Separation cut-offs for birth and death chains

P. Diaconis
Department of mathematics
Stanford University

L. Saloff-Coste*
Department of mathematics
Cornell University

May 27, 2005

Abstract

We give a necessary and sufficient spectral condition for finite continuous time birth and death chains started at 0 to present a separation cut-off.

1 Introduction

Some ergodic Markov chains show a sharp transition in convergence to stationarity. This was observed for random transpositions on the symmetric group by Diaconis and Shahshahani in [20]. The phenomenon was clearly identified in [2] where the term "cut-off phenomenon" was introduced (see, [2, Fig. 2]). Recently Yuval Peres observed that, for many examples, a cut-off occurs if and only if the product $\lambda\tau$ tends to infinity where λ is the spectral gap (i.e., 1 minus the second largest eigenvalue) and τ is the mixing time (i.e., the first time the distance to stationarity is less than $1/4$).

Our main theorem proves a precise version of this statement for all finite continuous time birth and death chains started at 0 when convergence is measured in separation distance. A detailed analysis of the cut-off window is also obtained. The proof uses the duality theory of [14] to convert convergence rates into first hitting time estimates and Keilson's representation of first hitting times as sums of independent exponentials with parameters related to the spectrum of the chain.

The paper is organized as follows. Section 2 discusses various distances and carefully defines the cut-off phenomenon. Section 3 gathers elementary remarks concerning the cut-off phenomenon in separation and total variation distance. Birth and death chains are introduced in Section 4 which reviews duality and Keilson's spectral representation of hitting times. The main results — Theorems 5.1, 5.2 — are stated and proved in Section 5. They give a characterization of the cut-off phenomenon for continuous time birth and death chains s tarted at 0

*Research partially supported by NSF grant DMS 0102126

(in discrete time, the result is restricted to chains satisfying a certain monotonicity condition). Section 6 gives, when it exists, a precise description of the shape of the separation cut-off. This shape may be Gaussian or not. It is Gaussian if and only if the size of the window is of an order of magnitude strictly larger than the relaxation time $1/\lambda$ (i.e., the inverse of the spectral gap). Section 7 gives detailed examples comparing cut-offs in separation, total variation and L^2-distance. These examples includes simple random walks, Metropolis chains, the Bernoulli-Laplace and Erhrenfest chains and simple random walk on distance transitive graphs.

Acknowledgment This work began at the ARCC workshop "Sharp Thresholds for Mixing Times" organised at the American Institute of Mathematics in Palo-Alto in December 2004. We thank the organizers, Amir Dembo, Yuval Peres, David Revelle, and the other participants of the workshop for their helpful comments.

2 Distances and cut-offs

Assume that, to any finite set Ω and any pair of probability measures μ, ν on Ω, is associated a real number $D(\mu, \nu)$ such that $D(\mu, \nu) \in [0, 1]$,

$$\max_{\Omega,\mu,\nu} D(\mu, \nu) = 1$$

and $D(\mu, \nu) = 0$ if and only if $\mu = \nu$. Examples of interest are the total variation distance

$$D(\mu, \nu) = \|\mu - \nu\|_{\mathrm{TV}} = \sup_{A \subset \Omega} \mu(A) - \nu(A),$$

or separation

$$D(\mu, \nu) = \mathrm{sep}(\mu, \nu) = \max_{\omega \in \Omega} \left\{ 1 - \frac{\mu(\omega)}{\nu(\omega)} \right\}.$$

Note that separation is not symmetric and is not a distance between probability measures. Separation was introduced in [2, 3] in the context of the study of convergence of ergodic finite Markov chains.

Consider a sequence of (finite) probability spaces (Ω_n, ν_n), $n = 1, 2, \ldots$, each equipped with a sequence of probability measures μ_n^k, $k = 0, 1, 2, \ldots$, such that

$$\lim_{k \to \infty} D(\mu_n^k, \nu_n) = 0.$$

Definition 2.1 *We say that a family $(\Omega_n, \nu_n, (\mu_n^k)_{k=0,1,\ldots})_{n=1,2,\ldots}$ presents a cut-off (more precisely, a D-cut-off) if there exists a sequence (t_n) of positive reals such that, for any $\epsilon \in (0, 1)$,*

(a) $\lim_{n \to \infty} D(\mu_n^{k_n}, \nu_n) = 0$ *if* $k_n > (1 + \epsilon) t_n$,

(b) $\lim_{n \to \infty} D(\mu_n^{k_n}, \nu_n) = 1$ *if* $k_n < (1 - \epsilon) t_n$.

The next definition introduces the notion of window size for the cut-off phenomenon.

Definition 2.2 *Given sequences $(t_n), (b_n)$ of positive reals, we say that the family*

$$(\Omega_n, \nu_n, (\mu_n^k)_{k=0,1,\ldots})_{n=1,2,\ldots}$$

presents a (t_n, b_n)-cut-off (more precisely, a (t_n, b_n)-D-cut-off) if b_n/t_n tends to zero and

(a) *$f_+(c) = \limsup_{n\to\infty} D(\mu_n^{\lceil t_n + cb_n\rceil}, \nu_n)$ satisfies $\lim_{c\to\infty} f_+(c) = 0$;*

(b) *$f_-(c) = \liminf_{c\to\infty} D(\mu_n^{\lfloor t_n - cb_n\rfloor}, \nu_n)$ satisfies $\lim_{c\to\infty} f_-(c) = 1$.*

Both definitions can be interpreted in an obvious way when the discrete family of measure μ_n^k, $k = 0, 1, \ldots$ is replaced by a continuous family μ_n^t, $t \geq 0$ (in this case, f_+ and f_- are defined using $t_n + cb_n$ without rounding to the next or previous integer). Versions of these definitions were introduced in [3, 13] in the case where the measures μ_n^k, $k = 0, 1\ldots$, are the marginals of a Markov chain on Ω_n with stationary probability ν_n. See also [4] and [30, Sect. 2.4].

Remarks 1. For simplicity we have restricted attention to the case where the maximum of the "distance" D is 1. However, Definitions 2.1 and 2.2 can easily be extended to the case where D is unbounded, e.g., the L^2 distance

$$D(\mu, \nu) = \left(\sum_{\omega\in\Omega} \left|\frac{\mu(\omega)}{\nu(\omega)} - 1\right|^2 \nu(\omega)\right)^{1/2}. \tag{2.1}$$

In this case, in part (b) of each of the two definitions above, simply replace the limit value 1 by ∞. See, e.g., [30, Sect. 2.4].

2. Note that the definitions above do not require that the sequence t_n tends to ∞ (this condition is required in [13, 30], probably for no good reasons. Thanks to Jim Fill for pointing this out to us). For instance, let $\Omega_n = \{1, \ldots, n\}$, $K_n(x, y) = 1/(n-1)$ for all $y \neq x \in \Omega_n$, $\nu_n \equiv 1/n$. This family presents a $(1, \epsilon_n)$ total variation cut-off for any sequence ϵ_n tending to 0. Indeed, $\|K_n^2 - \nu_n\|_{\mathrm{TV}} = 1/(n-1)$ whereas $\|K_n^0 - \nu_n\|_{\mathrm{TV}} = (n-1)/n$.

3. If a family (μ_n^t) with continuous parameter t has both a (t_n)-cut-off and an (s_n)-cut-off then $s_n \sim t_n$ (that is $\lim_{n\to\infty} s_n/t_n = 1$). This is also true for a discrete time family if one of t_n or s_n tends to infinity. However, for a discrete time family having both a (t_n) and an (s_n)-cut-off, the best that can be said in general is that the limit points of the sequence $|t_n - s_n|$ all belong to the interval $[0, 1]$. Because of this, cut-off sequences that do not tend to infinity have to be treated with some special care in discrete time.

Diaconis [13] discusses examples of finite Markov chain cut-offs in detail and asks the following questions: How widespread is the cut-off phenomenon for families of finite ergodic Markov chains and how can one recognize it?

It has been suggested by Yuval Peres that, in some generality, these questions could be answered simply in terms of two parameters, namely, the D-mixing time

$$\tau_n^D = \tau_n^D(\epsilon) = \inf\{k : D(\mu_n^k, \nu_n) \leq \epsilon\}, \tag{2.2}$$

and an appropriately defined notion of spectral gap. Here ϵ is a small fixed parameter (e.g., $\epsilon = 1/4$). In the special case when the chain is reversible (hence diagonalizable with real eigenvalues in $[-1, 1]$), set $\lambda_n = 1 - \beta_n$ where β_n is the second largest eigenvalue. Peres' suggestion is that a D-cut-off phenomenon occurs (say, in continuous time) if and only if the quantity $\lambda_n \tau_n^D$ tends to ∞. This is true in complete generality if D is the L^2 distance defined at (2.1), see [10, 30]. Examples due to David Aldous show that, if D is total variation, the condition $\lambda_n \tau_n^D \to \infty$ does not necessarily imply a cut-off. Still, it is quite possible that the condition that $\lambda_n \tau_n^D$ tends to infinity is sufficient under additional natural assumptions, e.g., for random walks on finite groups or for birth and death chains.

In this paper we consider the case of continuous time birth and death chains on $\{0, 1, \ldots, m_n\}$ started at 0. When D is separation, we show that there is a cut-off if and only if $\lambda_n \tau_n^D$ tends to infinity. This will follow from previous work of Diaconis and Fill [14] who produced optimal strong stationary times through the construction of strong stationary duals and works of Keilson [27] and Brown and Shao [8] linking spectral data to first passage times.

3 Remarks on total variation and separation cut-offs

For $n = 1, 2, \ldots$, let Ω_n be a finite set equipped with a Markov kernel $K_n(x, y)$ with stationary probability ν_n. Fix a starting point $x_n \in \Omega_n$ and consider the sequence of probability measures μ_n^k, $k = 0, 1, \ldots$ where μ_n^k is the distribution of the associated Markov chain started at x_n after k steps.

Let D stands for either the total variation distance or separation. One thing these two notions have in common is that, given the data above, there exists a sequence of real non-negative random variables T_n^D such that

$$D(\mu_n^k, \nu_n) = P(T_n^D > k).$$

When D is total variation, the T_n^D's are "optimal coupling times" whereas when D is separation the T_n^D's are "optimal strong stationary times". See [3] and the references therein. Let t_n and σ_n^2 be respectively the mean and variance of the random variable T_n^D. By a well-known form of Chebychev inequality (e.g., [22, (7.5), p. 152]), for all $a > 0$, we have

$$P(T_n^D > t_n + a\sigma_n) \leq \frac{1}{1 + a^2}, \quad P(T_n^D < t_n - a\sigma_n) \leq \frac{1}{1 + a^2}. \tag{3.1}$$

From these facts we can draw the following conclusions:

(a) The mixing time $\tau_n^D(\epsilon)$ defined at (2.2) satisfies

$$t_n - (\epsilon^{-1}-1)^{-1/2}\sigma_n \le \tau_n^D(\epsilon) \le t_n + (\epsilon^{-1}-1)^{1/2}\sigma_n.$$

(b) If there is a constant $c > 0$ such that $ct_n \ge \sigma_n$ and there is cut-off at time s_n with $\lim_{n\to\infty} s_n = \infty$ then $s_n \sim t_n$. When working in continuous time the conclusion $s_n \sim t_n$ holds true without having to assume that s_n tends to infinity.

(c) If $\sigma_n^{-1}t_n \to \infty$ then there is a (t_n, σ_n)-D-cut-off.

Only part (b) requires a little work and we treat only the continuous time case. Assume there is a cut-off at time s_n and fix $\eta, \epsilon \in (0,1)$. Then, for n large enough, we must have

$$(1-\eta)s_n \le \tau_n^D(\epsilon) \le (1+\eta)t_n.$$

Setting $\epsilon = (1+\eta^2)^{-1}$ and using the first bound in (a), we obtain

$$(1-\eta)s_n \le t_n + \eta\sigma_n \le (1+c\eta)t_n.$$

Using $\epsilon = (1+\eta^{-2})^{-1}$ and the second bound in (a) gives

$$(1-c\eta)t_n \le t_n - \eta\sigma_n \le (1+\eta)s_n.$$

This shows that $s_n \sim t_n$ as desired.

In general, little is known about the times T_n^D so these remarks have only theoretical value. In particular, we know of no non-trivial cases where an optimal coupling time has been constructed in a useful way. In contrast, there are several known examples of optimal strong stationary times to which the remarks above apply (e.g., top to random and riffle shuffles, see [12]).

In this context, the challenge posed by Peres' question is to relate the condition $\sigma_n^{-1}t_n \to \infty$ to spectral information. When that can be done the remarks above may yield useful results. This will be illustrated below.

4 Separation for birth and death chains

Let $\Omega = \{0,\ldots,m\}$. A birth and death chain is a Markov chain K on Ω such that $K(x,y) = 0$ unless $|x-y| \le 1$. Write

$$\begin{array}{rcll} q_x &=& K(x,x-1), & x = 1,\ldots,m \\ r_x &=& K(x,x), & x = 0,\ldots,m \\ p_x &=& K(x,x+1), & x = 0,\ldots,m-1, \end{array}$$

and, by convention, $q_0 = p_m = 0$. We will assume throughout that the chain is irreducible, i.e., that $q_x > 0$ for $0 < x \leq m$ and $p_x > 0$ for $0 \leq x < m$. Such chains have stationary probability

$$\nu(x) = c \prod_{y=1}^{x} \frac{p_{y-1}}{q_y}$$

with $c = \nu(0)$ a normalizing constant. Birth and death chains are in fact reversible, i.e., satisfy

$$\nu(x)K(x,y) = \nu(y)K(y,x).$$

It follows that the operator $K : L^2(\Omega,\nu) \to L^2(\Omega,\nu)$ defined by $f \mapsto Kf = \sum_y K(\cdot,y)f(y)$ is self-adjoint and thus diagonalizable with real eigenvalues in $[-1,1]$. Let λ_i, $i = 0,\dots,m$, be the eigenvalues of $I - K$ in non-decreasing order (I denotes the identity operator). Thus $\lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m \leq 2$. The irreducibility of the chain is reflected in the fact that $\lambda_1 > 0$. It is also well known that $\lambda_m = -1$ if and only if the chain is periodic (of period 2) which happens if and only if $r_x = 0$ for all x. In fact, because we are dealing here with irreducible birth and death chains, it is known that the λ_i's are all distinct (e.g., [8, 27]). Karlin and McGregor [25, 26] observed that the spectral analysis of any given birth and death chain can be treated as an orthogonal polynomial problem. This sometimes leads to the exact computation of the spectrum. See, e.g., [24, 25, 26, 31] and also [28] for a somewhat different approach based on continued fractions.

Given a birth and death chain as above, let μ^k be its distribution after k steps starting at 0. Let γ^t be the distribution at time $t \geq 0$ of the associated continuous time process started at 0, i.e.,

$$\gamma^t = e^{-t} \sum_0^\infty \frac{t^k}{k!}\mu^k.$$

In [14], Diaconis and Fill construct what they call a strong stationary dual for any discrete time birth and death chain satisfying the condition $p_x + q_{x+1} \leq 1$, $0 \leq x < m$, (such chains are called monotone chains). The dual chain is a birth and death chain with the same eigenvalues as the original chain. The first passage time at the extremity m for that dual chain is a strong stationary time for the original chain. The first passage time distribution is explicitly computed by Keilson and by Brown and Shao [27, 8] in terms of the spectral data. Fill [23] treats continuous time chains (the condition $p_x + q_{x+1} \leq 1$, $0 \leq x < m$, is not needed in that case). These works give the following result.

Theorem 4.1 ([14, 23]) *Let K be an irreducible birth and death chain as above.*

(a) For the associated continuous time process started at 0,*we have*

$$\operatorname{sep}(\gamma^t,\nu) = \max_{0\leq x\leq m}\left\{1 - \frac{\gamma^t(x)}{\nu(x)}\right\} = \sum_{i=1}^{m}\prod_{j\neq i}\frac{\lambda_j}{\lambda_i - \lambda_j}e^{-t\lambda_i}.$$

(b) For the discrete time chain, assuming that $p_x + q_{x+1} \le 1$, $0 \le x < m$, we have

$$\operatorname{sep}(\mu^k, \nu) = \max_{0 \le x \le m} \left\{ 1 - \frac{\mu^k(x)}{\nu(x)} \right\} = \sum_{i=1}^{m} \prod_{j \ne i} \frac{\lambda_j}{\lambda_i - \lambda_j} (1 - \lambda_i)^k.$$

Although these are beautiful formulas, it is not so obvious how to use them to derive explicit bounds. However, (a) has a very clear interpretation: its says that separation at time t is the tail of a sum of m independent exponential random variables with respective parameter λ_i, $1 \le i \le m$. Similarly, when all λ_i are in $[0, 1]$, (b) says that separation at time k is the tail of a sum of independent geometric random variables with respective parameter λ_i, $1 \le i \le m$. In particular, we have the following obvious corollary.

Corollary 4.2 *Let K be an irreducible birth and death chain as above.*

(a) For the associated continuous time process started at 0, we have

$$\operatorname{sep}(\gamma^t, \nu) = \max_{0 \le x \le m} \left\{ 1 - \frac{\gamma^t(x)}{\nu(x)} \right\} = P(T > t)$$

where $T = \sum_1^m S_i$ where each S_i is an exponential random variables with parameter λ_i and the S_i's are independent. In particular,

$$E(T) = \sum_1^m \lambda_i^{-1}, \quad \operatorname{Var}(T) = \sum_1^m \lambda_i^{-2}.$$

(b) For the discrete time chain, assuming that $p_x + q_{x+1} \le 1$, $0 \le x < m$, we have

$$\operatorname{sep}(\mu^k, \nu) = \max_{0 \le x \le m} \left\{ 1 - \frac{\mu^k(x)}{\nu(x)} \right\} = P(T > k)$$

where T is a random variable with

$$E(T) = \sum_1^m \lambda_i^{-1}, \quad \operatorname{Var}(T) = \sum_1^m (1 - \lambda_i) \lambda_i^{-2}.$$

The random variable T can be written as a sum $T = \sum_1^m S_i$ where the random variables S_i, $1 \le i \le m$, are independent and S_i is geometric with probability of success λ_i if $\lambda_i \in (0, 1]$ whereas S_i is a Bernoulli variable with parameter λ_i^{-1} if $\lambda_i > 1$.

Remarks 1. The times S_i have no known interpretation in terms of the underlying birth and death chain.

2. Of course the same results applies if the birth and death chain starts at the other extremity m. As the spectral data does not change, it follows that $\operatorname{sep}(\gamma^t(0, \cdot), \nu)$, i.e., the

separation starting from 0, and $\mathrm{sep}(\gamma^t(m,\cdot),\nu)$, i.e., the separation starting from m, are equal at all times! This is in sharp contrast with what happens in total variation distance for which starting at one or the other extremity can lead to very different behaviors.

3. In view of the above results, and from the viewpoint developed in the next few sections, it is interesting to note that for any set of m distinct positive numbers $0 < \lambda_1 < \ldots < \lambda_m \leq 1$, there is a birth and death chain as above with eigenvalues $(0, \lambda_1, \ldots, \lambda_m)$. See [29] and the references therein.

5 Separation cut-off for birth and death chains

We now describe what the previous section entails concerning the cut-off phenomenon. For $n = 1, 2, \ldots$, let $\Omega_n = \{0, 1, \ldots, m_n\}$ be equipped with an irreducible birth and death chain K_n having stationary measure ν_n. Let $q_{n,x}, r_{n,x}, p_{n,x}$ be the corresponding transition probabilities.

Let μ_n^k be the distribution of the associated chain at time k started at 0. Let γ_n^t be the distribution of the continuous time process at time t started 0. Let $\lambda_{n,i} \in [0,2]$, $0 \leq i \leq m_n$, be the corresponding eigenvalues. Set

$$\lambda_n = \lambda_{n,1}, \quad t_n = \sum_1^{m_n} \lambda_{n,i}^{-1}.$$

Finally, for any $\epsilon \in (0,1)$, consider the separation mixing time

$$\tau_n(\epsilon) = \inf\{t : \mathrm{sep}(\gamma_n^t, \nu_n) \leq \epsilon\}.$$

Theorem 5.1 *Referring to the setting and notation introduced above, the family*

$$(\Omega_n, \nu_n, (\gamma_n^t)_{t>0})_{n=1,2,\ldots}$$

has a separation cut-off if and only if $N_n = \lambda_n t_n$ tends to infinity. The separation bounds

$$\mathrm{sep}(\gamma^{(1+c)t_n}, \nu_n) \leq \frac{1}{1+c^2 N_n}, \quad \mathrm{sep}(\gamma^{(1-c)t_n}, \nu_n) \geq 1 - \frac{1}{1+c^2 N_n}, \quad c > 0, \tag{5.1}$$

always holds and, for any fixed $\epsilon \in (0,1)$, the condition $\lambda_n t_n \to \infty$ is equivalent to $\lambda_n \tau_n(\epsilon) \to \infty$.

Proof By Corollary 4.2, we have

$$\mathrm{sep}(\mu_n^t, \nu_n) = P(T_n > t)$$

where T_n has mean t_n and variance $\mathrm{Var}(T_n) = \sigma_n^2$ bounded by

$$\begin{aligned} \sigma_n^2 &= \sum_1^{m_n} \lambda_{n,i}^{-2} = \lambda_n^{-2} \sum_1^{m_n} (\lambda_n/\lambda_{n,i})^2 \\ &\leq \lambda_n^{-2} \left(\sum_1^{m_n} \lambda_n/\lambda_{n,i} \right) = \lambda_n^{-1} t_n. \end{aligned}$$

Here we have simply used the fact that $\lambda_n/\lambda_{n,i} \leq 1$ to obtain the middle inequality. Hence, we have

$$\sigma_n \leq t_n \text{ and, better, } \sigma_n \leq N_n^{-1/2} t_n \tag{5.2}$$

The separation bounds (5.1) follow directly from (5.2) and the Chebychev inequalities (3.1).

Assume that $N_n = \lambda_n t_n \to \infty$. By the second inequality in (5.2), it follows that $t_n/\sigma_n \to \infty$. By (c) of Section 3, there is a separation cut-off at time t_n and, even better, a (t_n, σ_n)-cut-off. Conversely, if there is a cut-off at time s_n then by (5.2) and (b) of Section 3, we must have $s_n \sim t_n$ and there must be a cut-off at time t_n. By (5.1), this imply that N_n tends to infinity.

Now, fix $\epsilon \in (0,1)$. By the upper bound in (a) of Section 3 and the first inequality in (5.2), we have

$$\tau_n(\epsilon) \leq t_n + (\epsilon^{-1} - 1)^{1/2}\sigma_n \leq (1 + (\epsilon^{-1} - 1)^{1/2}) t_n.$$

Hence $\lambda_n \tau_n(\epsilon) \to \infty$ implies $\lambda_n t_n \to \infty$. Conversely, if $t_n \lambda_n \to \infty$, then there is a cut-off at time t_n and by (5.2) and (a) of Section 3, $t_n \sim \tau_n(\epsilon)$. It follows that $\lambda_n \tau_n(\epsilon) \to \infty$. This ends the proof of Theorem 5.1.

Remarks 1. Theorem 5.1 shows that, for continuous time birth and death chains started at 0, a separation cut-off can occurs only if m_n tends to infinity.

2. For D and $(\Omega_n, \nu_n, (\mu_n^k)_{k=0,1,\dots})$ as in Definition 2.1, say that there is a D-precut-off at time s_n if there are constants $0 < c \leq 1 \leq C < \infty$ such that

$$\lim_{n\to\infty} D(\mu_n^{k_n}, \nu_n) \to \begin{cases} 0 & \text{if } k_n \geq C s_n \\ 1 & \text{if } k_n \leq c s_n. \end{cases}$$

Obvious modifications apply in continuous time. Theorem 5.1 shows that there cannot be a separation precut-off if $\lambda_n t_n$ is bounded. Hence, for continuous birth and death chains started at 0 the existence of a separation precut-off is equivalent to the existence of a separation cut-off.

The next result is the discrete time version of Theorem 5.1. It requires the "monotonicity" assumption $p_x + q_{x+1} \leq 1$. Given Theorem 4.2(b), the proof is similar and is omitted.

Theorem 5.2 *Referring to the setting and notation introduced above, assume that for each n and each $x \in \{0, \dots, m_n - 1\}$, we have*

$$p_{n,x} + q_{n,x+1} \leq 1.$$

Then the family

$$(\Omega_n, \nu_n, (\mu_n^k)_{k=0,1,\dots})_{n=1,2,\dots}$$

has a separation cut-off if and only if $N_n = \lambda_n t_n$ tends to infinity.

The so-called monotonicity condition $p_{n,x} + q_{n,x+1} < 1$ implies easily that $r_{n,0} > 0$ thus insuring aperiodicity. It is however a little surprising that negative eigenvalues of K (i.e., $1 - \lambda_i$ with $\lambda_i > 1$) play no role what so ever in Theorem 5.2. As in the continous case, for any fixed $\epsilon \in (0,1)$, the theorem above can be stated using $\tau_n(\epsilon) = \inf\{k : \mathrm{sep}(\mu_n^k, \nu_n) \leq \epsilon\}$ instead of t_n.

6 The shape of the cut-off

When a cut-off is determined, say at time s_n, the next task is to look at the window size. If one is able to establish an (s_n, b_n)-cut-off (possibly adjusting the sequence (s_n)), then the question of the optimality of the window size b_n is posed. One way to answer this question is to obtain the shape of the cut-off, that is, to determine the functions $f_\pm$ of Definition 2.2. If f_+ and $1-f_-$ are non-zero in neighborhoods of $\pm\infty$, then the sequence (b_n) is optimal and the functions $f_\pm$ describe the shape of the cut-off. Only a small number of such results have been established. See, e.g., [13]. In the cases of interest to us in this paper, Corollary 4.2 easily allows us to obtain the shape of the cut-off.

Theorem 6.1 *Referring to a family of the birth and death chains as in Section 5 and using the notation introduced there, assume that $\lambda_n t_n \to \infty$ and set*

$$\sigma_n^2 = \sum_1^{m_n} \lambda_{n,i}^{-2}.$$

(a) Assume that $\lambda_n \sigma_n \to \infty$. Then, for any real c,

$$\lim_{n\to\infty} \mathrm{sep}(\gamma_n^{t_n+c\sigma_n}, \nu_n) = 1 - \Phi(c) \quad \textit{where} \quad \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx.$$

In particular there is a (t_n, σ_n)-cut-off but no (t_n, λ_n^{-1})-cut-off.

(b) Assume that $\lambda_n \sigma_n$ is bounded. Then there is a (t_n, σ_n)-cut-off (equivalently a (t_n, λ_n^{-1})-cut-off) and, for any real $c > 0$, we have

$$\liminf_{n\to\infty} \mathrm{sep}(\gamma_n^{t_n+c\sigma_n}, \nu_n) > 0$$

whereas, for any real $c < 0$,

$$\limsup_{n\to\infty} \mathrm{sep}(\gamma_n^{t_n+c\sigma_n}, \nu_n) < 1$$

Proof We have $\mathrm{sep}(\gamma_n^t, \nu_n) = P(T_n > t)$. Consider the moment generating function

$$M_n(t) = E(e^{t(T_n - t_n)/\sigma_n}).$$

As T_n is a sum of m_n independent exponential random variables with respective parameter $\lambda_{n,i}$, $i = 1, \dots, m_n$, we have

$$M_n(t) = e^{-tt_n/\sigma_n} \prod_{i=1}^{m_n} \frac{\lambda_{n,i}}{\lambda_{n,i} - t/\sigma_n} = e^{F_n(t)}$$

where

$$F_n(t) = -t_n\sigma_n^{-1}t - \sum_{i=1}^{m_n} \log\left(1 - \lambda_{n,i}^{-1}\sigma_n^{-1}t\right) = \sum_{k=2}^{\infty} \frac{\theta_k(n)}{k\theta_2(n)^{k/2}} t^k$$

with

$$\theta_k(n) = \sum_{i=1}^{m_n} (\lambda_{n,1}/\lambda_{n,i})^k.$$

As $\lambda_{n,1}/\lambda_{n,,i} \leq 1$, we have $1 \leq \theta_k(n) \leq \theta_2(n), \;\; k \geq 2$. Hence the series above converges at least for $t \in (-1, 1)$ and

$$0 \leq F_n(t) - t^2/2 \leq \sum_{k=3}^{\infty} \frac{t^k}{k\theta_2(n)^{(k-2)/2}}.$$

If $\lambda_n\sigma_n = \theta_2(n)^{1/2} \to \infty$, then $M_n(t)$ tends to $e^{t^2/2}$ for any real t and it follows that $\sigma_n^{-1}(T_n - t_n)$ is asymptotically distributed as a standard normal random variable. This proves part (a) of Theorem 6.1.

Assume now that $\lambda_n\sigma_n$ is bounded, say $\lambda_n\sigma_n \leq A$. Hence

$$\lambda_n^{-1} \leq \sigma_n \leq A\lambda_n^{-1}$$

and, for any $k = 2, 3, \ldots$,

$$1 \leq \theta_k(n) \leq A.$$

Obviously (e.g., by Chebychev inequality), the distributions of $\sigma_n^{-1}(T_n - t_n)$, $n = 1, 2, \ldots$, form a tight family. Given any subsequence (n_j), we can extract from it a subsequence (n_{j_ℓ}) such that, along that subsequence, $P(T_n > t_n + c\sigma_n)$ converges to $P(T > c)$, $c \in \mathbb{R}$, for some random variable T. Now, from the previous moment generating function computation, it follows that, along (n_{j_ℓ}), the limit

$$\lim_{\ell\to\infty} \theta_k(n_{j_\ell}) = \theta_k \in [1, A]$$

exists, for each $k \geq 2$ and T has moment generating function

$$\exp\left(\frac{t^2}{2} + \sum_{k\geq 3} \frac{\theta_k t^k}{k\theta_2^{k/2}}\right) \tag{6.1}$$

for all $t \in (-\theta_2^{1/2}, \theta_2^{1/2})$. As the variables $\sigma_n^{-1}(T_n - t_n)$ are infinitely divisible, T is also infinitely divisible. By (6.1), the normal component of T is non-trivial and it follows that T has a smooth positive density. Obviously, this implies part (b) of Theorem 6.1. Note that no limit points of the sequence $(T_n - t_n)/\sigma_n$ can be normal. This ends the proof of Theorem 6.1.

Let us observe that the first part of Theorem 6.1(b), i.e., the fact that, for any real $c > 0$,

$$\liminf_{n\to\infty} \operatorname{sep}(\gamma_n^{t_n+c\sigma_n}, \nu_n) > 0,$$

can be proved by a very elementary argument. To bound $P(T_n > t_n + c\sigma_n)$ from below, write

$$\begin{aligned} P(T_n > t_n + c\sigma_n) &\geq P\left(S_{n,1} > \lambda_n^{-1} + (c+1)\sigma_n; \sum_{i=2}^{m_n} S_{n,i} > \sum_{i=2}^{m_n} \lambda_{n,i}^{-1} - \sigma_n\right) \\ &\geq P(S_{n,1} > \lambda_n^{-1} + (c+1)\sigma_n) P\left(\sum_{i=2}^{m_n} S_{n,i} > \sum_{i=2}^{m_n} \lambda_{n,i}^{-1} - \sigma_n\right). \end{aligned}$$

By the second inequality in (3.1), and the fact that $\sigma_n^2 \geq \mathrm{Var}(\sum_{i=2}^{m_n} S_{n,i})$, we have

$$P\left(\sum_{i=2}^{m_n} S_{n,i} > \sum_{i=2}^{m_n} \lambda_{n,i}^{-1} - \sigma_n\right) \geq \frac{1}{2}.$$

It follows that

$$P(T_n > t_n + c\sigma_n) \geq \frac{1}{2} e^{-(\lambda_n^{-1} + (c+1)\sigma_n)\lambda_n} \geq \frac{1}{2} e^{-(1+(c+1)A)}.$$

In contrast, the inequality concerning the lower tail seems harder to prove without the sophisticated tools of infinitely divisible distributions.

Remark In part (b) of Theorem 6.1, assume further that for each k,

$$\theta_k = \lim_{n\to\infty} \left(\frac{\lambda_n}{\lambda_{n,i}}\right)^k < \infty$$

exists. Then for any $c \in \mathbb{R}$ we have

$$\lim_{n \to infty} \mathrm{sep}(\gamma_n^{t_n + c\sigma_n}, \nu_n) = 1 - F(c)$$

where $F(t)$ is the distribution function of an infinitely divisible law whose moment generating function is given in $(-\theta_2^{-1/2}, \theta_2^{-1/2})$ by (6.1). In particular, $0 < F(c) < 1$ for all $c \in \mathbb{R}$. For instance, the Bernoulli-Laplace example in Section 7 has a non-normal cut-off shape in separation.

Remark Continuous time Markov chains offer the freedom to choose a time scale. Starting with a Markov kernel $K(x, y)$ on a countable space, we can consider the continuous time Markov chain generated by the $K - I$ where I denotes the identity matrix. Starting at x, this continuous time chain has probability distribution at time t given by

$$\gamma^t(x, \cdot) = e^{-t} \sum \frac{t^n}{n!} K^n(x, \cdot).$$

However, the most general and natural definition of a countable continuous time Markov chain involves a matrix $Q(x, y)$ (representing the generator) satisfying $\sum_y Q(x, y) = 0$ and $Q(x, y) \geq 0$ if $x \neq y$. In that generality, the quantity $\sum_{y \neq x} Q(x, y)$ does not have to be uniformly bounded (and explosion in finite time is possible). On a finite state space, we can always set $q = \max_x\{-Q(x,x)\}$ and consider the (discrete time) chain with kernel $K(x,y) = I(x,y) + q^{-1}Q(x,y)$. Let $\gamma^t_Q(x,\cdot)$ be the probability distribution of the continuous time Markov chain with generator Q started at x, we have

$$\gamma^t_Q(x,\cdot) = \gamma^{qt}(x,\cdot).$$

Let us now consider a family of continuous time (finite state space) ergodic Markov chains $(\Omega_n, \gamma^t_{Q_n}(x_n,\cdot), \nu_n)$ and consider whether or not this family presents a D-cut-off. The answer to this question is independent of the chosen time scale. Indeed, using the notation introduced above,

$$\gamma^t_{Q_n}(x_n,\cdot) = \gamma_n^{q_n t}(x_n,\cdot).$$

It follows that $(\Omega_n, \gamma^t_n(x_n,\cdot), \nu_n)$ presents a D-cut-off at time t_n (resp. a D-cut-off of type (t_n, b_n)) if and only if $(\Omega_n, \gamma^t_{Q_n}(x_n,\cdot), \nu_n)$ presents a D-cut-off at time t_n/q_n, (resp. a D-cut-off of type $(t_n/q_n, b_n/q_n)$). This remark would not be valid if we had required that (t_n) must tend to infinity in the definition of a cut-off.

7 Examples

This section illustrates our results by looking at various explicit (or not so explicit) families of birth and death chains.

Simple random walk

Consider the simple random walk on $\{0,\dots,n\}$ with $r_0 = r_n = p_0 = p_j = q_j = q_n = 1/2$, $j \in \{1,\dots,n-1\}$. In this case,

$$\lambda_{n,j} = 1 - \cos\frac{\pi j}{n+1}, \quad j = 0,\dots,n.$$

As $\lambda_{n,j} \geq (j/(n+1))^2$ (and this is optimal, up to a multiplicative constant), we see that $\lambda_n^{-1} = \lambda_{n,1}^{-1}$, $t_n = \sum \lambda_{n,j}^{-1}$ and $\sigma_n = (\sum \lambda_{n,j}^{-2})^{1/2}$ are all of order n^2. By Theorems 5.1 and 5.2 there is no separation cut-off (either in continuous or discrete time). Of course, this is well known!

Suppose instead that $r_0 = p_j = p$, $0 \leq j \leq n-1$ and $q_j = r_n = q$, $1 \leq j \leq n$, $p + q = 1$, $0 \leq q \leq p \leq 1$. Then the eigenvalues are

$$\lambda_{n,j} = 1 - 2\sqrt{pq}\cos\frac{\pi j}{n+1}, \quad j = 0,\dots,n.$$

Hence, if p,q are fixed with $0 < q < p < 1$, we have $\lambda_n = 1 - 2\sqrt{pq}$ whereas $t_n = \sum_1^n \lambda_{n,j}^{-1}$ is greater than $n/2$ and σ_n is greater than $\sqrt{n/2}$. Theorems 5.1, 5.2, 6.1 prove the existence of a

(t_n, σ_n) separation cut-off with a normal shape. Observe that

$$t_n = an + O(1), \quad a = \int_0^1 (1 - 2\sqrt{pq}\cos(\pi x))^{-1} dx = \frac{1}{\sqrt{1-4pq}}$$

and

$$\sigma_n^2 = bn + O(1), \quad b = \int_0^1 (1 - 2\sqrt{pq}\cos(\pi x))^{-2} dx.$$

Note that the window of the separation cut-off is not given by $\lambda_n^{-1} \simeq 1$ in this case.

Using diagonalization (e.g., [23, XVI.4]), one finds that (starting at 0) this chain has a $(2cn, 1)$-L^∞ cut-off and a $(cn, 1)$-L^2 cut-off with

$$c = \frac{\log(p/q)}{2(1 - 2\sqrt{pq})}.$$

It is a calculus exercise to check that $c > a$, i.e.,

$$\log(p/q) \geq \frac{2(1-2\sqrt{pq})}{\sqrt{1-4pq}}.$$

Indeed, writing $p = (1+u)/2$, $q = (1-u)/2$, $u \in (0,1)$, we get $1 - 4pq = u^2$ and the above inequality boils down to

$$\frac{\log(1+u) - \log(1-u)}{2u} \geq \frac{1 - \sqrt{1-u^2}}{u^2}$$

which holds true for $u \in (0,1)$ because the left hand side is at least 1 whereas the right hand side is at most 1. Thus the L^2-cut-off occurs later than the separation cut-off (this is not always true: there are many examples where the separation cut-off is twice the L^2-cut-off).

Belsley [6, Chap. V] studies various versions of this chain in detail (in discrete time) and shows that there is a cut-off in total variation at time an with an optimal window of size $\sqrt{n}$. The fact that the total variation cut-off is the same as the separation cut-off can be explained as follows. If the chain starts from the top point n, it is not hard to use the available spectral information to show that it converges in a constant number of steps (most of the mass is around n). To understand total variation starting from the bottom point 0, it will thus be enough to analyse the first time one hits n. By [8, 27], this first hitting time is equal in law to the optimal strong stationary time of Corollary 4.2(a).

Note the very different window sizes, namely, of order 1 for the L^2 and L^∞ cut-offs and of order $\sqrt{n}$ for the separation and total variation cut-off.

Metropolis chains

Now consider an arbitrary probability distribution ν on $\{0, \dots, n\}$ with $\nu(j) > 0$, $j \in \{0, \dots, n\}$. Use the Metropolis algorithm with base chain the simple symmetric random walk above to obtain a birth and death chain with stationary measure ν (see, e.g., [18]). By construction, this chain satisfies the monotonicity condition $p_x + q_{x+1} \leq 1$ and Theorems 5.1, 5.2, 6.1 apply.

For instance, if $\nu(j) = a(1+j)^d$ then the results in [18] show that there is no cut-off (in total variation or separation) and that λ_n is of order n^{-2}. Hence it follows from Theorem 5.1 that $t_n = \sum_1^n \lambda_{n,i}^{-1}$ must be of order n^2, i.e., the eigenvalues λ_i must grow rapidly enough from their minimum of order n^{-2}. We do not know if this can be checked easily by bounding higher eigenvalues. For instance, it does not follows from the (rather sophisticated) eigenvalue bound $\lambda_{n,i} \geq ci^{2/d}n^{-2}$ given by [30, Theorem 3.4.4] and [18].

As a second example, take $\nu(j) = 2^{-n}\binom{n}{j}$. It is proved in ([19]) that, for this example, λ_n is of order $1/n$ and it follows from the proof that in fact $\lambda_{n,i}$ is of order i/n. Hence t_n is of order $n \log n$, $\lambda_n t_n \to \infty$ and $\lambda_n \sigma_n$ is bounded. By Theorems 5.1, 5.2, there is a (t_n, λ_n^{-1}) cut-off in separation (starting form 0). The exact asymptotic behavior of the cut-off time t_n is not known.

Bernoulli-Laplace models

Consider two urns, the left containing r red balls, the right $(n-r)$ black balls, with $0 < 2r \leq n$. At each step, a ball is picked uniformly at random in each urn and the two balls are switched. The process is completely determined by the number of red balls in the right urn and this is a birth and death chain on $\{0, \dots, r\}$. The stationary distribution is

$$\nu_{n,r}(j) = \frac{\binom{r}{j}\binom{n-r}{r-j}}{\binom{n}{r}}$$

and, for $x = \{0, \dots, r\}$, the rates are given by

$$p_x = \frac{(r-x)(n-r-x)}{r(n-r)}, \quad q_x = \frac{x^2}{r(n-r)}, \quad r_x = 1 - p_x - q_x.$$

The eigenvalues of this chain are well known (this goes back at least to Karlin and McGregor [25]; see, e.g., [21]) and given by (with an obvious change in notation)

$$\lambda_{n,r,i} = \frac{i(n-i+1)}{r(n-r)}.$$

Hence, the smallest non-zero eigenvalue is

$$\lambda_{n,r} = \frac{n}{r(n-r)}$$

and we have

$$\begin{aligned} t_{n,r} &= r(n-r)\sum_1^r \frac{1}{i(n-i+1)} = \frac{r(n-r)}{n}\sum_1^r \left(\frac{1}{i} + \frac{1+1/i}{n(1-(i-1)/n)}\right) \\ &= \frac{r(n-r)}{n}(\log r \ + O(1)) \end{aligned}$$

and

$$\theta_2(n,r) = \lambda_{n,r}^{-2}\sigma_{n,r}^2 = \sum_1^r \left(\frac{n}{i(n-i+1)}\right)^2 = O(1).$$

In both cases the $O(1)$ is uniform for all $r \le n/2$ as r tends to infinity. Given this data, Theorem 5.1 shows that, for any sequence (r_ℓ, n_ℓ) with $r_\ell \to \infty$ and $r_\ell < n_\ell/2$, the associated continuous time chain has a (s_ℓ, ξ_ℓ)-separation cut-off with

$$s_\ell = (1 - r_\ell/n_\ell) r_\ell \log r_\ell, \;\; \xi_\ell = (1 - r_\ell/n_\ell) r_\ell.$$

If r_ℓ/n_ℓ tends to zero then

$$t_{n_\ell, r_\ell} = r_\ell \left(\log r_\ell \; + \gamma + o(1)\right) \text{ and } \quad \theta_2(n_\ell, r_\ell) = \frac{\pi^2}{6} + o(1)$$

where γ denotes the well-known γ constant). In this case, a slight variation on Theorem 6.1 shows that the limit shape for the (s_ℓ, ξ_ℓ) separation cut-off is given by the Gumbel distribution (density $\exp(-(x + e^{-x}))$ on $\mathbb{R}$).

The results above should be compared with those of [21, Theorem 2] where an L^2-cut-off of type (ζ_ℓ, ξ_ℓ) is prove with ξ_ℓ as above and

$$\zeta_\ell = \tfrac{1}{2}(1 - r_\ell/n_\ell) r_\ell \log n_\ell = \left(\frac{\log n_\ell}{2 \log r_\ell}\right) s_\ell.$$

We now describe what happens in total variation for this family. This is briefly discussed in [7, Sect. 1.5]. Recall that the total variation distance is easily bounded by both separation and L^2 (see, e.g., [3, 4]). Hence, if there is a total variation cut-off, it is bounded above by

$$\rho_\ell = \min\{s_\ell, \zeta_\ell\} = \begin{cases} s_\ell & \text{if } \; n_\ell \ge r_\ell^2, \\ \zeta_\ell & \text{if } \; n_\ell \le r_\ell^2 \le n_\ell^2/4, \end{cases}$$

i.e, the minimum of the separation cut-off time s_ℓ and the L^2 cut-off time ζ_ℓ. It turns out that this upper bound is sharp and that there is, in fact, a (ρ_ℓ, ξ_ℓ)-cut-off in total variation. We find this phenomenon quite interesting and surprising: For this natural family of chains, separation and L^2 cut-off times cross each other (as functions of the parameters (n,r)) and the total variation cut-off time is given by the minimum.

To prove this we only need a lower bound on total variation matching the uper bound provided by the separation and L^2 results. This lower bound can be obtained by the method introduced in [21] and used there in the case $r = n/2$. Namely, to lower bound total variation between $\gamma_{n,r}^t$ and its stationary measure $\nu_{n,r}$, use a set of the form $A = \{\phi_1 \le \alpha\}$ where ϕ_1 is an eigenfunction associated with the lowest non-zero eigenvalue $\lambda_{n,r} = \lambda_{n,r,1}$. A complete set of eigenfunctions $\{\phi_i : i = 0, \ldots r\}$ ($\phi_i = \phi_{n,r,i}$ associated with $\lambda_{n,r,i}$) is described in [21]. In particular, we can take

$$\phi_0(x) = 1, \;\; \phi_1(x) = 1 - \frac{xn}{n(n-r)}$$

and

$$\phi_2(x) = 1 - \frac{2x(n-1)}{n(n-r)} + \frac{(n-1)(n-2)x(x-1)}{r(n-r)(n-r-1)(r-1)},$$

where $x \in \{0,\dots,r\}$. Given this data, one checks that

$$\phi_1^2 = \frac{1}{n-1}\phi_0 + \frac{n^2-4r(n-r)}{r(n-r)(n-2)}\phi_1 + \frac{n^2(n-r-1)(r-1)}{r(n-r)(n-1)(n-2)}\phi_2.$$

This formula allows us to compute the variance of ϕ_1 under $\gamma^t_{n,r}$ (the variance of ϕ_1 under the stationary measure $\nu_{n,r}$ is $(n-1)^{-1}$). Namely,

$$\begin{aligned}
\mathrm{Var}_{\gamma^t_{n,r}}(\phi_1) &= \frac{1}{n-1} + \frac{n^2-4r(n-r)}{r(n-r)(n-2)}e^{-t\lambda_{n,r,1}} + \frac{n^2(n-r-1)(r-1)}{r(n-r)(n-1)(n-2)}e^{-t\lambda_{n,r,2}} - e^{-2t\lambda_{n,r,1}}\\
&= \frac{1}{n-1} + \frac{n^2-4r(n-r)}{r(n-r)(n-2)}e^{-t\lambda_{n,r,1}}\\
&\quad + \left(\frac{n^2(n-r-1)(r-1)}{r(n-r)(n-1)(n-2)} - 1\right)e^{-t\lambda_{n,r,2}} + (1-e^{-t(2\lambda_{n,r,1}-\lambda_{n,r,2}})e^{-t\lambda_{n,r,2}}\\
&= \frac{1}{n-1} + \frac{n^2-4r(n-r)}{r(n-r)(n-2)}e^{-t\lambda_{n,r,1}}\\
&\quad + \frac{r(n-r)(2-3n)+n^2(n-1)}{r(n-r)(n-1)(n-2)}e^{-t\lambda_{n,r,2}} + (1-e^{-t(2\lambda_{n,r,1}-\lambda_{n,r,2})})e^{-t\lambda_{n,r,2}}\\
&\le C\left(\frac{1}{n-1} + \frac{1}{r}e^{-t\lambda_{n,r,1}} + \frac{t}{r(n-r)}e^{-t\lambda_{n,r,2}}\right).
\end{aligned}$$

For the last term, we have used $1-e^{-u} \le u$, $u \ge 0$, and $2\lambda_{n,r,1} - \lambda_{n,r,2} = \frac{1}{r(n-r)}$. That the difference between $2\lambda_{n,r,1}$ and $\lambda_{n,r,2}$ is small is what makes this proof works.

Now we consider the two cases $r^2 \le n$ and $n \le r^2 \le n^2/4$. In the first case ($r^2 \le n$), set

$$t = \frac{1}{\lambda_{n,r,1}}(\log r \ - c), \quad 0 < c < \log r,$$

and consider $A = \{\phi_1 \le \alpha\}$ with $2\alpha = e^{-t\lambda_{n,r,1}}$. Chebyshev's inequality gives

$$\begin{aligned}
\nu_{n,r}(A) - \gamma^t_{n,r}(A) &\ge 1 - \frac{r^2e^{-2c}}{n-1} - C'\left(\frac{1}{n-1} + \frac{1}{r^2}e^c + \frac{\log r}{r^2 n}e^{2c}\right)r^2e^{-2c}\\
&\ge 1 - C''\left(e^{-c} + \frac{\log n}{n}\right).
\end{aligned}$$

This, together with the earlier separation result, proves the existence of a (s_ℓ, ξ_ℓ) total variation cut-off when $r_\ell^2 \le n_\ell^2$ and r_ℓ tends to infinity.

In the second case ($n \le r^2 \le n^2/4$), set

$$t = \frac{1}{2\lambda_{n,r,1}}(\log n \ - c), \quad 0 < c < \log n,$$

and consider $A = \{\phi_1 \le \alpha\}$ with, again, $2\alpha = e^{-t\lambda_{n,r,1}}$. Now, Chebyshev's inequality gives

$$\begin{aligned} \nu_{n,r}(A) - \gamma^t_{n,r}(A) &\ge 1 - \frac{ne^{-2c}}{n-1} - C'\left(\frac{1}{n-1} + \frac{1}{r\sqrt{n}}e^c + \frac{\log n}{n^2}e^{2c}\right) ne^{-2c} \\ &\ge 1 - C''\left(e^{-c} + \frac{\log n}{n}\right). \end{aligned}$$

This, together with the L^2 result, proves the existence of a (ζ_ℓ, ξ_ℓ) total variation cut-off when $n_\ell \le r_\ell^2 \le n_\ell^2/4$ and r_ℓ tends to infinity.

The hypercube and Hamming chains

Consider the set $\{1, \dots, n\}^r$ and the Markov chain that picks one of the r coordinates uniformly at random and change this coordinate to one of $n-1$ distinct possible values picked uniform at random. Starting from the 0 vector, the number of non-zero coordinates evolves as a birth and death chain on $\{0, \dots, r\}$ with

$$p_x = \frac{(r-x)}{r}, \ q_x = \frac{x}{r(n-1)}, \ r_x = \frac{x(n-2)}{r(n-1)}.$$

The stationary distribution is

$$\nu_{n,r}(x) = \binom{r}{x}(n-1)^x n^{-r}.$$

This chain has eigenvalues (see, e.g., [17, Section 5] for an elementary argument)

$$\lambda_{n,r,i} = \frac{in}{r(n-1)}.$$

Hence,

$$\lambda_{n,r} = \frac{n}{r(n-1)}, \quad t_{n,r} = \frac{r(n-1)}{n}\sum_1^r \frac{1}{i}, \quad \theta_2(n,r) = \lambda_{n,r}^{-2}\sigma_{n,r}^2 = \sum_1^r \frac{1}{i^2}.$$

Fix a sequence (n_ℓ, r_ℓ) with $r_\ell \to \infty$. Then, by Theorem 5.1, the associated continuous time chain has a (s_ℓ, ξ_ℓ) separation cut-off with

$$s_\ell = (1 - 1/n_\ell) r_\ell \log r_\ell, \quad \xi_\ell = (1 - 1/n_\ell) r_\ell.$$

The shape is given by the Gumbell distribution.

As a variation, consider the birth and death chain on $\{0,\dots,r\}$ with

$$p_x = \frac{r-x}{r}, \quad q_x = \frac{x}{r}\theta, \quad r_x = \frac{x}{r}(1-\theta), \quad \theta \in (0,1). \tag{7.1}$$

This has stationary distribution

$$\nu_{\theta,r}(x) = \binom{r}{x}\theta^{r-x}(1+\theta)^{-r}.$$

This is the projection (under the natural action of the symmetric group S_r) of the probability measure

$$\overline{\nu}_{\theta,r}(\overline{x}) = \frac{\theta^{|\overline{x}|}}{(1+\theta)^r}$$

on the hypercube $\{0,1\}^r$ with $\overline{x} = (x_i)_1^r$, $x_i \in \{0,1\}$ and $|\overline{x}| = \sum x_i$. The birth and death chain on $\{0,\dots,r\}$ with rates (7.1) is the projection of a chain K on the hypercube with $K(\overline{x},\overline{x}) = |\overline{x}|(1-\theta)/r$,

$$K(\overline{x},\overline{y}) = \begin{cases} 1/r & \text{if } |\overline{x}| = |\overline{y}|+1 \\ \theta/r & \text{if } |\overline{x}| = |\overline{y}|-1 \end{cases}$$

and $K(\overline{x},\overline{y}) = 0$ otherwise. (see, e.g., [15, Sect. 3]). The eigenvalues are

$$\lambda_{\theta,r,i} = \frac{i}{r}(1+\theta).$$

Hence,

$$\lambda_{\theta,r} = \frac{1+\theta}{r}, \quad t_{\theta,r} = \frac{r}{1+\theta}\sum_1^r \frac{1}{i}, \quad \theta_2(n,r) = \lambda_{\theta,r}^{-2}\sigma_{\theta,r}^2 = \sum_1^r \frac{1}{i^2}.$$

Consider a sequence (θ_ℓ, r_ℓ) with r_ℓ tending to infinity. By Theorem 5.1, the associated continuous time chain has a separation cut-off of type

$$((1+\theta_\ell)^{-1} r_\ell \log r_\ell, r_\ell)$$

(Theorem 5.1 gives a window of size $(1+\theta_\ell)^{-1}r_\ell$ but this is essentially equivalent since $\theta_\ell \in (0,1)$). This chains is studied in [15] and [16, Section 5] (as a certain Metropolis chain on the hypercube, in discrete time). There it is proved that the chain has an L^2 cut-off of type

$$((1+\theta_\ell)^{-1} r_\ell \log\sqrt{r_\ell/\theta_\ell}\,, r_\ell).$$

The reference [11] also proves that, for a fixed θ, there is a $((1+\theta)^{-1} r\log\sqrt{r/\theta}\,, r)$ total variation cut-off has r tends to infinity. Note however that this last result cannot hold true if θ is allowed to vary and tend to 0. In general, for a sequence (θ_ℓ, r_ℓ) with r_ℓ tending to infinity, there is a total variation cut-off of type (ρ_ℓ, r_ℓ) with

$$\rho_\ell = (1+\theta_\ell)^{-1} r_\ell \min\left\{\log r_\ell, \log\sqrt{r_\ell/\theta_\ell}\right\}.$$

As for the Bernoulli-Laplace models, we only need to prove a total variation lower bound matching the upper bound given by the separation and L^2 results. Such a total variation lower bound is easily derived using the data and method of [11, p. 179].

Distance regular graphs

A finite graph is distance transitive if the automorphism group of that graph acts transitively on the set of vertex pairs (x, y) with $d(x, y) = k$, for any k. Distance regular graphs generalize this notion without requiring a group action. See, e.g., [5, 7, 9]. One well-known basic result is that random walk on a distance regular graph can be studied by collapsing to a birth and death chain on $\{0, \ldots, m\}$ started at 0. Here m is the diameter of the graph. In this collapse the set of all eigenvalues (without multiplicities) is conserved and separation, total variation or L^p distance to the stationary measure for any fixed p are conserved. Thus Theorem 5.1 yields the following result.

Theorem 7.1 *Let $\mathcal{G}_n$ be a family of distance regular graphs with diameter m_n tending to infinity as n tends to infinity. On the vertex set V_n of $\mathcal{G}_n$, consider the continuous time simple random walk and let γ^t_{n,x_n} be the law of that process started at a fixed arbitrary point $x_n \in V_n$. Let ν_n be the uniform probability measure on (V_n). Let λ_n be the associated spectral gap and, for any fixed $\epsilon \in (0, 1)$ set,*

$$\tau_n = \inf\{t > 0 : \ \mathrm{sep}(\gamma^t_{n,x_n}, \nu_n) \leq \epsilon\}.$$

Then the family $(V_n, \gamma^t_{n,x_n}, \nu_n)$ has a separation cut-off if and only if $\lambda_n \tau_n \to \infty$.

Remark 1. The separation $\mathrm{sep}(\gamma^t_{n,x_n}, \nu_n)$ does not depend on the starting point x_n so this result can be read as a max-separation result.

2. Note that the family $\mathcal{G}_n$ does not need to be "natural" in any way. It can mix elements from the various natural families described below.

Theorem 5.1 leads to the computation of the cut-off time, when it exists, in terms of the spectrum. It is conjectured by experts that distance regular graphs have been classified. They include the following examples:

(i) The finite circle $\mathbb{Z}/n\mathbb{Z}$ with an edge from x to y if and only if $|x - y| = 1$. This familly has no cut-off.

(ii) Hamming distance graphs such as the hypercube. These have been discussed above.

(iii) The natural graph on r-sets of an n-set with an edge from x to y if $\#(x \cap y) = k - 1$. This is equivalent to the Bernoulli-Laplace models discussed above.

(iv) q-Families: These are described in some detail in [7] with data that is useful for our purpose. These families are all related to certain types of vector-subspaces of a finite dimensional vector space over a finite field $\mathbb{F}_q$ (q a prime power), hence the name. The simplest example is the set of all m dimensional vector subspaces of an n dimensional vector space discussed below.

Regarding the known distance regular graphs from the q-Families (q-DRG for short) listed in [7], we can state the following theorem.

Theorem 7.2 *Referring to the setting of* Theorem 7.1*, assume that the family $\mathcal{G}_n$ is made of some of the known q-DRGs listed in [7]. Then there is a separation cut-off if and only if the diameter m_n of $\mathcal{G}_n$ tends to infinity. Moreover, if m_n tends to infinity then there is a $(m_n, \sqrt{m_n})$ separation cut-off with a non-degenerate normal shape.*

This immediately follows from the data reviewed in [7] and from Theorem 5.1. Instead of going into the details we will illustrate the result with the simplest case of a natural q-DRG family.

Fixed dimension subspaces of a q vector space

Let q be a prime power and $\mathbb{F}_q$ be a finite field of order q. Let E_n be an n-dimensional vector space over $\mathbb{F}_q$. For $m \leq n/2$, let $V_{q,n,m}$ be the set of all m-dimensional vector subspaces of E_n. The finite set $V_{q,n,m}$ is equipped with the distance

$$d(x, y) = m - \dim(x \cap y)$$

and the graph structure for which (x, y) is an edge if and only if $d(x, y) = 1$. The induced graph distance is the distance d and this graph has diameter $m \leq n/2$. The action of $GL_n(\mathbb{F}_q)$ on vector subspaces shows that this is a distance transitive graph hence a distance regular graph. We consider the simple random walk on this graph (started at an arbitrary subspace) in continuous time. Let $\gamma^t_{q,n,m}$ be its law at time t and by $\nu_{q,n.m}$ its (uniform) stationary measure. Details (dealing with the discrete time version) can be found in [7, 11]. As explained in [7, 11] and briefly above, this process can be studied through a birth and death chain on $\{0, \ldots, m\}$ which is simply the associated "distance process". Known computations involving relevant families of orthogonal polynomials give the eigenvalues $\lambda_{q,n,m,i}$ of this chain as

$$\lambda_{q,n,m,i} = \frac{(1 - q^{-i})(1 - q^{i-n-1})}{(1 - q^{m-n})(1 - q^{-m})}, \quad 0 \leq i \leq m.$$

Hence the smallest non-zero eigenvalue is

$$\frac{(1 - q^{-1})(1 - q^{-n})}{(1 - q^{m-n})(1 - q^{-m})}$$

and, when m (hence also n) tends to infinity, we have

$$\sum_{i=1}^{m} \lambda^{-1}_{q,n,m,i} = m + O(1), \quad \sum_{i=1}^{m} \lambda^{-2}_{q,n,m,i} = m + O(1).$$

If we now choose an arbitrary sequence (q_ℓ, n_ℓ, m_ℓ) and consider the familly $(\Omega_\ell, \gamma^t_\ell, \nu_\ell)$ where

$$\Omega_\ell = V_{q_\ell,n_\ell,m_\ell}, \quad \gamma^t_\ell = \gamma^t_{q_\ell,n_\ell,m_\ell}, \quad \nu_\ell = \nu_{q_\ell,n_\ell,m_\ell}$$

then Theorem 5.1 shows that this family has a separation cut-off if and only if m_ℓ tends to infinity. Assuming that m_ℓ tends to infinity, Theorem 6.1 shows that there is a $(m_\ell, \sqrt{m_\ell})$-separation cut-off and that the window size $\sqrt{m_\ell}$ is optimal.

The references [7, 11] (translated into continuous time: in the present case, there are significant differences between discrete and continuous time) give a total variation cut-off of type $(m_\ell, \sqrt{m_\ell})$ and an L^2 cut-off of type $(s_\ell, 1)$ with

$$s_\ell = \tfrac{1}{2} m_\ell (n_\ell - m_\ell) \log q_\ell.$$

References

[1] Aldous, D. *Random walks on finite groups and rapidly mixing Markov chains.* Seminar on probability, XVII, 243–297, Lecture Notes in Math., 986, Springer, Berlin, 1983.

[2] Aldous, D. and Diaconis, P. *Shuffling cards and stopping times.* Amer. Math. Monthly 93, 333–348, 1986.

[3] Aldous, D. and Diaconis, P. *Strong uniform times and finite random walks.* Adv. in Appl. Math. 8, 69–97, 1986.

[4] Aldous, D. and Fill, J. *Reversible Markov chains and random walks on graphs.* Book project. http://www.stat.berkeley.edu/users/aldous/RWG/book.html

[5] Bannai, E and Ito, T. *Algebraic Combinatorics I: Association Schemes.* Benjamin, Menlo Park, CA. 1987.

[6] Belsley, E. *Rates of convergence of Markov chains related to association schemes.* Ph.D. Thesis, Harvard University, 1993.

[7] Belsley, E. *Rates of convergences of random walk on distance regular graphs.* Probab. Theory Relat. Fields 112, 493–533, 1998.

[8] Brown, M. and Shao, Y-S. *Identifying coefficients in the spectral representation for first passage time distributions.* Probab. Engrg. Inform. Sci. 1,69–74, 1987.

[9] Brouwer, A. E.; Cohen, A. M.; Neumaier, A. *Distance-regular graphs.* Ergebnisse der Mathematik und ihrer Grenzgebiete (3), 18. Springer-Verlag, Berlin, 1989.

[10] Chen, G-Y. and Saloff-Coste, L. *The cut-off phenomenon for finite Markov chains.* Preprint. 2005.

[11] D'Aristotle, A. *The nearest neighbor random walk on subspaces of a vector space and rate of convergence.* J. Theoret. Probab. 8-2, 321–346, 1993.

[12] Diaconis, P. *Group representations in probability and statistics.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.

[13] Diaconis, P. *The cutoff phenomenon in finite Markov chains.* Proc. Nat. Acad. Sci. U.S.A. 93-4, 1659–1664, 1996

[14] Diaconis, P. and Fill, J. *Strong stationary times via a new form of duality.* Ann. Probab. 18-4, 1483–1522, 1990.

[15] Diaconis, P. and Hanlon, P. *Eigen analysis for some examples of the Metropolis algorithm.* In "Hypergeometric functions on domains of positivity, Jack polynomials, and applications" (Tampa, FL, 1991), 99–117, Contemp. Math., 138, Amer. Math. Soc., 1992.

[16] Diaconis, P. and Ram, A. *Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques.* Michigan Math. J. 48, 157–189, 2000.

[17] Diaconis, P. and Saloff-Coste L. *Comparison techniques for random walk on finite groups.* Ann. Probab. 21-4, 2131–2156, 1993.

[18] Diaconis, P.and Saloff-Coste, L. *What do we know about the Metropolis algorithm?* 27th Annual ACM Symposium on the Theory of Computing (STOC'95) (Las Vegas, NV). J. Comput. System Sci. 57-1, 20–36, 1998.

[19] Diaconis, P.and Saloff-Coste, L. *Logarithmic Sobolev inequalities for finite Markov chains.* Ann. Appl. Probab. 6-3, 695–750, 1996.

[20] Diaconis, P. and Shahshahani, M. *Generating a random permutation with random transpositions.* Z. Wahrsch. Verw. Gebiete 57-2, 159–179, 1981.

[21] Diaconis, P. and Shahshahani, M. *Time to reach stationarity in the Bernoulli-Laplace diffusion model.* SIAM J. Math. Anal. 18-1, 208–218, 1987.

[22] Feller, W. *An introduction to probability theory and its applications. Volume II* (Sec. Ed.), Wiley & Sons, 1966,1971.

[23] Fill, J. *Strong Stationary dualiy for continuous-time Markov chains. Part I: Theory.* J. Theoret. Probab. 5-1, 45–70, 1992.

[24] Ismail, M.; Masson, D.; Letessier, J.; Valent, G. *Birth and death processes and orthogonal polynomials.* In "Orthogonal polynomials (Columbus, OH, 1989)", 229–255, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 294, Kluwer Acad. Publ., Dordrecht, 1990.

[25] Karlin, S. and McGregor, J. *The Hahn polynomials, formulas and and application.* Scripta Math. 26, 33–46, 1961.

[26] Karlin, S. and McGregor, J. *Ehrenfest urn models.* J. Appl. Probability 2, 352–376, 1965.

[27] Keilson, J. *Markov chain models—rarity and exponentiality.* Applied Mathematical Sciences, 28. Springer-Verlag, New York-Berlin, 1979.

[28] Parthasarathy, P.R. and Lenin R.B. *On the exact transient solution of finite birth and death processes with specific quadratic rates.* Math. Sci. 22, 92–105, 1997.

[29] Parthasarathy, P.R. and Lenin R.B. *An inverse problem in birth and death processes.* Comput. Math. Appl. 38, 33–40, 1999.

[30] Saloff-Coste, L. *Lectures on finite Markov chains.* Lectures on probability theory and statistics (Saint-Flour, 1996), 301–413, Lecture Notes in Math., 1665, Springer, Berlin, 1997.

[31] Van Assche, W.; Parthasarathy, P.R.; Lenin, R.B. *Spectral representation of four finite birth and death processes.* Math. Sci. 24, 105–112, 1999.