-\lambda\tau} = \left(\pi_0 + \sum_{k=1}^n \pi_k(\mathbb{E}e^{-\lambda\tau_k})^{-1}\right)^{-1}, \quad \lambda \geqslant 0. \tag{3.15}$$

Proof By the skip-free property of birth and death processes, we have

$$\xi_0(\lambda) = \mathbb{E}[e^{-\lambda\tau_i} \mid X_0 = 0]\,\mathbb{E}[e^{-\lambda\tau_m} \mid X_0 = i] = \mathbb{E}[e^{-\lambda\tau_i} \mid X_0 = 0]\xi_i(\lambda).$$

Since

$$\xi_m(\lambda) = 1, \quad \phi_{0m}(\lambda) = 0,$$

it follows from (3.14) that

$$\psi_{0m}(\lambda) = \frac{\xi_0(\lambda)\pi_m}{\lambda\sum_{i=0}^m \pi_i\xi_i(\lambda)} = \frac{\pi_m}{\lambda}\frac{1}{\sum_{i=0}^m \pi_i(\mathbb{E}[e^{-\lambda\tau_i} \mid X_0 = 0])^{-1}}.$$

This plus (3.13) gives (3.15). □

From (3.15), we can obtain the moments of FSST τ from that of hitting time τ_k, and the moments of hitting time are explicitly known. By taking derivatives in (3.15) twice, we have

$$\mathbb{E}\tau = \sum_{i=0}^m \pi_i\mathbb{E}\tau_i, \quad \mathbb{E}\tau^2 = 2(\mathbb{E}\tau)^2 + \sum_{i=0}^m \pi_i\mathbb{E}\tau_i^2 - 2\sum_{i=0}^m \pi_i(\mathbb{E}\tau_i)^2. \tag{3.16}$$

And it is known [15] that

$$\mathbb{E}\tau_i = \sum_{j=0}^{i-1}\frac{1}{\pi_j b_j}\sum_{k=0}^j \pi_k, \quad \mathbb{E}\tau_i^2 = 2(\mathbb{E}\tau_i)^2 - 2\sum_{j=0}^{i-1}\frac{1}{\pi_j b_j}\sum_{k=0}^j \pi_k\mathbb{E}\tau_k. \tag{3.17}$$

Combining these formula, we have

$$\mathbb{E}\tau = \sum_{i=0}^{m} \frac{1}{\pi_i b_i} \sum_{j=0}^{i} \pi_j \sum_{j=i+1}^{m} \pi_j, \quad \mathbb{E}\tau^2 = 2(\mathbb{E}\tau)^2 - 2\sum_{i=0}^{m} \pi_i \sum_{j=0}^{i-1} \frac{1}{\pi_j b_j} \sum_{k=0}^{j} \pi_k \mathbb{E}\tau_k.$$

If let

$$S = (\mathbb{E}\tau)^2 - \frac{1}{2}\,\mathbb{E}\tau^2 = \sum_{i=0}^{m} \pi_i \sum_{j=0}^{i-1} \frac{1}{\pi_j b_j} \sum_{k=0}^{j} \pi_k \sum_{u=0}^{k-1} \frac{1}{\pi_u b_u} \sum_{v=0}^{u} \pi_v,$$

and exchange the first summation and the second one, then exchange the third one and the forth one, one can see that S is the same as in (3.6). Again, we derive Theorem 3.3 in case of the continuous time birth and death processes.

As for the discrete time case, in the place of Lemma 3.4, we have the following construction result. The proof is almost the same as in case of the continuous time, thus is omitted here.

Lemma 3.8 *For the discrete time birth death process X_n, let*

$$\phi_{im}(s) = 0, \quad \phi_{ij}(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}[X_n = i, t < \tau_m \mid X_0 = i], \quad 0 \leqslant i, j < m,$$

and for $0 \leqslant i, j \leqslant m$,

$$\psi_{ij}(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}[X_n = i \mid X_0 = i].$$

It holds that for $0 \leqslant i, j \leqslant m$,

$$\psi_{ij}(s) = \phi_{ij}(s) + \frac{\xi_i(s)\pi_j\xi_j(s)}{(1-s)\sum_{i=0}^{m} \pi_i \xi_i(s)}, \tag{3.18}$$

where

$$\xi_i(s) = \mathbb{E}[s^{\tau_m} \mid X_0 = i].$$

From Lemma 3.8, we have the following relationship.

Theorem 3.9 *For the stochastically monotone discrete-time birth and death chain starting at 0, the FSST τ satisfies*

$$\mathbb{E}s^{\tau} = \left(\pi_0 + \sum_{k=1}^{n} \pi_k (\mathbb{E}_0 s^{\tau_k})^{-1}\right)^{-1}, \quad s \in [0, 1). \tag{3.19}$$

By taking derivatives at $s = 1$ twice, we can also get similar relations as in (3.16). Therefore, we deduce that Theorem 3.3 holds in case of discrete time stochastically monotone birth and death chains.

4 Applications and examples

In this section, we will give two applications and some examples. First, we apply [9] to the family of restricted chains from a birth and death chain on the half line. We show that the separation cutoff is closely related to the exponential ergodicity and strong ergodicity. Second, we use the explicit formula in Theorem 3.3 to detect the separation cutoff for the Metropolis algorithm based on the simple symmetric random walk. These results can be comparable with that in [9], where the arguments rely heavily on the exact estimates of the eigenvalues.

4.1 Restricted chains

In this section, we will use the explicit criteria to study a special case: the family of the birth and death chains are the restricted chains taking from a birth and death chain on a countable state space.

Let a_i $(i \geqslant 1)$ and b_i $(i \geqslant 0)$ be the death and birth rates, respectively, for the continuous time birth and death processes on $\{0, 1, \dots\}$. For each $n = 1, 2, \dots$, a restricted process $X_t^{(n)}$ on $\{0, 1, \dots, n\}$ is referred to a birth and death process with birth rates $b_i^{(n)} = b_i$ for $0 \leqslant i \leqslant n-1$ and $b_n^{(n)} = 0$, and death rates $a_i^{(n)} = a_i$ for $1 \leqslant i \leqslant n$. Then $X_t^{(n)}$ is ergodic with reflecting boundary at n. In the discrete time case, we can define the ergodic restricted chains in a similar way.

Corollary 4.1 *Suppose that X_t is a continuous time birth and death process on $E = \{0, 1, 2, \dots\}$. For each $n = 1, 2, \dots$, let $X_t^{(n)}$ be the restricted chain on $E_n = \{0, 1, \dots, n\}$ of X_t with reflecting boundary at n. Assume that X_t is exponential ergodic. Then $X_t^{(n)}$ has separation cutoff if and only if X_t is not strongly ergodic.*

Proof Let λ_1 be the spectral gap for X_t and $\lambda_1 > 0$ by exponential ergodicity. Let $\lambda_1^{(n)}$ as before. Then by [4; Theorem 9.21], $\lim_{n\to\infty} \lambda_1^{(n)} = \lambda_1$. On the other hand, it is easy to see that

$$\begin{aligned}
\sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \sum_{j=i+1}^{\infty} \pi_j &\geqslant \lim_{n\to\infty} T^{(n)} \\
&= T \\
&= \sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \sum_{j=0}^{i} \pi_j \sum_{j=i+1}^{\infty} \pi_j \\
&\geqslant \pi_0 \sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \sum_{j=i+1}^{\infty} \pi_j.
\end{aligned}$$

But the birth and death process is not strongly ergodic if and only if

$$\sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \sum_{j=i+1}^{\infty} \pi_j = \infty.$$

Therefore, the non-strong ergodicity is equivalent to

$$\lim_{n\to\infty} \lambda_1^{(n)} T^{(n)} = \infty,$$

which completes the proof according to [9]. □

Since the strong ergodicity implies the exponential ergodicity, the family of restricted chains of a strongly ergodic birth and death process cannot have the separation cutoff.

The discrete time case is somewhat different and simpler.

Corollary 4.2 *Suppose that X_t is a discrete time birth and death chain on $E = \{0, 1, 2, \dots\}$, which is stochastically monotone and geometrically ergodic. For each $n = 1, 2, \dots$, let $X_t^{(n)}$ be the restricted chain with reflecting boundary at n. Then $X_t^{(n)}$ has separation cutoff.*

Proof From [17], we know that for X_n, the spectral gap $\lambda_1 > 0$ if and only if it is geometrically ergodic. As in the continuous time case,

$$\lim_{n\to\infty} \lambda_1^{(n)} = \lambda_1, \quad \lim_{n\to\infty} T^{(n)} = T.$$

But noting $a_i \leqslant 1$, we get

$$\begin{aligned} T &= \sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \sum_{j=0}^{i} \pi_j \sum_{j=i+1}^{\infty} \pi_j \\ &\geqslant \pi_0 \sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \sum_{j=i+1}^{\infty} \pi_j \\ &\geqslant \pi_0 \sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \pi_{i+1} \\ &= \pi_0 \sum_{i=0}^{\infty} \frac{1}{a_{i+1}} \\ &= \infty. \end{aligned}$$

The proof is finished. □

Let us present some examples.

Example 4.3 For the continuous time case, let $a_i = a$, $b_i = b$, or for discrete time cases, let $a_i = a$, $b_i = b$ with $a + b \leqslant 1$.

(1) If $a > b > 0$, then

$$\lambda_1 = (\sqrt{a} - \sqrt{b}\,)^2.$$

Thus, this process is exponentially ergodic, but not strongly ergodic, so that by Corollary 4.1, there is separation cutoff for the restricted chains.

(2) If $b \geqslant a > 0$, then the process is transient. In this case, we cannot apply Corollary 4.1. Instead, we apply Theorem 3.3. For $\beta := b/a > 1$, we have

$$Z = \frac{\beta^{m+1} - 1}{\beta - 1}, \quad \pi_i = \frac{\beta^i}{Z}.$$

A direct and tedious calculation implies that the separation cutoff occurs.

(3) For $a = b$, we have $\pi_i = 1/m$. Again, an easy calculation shows that there is no separation cutoff.

Example 4.4 For the continuous time, let $a_i = ai^\gamma$, $b_i = b$ with $\gamma \geqslant 0$, $a, b > 0$. Then there is separation cutoff for the restricted chains if and only if $\gamma \in [0, 1]$.

Proof As the death rates $a_i > b$ for i large enough, by comparing with the processes with constant birth and death rates as in Example 4.3, we see that the process is exponentially ergodic. To derive the assertion, we will show that $T = \infty$ if and only if $\gamma \in [0, 1]$. For this, note that

$$\pi_i = \frac{1}{Z} \frac{\lambda^i}{(i!)^\gamma}, \quad Z = \sum_{i=0}^{\infty} \frac{\lambda^i}{(i!)^\gamma},$$

where $\lambda = b/a$. It is easy to see that

$$\sum_{j \geqslant i+1} \pi_j \approx \pi_{i+1}$$

($x \approx y$ means $\exists\, c, C$, $0 < c \leqslant C < \infty$, such that $cy \leqslant x \leqslant Cy$). Thus,

$$\begin{aligned} T &= \sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \sum_{j=1}^{i} \pi_j \sum_{j=i+1}^{\infty} \pi_j \\ &\approx \sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \sum_{j=i+1}^{\infty} \pi_j \\ &\approx \sum_{i=0}^{\infty} \frac{1}{\pi_i b_i} \pi_{i+1} \\ &= \frac{1}{\lambda b} \sum_{i=0}^{\infty} (i+1)^{-\gamma}, \end{aligned}$$

which is infinity if and only if $\gamma \in [0, 1]$. □

4.2 Metropolis algorithm

Another application is the Metropolis algorithm. The base chain is chosen to be the simple symmetric random walk. See [8] for example. For each $m = 1, 2, \dots,$ a Metropolis chain on $\{0, 1, \dots, m\}$ is formulated as follows:

$$a_i = \begin{cases} \dfrac{1}{2}, & \dfrac{\pi_{i-1}}{\pi_i} \geqslant 1, \\ \dfrac{\pi_{i-1}}{\pi_i}, & \dfrac{\pi_{i-1}}{\pi_i} < 1, \end{cases} \quad 1 \leqslant i \leqslant m,$$

$$b_i = \begin{cases} \dfrac{1}{2}, & \dfrac{\pi_{i+1}}{\pi_i} \geqslant 1, \\ \dfrac{\pi_{i+1}}{\pi_i}, & \dfrac{\pi_{i+1}}{\pi_i} < 1, \end{cases} \quad 0 \leqslant i \leqslant m-1.$$

In the following, for positive sequences $\{x_m\}$ and $\{y_m\}$, $x_m \sim y_m$ means

$$\lim_{m\to\infty} \frac{x_m}{y_m} = 1.$$

Corollary 4.5 *Let $\pi_i = (i+1)^d/Z$ with $d \in \mathbb{R}$, where Z makes π a probability measure. For a family of Metropolis chains on $\{0, 1, \dots, m\}$ defined as above, there is no separation cutoff.*

Proof **Case 1** $d \geqslant 0$.

Since π_i is increasing in i, $b_i = 1/2$ for $0 \leqslant i \leqslant m-1$. Note that

$$\sum_{k=1}^{i} k^d \sim \frac{i^{d+1}}{d+1}, \quad i \to \infty.$$

Thus, by definitions of T and S, we have

$$T \sim \frac{m^2}{d+3}, \quad S \sim \frac{m^4}{2(d+3)(d+5)},$$

so that

$$\frac{S}{T^2} \sim \frac{d+3}{2(d+5)} \neq \frac{1}{2}.$$

Case 2 $-1 < d < 0$.

As π_i is decreasing in i, $a_i = 1/2$ for $1 \leqslant i \leqslant m$, using facts

$$\pi_i b_i = \pi_{i+1} a_{i+1} \ (0 \leqslant i \leqslant m-1), \quad \sum_{k=1}^{i} k^d \sim \frac{i^{d+1}}{d+1},$$

we have

$$T = \sum_{i=1}^{m} \frac{H_{i-1}(1-H_{i-1})}{\pi_i a_i} \sim \frac{m^2}{d+3}, \quad S \sim \frac{m^4}{2(d+3)(d+5)},$$

so that

$$\frac{S}{T^2} \not\to \frac{1}{2}.$$

Case 3 $d = -1$.

As in Case 2, we have

$$T = \sum_{i=1}^{m} \frac{H_{i-1}(1 - H_{i-1})}{\pi_i a_i}.$$

Now, using facts

$$\sum_{i=1}^{m} i^{-1} \sim \log m, \quad \sum_{i=1}^{m} i^p (\log i)^q = \frac{m^{p+1} (\log m)^q}{p+1} - \frac{q}{p+1} \sum_{i=1}^{m} i^{p-1} (\log i)^{q-1},$$

we can inductively get

$$T \sim \frac{m^2}{d+3}, \quad S \sim \frac{m^4}{2(d+3)(d+5)},$$

as in Case 1.

Case 4 $d < -1$.

Using the fact

$$\sum_{i=1}^{m} i^d \sim \frac{c - m^{d+1}}{|d+1|} \quad \text{(for some } c = c(d) > 0),$$

we get

$$T \sim \frac{m^2}{1-d}, \quad S \sim \frac{m^4}{4(3-d)(1-d)},$$

so that

$$\frac{S}{T^2} \not\to \frac{1}{2}. \qquad \square$$

From Example 4.3, we obtain the following results for geometric measures.

Corollary 4.6 *Let* $\pi_i = \beta^i / Z$ *with* $\beta > 0$, *where* Z *makes* π *a probability measure. For a family of Metropolis chains on* $\{0, 1, \ldots, m\}$ *defined as above, there is separation cutoff except* $\beta = 1$.

Acknowledgements This work was supported in part by 985 Project, 973 Project (No. 2011CB808000), the National Natural Science Foundation of China (Grant No. 11131003), the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20100003110005), and the Fundamental Research Funds for the Central Universities.

References

1. Aldous D, Diaconis P. Strong uniform times and finite random walks. Adv in Appl Math, 1987, 8(1): 69–97
2. Aldous D, Fill J A. Reversible Markov Chains and Random Walks on Graphs. www.berkeley.edu/users/aldous/book.html, 2003
3. Billingsley P. Probability and Measures. 3rd ed. New York: Wiley, 1995
4. Chen M F. From Markov Chains to Non-Equilibrium Particle Systems. 2nd ed. Singapore: World Scientific, 2004
5. Chen M F. Speed of stability for birth-death processes. Front Math China, 2010, 5(3): 379–515
6. Diaconis P. The cutoff phenomenon in finite Markov chains. Proc Natl Acad Sci USA, 1996, 93(4): 1659–1664
7. Diaconis P, Fill J A. Strong stationary times via a new form of duality. Ann Probab, 1990, 18: 1483–1522
8. Diaconis P, Saloff-Coste L. What do we know about the Metropolis algorithm. J Comput System Sci, 1998, 57: 20–36
9. Diaconis P, Saloff-Coste L. Separation cut-offs for birth and death chains. Ann Appl Probab, 2006, 16(4): 2098–2122
10. Feller W. An Introduction to Probability Theory and Its Applications, II. 2nd ed. New York: Wiley, 1971
11. Fill J A. Strong stationary duality for continuous-time Markov chains, Part I: theory. J Theoret Probab, 1992, 5(1): 45–69
12. Gong Y, Mao Y H, Zhang C. Hitting time distribution for a denumerable birth-death process. J Theoret Probab, 2012, 25: 950–980
13. Levin D A, Peres Y, Wilmer E L. Markov Chains and Mixing Times. Providence: Amer Math Sos, 2008
14. Mao Y H. Algebraic convergence for discrete-time ergodic Markov chains. Science in China (Series A), 2003, 46(5): 621–630
15. Mao Y H. Ergodic degrees for continuous-time Markov chains. Sci China Ser A, 2004, 47(2): 161–174
16. Mao Y H. The eigentime identity for ergodic Markov chains. J Appl Probab, 2004, 41(4): 1071–1080
17. Mao Y H. Convergence rates for reversible Markov chains without the assumption of nonnegative definite matrices. Sci China Math, 2010, 53(8): 1979–1988