



HAL
open science

Optimal stopping time and halting set for total variation distance

Agnes Coquio

► **To cite this version:**

| Agnes Coquio. Optimal stopping time and halting set for total variation distance. 2015. hal-01136083

HAL Id: hal-01136083

<https://hal.science/hal-01136083v1>

Preprint submitted on 31 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal stopping time and halting set for total variation distance

March 31, 2015

Abstract

An aperiodic and irreducible Markov chain on a finite state space converges to its stationary distribution. When convergence to equilibrium is measured by total variation distance, there exists an optimal coupling and a maximal coupling time. In this article, the maximal coupling time is compared to the hitting time of a specific state or set. Such sets, named halting sets, are studied in the case of symmetric birth-and-death chains and in some other examples. Some applications to the cutoff phenomenon are given. These results yield new methods to calculate cutoff times for some monotone birth-and death chains without the lazy hypothesis .

Key words. Markov chains, total variation distance, birth-and-death chains, stochastic monotonicity, passage time, eigenvalues

1 Introduction

Let \mathbb{X} be a finite space, $(X_t)_{t \geq 0}$ a discrete or continuous time irreducible Markov chain on \mathbb{X} . Let π be its stationary distribution . For $t \geq 0$, denote by π_t the distribution of X_t . Let d be a distance between probability distributions on \mathbb{X} . For $t \geq 0$, denote by $d(t)$ the distance between the distribution of X_t and π :

$$d(t) = d(\pi_t, \pi)$$

For $A \subset \mathbb{X}$, denote by T_A the hitting time of A .

Definition 1.1. *A set $A \subset \mathbb{X}$ is a halting set for the distance d and the Markov chain $(X_t)_{t \geq 0}$ if for all $t \geq 0$,*

$$d(t) \leq \mathbb{P}(T_A > t)$$

If the Markov chain is also aperiodic in the discrete time case, $d(t)$ goes to 0 when t goes to infinity. The question of non-asymptotic behavior is hard. There are several techniques for obtaining explicit bounds on $d(t)$: Fourier analysis, coupling, strong stationary times, . . .

Several of them use hitting times.

Strong stationary times are allied with separation distance, $s(t) = \max_{x \in X} (1 - \frac{\pi_t(x)}{\pi(x)})$. T a randomized stopping time is a strong stationary time if T and X_T are independent and if the law of X_T is given by the stationary distribution π . In [2], Aldous and Diaconis prove that there exists a strong stationary time T such that for all $t \geq 0$, $s(t) = \mathbb{P}(T > t)$. Furthermore, in [7], Diaconis and Fill construct an absorbing dual Markov chain X^* such that T is the hitting time

of an absorbing state for X^* . This is used by Diaconis and Saloff-Coste in [9] for birth-and-death chain to prove some cut-off results.

Let T be a strong stationary time. A state y^* is a halting state for T if $X_t = y^*$ implies $T \leq t$ or equivalently if $T \leq T_{y^*}$. Furthermore, if there exists a halting state for T , then $s(t) = \mathbb{P}(T > t)$ for all t . This is an example of the use of hitting times to obtain bounds on a distance.

In coupling methods, hitting times are very often used to bound $\mathbb{P}(T > t)$ where T is the coupling time. Several examples can be found in the book of Levin, Peres and Wilmer [17] p.65.

In [19], Martinez and Ycart also use hitting times to obtain bounds on the total variation distance. They prove that in general, the access time to equilibrium and the hitting times tend to be equivalent, if the process starts “far away”. But the context is different, they consider continuous time Markov chain on a countable set I and they study $\|\delta_a P_t - \pi\|_{TV}$ when a goes to infinity.

In this article, mainly the total variation distance is used, that is:

$$d(t) = \|\pi_t - \pi\|_{TV} := \sup_{A \subset X} |\pi_t(A) - \pi(A)| = \frac{1}{2} \sum_{x \in X} |\pi_t(x) - \pi(x)|$$

Find some bounds on $d(t)$ is equivalent to bound the mixing time of the chain. For $\epsilon \in]0, 1[$, denote by $t_{mix}(\epsilon)$, defined by

$$t_{mix}(\epsilon) = \min\{t \geq 0 / d(t) \leq \epsilon\}.$$

In [17] (theorem 10.14 p 134), the authors established the following theorem for a lazy chain :

$$t_{mix}\left(\frac{1}{4}\right) \leq 2 \max_{x \in X} \mathbb{E}_\pi[T_x] + 1$$

In fact, they prove that $d(t) \leq \sqrt{\frac{\mathbb{E}_\pi[T_x]}{8t}}$ where $\pi_0 = \delta_x$.

In [10], Ding, Lubetzky and Peres prove the following result for a lazy birth-and-death chain on $\{0, \dots, N\}$ started at 0,

$$\forall 0 \leq l \leq N, \forall n \in \mathbb{N}, \|\pi_n - \pi\|_{TV} \leq \mathbb{P}(T_l > n) + \pi(\{l+1, \dots, N\}) \quad (1)$$

So it is a natural idea and often used method to compare the access time to equilibrium with the hitting time of a given state. This state will depend on the initial state.

One object of this article is to look at “minimal” coupling time for the total variation distance, mimicking in some meaning what is done for the separation distance. The construction of Aldous and Diaconis of a randomized stopping time satisfying $s(n) = \mathbb{P}(T > n)$ for all n leads naturally to a strong stationary stopping time. So rewriting what can be found in [14] and [21], a randomized stopping time T is constructed satisfying $d(t) = \mathbb{P}(T > t)$ for all t and the properties of the law of (X_T, T) lead to the following result:

Theorem 1.2. *Let $\mathcal{M} = \{x \in X / \forall t \geq 0, \pi_t(x) \leq \pi(x)\}$, then \mathcal{M} is a halting set for the total variation distance and the chain $(X_t)_{t \geq 0}$.*

In the following, we omit to say *for the total variation distance and the chain $(X_t)_{t \geq 0}$* . Of course if $x^* \in \mathcal{M}$, $\{x^*\}$ is a halting set and so for all $t \geq 0$, $d(t) \leq \mathbb{P}(T_{x^*} > t)$, a property to be compared to (1). A such state is called a halting state.

The difficulty then will be to find halting sets. The first fact is that if we have some property of monotonicity, a halting set must exist (proposition (3.2).

Halting sets can be used to prove some results about cutoff. This concepts was introduced by Aldous and Diaconis in [1] to describe the fact that many ergodic Markov chains converge abruptly to their stationary distribution.

Consider a family $(X_n^{(N)})_{n \geq 0}$ of aperiodic irreducible Markov chains on a finite state space $X^{(N)}$, each with its stationary distribution $\pi^{(N)}$ and its distance from stationary $d_n^{(N)}$. t_N is a cutoff if

$$\begin{aligned} d^{(N)}(\lceil ct_N \rceil) &\longrightarrow 0 & \text{if } c > 1 \\ d^{(N)}(\lceil ct_N \rceil) &\longrightarrow 1 & \text{if } c < 1 \end{aligned}$$

Diaconis and Saloff-Coste in [9] verified this conjecture for continuous-time birth-and-death chains, started at an endpoint, with convergence measured in separation.

Ding, Lubetzky and Peres in [10] proved it for continuous-time birth-and-death chains and lazy discrete-time birth-and-death chains, started at an endpoint, with convergence measured in total variation. For such chains, the fact that the product of the mixing time and the spectral gap tends to infinity is equivalent to the fact that the product of the expected value of the hitting time of the median of the stationary distribution and the spectral gap tends to infinity.

The cutoff depends of the distance. For example, the Ehrenfest process starting at 0 on $\{0, \dots, N\}$ has a $\frac{N \ln N}{2}$ separation cutoff but a $\frac{N \ln N}{4}$ total variation and L^2 cutoff. The biased $(p; q)$ -random walk starting at 0 on $\{0, \dots, N\}$ has a $\frac{N}{p-q}$ total variation and separation cutoff but a $\frac{\ln(\frac{p}{q})}{2(1-2\sqrt{pq})} N$ L^2 cutoff.

Recently, Basu, Hermon and Peres in [4], study the link between cutoff and some concentration of hitting time of “worst” sets of stationary measure at least α , for some $\alpha \in]0, 1[$. They prove that in the case of a lazy reversible irreducible Markov chain, cutoff is equivalent to a notion of cutoff for hitting times, denoted hit_α -cutoff. Here, we use explicit hitting time, the hitting time of a halting state for total variation.

Recall the definition of a cutoff window :

If (w_N) and (t_N) are two sequences such that $w_N = o(t_N)$, one may define that the family of chain $(X_n^{(N)})_{n \geq 0}$ exhibits a cutoff at t_N with window w_N if

$$\begin{aligned} \lim_{\gamma \rightarrow +\infty} \underline{\lim}_{N \rightarrow +\infty} d^{(N)}(\lceil t_N - \gamma w_N \rceil) &= 1 \\ \lim_{\gamma \rightarrow +\infty} \overline{\lim}_{N \rightarrow +\infty} d^{(N)}(\lceil t_N + \gamma w_N \rceil) &= 0 \end{aligned}$$

A general result can be the following in the discrete time:

Proposition 1.3. *Let $A^{(N)}$ be a halting set such that $\sigma(T_{A^{(N)}}) = o(\mathbb{E}[T_{A^{(N)}}])$. Let $\lambda_{A^{(N)}}$ be the largest eigenvalue of the restriction of the transition kernel of the chain to $\mathbb{X} \setminus A^{(N)}$.*

If $(1 - \lambda_{A^{(N)}})\mathbb{E}[T_{A^{(N)}}]$ goes to infinity when N goes to infinity, then the chain has a cutoff at time $\mathbb{E}[T_{A^{(N)}}]$ for total variation.

If furthermore, $\underline{\lim}(1 - \lambda_{A^{(N)}})\sigma(T_{A^{(N)}}) > 0$, then the chain has a cutoff at time $\mathbb{E}[T_{A^{(N)}}]$ with window $\sigma(T_{A^{(N)}})$ for total variation.

The rest of the paper is to determine halting set in several examples, essentially in the case of birth-and-death chains. A Markov chain on $\{0, \dots, 2N+1\}$ on $\{0, \dots, N\}$ with transitions given by $K = (k(x, y))_{x, y \in X}$ is symmetric if for all $x, y \in X$, $k(N-x, N-y) = k(x, y)$. An example of result is the following: In the case of a monotone and symmetric birth-and-death chain on $\{0, \dots, 2N+1\}$, $N+1$ is a halting state for total variation. Furthermore, if $1 = \lambda_0 > \lambda_1^{(N)} >$

$\dots > \lambda_{2N+1}^{(N)}$ are the eigenvalues of the transition kernel and if $(1 - \lambda_1^{(N)}) \sum_{k=1}^{2N+1} \frac{1}{1 - \lambda_k^{(N)}}$ goes to

infinity when N goes to infinity, then the chain has a cutoff at time $\frac{1}{2} \sum_{k=1}^{2N+1} \frac{1}{1 - \lambda_k^{(N)}}$ for total variation.

The paper is organized as follows. In section 2, the construction of an optimal randomized stopping time for the total variation distance for discrete and continuous times markov chains is given. Section 3 exposes the notion of a halting set for total variation for a chain with an initial distribution and gives applications to cutoff. In section 4, general results about monotone birth-and-death Markov chain are given and the case of symmetric monotone birth-and-death Markov chain is studied. These results are used to prove cutoff in several examples.

2 Optimal stopping time

2.1 Construction of an optimal stopping time in discrete time

This construction is not new but the presentation made here is different from Lindvall's proof in [21].

Let $(X_n)_{n \geq 0}$ be a Markov chain on a finite space \mathbb{X} with a stationary probability π such that for all $x \in X$, $\pi(x) > 0$.

The transition kernel is denoted by $K = (k(x, y))_{x, y \in X}$. Let π_0 be the distribution of X_0 and for $n \geq 1$, $\pi_n = \pi_0 K^n$ be the distribution of X_n . Let $d_n = \|\pi_n - \pi\|_{TV}$ be the total variation distance between π_n and π .

We define for $n \geq 0$, $\gamma_n : X \rightarrow \mathbb{R}_+$ by

$$\gamma_0(x) = \pi_0 \wedge \pi(x) \quad \text{and for } n \geq 1, \gamma_n(x) = (\pi_n \wedge \pi)(x) - (\pi_{n-1} \wedge \pi)K(x)$$

$$\text{As } d(n) = 1 - \sum_{x \in X} (\pi_n(x) \wedge \pi(x)), \text{ for all } n \geq 1, d(n-1) - d(n) = \sum_{x \in X} \gamma_n(x).$$

By consequence, we are looking for a randomized stopping time T which satisfies for all $n \geq 0$, for all $x \in X$, $\mathbb{P}(X_n = x, T = n) = \gamma_n(x)$.

Proposition 2.1. *Let Δ_n be defined by*

$$\begin{cases} \Delta_0 = \pi_0 \\ \Delta_{n+1} = \pi_{n+1} - (\pi_n \wedge \pi)K \end{cases} \quad (2)$$

Let Ψ_n be defined for all $n \geq 0$ by

$$\psi_n(x) = \begin{cases} \frac{\gamma_n(x)}{\Delta_n(x)} & \text{if } \Delta_n(x) \neq 0 \\ 1 & \text{if } \Delta_n(x) = 0 \end{cases} \quad (3)$$

For all $n \geq 0$, ψ_n takes value in $[0, 1]$.

Let $(U_n)_{n \geq 0}$ be independent variables uniformly distributed on $[0, 1]$, independent of $(X_n)_{n \geq 0}$.

Let T be the randomized stopping time defined by $T = \inf\{n \geq 0 / U_n \leq \psi_n(X_n)\}$.

Then $\forall n \geq 1, \forall x \in \mathbb{X}$,

$$\mathbb{P}(X_n = x, T = n) = \gamma_n(x) \quad (4)$$

$$\mathbb{P}(X_n = x, T \geq n) = \Delta_n(x), \mathbb{P}(X_n = x, T > n) = \pi_n(x) - (\pi_n \wedge \pi)(x) \quad (5)$$

$$\mathbb{P}(T > n) = d(n) \quad (6)$$

Proof. As $0 \leq \gamma_n \leq \Delta_n$, ψ_n takes value in $[0, 1]$ and T is well defined.

We prove (5) inductively :

$$\begin{aligned} \Delta_0(x) &= \pi_0(x) = \mathbb{P}(X_0 = x, T \geq 0) \\ \mathbb{P}(X_{n+1} = x, T \geq n+1) &= \sum_{y \in \mathbb{X}} \mathbb{P}(X_n = y, X_{n+1} = x, T \geq n, U_n > \psi_n(y)) \\ &= \sum_{y \in \mathbb{X}} (1 - \psi_n(y)) \mathbb{P}(X_n = y, T \geq n) k(y, x) \\ &= ((1 - \psi_n) \Delta_n) K(x) = (\Delta_n - \psi_n \Delta_n) K(x) \\ &= (\Delta_n - \gamma_n) K(x) = \Delta_{n+1}(x) \end{aligned}$$

The second equality comes from the Markov property and definition of T . (4) and (6) are easy consequences of (5). □

Remark 2.2. ψ_n is given by the following formulae:

$$\psi_n(x) = \begin{cases} \frac{\gamma_n(x)}{\Delta_n(x)} = \frac{\pi_n \wedge \pi(x) - (\pi_{n-1} \wedge \pi) K(x)}{\pi_n(x) - (\pi_{n-1} \wedge \pi) K(x)} & \text{if } \Delta_n(x) \neq 0 \\ 1 & \text{if } \Delta_n(x) = 0 \end{cases} \quad (7)$$

Remark 2.3. An alternative definition can be made for the randomized stopping time T .

We define for all $n \geq 0$

$$J_n = \prod_{k=0}^n (1 - \psi_k(X_k)) \quad (8)$$

and $T = \inf\{n \geq 0 / U \geq J_n\}$ where U is a random variable uniformly distributed $[0, 1]$ independent of $(X_n)_{n \geq 0}$.

Remark 2.4. To study distance from stationarity for Markov chains, separation distance has good properties.

The separation distance is defined by

$$s(n) := \text{sep}(\pi_n, \pi) = \sup_{y \in \mathbb{X}} s(n, y) \quad \text{where} \quad s(n, y) = 1 - \frac{\pi_n(y)}{\pi(y)}$$

and satisfies $d(n) \leq s(n)$.

The following result was established by Aldous and Diaconis (1987) :

- If T is a strong stationary time, for all $n \in \mathbb{N}$, $s(n) \leq \mathbb{P}(T > n)$.
- Conversely, there exists a strong stationary time T such that there is equality for all $n \in \mathbb{N}$.

The construction of the stochastically optimal strong stationary time is the following :

$$\gamma_n(x) = (s(n-1) - s(n))\pi(x), \text{ so as before } \sum_{x \in \mathbb{X}} \gamma_n(x) = s(n-1) - s(n)$$

With the method of proposition 2.1 with $\psi_0(x) = (1 - s(0)) \frac{\pi(x)}{\pi_0(x)}$ on the support of π_0 , we obtain $\mathbb{P}(X_n = x, T = n) = (s(n-1) - s(n))\pi(x)$. So T is an optimal strong stationary time.

Remark 2.5. We see in the preceding remark that in the cases of separation and total variation, we can construct a randomized stopping time T such that $d(n) = \mathbb{P}(T > n)$ for all $n \in \mathbb{N}$, the construction depends of the choice of γ_n such that $\sum_{x \in \mathbb{X}} \gamma_n(x) = d(n-1) - d(n)$.

For example, if for the total variation distance, we take $\gamma_n(x) = (d(n-1) - d(n))\pi(x)$ as for separation, we can prove that it does not exist ψ_n with values in $[0, 1]$ which satisfies $\Delta_n \psi_n = \gamma_n$.

We can deal with others distances and the first condition is to have $n \rightarrow d(n)$ decreasing.

For any convex function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$, one may define the f -divergence distance on $\mathcal{P}(E)$ by

$$d_f(\mu, \pi) = \sum_{x \in \mathbb{X}} \pi(x) f\left(\frac{\mu(x)}{\pi(x)}\right).$$

These measures are studied in [18] by Liese-Vajda.

We suppose that $f(1) = 0$, so $d_f(\mu, \pi) \geq 0$.

Some examples :

- The total variation distance given by $f(x) = \frac{1}{2}|x - 1|$.
- The relative entropy given by $f(x) = x \ln(x)$.
- The Hellinger distance given by $f(x) = \frac{1}{2}(1 - \sqrt{x})^2$.
- The χ^2 distance given by $f(x) = \frac{1}{2}(1 - x)^2$.

We note $d_f(n) = d_f(\pi_n, \pi)$.

Let $\gamma_n : \mathbb{X} \rightarrow \mathbb{R}$ be defined by $\gamma_n(x) = (\pi f(\frac{\pi_{n-1}}{\pi}))K - f(\frac{\pi_n}{\pi})\pi$. So $d_f(n-1) - d_f(n) = \sum_{x \in \mathbb{X}} \gamma_n(x)$. The convexity of f implies that $\gamma_n(x) \geq 0$.

In all generality, $d_f(n)$ can be superior to 1 and so it will be impossible to find a randomized stopping time T such that for all $n \geq 0$, $\mathbb{P}(T > n) = d(n)$.

So a constant $c > 0$ and a function $\psi_0 : \mathbb{X} \rightarrow [0, 1]$ have to be found such that for all $n \geq 0$, $\mathbb{P}(X_n = x, T = n) = \frac{\gamma_n(x)}{c} = \psi_n(x)\Delta_n(x)$.

If we want $\mathbb{P}(T < +\infty) = 1$, it is necessary to choose ψ_0 such that

$$1 = \mathbb{P}(T = 0) + \sum_{n \geq 1} \mathbb{P}(T = n) = \sum_{x \in \mathbb{X}} \psi_0(x)\pi_0(x) + \frac{1}{c}d_f(0)$$

By mimicking the preceding calculus, for all $n \geq 1$,

$$\Delta_n(x) = \pi_n(x) - \left[(\psi_0 \pi_0) + \frac{1}{c}f\left(\frac{\pi_0}{\pi}\right)\pi \right] K^n(x) + \frac{1}{c}f\left(\frac{\pi_n}{\pi}\right)\pi(x) + \frac{1}{c}\gamma_n(x)$$

So the condition $\Delta_n \geq \frac{1}{c}\gamma_n$ implies that

$$f\left(\frac{\pi_n}{\pi}\right)\pi \geq -c\pi_n + \left[c(\psi_0 \pi_0) + f\left(\frac{\pi_0}{\pi}\right)\pi \right] K^n$$

In the case where $\pi_0 = \delta_{x_0}$, this condition becomes $f\left(\frac{\pi_n}{\pi}\right) \geq f(0)\left(1 - \frac{\pi_n}{\pi}\right)$.

But as $f(1) = 0$ and f is convex, if $\frac{\pi_n(x)}{\pi(x)} \leq 1$, $f\left(\frac{\pi_n(x)}{\pi(x)}\right) \leq \left(1 - \frac{\pi_n(x)}{\pi(x)}\right)f(0)$.

This implies that $f/[0, 1]$ is affine.

In our examples, the only one is the total variation distance.

2.2 Construction of an optimal stopping time in continuous time

Let $Q = (q(x, y))_{(x, y) \in \mathbb{X} \times \mathbb{X}}$ be the generator of an irreducible positive recurrent Markov chain on the finite space \mathbb{X} .

As before, π_0 is a probability measure on \mathbb{X} . $(X_t)_{t \geq 0}$ is a Markov chain with initial distribution π_0 and generator Q .

We write $\pi_t = \pi_0 \exp(tQ)$ for the distribution of X_t .

Let π be the unique stationary distribution. π is the unique distribution satisfying $\pi Q = 0$.

We write $d(t) = \|\pi_t - \pi\|_{TV}$.

Let $(J_t)_{t \geq 0}$ be defined by

$$J_t = \begin{cases} 0 & \text{if } \exists s \in [0, t], \pi_s(X_s) \leq \pi(X_s) \\ (1 - \psi_0(X_0)) \exp\left(\int_0^t \frac{(\pi_s \wedge \pi)Q(X_s)}{\pi_s(X_s) - \pi(X_s)} ds\right) & \text{else} \end{cases} \quad (9)$$

where ψ_0 is defined in (3). Let U be a uniformly distributed on $[0, 1]$ random variable independent of $(X_t)_{t \geq 0}$.

A randomized stopping time T is defined by $T = \inf\{t \geq 0 / U \geq J_t\}$.

Proposition 2.6.

$$\forall t \geq 0, d(t) = \mathbb{P}(T > t) \quad (10)$$

$$\forall t \geq 0, \mathbb{P}(T \leq t, X_t = x) = (\pi_t \wedge \pi)(x) \quad (11)$$

$$\forall t \geq 0, \mathbb{P}(T \leq t, X_T = x) = (\pi_t \wedge \pi)(x) - \int_0^t (\pi_s \wedge \pi)Q(x) ds \quad (12)$$

The distribution ν of (T, X_T) on $(\mathbb{R}_+ \cup \{+\infty\}) \times \mathbb{X}$ is given by

$$\nu(ds, x) = (\pi_0 \wedge \pi)(x) \delta_0(s) + (\mathbf{1}_{\pi_s(x) < \pi(x)} \pi_s Q(x) - (\pi_s \wedge \pi)Q(x)) ds \quad (13)$$

The proof of the proposition (2.6) can be found in part (5.2) of the appendix.

2.3 Some remarks about weak and maximal coupling

In the preceding sections, the construction of an optimal randomized stopping time T such that for all n , $\mathbb{P}(T > n) = d(n)$ is given. It is easy to see that by the same way, a randomized stopping time \tilde{T} for a Markov chain $(Y_n)_{n \geq 0}$ whose initial distribution given by π , can be defined such that $\mathbb{P}(\tilde{T} > n) = d(n)$ and $\mathbb{P}(X_n = x, T = n) = \mathbb{P}(Y_n = x, \tilde{T} = n) = \gamma_n(x)$.

So $((X_n)_{n \geq 0}, T)$ and $((Y_n)_{n \geq 0}, \tilde{T})$ is a weak optimal coupling as defined in [21]. Then Lindvall prove that it exists (\hat{X}, \hat{Y}) a coupling of (X, Y) such that the coupling time \hat{T} of (\hat{X}, \hat{Y}) has the same distribution as T .

Furthermore, the coupling process (\hat{X}, \hat{Y}) can be chosen such that it is a time inhomogeneous Markov chain. The transitions can be computed but are quite complicated. It is not necessarily a Markovian coupling. By definition a Markovian coupling $((X_n, Y_n)_{n \geq 0})$ of $(\pi_0, K), (\pi, K)$ satisfies the following property :

$$\forall N, \text{ under } \mathbb{P}(\cdot | (X_k, Y_k)_{0 \leq k \leq N}), (X_{N+n}, Y_{N+n})_{n \geq 0} \text{ is a coupling of } (\pi_N, K), (\pi, K)$$

This is sometimes called causal or co-adapted coupling. These ones do not always exist.

Furthermore, it can be proved that $\tilde{T} = \min(S, \tilde{S})$ where $S = \inf\{n \geq 0 / \pi_n(\tilde{X}_n) \leq \pi(\tilde{X}_n)\}$ and $\tilde{S} = \inf\{n \geq 0 / \pi_n(\tilde{Y}_n) \geq \pi(\tilde{Y}_n)\}$.

Here is an example of what is obtained for the two-state chain in continuous time.

$$\mathbb{X} = \{0, 1\}, Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}, \theta = \lambda + \mu \text{ and } \pi_0 = p\delta_0 + q\delta_1.$$

The stationary distribution π is given by $\pi = \frac{\mu}{\theta}\delta_0 + \frac{\lambda}{\theta}\delta_1$ and suppose that $p > \frac{\mu}{\theta}$.

We have $\pi_t = (\frac{\mu}{\theta} + \frac{p\lambda - \mu q}{\theta} e^{-t\theta})\delta_0 + (\frac{\lambda}{\theta} - \frac{p\lambda - \mu q}{\theta} e^{-t\theta})\delta_1$ and $d(t) = \frac{p\lambda - \mu q}{\theta} e^{-t\theta}$.

The initial law of the coupling is given by what is called the γ -coupling of (π_0, π) in [21],

$$\mathbb{P}((X_0, Y_0) = (0, 0)) = \pi(0), \mathbb{P}((X_0, Y_0) = (1, 1)) = \pi(1), \mathbb{P}((X_0, Y_0) = (0, 1)) = p - \pi(0)$$

We find in this case an homogeneous Markov chain on $\mathbb{X} \times \mathbb{X}$ with generator given by

$$q((0, 1), (1, 1)) = \lambda, q((0, 1), (0, 0)) = \mu, q((1, 1), (0, 0)) = \mu, q((0, 0), (1, 1)) = \lambda$$

The distribution of the coupling time T^* is $(q + \frac{\mu}{\theta})\delta_0 + (p - \frac{\mu}{\theta})\mathcal{E}(\theta)$.

In [12], Fill finds that in this example, the optimal coupling time for separation is given by an exponential random variable with parameter θ .

3 Halting set for total variation distance

3.1 Proof of theorem 1.2, existence of halting set

By the preceding constructions, for all $t \geq 0$, $d(t) = \mathbb{P}(T > t) = \mathbb{P}(U < J_t)$. By definition of J_t (9), (8), $\{U < J_t\} \subset \{T_{\mathcal{M}} > t\}$. So $T \leq T_{\mathcal{M}}$.

Remark 3.1. *How can we find a halting state ?*

Suppose that our Markov chain is a reversible aperiodic Markov chain.

Let $|\mathbb{X}| = N + 1$ and let $1 = \beta_0 > \beta_1 \geq \dots \geq \beta_N > -1$ be the eigenvalues of K with $L^2(\pi)$ -normalized eigenvectors V_0, \dots, V_N . The spectral decomposition gives:

$$\frac{\pi_n(x)}{\pi(x)} - 1 = \sum_{k=1}^N \beta_k^n \left(\sum_{y \in \mathbb{X}} \pi_0(y) V_k(y) V_k(x) \right)$$

Denote by $\rho = \max(|\beta_i|, 1 \leq i \leq N)$. So if an halting state x^ exists, it must satisfy*

$$\sum_{k/|\beta_k|=\rho} \sum_{y \in \mathbb{X}} \pi_0(y) V_k(y) V_k(x_*) \leq 0 \text{ and } \sum_{k/|\beta_k|=\rho} \text{sign}(\beta_k) \sum_{y \in \mathbb{X}} \pi_0(y) V_k(y) V_k(x_*) \leq 0 \quad (14)$$

Of course, the set \mathcal{M} can be empty. Property of monotonicity can imply the existence of halting state, as proved in the following proposition.

Let \prec be a partial order on \mathbb{X} .

A Markov chain on \mathbb{X} with transition kernel given by $K = (k(x, y))_{x, y \in \mathbb{X}}$ is monotone if for all $x \prec y$, the probability $K(x, \cdot)$ is stochastically smaller than $K(y, \cdot)$. That means that for an increasing function f , Kf is increasing.

Proposition 3.2. *Let $K = (k(x, y))_{x, y \in \mathbb{X}}$ a monotone kernel for a partial order on \mathbb{X} . Suppose furthermore that there exists a smallest element denoted by $\mathbf{0}$ and a largest element denoted by $\mathbf{1}$.*

Then $\mathbf{1}$ is a halting state for the Markov chain started at $\mathbf{0}$.

Proof. $\frac{\pi_0}{\pi}$ is decreasing. We prove inductively that for all $n \geq 0$, $\frac{\pi_n}{\pi}$ is decreasing.

It comes from the relation $\frac{\pi_{n+1}}{\pi} = K\left(\frac{\pi_n}{\pi}\right)$.

So if $\frac{\pi_n}{\pi}(\mathbf{1}) > 1$, for all $x \in \mathbb{X}$, $\pi_n(x) > \pi(x)$ that is false. \square

Remark 3.3. *In the continuous case, the hypothesis is changed to the following, for all $t \geq 0$, the kernel $P_t = e^{tQ}$ is monotone.*

3.2 One example: The riffle shuffle

The well-known riffle-shuffle is a method of shuffling cards. Its mathematical description was made by Gilbert and Shannon (see [13]) and independently by Reeds [20]. A sharp mathematical analysis for the riffle shuffling was carried out by Bayer and Diaconis (1992) [5].

Denote by $(\sigma_n)_{n \geq 0}$ this Markov chain on the symmetric group \mathfrak{S}_N . σ_0 is the identity and the stationary measure is given by the uniform one.

Let $f(\sigma)$ be the number of rising sequence of σ . In [5], corollary 2. it is proved that $(f(\sigma_n))_{n \geq 0}$ is a Markov chain on $\mathbb{X} = \{1, \dots, N\}$. Its stationary probability is given by $\nu(r) = \frac{A_{N,r}}{N!}$ where $A_{N,r}$, the Eulerian number, is the number of permutations with r rising sequences.

If we denote by ν_n the law of $f(\sigma_n)$, we have $\|\pi_n - \pi\|_{TV} = \|\nu_n - \nu\|_{TV}$.

In [5], theorem 3, it is proved that $\mathbb{P}(\sigma_n = \sigma) = \frac{\binom{N+2^n-r}{2^n}}$ if $f(\sigma) = r$.

So a halting state for total variation for the chain $(f(\sigma_n))_{n \geq 0}$ is a state r such that for all $n \geq 0$, $\frac{\binom{N+2^n-r}{2^n}}{2^{nN}} \leq \frac{1}{N!}$.

We can see easily that $\lfloor \frac{N+1}{2} \rfloor$ is a halting state for total variation.

Furthermore $(f(\sigma_n) - 1)_{n \geq 0}$ is a Markov chain on $\{0, \dots, N-1\}$ which is symmetric, i.e. $k(x, y) = k(N-1-x, N-1-y)$. We shall see later that for symmetric birth-and-death chain, the middle element is often a halting state.

If \mathfrak{M} is the set of permutations with more than $\lfloor \frac{N+1}{2} \rfloor$ rising sequence then we have $\|\pi_n - \pi\|_{TV} \leq \mathbb{P}(T_{\mathfrak{M}} > n)$.

We deduce using Theorem 4 of ([5]) that $\mathbb{P}(T_{\mathfrak{M}} > n) \geq 1 - 2\Phi(\frac{-1}{4c\sqrt{3}}) + O_c(\frac{1}{N^4})$ where $0 < c$,

$$n = \log_2(N^{\frac{3}{2}}c) \text{ and } \Phi(x) = \int_{-\infty}^x e^{-\frac{1}{t^2}} \frac{dt}{\sqrt{2\pi}}.$$

So for example, $\underline{\lim}_{\log_2(N^{\frac{3}{2}})} \frac{\mathbb{E}[T_{\mathfrak{M}}]}{\log_2(N^{\frac{3}{2}})} \geq 1$.

3.3 Applications to the bounding of the total variation distance and to problems of cutoffs

To prove the proposition 1.3, the following results found in the book of Aldous and Fill [3] are used.

They are given in the discrete time.

Let P be an irreducible transition matrix on a finite space \mathbb{X} with a reversible distribution π .

Let $A \subset \mathbb{X}$ and $P^A = P/A^c \times A^c$. Let λ_A be the largest eigenvalue of P^A and λ_1 the second eigenvalue of P . P^A is supposed to be irreducible and aperiodic, then:

$$1 - \lambda_A \geq \pi(A)(1 - \lambda_1) \tag{15}$$

The proof can be found in [3], Theorem 33, Corollary 34, Chapter 3 Reversible Markov Chains.

For all $n \geq 0$,

$$\mathbb{P}_\pi(T_A > n) \leq \pi(A^c)\lambda_A^n \tag{16}$$

The proof can be found in [3], Proposition 21, Chapter 3 Reversible Markov Chains. This is an application of the Perron-Froebenius theorem to P^A .

These results are also used in [10] and [4].

The third useful result, seen in [4] is the following :

$$\forall x \in \mathbb{X}, \forall n, m, \mathbb{P}_x(T_A > n + m) \leq \|\delta_x P^n - \pi\|_{TV} + \mathbb{P}_\pi(T_A > m) \tag{17}$$

We prove this result for clarity:

Proof. Let $(X_n)_{n \geq 0}$ be (δ_x, P) Markov chain. Let $n \geq 0$.

We can find a random variable U independent of $(X_n)_{n \geq 0}$ and a function of X_n and U denoted by Y_n such that the law of Y_n is π and $\|\delta_x P^n - \pi\|_{TV} = \mathbb{P}(X_n \neq Y_n)$. It is a standard argument of coupling.

We have $\mathbb{P}_x(T_A > n + m) \leq \mathbb{P}(X_n \neq Y_n) + \mathbb{P}(T_A > n + m, X_n = Y_n)$. Using the Markov property, we have

$$\begin{aligned} \mathbb{P}(T_A > n + m, X_n = Y_n) &= \mathbb{E}[\mathbf{1}_{T_A > n, X_n = Y_n} \mathbb{P}(T_A(X) > n + m | \sigma(X_k, k \leq n) \vee \sigma(U))] \\ &= \mathbb{E}[\mathbf{1}_{T_A > n, X_n = Y_n} \mathbb{P}_{X_n}(T_A(X) > m)] \end{aligned}$$

$$\text{So } \mathbb{P}_x(T_A > n + m) \leq \|\delta_x P^n - \pi\|_{TV} + \mathbb{E}[\mathbb{P}_{Y_n}(T_A(X) > m)]. \quad \square$$

Proof of proposition 1.3

The proof is similar to these found in the paper of Ding, Lubetzky, Peres ([10]) and of Diaconis and Saloff-Coste ([9]) or in [4].

Bounding the total variation from above For all $n \geq 0$, $d^{(N)}(n) \leq \mathbb{P}(T_{A^{(N)}} > n)$. So for all $\gamma > 0$,

$$d^{(N)}(\lceil \mathbb{E}[T_{A^{(N)}}] + \gamma \sigma(T_{A^{(N)}}) \rceil) \leq \mathbb{P}(T_{A^{(N)}} - \mathbb{E}[T_{A^{(N)}}] > \gamma \sigma(T_{A^{(N)}})) \leq \frac{1}{1 + \gamma^2}$$

Bounding the total variation from below By (17),

$$d^{(N)}(n) \geq \mathbb{P}(T_{A^{(N)}} > n + m) - \mathbb{P}_\pi(T_{A^{(N)}} > m)$$

Let $0 < \epsilon < 1$ and $n = \lceil (1 - \epsilon) \mathbb{E}[T_{A^{(N)}}] \rceil$ and $m = \lceil \frac{\epsilon}{2} \mathbb{E}[T_{A^{(N)}}] \rceil$. Then by (16),

$$\begin{aligned} \mathbb{P}_\pi(T_{A^{(N)}} > m) &\leq (1 - \pi(A^{(N)})) e^{-m(1 - \lambda_{A^{(N)}})} \\ &\leq e e^{-\frac{\epsilon}{2} (1 - \lambda_{A^{(N)}}) \mathbb{E}[T_{A^{(N)}}]} \end{aligned}$$

So

$$\begin{aligned} \mathbb{P}(T_{A^{(N)}} > n + m) &\geq \mathbb{P}(T_{A^{(N)}} - \mathbb{E}[T_{A^{(N)}}] \geq -\frac{\epsilon}{2} \mathbb{E}[T_{A^{(N)}}]) \\ &\geq 1 - \frac{\sigma(T_{A^{(N)}})^2}{(\frac{\epsilon}{2} \mathbb{E}[T_{A^{(N)}}])^2} \end{aligned}$$

D'où $d^{(N)}(\lceil (1 - \epsilon) \mathbb{E}[T_{A^{(N)}}] \rceil) \geq 1 - \left(\frac{\sigma(T_{A^{(N)}})^2}{(\frac{\epsilon}{2} \mathbb{E}[T_{A^{(N)}}])^2} \right) - e e^{-\frac{\epsilon}{2} (1 - \lambda_{A^{(N)}}) \mathbb{E}[T_{A^{(N)}}]}$

If $n = \lceil \mathbb{E}[T_{A^{(N)}}] - \gamma \sigma(T_{A^{(N)}}) \rceil$, we obtain using the same inequalities,

$$d^{(N)}(n) \geq 1 - \frac{4}{\gamma^2} - e^{-\frac{\gamma}{2} (1 - \lambda_{A^{(N)}}) \sigma(T_{A^{(N)}})}$$

Remark 3.4. If $\liminf \pi(A^{(N)}) > 0$, the condition $(1 - \lambda_{A^{(N)}}) \mathbb{E}[T_{A^{(N)}}] \rightarrow +\infty$ can be changed to $(1 - \lambda_1^{(N)}) \mathbb{E}[T_{A^{(N)}}] \rightarrow +\infty$ by the property (15).

4 Birth-and-death chain on $\mathbb{X}^{(N)} = \{0, \dots, N\}$ started at 0

4.1 General results

For every birth-and-death chain on $\{0, \dots, N\}$, we will denote

$$\begin{aligned} p_x &= k(x, x+1) & \text{if } 0 \leq x \leq N-1 \\ q_x &= k(x, x-1) & \text{if } 1 \leq x \leq N \\ r_x &= k(x, x) & \text{if } 0 \leq x \leq N \end{aligned}$$

We have for all $0 \leq x \leq N$, $p_x + r_x + q_x = 1$. We suppose the Markov chain irreducible, so for all $0 \leq x \leq N-1$, $p_x > 0$ and for all $1 \leq x \leq N$, $q_x > 0$.

The chain is started at 0. For $x, y \in \mathbb{X}^{(N)}$, T_x is the hitting time of the state x and $\tau_{x,y}$ is the hitting time of y by the chain started at x .

To use proposition (1.3), it is necessary to know the spectrum of the transition matrix. So below, is given a result without eigenvalues.

Proposition 4.1. *Suppose that x_N^* is a halting state for total variation for the chain started at 0 such that $\mathbb{E}[T_{x_N^*}]$ goes to $+\infty$ when N goes to $+\infty$ and such that $\sigma(T_{x_N^*}) = o(\mathbb{E}[T_{x_N^*}])$.*

Suppose also that there exists $y < x_N^$ such that $\mathbb{E}[T_{x_N^*}] \sim \mathbb{E}[T_y]$ and $\pi^{(N)}(\{0, \dots, y\}) \rightarrow 0$;*

Then the chain has a cutoff at time $\mathbb{E}[T_{x_N^}]$ for total variation.*

If furthermore $\mathbb{E}[\tau_{y, x_N^}] = O(\sigma(T_{x_N^*}))$, then there is a cutoff at time $\mathbb{E}[T_{x_N^*}]$ with window $\sigma(T_{x_N^*})$ for total variation.*

Proof. To bound the total variation distance from below, the following inequality is used :

For all $n \geq 0$, $d^{(N)}(n) \geq \mathbb{P}(T_y \geq n) - \pi^{(N)}(\{0, \dots, y\})$.

If $n = [(1 - \epsilon)\mathbb{E}[T_{x_N^*}]]$,

$$\begin{aligned} \mathbb{P}(T_y \geq n) &= 1 - \mathbb{P}\left(\mathbb{E}[T_y] - T_y > \mathbb{E}[T_{x_N^*}](\epsilon - (1 - \frac{\mathbb{E}[T_y]}{\mathbb{E}[T_{x_N^*}]})\right) \\ &\geq 1 - \frac{\sigma(T_y)^2}{\mathbb{E}[T_{x_N^*}]^2(\epsilon - (1 - \frac{\mathbb{E}[T_y]}{\mathbb{E}[T_{x_N^*}]})^2)} \\ &\geq 1 - \frac{\sigma(T_{x_N^*})^2}{\mathbb{E}[T_{x_N^*}]^2(\epsilon - (1 - \frac{\mathbb{E}[T_y]}{\mathbb{E}[T_{x_N^*}]})^2)} \end{aligned}$$

So $d^{(N)}(n) \geq 1 - \frac{\sigma(T_{x_N^*})^2}{\mathbb{E}[T_{x_N^*}]^2(\epsilon - (1 - \frac{\mathbb{E}[T_y]}{\mathbb{E}[T_{x_N^*}]})^2)} - \pi^{(N)}(\{0, \dots, y\})$.

According to the hypothesis, it is clear that $\lim_{N \rightarrow +\infty} d^{(N)}([(1 - \epsilon)\mathbb{E}[T_{x_N^*}]]) = 1$.

If furthermore $\mathbb{E}[\tau_{y, x_N^*}] = O(\sigma(T_{x_N^*}))$, if γ big enough,

$$d^{(N)}(\mathbb{E}[T_{x_N^*}] - \gamma\sigma(T_{x_N^*})) \geq 1 - \frac{1}{(\gamma - \frac{\mathbb{E}[\tau_{y, x_N^*}]}{\sigma(T_{x_N^*})})^2} - \pi^{(N)}(\{0, \dots, y\}) \quad (18)$$

□

There are several methods to calculate $\mathbb{E}[T_x]$ and $\sigma(T_x)$ for a birth-and-death chain. The first one uses the spectrum with the following formulae, see [16] :

$$\mathbb{E}[T_x] = \sum_{k=0}^{x-1} \frac{1}{1 - \alpha_k}, \quad \sigma(T_x)^2 = \sum_{k=0}^{x-1} \frac{\alpha_k}{(1 - \alpha_k)^2} \quad (19)$$

where $\alpha_0 > \dots > \alpha_{x-1}$ are the eigenvalues of $P_{/\{0, \dots, x-1\}}$. This implies the following inequality:

$$\sigma(T_x)^2 \leq \frac{1}{1 - \alpha_0} \mathbb{E}[T_x] \quad (20)$$

The second one uses the following formulae :

$$\mathbb{E}[T_x] = \sum_{y=0}^{x-1} \mathbb{E}[\tau_{y,y+1}] \quad \text{and} \quad \sigma(T_x)^2 = \sum_{y=0}^{x-1} \sigma(\tau_{y,y+1})^2 \quad (21)$$

$$\mathbb{E}[\tau_{x,x+1}] = \frac{\pi(\{0, \dots, x\})}{p_x \pi(x)}, \quad \sigma(\tau_{x,x+1})^2 = \frac{2}{\pi(x)p_x} \sum_{y=0}^{x-1} \frac{\pi(\{0, \dots, y\})^2}{p_y \pi(y)} + \mathbb{E}[\tau_{x,x+1}]^2 - \mathbb{E}[\tau_{x,x+1}] \quad (22)$$

Sometimes, it is possible to calculate $\mathbb{E}[T_x]$ by using martingale.

4.2 Monotone birth-and-death chain on $\mathbb{X}^{(N)} = \{0, \dots, N\}$ started at 0

A birth-and-death chain is monotone if for all $0 \leq x \leq N-1$, $p_x + q_{x+1} \leq 1$.

In this case if $\pi_0 = \delta_0$ and as usual, $\pi_n = \pi_0 K^n$, then for all $n \geq 0$, $x \mapsto \frac{\pi_n(x)}{\pi(x)}$ is decreasing and by consequence N is a halting state for total variation as proved in proposition (3.2).

An example of monotone birth-and-death chain is given by birth-and-death chain with positive spectrum and in particular by lazy chain which are chain with $r_x \geq \frac{1}{2}$ for all x .

Remark 4.2. *With the notation of the section 2, it can be proved that in the case of a monotone birth-and-death chain on $\{0, \dots, N\}$ started at 0, the support of the distribution of X_T is $\{0, \dots, x^*\}$ where x^* is the smallest halting state for total variation.*

$$\text{Indeed, } \mathbb{P}(X_T = x) = \sum_{n=0}^{+\infty} \gamma_n(x).$$

$$\text{So } \mathbb{P}(X_T = 0) \geq \gamma_0(0) > 0, \quad \mathbb{P}(X_T = 1) \geq \gamma_1(1) = p_0 \wedge \pi(1) - p_0 \pi(0) > 0$$

$$\text{By iteration and by using monotonicity of } x \mapsto \frac{\pi_n(x)}{\pi(x)},$$

$$\forall n \geq 0, \quad \gamma_n(x) = 0 \Leftrightarrow \forall n \geq 0, \quad \pi_n(x-1) \leq \pi(x-1)$$

By consequence, the support of the distribution of X_T is $\{0, \dots, x^*\}$ where for all $n \geq 0$, $\pi_n(x^*) \leq \pi(x^*)$ and there exists $n \geq 0$ with $\pi_n(x^* - 1) > \pi(x^* - 1)$.

So x^* is the smallest halting state for total variation.

The proof is the same in continuous time using (2.6).

There is an interesting property satisfied by monotone birth-and-death chain. With the notations of (3.1)

Lemma 4.1. *If P is a monotone irreducible birth-and-death matrix with eigenvalues given by $1 = \beta_0 > \beta_1 \geq \dots \geq \beta_N > -1$, then $|\beta_N| \leq \beta_1$.*

Proof. The spectrum of an irreducible birth-and-death chain satisfies $1 = \beta_0 > \beta_1 > \dots > \beta_N$, so $\rho \in \{\beta_1, |\beta_N|\}$

We have seen by monotony that for all $n \geq 0$, $\frac{\pi_n(N)}{\pi(N)} \leq 1$. But

$$\begin{aligned} \frac{\pi_n(N)}{\pi(N)} - 1 &= \sum_{k=1}^N \beta_k^n V_k(0) V_k(N) \\ &= \beta_1^n V_1(0) V_1(N) + \beta_N^n V_N(0) V_N(N) + \sum_{k=2}^{N-1} \beta_k^n V_k(0) V_k(N) \end{aligned}$$

But by using property (30) of the appendix on the eigenvectors of irreducible birth-and-death chain, we have that $V_1(0)V_1(N) < 0$ and $(-1)^N V_N(0)V_N(N) > 0$.

So if $|\beta_N| > \beta_1$, $\frac{\pi_n(N)}{\pi(N)} - 1 \sim \beta_N^n V_N(0)V_N(N)$ and we have a contradiction with the fact that for all $n \geq 0$, $\frac{\pi_n(N)}{\pi(N)} - 1 \leq 0$.

By consequence $|\beta_N| \leq \beta_1$. \square

Example 4.3. Metropolis chains

Let π be a probability on $\mathbb{X}^{(N)} = \{0, \dots, N\}$ with $\pi(x) > 0$ for all $x \in \mathbb{X}^{(N)}$. Use the Metropolis algorithm with base chain the simple symmetric random walk to obtain a birth-and- death chain with stationary measure π (see e.g. ([8])). By construction, it is a monotone chain and so N is a halting state.

Example 4.4. Simple random walk

Consider the birth-and-death chain on $\mathbb{X}^{(N)} = \{0, \dots, N\}$ started at 0 with $p_x = p$, $q_x = q$, $r_0 = q$ and $r_N = p$. We suppose $0 \leq q < p \leq 1$ and $p + q = 1$. It is a monotone chain, so N is a halting state.

The spectrum of this kernel is known, see [11] for example.

The eigenvalues are $\beta_k = 2\sqrt{pq} \cos(\frac{\pi k}{N+1})$ for $1 \leq k \leq N$ and V_k is proportional to $\begin{pmatrix} x_{k,0} \\ \vdots \\ x_{k,N} \end{pmatrix}$

where $x_{k,l} = \sqrt{p} \sin(\frac{\pi k(l+1)}{N+1}) - \sqrt{q} \sin(\frac{\pi kl}{N+1})$.

Some calculus prove that (14) is satisfied if $\cot(\frac{\pi x^*}{N+1}) \leq -(1-2q) \cot(\frac{\pi}{N+1})$.

If x_N is the smallest integer such that $x_N > N + 1 - \frac{N+1}{\pi} \cot^{-1}((1-2q) \cot(\frac{\pi}{N+1}))$, we have $x_N \simeq N + 1 - \frac{1}{1-2q}$ which is close to N for N large enough.

It can be proved, using methods of martingales that for all $0 \leq x \leq N$,

$$\mathbb{E}[T_x] = \frac{x}{p-q} - \frac{q}{(p-q)^2} (1 - (\frac{q}{p})^x) \quad \sigma^2(T_x) = Ax + B + C(\frac{q}{p})^x + D(\frac{q}{p})^{2x} + Ex(\frac{q}{p})^x$$

with $A = \frac{4pq}{(p-q)^3}$ and B, C, D, E some constants depending of p .

So proposition(4.1) implies that there is a cutoff at time t_N with a window σ_N , where $t_N = \frac{N}{p-q}$ and $\sigma_N = \sqrt{N}$.

To prove this, one can take $y = N - \sqrt{N}$.

If $p = q = \frac{1}{2}$, we have $\mathbb{E}[T_x] = x(x+1)$ and $\sigma^2(T_x) = \frac{x(x+1)(2x^2+2x-1)}{3}$.

It is well known that in this case, there is no cutoff.

We can prove that in this case $x_N^* = \lfloor \frac{N}{2} \rfloor + 1$ is the smallest halting state for total variation for the chain started at 0, but the condition $\sigma(T_{x_N^*}) = o(\mathbb{E}[T_{x_N^*}])$ is not satisfied.

Remark 4.5. We can suppose that p depends of N .

As we have seen above, the smallest halting state for total variation for the chain started at 0 is bigger than $N + 1 - \frac{N+1}{\pi} \cot^{-1}((1-2q) \frac{\cos(\frac{\pi}{N+1})}{\sin(\frac{\pi}{N+1})})$.

Let $\delta > 0$ and $p_N = \frac{1}{2} + \frac{\delta}{N}$ and $q_N = \frac{1}{2} - \frac{\delta}{N}$.

So the smallest halting state for total variation for the chain started at 0 belongs to $\{\lfloor \frac{N}{2} \rfloor, \dots, N\}$. Denote it by x_N^* . As in the case above, the condition $\sigma(T_{x_N^*}) = o(\mathbb{E}[T_{x_N^*}])$ is not satisfied.

Remark 4.6. If we suppose that $p_N = \frac{1}{2} + \epsilon_N$ and $q_N = \frac{1}{2} - \epsilon_N$ with $\epsilon_N > 0$, $\epsilon_N = o(1)$ and $N\epsilon_N \rightarrow +\infty$.

We have $\sigma^2(T_N) \sim \frac{N}{2\epsilon_N^3}$ and $\mathbb{E}[T_N] \sim \frac{N}{2\epsilon_N}$. So $\sigma(T_N) = o(\mathbb{E}[T_N])$.

Furthermore it is easy to see that $y = N - \sqrt{\frac{N}{4\epsilon_N}}$ satisfies the hypothesis of proposition(4.1), so we have a $\frac{N}{2\epsilon_N}$ cut-off.

4.3 The case of symmetric birth-and-death chain

Definition 4.7. A Markov chain on $\mathbb{X} = \{0, \dots, N\}$ with transitions given by $K = (k(x, y))_{x, y \in \mathbb{X}}$ is symmetric if for all $x \in \mathbb{X}, y \in \mathbb{X}$ then $k(N - x, N - y) = k(x, y)$.

Two different cases can occur, N odd or N even. The spectral analysis differs in these two cases.

4.3.1 The case of symmetric birth-and-death chain on $\{0, \dots, 2N + 1\}$

Denote by P the transition matrix of the birth-and-death chain, by Q the restriction of P to $\{0, \dots, N\} \times \{0, \dots, N\}$.

Let L be the $(N + 1) \times (N + 1)$ matrix define by $L_{i,j} = 1$ if $i + j = N$ and $L_{i,j} = 0$ if not.

$$\text{So } P \text{ is the tridiagonal matrix given by } P = \left(\begin{array}{c|c} Q & \\ \hline & p_N \\ \hline q_{N+1} & \\ & LQL \end{array} \right)$$

Denote by Q_1 the matrix equal to Q unless the entry $Q_1(N, N)$ that is $Q_1(N, N) = r_N + p_N$. So Q_1 is a stochastic matrix.

Proposition 4.8. If P is monotone, then $N + 1$ is a halting state for total variation distance for the chain started at 0.

Proof. If $(X_n)_{n \geq 0}$ is a birth-and-death symmetric chain on $\{0, \dots, 2N + 1\}$ with transition matrix given by P then the process $(Z_n)_{n \geq 0}$ given by $Z_n = N + \frac{1}{2} - |X_n - (N + \frac{1}{2})|$ is a birth-and-death chain on $\{0, \dots, N\}$ with transition matrix given by Q_1 . Its stationary probability is given by $\tilde{\pi}$ with $\tilde{\pi}(x) = 2\pi(x)$.

And if P is monotone then Q_1 is monotone too. So for all $n \geq 0$, $\frac{Q_1^n(0, N)}{\tilde{\pi}(N)} \leq 1$.

Also for all $n \geq 0$, $x \mapsto \frac{P^n(0, x)}{\pi(x)}$ is decreasing. So

$$\begin{aligned} 2 \frac{P^n(0, N+1)}{\pi(N+1)} &\leq \frac{P^n(0, N)}{\pi(N)} + \frac{P^n(0, N+1)}{\pi(N+1)} \\ &\leq \frac{2}{\pi(N)} (\mathbb{P}(X_n = N) + \mathbb{P}(X_n = N + 1)) \\ &\leq \frac{2}{\pi(N)} \mathbb{P}(Z_n = N) = 2 \frac{Q_1^n(0, N)}{\tilde{\pi}(N)} \leq 2 \end{aligned}$$

□

Proposition 4.9. If for all $x \in \{0, \dots, 2N + 1\}$, $r_x = 0$ then for all $n \geq 0$, for all $k \geq 0$,

$$\frac{P^n(0, N + 1)}{2\pi(N + 1)} \leq 1 \quad \text{and} \quad \frac{P^n(0, N + 2)}{2\pi(N + 2)} \leq 1$$

Proof. Denote by Q_2 the matrix equal to Q unless the entry $Q_2(N, N)$ that is $Q_2(N, N) = r_N - p_N$.

We have the following easy lemma :

Lemma 4.2. *If λ is an eigenvalue of Q_1 associated with the eigenvector v , then λ is an eigenvalue of P associated with the eigenvector $\begin{pmatrix} v \\ Lv \end{pmatrix}$.*

If λ is an eigenvalue of Q_2 associated with the eigenvector v , then λ is an eigenvalue of P associated with the eigenvector $\begin{pmatrix} v \\ -Lv \end{pmatrix}$.

Denote by $1 = \beta_0 > \beta_1 > \dots > \beta_N$ the eigenvalues of Q_1 .

Suppose $N = 2a$, the proof is similar if N is odd.

The formula (33) of the appendix tells us :

$$P^{2n+1}(0, N+1) = 2\pi(N+1)\left(1 + \sum_{k=1}^N \beta_k^{2n+1} (-1)^k \nu_k\right) \text{ with } \nu_k > 0$$

If we denote for $0 \leq k \leq a$, $\lambda_{2k} = \beta_k$ and for $0 \leq k \leq a-1$, $\lambda_{2k+1} = |\beta_{N-k}|$, the sequence $(\lambda_k)_{0 \leq k \leq N}$ by property (32) of the appendix, is decreasing and

$$P^{2n+1}(0, N+1) = 2\pi(N+1)\left(1 + \sum_{k=1}^N \lambda_k^{2n} B_k\right) \text{ where } B_{2k} = (-1)^k \lambda_{2k} \nu_k \text{ and } B_{2k+1} = (-1)^{k+1} \lambda_{2k+1} \nu_{N-k}$$

So $B_{2k} B_{2k+1} < 0$ and $B_{2k+1} B_{2k+2} > 0$. The following lemma gives the result.

Lemma 4.3. *Let $\Gamma_n = \sum_{k=0}^{2d} \alpha_k^n B_k$ where*

$$B_0 > 0, \forall i B_{2i} B_{2i+1} < 0, B_{2i+1} B_{2i+2} > 0, \alpha_0 = 1 > \alpha_1 > \alpha_2 > \dots > \alpha_{2d} > 0 \quad (23)$$

We suppose that $\Gamma_0 = \Gamma_1 = \dots = \Gamma_{d-1} = 0$. Then $n \mapsto \Gamma_n$ is strictly increasing on $\{d, d+1, \dots\}$.

The proof is in part (5.3) of the appendix □

4.3.2 The case of symmetric birth-and-death chain on $\{0, \dots, 2N\}$

Denote by Q_+ the stochastic matrix on $\{0, \dots, N\} \times \{0, \dots, N\}$ defined by :

$$\begin{cases} Q_+(x, y) = P(x, y) \text{ if } 0 \leq x, y \leq N-1 \\ Q_+(N, N-1) = 2p_N, Q_+(N, N) = r_N \end{cases}$$

Proposition 4.10. *If P is monotone and $p_{N-1} \leq r_N$ then N is a halting state for total variation distance for the chain started at 0.*

Proof. If $(X_n)_{n \geq 0}$ is a birth-and-death symmetric chain on $\{0, \dots, 2N\}$ with transition matrix given by P then the process $(Z_n)_{n \geq 0}$ given by $Z_n = N - |X_n - N|$ is a birth-and-death chain on $\{0, \dots, N\}$ with transition matrix given by Q_+ . Its stationary probability is given by $\tilde{\pi}$ with $\tilde{\pi}(x) = 2\pi(x)$ if $x \leq N-1$ and $\tilde{\pi}(N) = \pi(N)$.

And if P is monotone and $p_{N-1} \leq r_N$ then Q_+ is monotone. So for all $n \geq 0$, $\frac{Q_+^n(0, N)}{\tilde{\pi}(N)} \leq 1$.

But $Q_+^n(0, N) = \mathbb{P}(Z_n = N) = \mathbb{P}(X_n = N)$ which gives the result. □

Lemma 4.4. *It exists $\gamma_1, \dots, \gamma_N$ such that for all $n \geq 0$, $\frac{P^n(0, N)}{\pi(N)} = 1 + \sum_{k=1}^N \beta_k^n \gamma_k$ where $1 > \beta_1 > \dots > \beta_N$ are the eigenvalues of Q_+ .*

Proof. If λ is an eigenvalue of Q associated with the eigenvector v , then λ is an eigenvalue of P associated with the eigenvector $\begin{pmatrix} v \\ 0 \\ -Lv \end{pmatrix}$.

If λ is an eigenvalue of Q_+ associated with the eigenvector $\begin{pmatrix} v \\ x \end{pmatrix}$, then λ is an eigenvalue of P associated with the eigenvector $\begin{pmatrix} v \\ x \\ Lv \end{pmatrix}$.

Spectral properties give the result. □

Proposition 4.11. *If all the eigenvalues of P are positive then N is a halting state for total variation distance for the chain started at 0.*

Proof. By using the lemma above, for all $n \geq 0$, $\frac{P^n(0, N)}{\pi(N)} = 1 + \sum_{k=1}^N \beta_k^n \gamma_k$.

So the result comes from the following lemma (4.5) proved in the appendix with $\Gamma_n = \frac{P^n(0, N)}{\pi(N)}$.

Lemma 4.5. *Let $1 = \lambda_0 > \lambda_1 > \dots > \lambda_d > 0$ and $\Gamma_n = \sum_{i=0}^d A_i \lambda_i^n$, $A_0 \geq 0$*

If for $0 \leq i \leq d-1$, $\Gamma_i = 0$, then $n \rightarrow \Gamma_n$ is increasing and so for all $n \geq 0$, $\Gamma_n \leq A_0$. □

Proposition 4.12. *If for all $x \in \{0, \dots, 2N\}$, $r_x = 0$ then for all $n \geq 0$,*

$$\text{If } 2k \geq N, \quad \frac{P^{2n}(0, 2k)}{2\pi(2k)} \leq 1$$

$$\text{If } 2k+1 \geq N, \quad \frac{P^{2n+1}(0, 2k+1)}{2\pi(2k+1)} \leq 1$$

Proof. We define two stochastic matrices Q_e and Q_o respectively on $\{0, \dots, N\} \times \{0, \dots, N\}$ and on $\{0, \dots, N-1\} \times \{0, \dots, N-1\}$ by

$$Q_e(x, y) = P^2(2x, 2y) \quad \text{and} \quad Q_o(x, y) = P^2(2x+1, 2y+1)$$

These stochastic matrices have positive eigenvalues, so they are monotone. Their stationary measures are given respectively by $\pi_e(x) = 2\pi(2x)$ and $\pi_o(x) = 2\pi(2x+1)$. They are symmetric so we can apply proposition (4.8) and proposition(4.11).

The relations $P^{2n}(0, 2x) = Q_e^n(0, x)$ and $P^{2n+1}(0, 2x+1) = Q_o^n(0, x)$ give the results. □

Proposition 4.13. *Let P be a transition matrix of a birth-and-death Markov chain on $\mathbb{X} = \{0, \dots, 2N\}$ be given by $P = rI + (1-r)\tilde{P}$ where for all $x \in \mathbb{X}$, $\tilde{P}(x, x) = 0$. We suppose $0 < r < \frac{1}{2}$ and P monotone. Then for all $n \geq 0$, $P^n(0, N+1) \leq \pi(N+1)$.*

Proof. It is an easy consequence of the above proposition (4.12).

For all $n \geq 0$, $P^n = \sum_{k=0}^n \binom{n}{k} r^{n-k} (1-r)^k \tilde{P}^k$. Let $2x+1 \in \mathbb{X}$, $2x \geq N$.

Then $P^{2n}(0, 2x+1) = \sum_{k=0}^{n-1} \binom{2n}{2k+1} r^{2n-2k-1} (1-r)^{2k+1} \tilde{P}^{2k+1}(0, 2x+1)$.

So $P^{2n}(0, 2x+1) \leq \pi(2x+1)(1-(2r-1)^{2n}) \leq \pi(2x+1)$

The same method gives that for $2x \in \mathbb{X}$, $2x \geq N$, $P^{2n+1}(0, 2x) \leq \pi(2x)$.

Using monotonicity, we have the result. \square

4.3.3 Continuous time symmetric birth-and-death chain on $\mathbb{X} = \{0, \dots, N\}$

Proposition 4.14. $[\frac{N+1}{2}]$ is a halting state for total variation for the continuous time Markov chain started at 0.

Proof. Let Q be the generator of the markov chain. Q is supposed symmetric, $Q(N-x, N-y) = Q(x, y)$.

Let $q \geq 2 \max\{q(x), x \in \mathbb{X}\}$ and $P = I + \frac{1}{q}Q$.

P is the transition of a symmetric birth-and-death chain on \mathbb{X} which satisfies $r_x = \mathbb{P}(x, x) = 1 - \frac{q(x)}{q} \geq \frac{1}{2}$. So this chain is lazy and we can apply proposition(4.8) and proposition(4.11) which say that for all $n \geq 0$, $P^n(0, [\frac{N+1}{2}]) \leq \pi([\frac{N+1}{2}])$.

But for all $t \geq 0$, $\pi_t([\frac{N+1}{2}]) = e^{tQ}(0, [\frac{N+1}{2}]) = e^{-qt} \sum_{n=0}^{+\infty} \frac{(tq)^n}{n!} P^n(0, [\frac{N+1}{2}]) \leq \pi([\frac{N+1}{2}])$. \square

4.3.4 Applications to cutoffs for symmetric birth-and-death chain

Proposition 4.15. Under the hypothesis that the birth-and-death is symmetric on $\{0, \dots, N\}$ such that $x_N^* = [\frac{N}{2}] + 1$ is a halting state and $(1 - \lambda_1) \sum_{k=1}^N \frac{1}{1 - \lambda_k} \rightarrow +\infty$, then the chain has a $\mathbb{E}[T_{x_N^*}]$ cutoff for total variation.

Furthermore $\mathbb{E}[T_{x_N^*}] \sim \frac{1}{2} \sum_{k=1}^N \frac{1}{1 - \lambda_k}$

If $\liminf (1 - \lambda_1)^2 \sum_{k=1}^N \frac{\lambda_k}{(1 - \lambda_k)^2} > 0$, then the chain exhibits a cutoff at time $\mathbb{E}[T_{x_N^*}]$ with a window $\sigma(T_{x_N^*})$.

Furthermore in this case, the chain exhibits a cutoff at time $\frac{1}{2} \sum_{k=1}^N \frac{1}{1 - \lambda_k}$ with a window

$$\frac{1}{2} \sqrt{\sum_{k=1}^N \frac{\lambda_k}{(1 - \lambda_k)^2}}$$

Here $1 = \lambda_0 > \lambda_1 > \dots > \lambda_N$ are the eigenvalues of the chain.

Proof. One has to prove that the hypothesis of this proposition imply these of the proposition (1.3).

By (19) and using the notation of the appendix,

$$\mathbb{E}[T_{N+1}] = \sum_{k=0}^N \frac{1}{1 - \gamma_k}$$

But as $\pi(\{N+1, \dots, 2N+1\}) = \frac{1}{2}$, by (15) $\frac{1-\lambda_1}{2} \leq 1 - \gamma_0 \leq 1 - \lambda_1$.
 By lemma (4.2) and (31), for all $0 \leq k \leq N$, $\lambda_{2k} = \beta_k$ and $\lambda_{2k+1} = \alpha_k$.
 So by (31), $(1 - \lambda_1)\mathbb{E}[T_{N+1}]$ tends to $+\infty$.

The proof is almost the same in the case of a symmetric birth-and-death chain on $\{0, \dots, 2N\}$. \square

4.3.5 Two examples

The Ehrenfest process The Markov kernel of the Ehrenfest chain on $\{0, \dots, N\}$ is defined by

$$p_x = \frac{N-x}{N+1}, \quad r_x = \frac{1}{N+1}, \quad q_x = \frac{x}{N+1}$$

The stationary distribution is given by $\pi(x) = \binom{N}{x} 2^{-N}$ and the eigenvalues are $\lambda_k = 1 - \frac{2k}{N+1}$, $0 \leq k \leq N$ with $L^2(\pi)$ -normalized eigenvectors

$$V_k(x) = \binom{N}{k}^{-\frac{1}{2}} \sum_{i=0}^k (-1)^i \binom{x}{i} \binom{N-x}{k-i}$$

In this case $|\lambda_N| = \lambda_1$. It was proved in lemma (4.1) that for a monotone chain, $|\lambda_N| \leq \lambda_1$. We can search candidates to be halting states by (14) for the chain started at 0 :

$$V_1(0)V_1(x) \pm V_N(0)V_N(x) \leq 0 \iff 2x \geq N+1$$

So the smallest candidate is $\lceil \frac{N}{2} \rceil + 1$.

If N is odd, we can apply proposition(4.8) and so $\frac{N+1}{2}$ is a halting state for total variation for the chain started at 0.

If $N = 2N'$ is even, we haven't $q_{N'+1} \leq r_{N'}$ so we can't apply proposition(4.10).

But $P = rI + (1-r)\tilde{P}$ with $r = \frac{1}{N+1}$ and $\tilde{P}(x, x) = 0$ for all x . So we can apply proposition(4.13) and so $N'+1$ is a halting state for total variation for the chain started at 0.

So $x_N = \lceil \frac{N}{2} \rceil + 1$ is the smallest halting state for total variation for the chain started at 0.

As $(1 - \lambda_1) \sum_{k=1}^N \frac{1}{1 - \lambda_k} = \sum_{k=1}^N \frac{1}{k} \rightarrow +\infty$ and $(1 - \lambda_1)^2 \sum_{k=1}^N \frac{\lambda_k}{(1 - \lambda_k)^2} = -\frac{2}{N+1} \sum_{k=1}^N \frac{1}{k} + \sum_{k=1}^N \frac{1}{k^2} \rightarrow \frac{\pi^2}{6}$, proposition (4.15) gives the existence of a cutoff at time $\frac{1}{2} \sum_{k=1}^N \frac{1}{1 - \lambda_k} \sim \frac{N \ln N}{4}$ with a window N .

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 $\mathbb{X} = \{0, \dots, r\}$. The stationary distribution is

$$\pi_{N,r}(j) = \frac{\binom{r}{j} \binom{N-r}{r-j}}{\binom{N}{r}}$$

and for $x \in \mathbb{X}$, the rates are given by

$$p_x = \frac{(r-x)(N-r-x)}{r(N-r)}, \quad q_x = \frac{x^2}{r(N-r)}$$

The eigenvalues of this chain are well known (see [15] or [6]) and are given by

$$\beta_{N,r,i} = 1 - \frac{i(N-i+1)}{r(N-r)}, \quad 0 \leq i \leq r \quad (24)$$

The first eigenvectors are

$$V_0 \equiv 1, \quad V_1(x) = \left(1 - \frac{Nx}{r(N-r)}\right) C_1 \text{ where } C_1 \text{ is defined by the condition } \|V_1\|_{L^2(\pi_{N,r})} = 1.$$

The sign of $V_1(0)V_1(x)$ is that of $1 - \frac{Nx}{r(N-r)}$ so the halting states belong to $\{x/x \geq \frac{r(N-r)}{N}\}$.

In the symmetric case where $r = \frac{N}{2}$, a halting state must be $\geq \frac{r}{2}$.

Proposition 4.16. $[\frac{N+1}{2}]$ is the smallest halting state for total variation.

Proof. Unfortunately, this chain is not monotone ($p_0 + q_1 > 1$) so the proposition(4.10) cannot be applied.

Consider the case of N even, and change N to $2N$.

We use the notation of part (3.3.2) and of proposition (4.10). The result will be true if Q_+^2 is monotone and $x \mapsto \frac{Q_+^3(0,x)}{\tilde{\pi}(x)}$ is decreasing.

Indeed, for all $n \geq 0$, then $x \mapsto \frac{(Q_+^3(0,\cdot)Q_+^{2n})(x)}{\tilde{\pi}(x)} = \frac{Q_+^{2n+3}(0,x)}{\tilde{\pi}(x)}$ is decreasing and so $P^{2n+3}(0, N) = Q_+^{2n+3}(0, N) \leq \tilde{\pi}(N) = \pi(N)$.

In the same way, using decreasing of $x \mapsto \frac{\delta_0}{\tilde{\pi}}(x)$, for all $n \geq 0$, $P^{2n}(0, N) \leq \pi(N)$.

So first, we have to prove that for all x , for all y ,

$$Q_+^2(x, \{0, \dots, y\}) \geq Q_+^2(x+1, \{0, \dots, y\}) \quad (25)$$

We have that for all $N-2 \geq x \geq 1$, $p_x + q_{x+1} \leq 1$ and $p_{N-1} + 2q_N \leq 1$, so if $x \geq 1$, $Q_+(x, \{0, \dots, y\}) \geq Q_+(x+1, \{0, \dots, y\})$. This inequality is also true if $x = 0$ and $y \geq 1$.

So if $x \geq 2$,

$$\begin{aligned} Q_+^2(x, \{0, \dots, y\}) &= q_x(Q_+(x-1, \{0, \dots, y\}) - Q_+(x, \{0, \dots, y\})) \\ &\quad + (1-p_x)(Q_+(x, \{0, \dots, y\}) - Q_+(x+1, \{0, \dots, y\})) \\ &\quad + Q_+(x+1, \{0, \dots, y\}) - Q_+(x+2, \{0, \dots, y\}) \\ &\quad + Q_+(x+2, \{0, \dots, y\}) \\ &\geq q_{x+1}(Q_+(x, \{0, \dots, y\}) - Q_+(x+1, \{0, \dots, y\})) \\ &\quad + (1-p_{x+1})Q_+(x+1, \{0, \dots, y\}) - Q_+(x+2, \{0, \dots, y\}) \\ &\quad + Q_+(x+2, \{0, \dots, y\}) = Q_+(x+1, \{0, \dots, y\}) \end{aligned}$$

It remains to prove (25) for $x = 0$ and $x = 1, y = 0$.

$$\begin{aligned} Q_+^2(0, \{0\}) &= p_0 q_1 + r_0^2 = q_1 \geq r_1 q_1 = Q_+^2(1, \{0\}) \\ Q_+^2(0, \{0, 1\}) &= q_1 + r_1 \\ Q_+^2(1, \{0, 1\}) &= q_1 + r_1(1 - p_1) + p_1 q_2 = q_1 + r_1 + p_1(q_2 - r_1) \\ Q_+^2(1, \{0\}) &= r_1 q_1 \\ Q_+^2(2, \{0\}) &= q_1 q_2 \end{aligned}$$

So the condition is satisfied if $q_2 \leq r_1$ and this the case for the Bernoulli-Laplace chain.

Now it remains to prove that $x \mapsto \frac{Q_+^3(0, x)}{\tilde{\pi}(x)}$ is decreasing.

Denote $\tilde{\pi}_n = Q_+^n(0, \cdot)$.

We have by reversibility that $\frac{\tilde{\pi}_{n+1}}{\tilde{\pi}}(x) = \sum_y \frac{\tilde{\pi}_n}{\tilde{\pi}}(y) Q_+(x, y)$.

So by some calculus, we find

$$\frac{\tilde{\pi}_3}{\tilde{\pi}}(0) = \frac{r_1}{\tilde{\pi}_1(1)} \frac{\tilde{\pi}_3}{\tilde{\pi}}(1) = \frac{q_1 + r_1^2 + p_1 q_2}{\tilde{\pi}_1(1)} \frac{\tilde{\pi}_3}{\tilde{\pi}}(2) = \frac{r_1 q_2 + q_2 r_2}{\tilde{\pi}_1(1)} \frac{\tilde{\pi}_3}{\tilde{\pi}}(3) = \frac{q_2 q_3}{\tilde{\pi}_1(1)}$$

So we have to verify if $r_1 \geq q_1 + r_1^2 + p_1 q_2 \geq r_1 q_2 + q_2 r_2 \geq q_2 q_3$. It is true if N is big enough.

Consider now the case where N is odd and replace N by $2N + 1$.

By the same ideas, it suffices to show that Q_1^2 and P^2 are monotone where Q_1 is given in proposition (4.8). In fact $p_x + q_{x+1} \leq 1$ if $1 \leq x \leq 2N - 1$, but using symetry, $P^2(2N + 1 - x, \{0, \dots, y\}) = 1 - P^2(x, \{0, \dots, 2N - y\})$, so there is no additional verifications to do. We conclude as in proposition (4.8). \square

By the same method as used for the Ehrenfest chain, using (4.16) and the knowledge of eigenvalues given by (24), the symmetric Bernoulli-Laplace birth-and-death chain on $\{0, \dots, N\}$ exhibits a cutoff at time $\frac{N \ln N}{4}$ with a window N .

4.3.6 Asymmetric continuous-time Erhenfest process

Let $Z_t = (Z_t^{(1)}, \dots, Z_t^{(N)})$ be a continuous-time Markov chain on $(\mathbb{Z}/2\mathbb{Z})^N$ where $\left((Z_t^{(i)})_{t \geq 0}\right)_{1 \leq i \leq N}$ are iid continuous-time Markov chain on $\mathbb{Z}/2\mathbb{Z}$ with infinitesimal generator given by $\begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$.

We have

$$\begin{aligned} \mathbb{P}_0(Z_t^1 = 0) &= p_t(0, 0) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-t(\lambda + \mu)} \\ \mathbb{P}_0(Z_t^1 = 1) &= p_t(0, 1) = \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} e^{-t(\lambda + \mu)} \\ \mathbb{P}_1(Z_t^1 = 0) &= p_t(1, 0) = \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-t(\lambda + \mu)} \\ \mathbb{P}_1(Z_t^1 = 1) &= p_t(1, 1) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-t(\lambda + \mu)} \end{aligned} \quad (26)$$

If $X_t = |Z_t| = \sum_{i=1}^N Z_t^{(i)}$, then $(X_t)_{t \geq 0}$ is the asymmetric continuous-time Ehrenfest Markov chain on $\{0, \dots, N\}$.

Furthermore if $0 \leq i, j \leq N$,

$$\mathbb{P}_i(X_t = j) = \sum_{k+l=j} \binom{i}{k} \binom{N-i}{l} p_t(1, 1)^k p_t(1, 0)^{i-k} p_t(0, 1)^l p_t(0, 0)^{N-i-l} \quad (27)$$

The stationary distribution $\pi_{\lambda,\mu}$ is the binomial distribution $\mathcal{B}(N, \frac{\lambda}{\lambda+\mu})$.

$$\text{So } \frac{\mathbb{P}(X_t=x)}{\pi(x)} = (1 + \frac{\lambda}{\mu} e^{-t(\lambda+\mu)})^{N-x} (1 - e^{-t(\lambda+\mu)})^x.$$

So it is easy to prove that the minimal halting state is $x_{\lambda,\mu}^* = \lceil \frac{\lambda}{\lambda+\mu} N \rceil$.

Proposition 4.17. *The asymmetric continuous-time Erhenfest process has a cutoff at time $\frac{\log(N)}{2(\lambda+\mu)}$.*

Proof. The smallest halting state is given by $x_{\lambda,\mu}^* = \lceil \frac{\lambda}{\lambda+\mu} N \rceil$. Let $y^* = N - x^*$.

We have to prove that the hypothesis of (1.3) are satisfied.

Denote by $Q_{\lambda,\mu}$ the generator of our chain and by A the restriction of $Q_{\lambda,\mu}$ to $\{0, \dots, x_{\lambda,\mu}^* - 1\}$ and by B the restriction of $Q_{\mu,\lambda}$ to $\{0, \dots, y_{\mu,\lambda}^* - 1\}$.

For all $0 \leq i \leq N$, $q_{\lambda,\mu}(i, i+1) = \lambda(N-i)$ and $q_{\lambda,\mu}(i, i-1) = \mu i$.

Let $\varphi_0^{\lambda,\mu}, \varphi_1^{\lambda,\mu}, \dots, \varphi_N^{\lambda,\mu}$ the sequence of orthogonal polynomials associated to $Q_{\lambda,\mu}$.

The eigenvalues of $-Q_{\lambda,\mu}$ are $t_k = k(\lambda + \mu)$, $0 \leq k \leq N$.

If L is the $(N+1) \times (N+1)$ matrix defined by $L_{i,j} = 1$ if $i+j = N$ and $L_{i,j} = 0$ if not, then $LQ_{\lambda,\mu}L = Q_{\mu,\lambda}$.

So by an argument of symmetry, for all $0 \leq k \leq N$, for all $0 \leq x \leq N$, $\varphi_{N-x}^{\lambda,\mu}(t_k) = \varphi_N^{\lambda,\mu}(t_k)\varphi_x^{\lambda,\mu}(t_k)$. Furthermore, by (30), $(-1)^k \varphi_N^{\lambda,\mu}(t_k) > 0$.

Let $V_A = 0 < \alpha_0 < \alpha_1 < \dots < \alpha_{x^*-1}$ be the eigenvalues of $-A$.

Let $V_B = 0 < \beta_0 < \beta_1 < \dots < \beta_{y^*-1}$ be the eigenvalues of $-B$.

Lemma 4.6. *For all $0 \leq i \leq N-1$, the interval $[t_i, t_{i+1}]$ contains a single element of $V_A \cup V_B$.*

Proof. It is a consequence of Sturm's theorem : The number of sign changes in the sequence $\varphi_0^{\lambda,\mu}(t), \varphi_1^{\lambda,\mu}(t), \dots, \varphi_N^{\lambda,\mu}(t)$ gives the number of eigenvalues $> t$. The convention is that if $\varphi_{k+1}(t) = 0$ and $\varphi_k(t) \neq 0$, the sign of $\varphi_{k+1}(t)$ is equal to the sign of $\varphi_k(t)$.

Apply this to $t = -t_1$:

The single change of sign can occur in the sequence $\varphi_0^{\lambda,\mu}(-t_1), \varphi_1^{\lambda,\mu}(-t_1), \dots, \varphi_{x^*}^{\lambda,\mu}(-t_1)$, and in this case, $-t_1 < -\alpha_0 < 0$

or can occur in the sequence $(\varphi_{x^*}^{\lambda,\mu}(-t_1), \dots, \varphi_N^{\lambda,\mu}(-t_1)) = -(\varphi_0^{\mu,\lambda}(-t_1), \dots, \varphi_{N-y^*}^{\mu,\lambda}(-t_1))$, and in this case $-t_1 < -\beta_0 < 0$.

This method is reiterated. \square

But $\mathbb{E}_0[T_{x^*}^{\lambda,\mu}] = \sum_{i=0}^{x^*-1} \frac{1}{\alpha_i}$ and $\mathbb{E}_0[T_{y^*}^{\mu,\lambda}] = \sum_{i=0}^{y^*-1} \frac{1}{\beta_i}$. Denote the first mean by U_N et the second by V_N .

The lemma implies that $\sum_{i=1}^N \frac{1}{t_i} \leq U_N + V_N \leq \frac{1}{\alpha_0} + \frac{1}{\beta_0} + \sum_{i=1}^N \frac{1}{t_i}$.

As in the discrete time, we have by (15), $\alpha_0 \geq \pi_{\lambda,\mu}(\{x^* + 1, \dots, N\})t_1$.

But $\pi_{\lambda,\mu}(\{x^* + 1, \dots, N\}) \sim \frac{1}{2}$, then $U_N + V_N \sim \sum_{i=1}^N \frac{1}{t_i} \sim \frac{\log(N)}{\lambda + \mu}$.

Lemma 4.7. *$(U_N - V_N)$ is bounded.*

Proof. By using the equality proved in [3], chapter 2, lemma 12,

$$\mathbb{E}_i[T_j] = \frac{Z_{j,j} - Z_{i,j}}{\pi(j)}, \text{ with } Z_{i,j} = \int_0^{+\infty} (\mathbb{P}_i(X_t = j) - \pi(j)) dt,$$

we have

$$U_N - V_N = \frac{1}{\pi_{\lambda, \mu}(x^*)} \int_0^{+\infty} (\mathbb{P}_N(X_t^{\lambda, \mu} = x^*) - \mathbb{P}_0(X_t^{\lambda, \mu} = x^*)) dt$$

By using formula (27),

$$U_N - V_N = \int_0^{+\infty} \left(1 + \frac{\mu}{\lambda} e^{-t(\lambda+\mu)}\right)^{x^*} (1 - e^{-t(\lambda+\mu)})^{y^*} - (1 - e^{-t(\lambda+\mu)})^{x^*} \left(1 + \frac{\lambda}{\mu} e^{-t(\lambda+\mu)}\right)^{y^*} dt$$

After the change of variables $u = e^{-t(\lambda+\mu)}$, we obtain

$$U_N - V_N = \frac{1}{\lambda + \mu} \int_0^1 \frac{\left(1 + \frac{\mu}{\lambda} u\right)^{x^*} (1 - u)^{y^*} - (1 - u)^{x^*} \left(1 + \frac{\lambda}{\mu} u\right)^{y^*}}{u} du$$

Some calculus gives the result. \square

Then as $U_N + V_N \sim \frac{\log(N)}{\lambda + \mu}$ and $U_N - V_N$ is bounded, $U_N \sim \frac{\log(N)}{2(\lambda + \mu)}$ and the result comes from proposition (1.3). \square

5 Appendix

5.1 Birth-and-death chain and orthogonal polynomials

A reference for orthogonal polynomial and birth and death chain can be found in Karlin and McGregor([16], [15]).

Denote by P the transition matrix on $\{0, \dots, N\}$ and by P_k the restriction of P to $\{0, \dots, k\}$ for $0 \leq k \leq N$.

For $1 \leq k \leq N$, denote by φ_k the polynomial given by $\varphi_k(t) = \frac{1}{p_0 \cdots p_{k-1}} \det(tI_k - P_{k-1})$. Let $\varphi_0(t) = 1$.

So $(\varphi_0, \dots, \varphi_N)$ is a family of polynomials which satisfy the following recurrence equation :

$$\varphi_k(t) = \frac{t - r_{k-1}}{p_{k-1}} \varphi_{k-1}(t) - \frac{q_{k-1}}{p_{k-1}} \varphi_{k-2}(t) \quad (28)$$

Let $1 = \lambda_0 > \lambda_1 > \dots > \lambda_N$ be the eigenvalues of P .

It exists $(\mu_0, \dots, \mu_N) \in \mathbb{R}_+^{N+1}$ such that $(\varphi_0, \dots, \varphi_N)$ are the orthogonal polynomials in $L^2([-1, 1], \mu)$ where $\mu = \sum_{k=0}^N \mu_k \delta_{\lambda_k}$.

Furthermore $\int_{-1}^1 \varphi_k(t)^2 \mu(dt) = \frac{\pi(0)}{\pi(k)} = \frac{q_1 \cdots q_k}{p_0 \cdots p_{k-1}}$, $\mu_0 = \pi(0)$.

We have for all $n \geq 0$, for all $x, y \in \mathbb{X}$,

$$P^n(x, y) = \frac{\pi(y)}{\pi(0)} \sum_{k=0}^N \lambda_k^n \varphi_x(\lambda_k) \varphi_y(\lambda_k) \mu_k \quad (29)$$

So the normalized in $L^2(\pi)$ eigenvector V_k associated to the eigenvalue λ_k is given by $V_k(x) = \sqrt{\frac{\mu_k}{\mu_0}} \varphi_x(\lambda_k)$.

The zeros of the orthogonal polynomials have the following interlacing property : if $m > n$, there is a zero of φ_m between any two zeros of φ_n .

So there is a zero of φ_N in each interval $] \lambda_{k+1}, \lambda_k [$ and as $\varphi_N(1) = 1$, we have :

$$\forall 0 \leq k \leq N, \quad (-1)^k \varphi_N(\lambda_k) > 0 \quad (\text{or } (-1)^k V_k(N) > 0) \quad (30)$$

The case of a symmetric birth-and-death chain on $\{0, \dots, 2N+1\}$ Denote by Q_1 the matrix equal to P_N unless the entry $Q_1(N, N)$ that is $r_N + p_N$ and by $C_1(t)$ the characteristic polynomial of Q_1 .

Denote by Q_2 the matrix equal to P_N unless the entry $Q_2(N, N)$ that is $r_N - p_N$ and by $C_2(t)$ the characteristic polynomial of Q_2 .

Let $1 = \beta_0 > \beta_1 > \dots > \beta_N$ be the eigenvalues of Q_1 .

Let $\alpha_0 > \alpha_1 > \dots > \alpha_N$ be the eigenvalues of Q_2 .

Let $\eta_0 > \eta_1 > \dots > \eta_{N-1}$ be the eigenvalues of P_{N-1} .

Let $\gamma_0 > \gamma_1 > \dots > \gamma_N$ be the eigenvalues of P_N .

So by the interlacing property, for all $0 \leq k \leq N-1$, $\beta_{k+1} < \eta_k < \beta_k$, $\alpha_{k+1} < \eta_k < \alpha_k$ and $\gamma_{k+1} < \eta_k < \gamma_k$.

We have $(p_0 \dots p_{N-1})^{-1}C_1(t) = (t - r_N - p_N)\varphi_N(t) - q_N\varphi_{N-1}(t)$, $(p_0 \dots p_{N-1})^{-1}C_2(t) = (t - r_N + p_N)\varphi_N(t) - q_N\varphi_{N-1}(t)$ and $\varphi_{N+1}(t) = (t - r_N)\varphi_N(t) - q_N\varphi_{N-1}(t)$.

So $(p_0 \dots p_{N-1})^{-1}C_1(\alpha_k) = -2p_N\varphi_N(\alpha_k)$. But as $\eta_k < \alpha_k < \eta_{k-1}$, $(-1)^k\varphi_N(\alpha_k) > 0$.

But $\alpha_k \in]\eta_k, \beta_k[$ or $\alpha_k \in]\beta_k, \eta_{k-1}[$ and $(-1)^k C_1 < 0$ on the first interval and $(-1)^k C_1 > 0$ on the second interval. So we have for all $0 \leq k \leq N$, $\alpha_k < \beta_k$.

As $(p_0 \dots p_{N-1})^{-1}(C_1(t) + C_2(t)) = 2\varphi_{N+1}(t)$, $C_1(\gamma_k) = -C_2(\gamma_k)$ so for all $0 \leq k \leq N-1$,

$$\eta_k < \alpha_k < \gamma_k < \beta_k < \eta_{k-1} \quad (31)$$

with $\eta_{-1} = +\infty$ et $\eta_N = -\infty$.

Suppose now that for all $x \in \mathbb{X}$, $r_x = 0$.

In this case if k is odd then φ_k is odd and if k is even, φ_k is even.

Furthermore $C_2(-t) = (-1)^{N-1}C_1(t)$ and so for all $0 \leq k \leq N$, $\alpha_k = -\beta_{N-k}$. So we have :

$$\begin{cases} \text{If } N = 2a + 1, & 0 < |\beta_{a+1}| < \beta_a < |\beta_{a+2}| < \dots < \beta_1 < |\beta_{2a+1}| < \beta_0 = 1 \\ \text{If } N = 2a, & 0 < \beta_a < |\beta_{a+1}| < \beta_{a-1} < |\beta_{a+2}| < \dots < \beta_1 < |\beta_{2a}| < \beta_0 = 1 \end{cases} \quad (32)$$

So by (29), we have in the case where N is even :

$$P^{2n+1}(0, N+1) = \frac{2\pi(N+1)}{\pi(0)} \sum_{k=0}^N \beta_k^{2n+1} \varphi_{N+1}(\beta_k) \mu_{2k} \quad (33)$$

and $\varphi_{N+1}(\beta_k)$ has the sign of $(-1)^k$.

5.2 Proof of results on continuous time

For each $0 \leq h$, consider the h -skeleton discrete time Markov chain $X^{(h)}(n) = X(nh)$, $n \in \mathbb{N}$.

$X^{(h)}$ is a Markov chain with transition matrix e^{hQ} , initial distribution π_0 and stationary distribution π .

We denote the corresponding quantities (7) and (8) by $\psi_n^{(h)}$ and $J_n^{(h)} = \prod_{k=0}^n (1 - \psi_k^{(h)}(X_k^{(h)}))$

Proposition 5.1. *For all $t \notin \mathcal{A}$, there exists $(h_n) \rightarrow 0$ such that a.e. on Ω*

$$J_{\lfloor \frac{t}{h_n} \rfloor}^{(h_n)} \rightarrow J_t$$

Proof. For all $x \in \mathbb{X}$, $t \mapsto \pi_t(x)$ is analytic with derivative $\pi_t Q(x)$.

By consequence $\mathcal{A}_x = \{t \geq 0 / \pi_t(x) = \pi(x)\}$ is either equal to \mathbb{R}_+ or contains only isolated points.

Let $\mathcal{T} = \{x \in \mathbb{X} / \forall t \geq 0, \pi_t(x) = \pi(x)\}$ and $\mathcal{A} = \cup_{x \notin \mathcal{T}} \mathcal{A}_x$.

For all $t \geq 0$, $\mathcal{A} \cap [0, t]$ is a finite set.

Let $t \notin \mathcal{A}$. We take $(h_n) \rightarrow 0$ such that there exists $\epsilon > 0$ such that for all $n \geq 0$, for all $a \in \mathcal{A} \cap [0, t]$, $d(a, h_n \mathbb{N}) \leq h_n^{1+\epsilon}$.

We choose ω such that for all $r \geq 1$, $T_r(\omega) \notin \mathcal{A} \cup \{t\}$ where $(T_r)_{r \geq 0}$ are the jump times of the continuous-time Markov chain $(X_t)_{t \geq 0}$.

First case : $\forall s \in [0, t], \pi_s(X_s(\omega)) > \pi(X_s(\omega))$

So for all $s \in [0, t]$, $X_s(\omega) \notin \mathcal{T}$ and as $T_r(\omega) \notin \mathcal{A} \cup \{t\}$, by continuity there exists $\alpha > 0$ such that for all $s \in [0, t]$, $\pi_s(X_s(\omega)) - \pi(X_s(\omega)) > \alpha$.

$$\ln J_n^{(h_n)}(\omega) = \ln \left(1 - \frac{\pi(X_0(\omega))}{\pi_0(X_0(\omega))} \right) + \sum_{k=1}^{\lfloor \frac{t}{h_n} \rfloor} \ln \left(1 - \psi_k^{(h_n)}(X_{kh_n}(\omega)) \right)$$

So, we have, using continuity and derivability of $t \mapsto \pi_t(x)$:

$$\ln J_n^{(h_n)}(\omega) = \ln \left(1 - \frac{\pi(X_0(\omega))}{\pi_0(X_0(\omega))} \right) + h_n \sum_{k=1}^{\lfloor \frac{t}{h_n} \rfloor} \frac{(\pi_{h_n k} \wedge \pi)Q(X_{kh_n}(\omega))}{\pi_{h_n k}(X_{kh_n}(\omega)) - \pi(X_{kh_n}(\omega))} + O(h_n)$$

That gives the result in this case

Second case : $\exists s \in [0, t], \pi_s(X_s(\omega)) \leq \pi(X_s(\omega))$

If there exists $s \in [0, t]$ such that $\pi_s(X_s(\omega)) < \pi(X_s(\omega))$ then by continuity, $J_n^{(h_n)}(\omega) = 0$ for n big enough.

So we suppose that there exists $s_0 \in]0, t[$ which satisfies $\pi_{s_0}(X_{s_0}(\omega)) = \pi(X_{s_0}(\omega))$ and for all $u \in [0, t]$, $\pi_u(X_u(\omega)) \geq \pi(X_u(\omega))$.

Let $x = X_{s_0}(\omega)$ and r such that $T_r(\omega) < s_0 < T_{r+1}(\omega)$.

We want to prove that $J_n^{(h_n)}(\omega) \rightarrow 0$.

Let $s_n \in h_n \mathbb{N}$ which satisfies $|s_n - s_0| \leq h_n^{1+\epsilon}$. For n big enough, $T_r(\omega) < s_n - h_n < s_n < T_{r+1}(\omega)$.

We have $J_n^{(h_n)}(\omega) \leq \frac{\pi_{s_n}(x) - \pi(x)}{\pi_{s_n}(x) - (\pi_{s_n - h_n} \wedge \pi)e^{h_n Q}(x)}$.

Define for $n \geq 1$,

$$R_n = \{x \in \mathbb{X} / \pi_{s_0}(x) = \pi(x), \pi_{s_0} Q^k(x) = 0 \text{ for } 1 \leq k \leq n-1, (\pi_{s_0} \wedge \pi) Q^k(x) = 0 \text{ for } 1 \leq k \leq n\}$$

Let $n \geq 2$. We prove by iteration that

$$x \in R_n \implies \forall y/q(y, x) \neq 0, y \in R_{n-1}, (\pi_{s_0} \wedge \pi) Q^n(y) \leq 0 \text{ and } \pi_{s_0} Q^{n-1}(y) \geq 0 \quad (34)$$

Let $n \geq 2$. We prove by iteration that

$$x \in R_n \text{ and } \pi_{s_0} Q^n(x) = 0 \implies \forall y/q(y, x) \neq 0, y \in R_{n-1} \text{ and } \pi_{s_0} Q^{n-1}(y) = 0 \quad (35)$$

For n big enough, $\pi_{s_n}(x) - (\pi_{s_n - h_n} \wedge \pi)e^{h_n Q}(x) = \pi_{s_n}(x) - \pi(x) + \sum_{k \geq 1} (\pi_{s_n - h_n} \wedge \pi) Q^k(x) \frac{h_n^k}{k!}$.

Define the following sets :

$$\mathcal{P} = \{y \in \mathbb{X}/\pi_{s_0}(y) > \pi(y)\}, \mathcal{M} = \{y \in \mathbb{X}/\pi_{s_0}(y) < \pi(y)\}$$

$$\mathcal{A}_+ = \{y \in \mathbb{X}/\pi_{s_0}(y) = \pi(y), (\pi_{s_0} - \pi)(y) > 0\}, \mathcal{A}_- = \{y \in \mathbb{X}/\pi_{s_0}(y) = \pi(y), (\pi_{s_0} - \pi)(y) < 0\}$$

So, if n is big enough and $s = s_n - h_n$,

$$(\pi_s \wedge \pi)Q^k(x) = \sum_{y \in \mathcal{T} \cup \mathcal{P} \cup \mathcal{A}_+} \pi(y)q_k(y, x) + \sum_{y \in \mathcal{M} \cup \mathcal{A}_-} \pi_s(y)q_k(y, x) = \sum_{y \in \mathcal{M} \cup \mathcal{A}_-} (\pi_s(y) - \pi(y))q_k(y, x)$$

So if we denote $\alpha_{l,k} = \sum_{y \in \mathcal{M} \cup \mathcal{A}_-} \pi_{s_0}Q^k(y)q_l(y, x)$, we have

$$(\pi_{s_n - h_n} \wedge \pi)Q^l(x) = (\pi_{s_0} \wedge \pi)Q^l(x) + \sum_{r \geq 1} \frac{(s_n - h_n - s_0)^r}{r!} \alpha_{l,r} \quad \text{and so}$$

$$\pi(x) - (\pi_{s_n - h_n} \wedge \pi)e^{h_n Q}(x) = - \sum_{l \geq 1} (\pi_{s_0} \wedge \pi)Q^l(x) \frac{h_n^l}{l!} - \sum_{l \geq 1} \frac{h_n^l}{l!} \sum_{r \geq 1} \frac{(s_n - h_n - s_0)^r}{r!} \alpha_{l,r} \quad (36)$$

We prove by induction that for $n \geq 2$,

$$x \in R_n \implies \begin{cases} \text{if } k + l \leq n - 1 & \text{then } \alpha_{k,l} = 0 \\ \text{if } k + l = n & \text{then } \alpha_{k,l} \geq 0 \end{cases} \quad (37)$$

We prove by induction that for $n \geq 2$,

$$x \in R_n \text{ and } \pi_{s_0}Q^n(x) = 0 \implies \begin{cases} \text{if } k + l \leq n & \text{then } \alpha_{k,l} = 0 \\ \text{if } k + l = n + 1 & \text{then } (-1)^l \alpha_{k,l} \leq 0 \end{cases} \quad (38)$$

Let $2k_0$ be the smallest integer k which satisfies $\pi_{s_0}Q^k(x) \neq 0$. We have $\pi_{s_0}Q^{2k_0}(x) > 0$.

Let k_1 be the smallest integer k which satisfies $(\pi_{s_0} \wedge \pi)Q^k(x) \neq 0$.

If $k_1 \leq 2k_0$ then by (36)

$$\pi(x) - (\pi_{s_n - h_n} \wedge \pi)e^{h_n Q}(x) = -(\pi_{s_0} \wedge \pi)Q^{k_1}(x) \frac{h_n^{k_1}}{k_1!} - \sum_{l, r \geq 1, l+r \leq k_1} \frac{h_n^l}{l!} \frac{(s_n - h_n - s_0)^r}{r!} \alpha_{l,r} + o(h_n^{k_1})$$

But as $x \in R_{k_1-1}$, using (34), we have $(\pi_{s_0} \wedge \pi)Q^{k_1-1}(x) < 0$ and using (38) $(-1)^r \alpha_{l,r}(x) \leq 0$ if $l+r = k_1$ and $\alpha_{l,r}(x) = 0$ if $l+r < k_1$, so there exists $C > 0$ such that

$$\pi(x) - (\pi_{s_n - h_n} \wedge \pi)e^{h_n Q}(x) = h_n^{k_1} C + o(h_n^{k_1})$$

If $k_1 > 2k_0$, then $x \in R_{2k_0}$. By (36), we have

$$\pi(x) - (\pi_{s_n - h_n} \wedge \pi)e^{h_n Q}(x) = h_n^{2k_0} \left(\sum_{l, r \geq 1, l+r=2k_0} (-1)^{r+1} \frac{\alpha_{l,r}}{l!r!} \right) + o(h_n^{2k_0})$$

But $\alpha_{1,2k_0-1}(x) = \sum_{y \in \mathcal{A}_-} (\pi_{s_0} Q^{2k_0-1})(y)q(y, x)$. By (34) and (35), if $q(y, x) > 0$, then

$$y \in R_{2k_0-2} \text{ and } \pi_{s_0} Q^{2k_0-1}(y) \geq 0.$$

Furthermore if $y \in \mathcal{A}_+$, then $\pi_{s_0} Q^{2k_0-1}(y) = 0$. So $\alpha_{1,2k_0-1}(x) = \pi_{s_0} Q^{2k_0}(x) > 0$.

And using again (38), there exists $C > 0$ such that

$$\pi(x) - (\pi_{s_n-h_n} \wedge \pi)e^{h_n Q}(x) = h_n^{2k_0} C + o(h_n^{2k_0})$$

The fact that $\pi_{s_n}(x) - \pi(x) = \pi_{s_0} Q^{2k_0}(x) \frac{(s_n-s_0)^{2k_0}}{(2k_0)!} + o((s_n-s_0)^{2k_0})$ and $|s_n - s_0| \leq h_n^{1+\epsilon}$ implies that $\frac{\pi_{s_n}(x) - \pi(x)}{\pi_{s_n}(x) - (\pi_{s_n-h_n} \wedge \pi)e^{h_n Q}(x)}$ goes to 0.

□

We give now the proof of proposition (2.6).

Proof. We define for $t \geq 0$ and $x \in \mathbb{X}$, $f(t, x) = \mathbb{P}(T \leq t, X_T = x)$ and $l(t, x) = \mathbb{P}(T \leq t, X_t = x)$. This proposition is the consequence of the two following lemma.

Lemma 5.1. $l(t, x) = (\pi_t \wedge \pi)(x)$

Proof. By using the proposition (5.1), we have

$$l(t, x) = \mathbb{P}(U \geq J_t, X_t = x) = \lim_{n \rightarrow +\infty} \mathbb{P}(U \geq J_{\lfloor \frac{t}{h_n} \rfloor}^{(h_n)}, X_t = x)$$

But $h_n \lfloor \frac{t}{h_n} \rfloor \leq t$ so,

$$\begin{aligned} \mathbb{P}(U \geq J_{\lfloor \frac{t}{h_n} \rfloor}^{(h_n)}, X_t = x) &= \sum_{y \in \mathbb{X}} \mathbb{P}(U \geq J_{\lfloor \frac{t}{h_n} \rfloor}^{(h_n)}, X_{h_n \lfloor \frac{t}{h_n} \rfloor} = y, X_t = x) \\ &= \sum_{y \in \mathbb{X}} \mathbb{P}(U \geq J_{\lfloor \frac{t}{h_n} \rfloor}^{(h_n)}, X_{h_n \lfloor \frac{t}{h_n} \rfloor} = y) e^{(t-h_n \lfloor \frac{t}{h_n} \rfloor)Q}(y, x) \\ &= \sum_{y \in \mathbb{X}} (\pi_{h_n \lfloor \frac{t}{h_n} \rfloor} \wedge \pi)(y) e^{(t-h_n \lfloor \frac{t}{h_n} \rfloor)Q}(y, x) \\ &\rightarrow (\pi_t \wedge \pi)(x) \end{aligned}$$

this by (4) and (5).

□

Lemma 5.2. $\frac{\partial f}{\partial t}(t, x) = \frac{\partial l}{\partial t}(t, x) - \sum_{y \in \mathbb{X}} l(t, y)q(y, x)$

It is true for all stopping times.

□

5.3 Proof of lemma (4.3)

Lemma 5.3. Let $\Gamma_n = \sum_{k=0}^{2d} \alpha_k^n B_k$ where

$$B_0 > 0, \forall i B_{2i} B_{2i+1} < 0, B_{2i+1} B_{2i+2} > 0, \alpha_0 = 1 > \alpha_1 > \alpha_2 > \dots > \alpha_{2d} > 0 \quad (39)$$

We suppose that $\Gamma_0 = \Gamma_1 = \dots = \Gamma_{d-1} = 0$. Then $n \mapsto \Gamma_n$ is strictly increasing on $\{d, d+1, \dots\}$.

Proof. We prove the result inductively by the study of the real function $\varphi(t) = B_0 + \sum_{k=1}^{2d} B_k e^{-\mu_k t}$

where $\mu_k = -\ln \alpha_k$.

We prove that if (39) is satisfied then

- If d is even, there exists $1 \leq k \leq \frac{d}{2}$, there exists $t_1 < \dots < t_{2k-1}$ such that
 - $\forall 1 \leq i \leq k$, φ is strictly increasing on $[t_{2i-1}, t_{2i}[$
 - $\forall 0 \leq i \leq k-1$, φ is strictly decreasing on $[t_{2i}, t_{2i+1}[$
 - where $t_0 = -\infty, t_{2k} = +\infty$.
- If d is odd, there exists $1 \leq k \leq \frac{d-1}{2}$, there exists $t_1 < \dots < t_{2k}$ such that
 - $\forall 0 \leq i \leq k$, φ is strictly increasing on $[t_{2i}, t_{2i+1}[$
 - $\forall 1 \leq i \leq k$, φ is strictly decreasing on $[t_{2i-1}, t_{2i}[$
 - where $t_0 = -\infty, t_{2k+1} = +\infty$.

It is easy to see that if $d = 1$, φ is strictly increasing.

If d is odd and $\varphi(t) = B_0 + \sum_{k=1}^{2(d+1)} B_k e^{-\mu_k t}$, we write $\varphi'(t) = -e^{-\mu_1 t} \psi(t)$ and $\psi'(t) = e^{-(\mu_2 - \mu_1)t} \theta(t)$.

θ satisfy the same hypothesis so there exists $1 \leq k \leq \frac{d-1}{2}$, there exists $t_1 < \dots < t_{2k}$ such that

- $\forall 0 \leq i \leq k$, θ is strictly increasing on $[t_{2i}, t_{2i+1}[$
- $\forall 1 \leq i \leq k$, θ is strictly decreasing on $[t_{2i-1}, t_{2i}[$
- where $t_0 = -\infty, t_{2k+1} = +\infty$.
- Furthermore $\lim_{t \rightarrow +\infty} \theta(t) = -B_2 \mu_2 (\mu_2 - \mu_1) > 0$.

The number of zeros of θ is odd and smaller than $2k + 1$. So there exists $l \leq k$, there exists $s_1 < \dots < s_{2l+1}$ such that

- $\theta < 0$ on $]s_{2i}, s_{2i+1}[$ for $0 \leq i \leq l, s_0 = -\infty$.
- $\theta > 0$ on $]s_{2i+1}, s_{2i+2}[$ for $0 \leq i \leq l, s_{2l+2} = +\infty$.
- Furthermore $\lim_{t \rightarrow +\infty} \psi(t) = B_1 \mu_1 < 0$.

So the number of zeros of ψ is odd and smaller than $2l + 1$. So there exists $m \leq l$, there exists $r_1 < \dots < r_{2m+1}$ such that

- $\psi > 0$ on $]r_{2i}, r_{2i+1}[$ for $0 \leq i \leq m, r_0 = -\infty$.
- $\psi < 0$ on $]r_{2i+1}, r_{2i+2}[$ for $0 \leq i \leq m, r_{2m+2} = +\infty$.
- As $m \leq l \leq k \leq \frac{d-1}{2}$, we have proved the result for φ .
- The case where d is even is the same.

The preceding result prove that $\varphi(t) = B_0 + \sum_{k=1}^{2d} B_k e^{-\mu_k t}$ has at most d zeros. So if $\varphi(0) = \varphi(1) = \dots = \varphi(d-1) = 0$, $t \mapsto \varphi(t)$ is strictly increasing on $]d-1, +\infty[$ and this proves the lemma. □

5.4 Proof of lemma (4.5)

Lemma 5.4. *Let $1 = \lambda_0 > \lambda_1 > \dots > \lambda_d > 0$ and $\Gamma_n = \sum_{i=0}^d A_i \lambda_i^n, A_0 \geq 0$*

If for $0 \leq i \leq d-1$, $\Gamma_i = 0$, then $n \rightarrow \Gamma_n$ is increasing and so for all $n \geq 0$, $\Gamma_n \leq A_0$.

Proof. We show the result inductively.

If $d = 1$, $\Gamma_n = A_0(1 - \lambda_1^n)$.

We suppose the result true for $d-1$. We can suppose that for all $0 \leq i \leq d$, $A_i \neq 0$ et denote by ϵ the sign of A_1 .

Then for all $n \geq 0$, $\Gamma_{n+1} - \Gamma_n = \epsilon \lambda_1^n \theta_n$, where

$$\theta_n = B_0 + \sum_{i=1}^{d-1} B_i \tilde{\lambda}_i^n, \quad B_i = \epsilon A_{i+1} (\lambda_{i+1} - 1), \quad \tilde{\lambda}_i = \frac{\lambda_{i+1}}{\lambda_1}$$

Furthermore, $\theta_0 = \dots = \theta_{d-2} = 0$, so by hypothesis, $n \rightarrow \theta_n$ is increasing, so as $\theta_0 = 0$, for all $n \geq 0$, $\theta_n \geq 0$ and $n \rightarrow \epsilon \Gamma_n$ is increasing.

So $0 = \epsilon \Gamma_0 \leq \epsilon \Gamma_n \leq \lim_{n \rightarrow +\infty} \epsilon \Gamma_n = \epsilon A_0$.

As $A_0 > 0$, then $\epsilon = 1$. this completes the proof of the lemma. \square

References

- [1] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, (5):333–348, 1986.
- [2] D. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, (8):69–97, 1987.
- [3] D. Aldous and James Allen Fill. *Reversible Markov chains and random walks on graphs. Book in preparation.* URL for draft <http://www.stat.berkeley.edu/alldous/RWG/book.html>.
- [4] R. Basu, J. Hermon, and Y. Peres. Characterisation of cutoff for reversible markov chains. *arXiv:1409.3250v3*.
- [5] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, (2):294–313, 1992.
- [6] P. Diaconis and M. Shahshahani. Time to reach stationarity in the bernoulli-laplace diffusion model. *SIAM J. Math. Anal.*, (18-1):208–218, 1987.
- [7] Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, (18(4)):1483–1522, 1990.
- [8] Persi Diaconis and Saloff-Coste Laurent. 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.
- [9] Persi Diaconis and Laurent Saloff-Coste. Separation-cut-offs for birth-and-death chain. *Ann. Appl. Probab.*, (16(4)):2098–2122, 2006.
- [10] J. Ding, E. Lubetzky, and Y. Peres. Total-variation cutoff in birth-and-death chains. *Probability Theory and Related Fields*, (146(1)):61–85, 2010.
- [11] W. Feller. *An introduction to probability theory and its applications. Volume II.* Wiley and Sons, 1971.

- [12] James Allen Fill. Time to stationarity for a continuous-time markov chain. *Probability in the Engineering and Informational Sciences*, (5):61–76, 1991.
- [13] E. Gilbert. *Theory of shuffling*. Technical memorandum, Bell Laboratories, 1955.
- [14] David Griffeath. A maximal coupling for markov chains. *Z.Wahrscheinlichkeitstheorie verw.Gebiete*, (31), 1975.
- [15] S. Karlin and J. McGregor. The hahn polynomials, formulas and application. *Scripta Math.*, (26):33–46, 1961.
- [16] S. Karling and J. McGregor. Linear growth, birth and death processes. *J. Math. Mech.*, (7):643–662, 1958.
- [17] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times*. AMS, 2009.
- [18] Friedrich Liese and Igor Vajda. *Convex Statistical distances*. Teubner Texts in Mathematics, 1987.
- [19] S. Martinez and B. Ycart. Decay rates and cutoff for convergence and hitting times of markov chains with countably infinite state space. *Adv. in Appl.Probab.*, (33(1)):188–205, 2001.
- [20] J. Reeds. *Unpublished manuscript*. 1981.
- [21] Lindvall Torgny. *Lectures on the coupling method*. John Wiley and Sons, 1992.