A technical report on hitting times, mixing and cutoff

Jonathan Hermon

University of Cambridge,
Centre for Mathematical Sciences,
Wilberforce Rd, Cambridge CB3 0WA
Cambridge, United Kingdom.
E-mail address: `jonathan.hermon@statslab.cam.ac.uk`

Abstract. Consider a sequence of continuous-time irreducible reversible Markov chains and a sequence of initial distributions, μ_n. Instead of performing a worst case analysis, one can study the rate of convergence to the stationary distribution starting from these initial distributions. The sequence is said to exhibit (total variation) μ_n-cutoff if the convergence to stationarity in total variation distance is abrupt, w.r.t. this sequence of initial distributions.

In this work we give a characterization of μ_n-cutoff (and also of total-variation mixing) for an arbitrary sequence of initial distributions μ_n (in the above setup). Our characterization is expressed in terms of hitting times of sets which are "worst" (in some sense) w.r.t. μ_n.

Consider a Markov chain on Ω whose stationary distribution is π. Let $t_{\mathrm{H}}(\alpha) := \max_{x \in \Omega, A \subseteq \Omega: \pi(A) \geqslant \alpha} \mathbb{E}_x[T_A]$ be the expected hitting time of the set of stationary probability at least α which is "worst in expectation" (starting from the worst starting state). The connection between $t_{\mathrm{H}}(\cdot)$ and the mixing time of the chain was previously studied by Aldous and later by Lovász and Winkler, and was recently refined by Peres and Sousi and independently by Oliveira. In this work we further refine this connection and show that μ_n-cutoff can be characterized in terms of concentration of hitting times (starting from μ_n) of sets which are worst in expectation w.r.t. μ_n. Conversely, we construct a counter-example which demonstrates that in general cutoff (as opposed to cutoff w.r.t. a certain sequence of initial distributions) cannot be characterized in this manner.

Finally, we also prove that there exists an absolute constant C such that for any Markov chain $\varepsilon(t_{\mathrm{H}}(\varepsilon) - t_{\mathrm{H}}(1 - \varepsilon)) \leqslant Ct_{\mathrm{rel}}|\log \varepsilon|$, for all $0 < \varepsilon < 1/2$, where t_{rel} is the inverse of the spectral gap of the additive symmetrization $\frac{1}{2}(P + P^*)$.

1. Introduction

This work is a continuation of Basu et al. (2017), in which Starr's maximal inequality was used to characterize the cutoff phenomenon for reversible Markov chains in terms of concentration of hitting times of sets which are "worst" in some sense. Here using the same technique we present several new related results.

The connection between mixing times and hitting times of sets which are "worst" in some sense goes back to the pioneer work of Aldous (1982) and the later body of work of Lovász and Winkler (1998) on stopping rules. Recently, this connection was substantially refined by Peres and Sousi (2015) and independently by Oliveira (2012). All of the aforementioned works considered sets whose hitting time is in some sense the "worst in expectation"[1]. The results in Basu et al. (2017) give a more refined connection between hitting times and mixing times.

We extend the results of Basu et al. (2017) to the case cutoff (resp. mixing) is considered only w.r.t. a certain sequence of initial distributions (resp. initial distribution). We show that in this setup one may interpret "worst" above as "maximizing the expected hitting time" w.r.t. the considered sequence of initial distributions over all sets whose stationary probability is at least some $\alpha \in (0,1)$. Conversely, we show that this may fail when considering cutoff in the usual sense (not only w.r.t. a certain sequence of initial distributions).

Generically, we shall denote the state space of a Markov chain by Ω and its stationary distribution by π (or Ω_n and π_n, respectively, for the n-th chain in a sequence of chains). We say that the chain is finite, whenever Ω is finite. Let $(Y_t)_{t=0}^\infty$ be an irreducible Markov chain on a finite state space Ω with transition matrix P and stationary distribution π. We denote such a chain by (Ω, P, π). The *time-reversal* of P is P^*, given by $P^*(x,y) := \pi(y)P(y,x)/\pi(x)$. A chain (Ω, P, π) is called **reversible** if $P = P^*$, i.e. $\pi(x)P(x,y) = \pi(y)P(y,x)$, for all $x, y \in \Omega$. The *additive symmetrization* of P is given by $Q := \frac{1}{2}(P + P^*)$.

Periodicity issues can be avoided by considering the continuous-time version of the chain, $(X_t)_{t \geqslant 0}$. This is a continuous time Markov chain whose heat kernel is defined by $H_t(x,y) := \sum_{k=o}^\infty \frac{e^{-t}t^k}{k!} P^t(x,y)$. It is a classic result of probability theory that for any initial condition the distribution of X_t converges to π when t goes to infinity. The object of the theory of Mixing time for Markov chain is to study the characteristic of this convergence (see Aldous and Fill, 2002; Levin et al., 2017 for self-contained introductions to the subject). Throughout, we shall consider only continuous time chains, although all our results can be stated also in discrete time, assuming $P(x,x) \geqslant \delta$ for some $\delta > 0$, for all $x \in \Omega$ (in fact, even if this fails, the results are still valid if one replaces the relaxation time by the absolute relaxation time; see Basu et al., 2017, Remark 1.8).

We denote by H_μ^t (H_μ) the distribution of X_t ($(X_t)_{t \geqslant 0}$), given that the initial distribution is μ. When $\mu = \delta_x$ (where $\delta_x(y) = 1_{x=y}$), for some $x \in \Omega$, we simply write H_x^t (H_x). We denote the set of probability distributions on a (finite) set B by $\mathscr{P}(B)$. For any pair of distributions $\mu, \nu \in \mathscr{P}(B)$, their **total-variation distance**

---

[1]In Oliveira (2012) and Peres and Sousi (2015) the parameter $t_H(\cdot)$ was considered and in Aldous (1982); Lovász and Winkler (1998) $\max_{\alpha \in (0,1)} \alpha t_H(\alpha)$ was considered, where $t_H(\cdot)$ is defined in Definition 1.4.

is defined to be

$$\|\mu - \nu\|_{\mathrm{TV}} := \frac{1}{2} \sum_x |\mu(x) - \nu(x)| = \max_{A \subset B} \mu(A) - \nu(A) = \sum_{x \in B : \mu(x) > \nu(x)} \mu(x) - \nu(x).$$

The worst-case total variation distance at time t is defined as $d(t) := \max_{x \in \Omega} \|\mathrm{H}_x^t - \pi\|_{\mathrm{TV}}$. For $\mu \in \mathscr{P}(\Omega)$ let $d_\mu(t) := \|\mathrm{H}_\mu^t - \pi\|_{\mathrm{TV}} = \frac{1}{2} \sum_{y \in \Omega} |\sum_{x \in \Omega} \mu(x) H_t(x,y) - \pi(y)|$. The ε-**mixing-time** (resp. w.r.t. a fixed initial distribution μ) is defined as

$$t_{\mathrm{mix}}(\varepsilon) := \inf \{t : d(t) \leqslant \varepsilon\}, \quad (\text{resp. } t_{\mathrm{mix},\mu}(\varepsilon) := \inf \{t : d_\mu(t) \leqslant \varepsilon\}).$$

When $\varepsilon = 1/4$ we simply write t_{mix} and $t_{\mathrm{mix},\mu}$.

Recall that if (Ω, P, π) is a finite irreducible chain, then the additive symmetrization Q is reversible and hence self-adjoint w.r.t. the inner-product induced by π (see Definition 2.1). Thus Q has $|\Omega|$ real eigenvalues. Throughout we shall denote them by $1 = \lambda_1 > \lambda_2 \geqslant \ldots \geqslant \lambda_{|\Omega|} \geqslant -1$ (where $\lambda_2 < 1$ by irreducibility). Define the **spectral-gap** and **relaxation-time** of P as $\lambda := (1 - \lambda_2)$ and $t_{\mathrm{rel}} := 1/\lambda$. The following general relation holds for reversible chains (see Levin et al., 2017 Lemmas 20.5 and 20.11),

$$t_{\mathrm{rel}} |\log (2\varepsilon)| \leqslant t_{\mathrm{mix}}(\varepsilon) \leqslant t_{\mathrm{rel}} |\log \left(\varepsilon \min_x \pi(x) \right)|. \tag{1.1}$$

Next, consider a sequence of such chains, $((\Omega_n, P_n, \pi_n) : n \in \mathbb{N})$, each with its corresponding worst-distance from stationarity $d_n(t)$, its mixing-time $t_{\mathrm{mix}}^{(n)}$, etc.. Loosely speaking, the (total variation) **cutoff phenomenon** occurs when over a negligible period of time, known as the **cutoff window**, the (worst-case) total variation distance (of a certain finite Markov chain from its stationary distribution) drops abruptly from a value close to 1 to near 0. In other words, one should run the n-th chain until time $(1 - o(1))t_{\mathrm{mix}}^{(n)}$ for it to even slightly mix in total variation, whereas running it any further after time $(1 + o(1))t_{\mathrm{mix}}^{(n)}$ is essentially redundant. Formally, we say that a sequence of chains exhibits a **cutoff** if the following sharp transition in its convergence to stationarity occurs:

$$\lim_{n \to \infty} t_{\mathrm{mix}}^{(n)}(\varepsilon)/t_{\mathrm{mix}}^{(n)}(1 - \varepsilon) = 1, \text{ for every } 0 < \varepsilon < 1.$$

Similarly, for a sequence of initial distributions $\mu_n \in \mathscr{P}(\Omega_n)$, we say that a sequence of chains exhibits a μ_n-**cutoff** if

$$\lim_{n \to \infty} t_{\mathrm{mix},\mu_n}^{(n)}(\varepsilon)/t_{\mathrm{mix},\mu_n}^{(n)}(1 - \varepsilon) = 1, \text{ for every } 0 < \varepsilon < 1.$$

We say that the sequence has a **cutoff window** (resp. μ_n-cutoff window) w_n, if $w_n = o(t_{\mathrm{mix}}^{(n)})$ (resp. $w_n = o(t_{\mathrm{mix},\mu_n}^{(n)})$) and for any $\varepsilon \in (0,1)$ there exists $C_\varepsilon > 0$ such that for all n

$$t_{\mathrm{mix}}^{(n)}(\varepsilon) - t_{\mathrm{mix}}^{(n)}(1 - \varepsilon) \leqslant C_\varepsilon w_n \quad (\text{resp. } t_{\mathrm{mix},\mu_n}^{(n)}(\varepsilon) - t_{\mathrm{mix},\mu_n}^{(n)}(1 - \varepsilon) \leqslant C_\varepsilon w_n).$$

We say that a family of chains satisfies the **product condition** if $\lambda^{(n)} t_{\mathrm{mix}}^{(n)} \to \infty$ as $n \to \infty$ (or equivalently, $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix}}^{(n)})$). The following well-known fact follows easily from the first inequality in (1.1) (see e.g. Levin et al., 2017, Proposition 18.4).

Fact 1.1. If a sequence of irreducible reversible chains exhibits a cutoff, then $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix}}^{(n)})$.

The following mixing parameter, introduced in Basu et al. (2017), shall play a key role in this work.

Definition 1.2. Let (Ω, P, π) be an irreducible chain. Let $\mu \in \mathscr{P}(\Omega)$, $\delta, \varepsilon \in (0, 1)$ and $t \geqslant 0$. We define $p_\mu(\delta, t) := \max_{B \subseteq \Omega: \pi(B) \geqslant \delta} \mathrm{H}_\mu[T_B > t]$, where $T_B := \inf\{t : X_t \in B\}$ is the **hitting time** of the set B. Set $p(\delta, t) := \max_{x \in \Omega} p_x(\delta, t)$. We define

$$\mathrm{hit}_{\delta,\mu}(\varepsilon) := \min\{t : p_\mu(\delta, t) \leqslant \varepsilon\} \text{ and } \mathrm{hit}_\delta(\varepsilon) := \min\{t : p(\delta, t) \leqslant \varepsilon\}.$$

Definition 1.3. Let (Ω_n, P_n, π_n) be a sequence of irreducible chains and let $\alpha \in (0, 1)$. We say that the sequence exhibits a hit_α-cutoff (resp. $\mathrm{hit}_{\alpha,\mu_n}$-cutoff), if for every $0 < \varepsilon < 1/4$,

$$\mathrm{hit}_\alpha^{(n)}(\varepsilon) - \mathrm{hit}_\alpha^{(n)}(1 - \varepsilon) = o\left(\mathrm{hit}_\alpha^{(n)}(1/4)\right)$$

$$(\text{respectively,} \quad \mathrm{hit}_{\alpha,\mu_n}^{(n)}(\varepsilon) - \mathrm{hit}_{\alpha,\mu_n}^{(n)}(1 - \varepsilon) = o\left(\mathrm{hit}_{\alpha,\mu_n}^{(n)}(1/4)\right)).$$

The main abstract result in Basu et al. (2017) (Theorem 3) is the following theorem.

Theorem A. *Let (Ω_n, P_n, π_n) be a sequence of reversible irreducible finite Markov chains. The following are equivalent:*

(1) *The sequence exhibits a cutoff.*
(2) *The sequence exhibits a hit_α-cutoff for some $\alpha \in (0, 1)$ and $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix}}^{(n)})$.*

Definition 1.4. Let $\mu \in \mathscr{P}(\Omega)$ and $0 < \alpha < 1$. Define

$$t_{\mathrm{H},\mu}(\alpha) := \max_{A \subseteq \Omega: \pi(A) \geqslant \alpha} \mathbb{E}_\mu[T_A] \quad \text{and}$$

$$t_{\mathrm{H}}(\alpha) := \max_x t_{\mathrm{H},x}(\alpha) = \max_{x \in \Omega, A \subseteq \Omega: \pi(A) \geqslant \alpha} \mathbb{E}_x[T_A].$$

This work was greatly motivated by the results of Peres and Sousi (2015). Similar results were obtained independently by Oliveira (2012). Both papers refined previous results of Aldous (1982) and of Lovász and Winkler (1998). Their results share the general theme of describing mixing-times in terms of hitting-times. Their approach relied on the theory of random times to stationarity combined with a certain "de-randomization" argument due to Aldous (1982), which shows that for any reversible irreducible finite chain and any stopping time T (possibly w.r.t. a larger filtration, allowing some external randomness) such that $X_T \sim \pi$, we have $t_{\mathrm{mix}} = O(\max_{x \in \Omega} \mathbb{E}_x[T])$. As a consequence, they showed that for any $0 < \alpha < 1/2$ (this was extended to $\alpha = 1/2$ in Griffiths et al., 2014), there exist some constants $c_\alpha, c'_\alpha > 0$ such that for every reversible irreducible finite chain

$$c'_\alpha t_{\mathrm{H}}(\alpha) \leqslant t_{\mathrm{mix}} \leqslant c_\alpha t_{\mathrm{H}}(\alpha).$$

It is natural to ask whether the more studied mixing parameter $t_{\mathrm{H}}(\alpha)$ could be used in Theorem A instead of the mixing parameter $\mathrm{hit}_\alpha(\cdot)$.

The following theorem extends Theorem A to arbitrary starting distributions $\mu_n \in \mathscr{P}(\Omega_n)$, such that $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix},\mu_n}^{(n)})$. In addition, it asserts that "cutoff" w.r.t. these initial distributions (i.e. μ_n-cutoff), is in fact equivalent to concentration of hitting times of sets which are "worst in expectation" w.r.t. these initial distributions (in the sense of Definition 1.4).

Theorem 1. *Let (Ω_n, P_n, π_n) be a sequence of finite irreducible reversible chains. Let $\mu_n \in \mathscr{P}(\Omega_n)$ be such that $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix},\mu_n}^{(n)}(\delta))$, for some $0 < \delta < 1$. Then the following are equivalent:*

    i) *The sequence exhibits a μ_n-cutoff.*

    ii) *There exists some $\alpha \in (0,1)$ such that the sequence exhibits a $\mathrm{hit}_{\alpha,\mu_n}$-cutoff.*

    iii) *There exist some $\alpha \in (0,1)$ and a sequence of sets $A_n \subseteq \Omega_n$ with $\pi_n(A_n) \geqslant \alpha$ satisfying that $\mathbb{E}_{\mu_n}[T_{A_n}] = t_{\mathrm{H},\mu_n}^{(n)}(\alpha)$, such that*

$$\lim_{n\to\infty} \mathrm{H}_{\mu_n}[|T_{A_n} - t_{\mathrm{mix},\mu_n}^{(n)}| < \varepsilon \mathbb{E}_{\mu_n}[T_{A_n}]] = 1, \quad \text{for every } \varepsilon > 0.$$

Remark 1.5. It was shown in Basu et al. (2017) that the occurrence of hit_α-cutoff for some $\alpha \in (0, 1/2]$ implies that the product condition holds. Example 8.2 in Basu et al. (2017) demonstrates that this cannot be improved and that hit_α-cutoff for some $\alpha \in (1/2, 1)$ need not imply cutoff. A small variation of that example can be used to show that in general $\mathrm{hit}_{\alpha,\mu_n}$-cutoff (even for small α) need not imply μ_n-cutoff. Thus the assumption that $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix},\mu_n}^{(n)}(\delta))$ for some $0 < \delta < 1$, in Theorem 1 cannot be removed.

We say that a Markov chain is *transitive* if for every $x, y \in \Omega$ there exists a bijection $\phi : \Omega \to \Omega$ such that (1) $\phi(x) = y$ and (2) for all $a, b \in \Omega$ we have that $P(a,b) = P(\phi(a), \phi(b))$.

Corollary 1.6. *Let (Ω_n, P_n, π_n) be a sequence of finite irreducible reversible transitive chains. Then the following are equivalent:*

    i) *The sequence exhibits a cutoff.*

    ii) *The sequence satisfies the product condition, and for some sequence $x_n \in \Omega_n$ and some $0 < \alpha < 1$, there exists a sequence of sets $A_n \subseteq \Omega_n$ with $\pi_n(A_n) \geqslant \alpha$ satisfying that $\mathbb{E}_{x_n}[T_{A_n}] = t_{\mathrm{H},x_n}^{(n)}(\alpha)$, such that*

$$\lim_{n\to\infty} \mathrm{H}_{x_n}[|T_{A_n} - t_{\mathrm{mix}}^{(n)}| < \varepsilon \mathbb{E}_{x_n}[T_{A_n}]] = 1, \text{ for every } \varepsilon > 0.$$

The following proposition asserts that in general cutoff (as opposed to cutoff w.r.t. some sequence of initial distributions) cannot be characterized in terms of the parameter $t_{\mathrm{H}}(\cdot)$.

Proposition 1.7. *There exists a sequence (Ω_n, P_n, π_n) of finite irreducible reversible chains satisfying the product condition such that the following holds:*

    · *There exist $A_n \subseteq \Omega_n$ and $x_n \in \Omega_n$ such that $\pi_n(A_n) \geqslant 1/2$ and $\mathbb{E}_{x_n}[T_{A_n}] = t_{\mathrm{H}}^{(n)}(1/2)$.*

    · *The distribution of the hitting times of A_n are concentrated w.r.t. the initial states x_n.*

    · *The sequence does not exhibit a $\mathrm{hit}_{1/2}$-cutoff.*

    · *The sequence does not exhibit a cutoff.*

In Example 5.1 we construct a sequence of chains which exhibits the behavior described in Proposition 1.7.

Remark 1.8. It was shown in Chen and Saloff-Coste (2013) that a sequence of finite continuous-time Markov chains exhibits a cutoff iff $t_L^{(n)}(\varepsilon) - t_L^{(n)}(1 - \varepsilon) = o(t_L^{(n)}(1/4))$, where $t_L^{(n)}(\varepsilon)$ is the ε-mixing-time of the associated lazy chain. They also showed that the same holds for a sequence of fixed initial distributions. Hence

in part (i) of Theorem 1 and of Corollary 1.6 we could have considered the lazy version of the chain, rather than its continuous-time version.

Recall that the relaxation-time t_{rel} of P is defined to be the inverse of the smallest non-zero eigenvalue of $I - \frac{1}{2}(P + P^*)$. The main ingredient in the proof of Theorem 1 is the following proposition.

Proposition 1.9. *Let (Ω, P, π) be a finite irreducible reversible Markov chain. Let $\mu \in \mathscr{P}(\Omega)$. Let $0 < \alpha, \varepsilon < 1$ and let $0 < \delta \leqslant \min\{\varepsilon, 1 - \varepsilon, 1 - \alpha\}$. Then*

$$\text{hit}_{\alpha,\mu}(\varepsilon + \delta) - \alpha^{-1} \log\left(\frac{1-\alpha}{\delta}\right) t_{\text{rel}} \leqslant t_{\text{mix},\mu}(\varepsilon) \leqslant \text{hit}_{\alpha,\mu}(\varepsilon - \delta) + \frac{3}{2} t_{\text{rel}} |\log \delta|. \quad (1.2)$$

The same holds when μ is omitted from the inequality. Moreover, even if the chain is non-reversible, we still have that

$$\begin{aligned}
\text{hit}_{\alpha,\mu}(\varepsilon + \delta) &- \alpha^{-1} \log(\frac{1-\alpha}{\delta}) t_{\text{rel}} \\
&\leqslant t_{\text{mix},\mu}(\varepsilon) \leqslant \text{hit}_{1-\delta,\mu}(\varepsilon - \delta) + \log(2/\delta^3) t_{\text{rel}} \log(t_{\text{rel}} \vee e).
\end{aligned} \quad (1.3)$$

Conjecture 1.10. *Let $t_* := \inf\{t : \mathbf{Var}_\pi H_t f \leqslant e^{-2} \mathbf{Var}_\pi f$ for all $f \in \mathbb{R}^\Omega\}$, where for $g \in \mathbb{R}^\Omega$ we define $\mathbf{Var}_\pi g := \|g - \mathbb{E}_\pi g\|_2^2$, $\mathbb{E}_\pi g := \sum_x \pi(x) g(x)$, $\|g\|_2^2 := \mathbb{E}_\pi g^2$ and $H_t g(x) := \sum_y H_t(x, y) g(y) = \mathbb{E}_x g(X_t)$. Then there exist some constants $c_\varepsilon, C_\delta > 0$ such that for every irreducible finite Markov chain, for every $\varepsilon \in (0, 1)$ and every $\delta \leqslant \min\{\varepsilon, 1 - \varepsilon\}$ we have that*

$$\max\{c_\varepsilon t_*, \text{hit}_{1-\delta}(\varepsilon + \delta)\} \leqslant t_{\text{mix}}(\varepsilon) \leqslant \text{hit}_{1-\delta}(\varepsilon - \delta) + C_\delta t_*.$$

For more details regarding this conjecture and how one might prove it, see Conjecture 2.5 and Remark 2.6.

The advantage in the definition of t_{rel} as the relaxation-time of the additive symmetrization is that in this fashion it still admits an extremal characterization and thus in applications can be bounded from above via various standard comparison techniques. The disadvantage is that without reversibility, in general it no longer provides a lower bound on t_{mix} nor on the rate of exponential decay of variances. To see this consider the following example due to Chen (2006) (similar examples can be found at Montenegro and Tetali, 2006, Section 6 and Aldous and Fill, 2002, Example 9.26). Let $\Omega = \{0, 1\}^n$ and $P((x_1, x_2, \ldots, x_n), (x_2, x_3, \ldots, x_n, \eta)) = \frac{1}{2}$ for all $(x_1, x_2, \ldots, x_n) \in \Omega$ and $\eta \in \{0, 1\}$. Observe that after n steps the discrete-time chain attains (exactly) the uniform distribution on Ω, while the relaxation-time of the additive-symmetrization is governed by that of simple random walk on the n-cycle and hence is of order n^2 (e.g. Levin et al., 2017, Section 12.3)

We now comment about t_* being a natural alternative extension of t_{rel} to the non-reversible setup, which is meant to eliminate the existence of such examples.

Remark 1.11. It is possible to replace in all of our results t_{rel} by t_* from the previous conjecture. Let $H_t^*(x, y) := \frac{\pi(y)}{\pi(x)} H_t(x, y)$. Since H_t^* is the dual operator of H_t w.r.t. the inner-product on $\mathbb{R}^\Omega$ given by $\langle f, g \rangle_\pi := \mathbb{E}_\pi[fg]$ we have that

$$t_* = \inf\{t : \langle H_t^* H_t f, f \rangle_\pi \leqslant e^{-2} \text{ for all } f \in \mathbb{R}^\Omega \text{ such that } \mathbb{E}_\pi f = 0 \text{ and } \mathbb{E}_\pi f^2 = 1\}$$

Observe that under reversibility we have that $t_* = t_{\text{rel}}$, while in general $t_* \leqslant t_{\text{rel}}$. Moreover, observe that for all $k \in \mathbb{N}$ we have that $\mathbf{Var}_\pi H_{kt_*} f \leqslant e^{-2k} \mathbf{Var}_\pi f$.

Since $H_t^* H_t$ is self-adjoint w.r.t. $\langle \bullet, \bullet \rangle_\pi$ it follows from the Courant-Fischer-Weyl min-max principle (and continuity w.r.t. t) that there exists some non-constant $f \in$

$\mathbb{R}^\Omega$ such that $H_{t_*}^* H_{t_*} f = e^{-2} f$ and $\mathbb{E}_\pi f = 0$. It follows from general considerations (cf. Levin et al., 2017, the proof of Theorem 12.4) that if $|f(x)| = \max_{y \in \Omega} |f(y)|$ then

$$\|H_{t_*}^* H_{t_*}(x, \bullet) - \pi(\bullet)\|_{\mathrm{TV}} = \sum_y \tfrac{1}{2} |H_{t_*}^* H_{t_*}(x, y) - \pi(y)| \geqslant \frac{1}{2e^2}.$$

It follows that $\|\mathrm{H}_\mu^{t_*} - \pi\|_{\mathrm{TV}} \geqslant \frac{1}{2e^2}$, where $\mu(z) := H_{t_*}^*(x, z)$, and so $t_{\mathrm{mix}}(\frac{1}{2e^2}) \geqslant t_*$.

Remark 1.12. Equation (1.2) is essentially Proposition 1.7 from Basu et al. (2017), written (and proved) in a neater manner. Equation (1.3) is new. It is not hard to extend the proof of Theorem 1 to the non-reversible setup if one replaces the assumption that $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix},\mu_n}^{(n)}(\delta))$ with the assumption that $t_{\mathrm{rel}}^{(n)} \log(t_{\mathrm{rel}}^{(n)} \vee e) = o(t_{\mathrm{mix},\mu_n}^{(n)}(\delta))$. Similarly, under the assumption that $t_{\mathrm{rel}}^{(n)} \log(t_{\mathrm{rel}}^{(n)} \vee e) = o(t_{\mathrm{mix}}(\delta))$, even without reversibility we have that cutoff is equivalent to the occurrence of hit_α-cutoff for some (and in fact, for all) $\alpha \in (0, 1)$.

In Griffiths et al. (2014) the following general inequality was proved (without a reversibility assumption).

Theorem B. *Fix $0 < \varepsilon < 1/2$. For any irreducible finite Markov chain*

$$\varepsilon t_{\mathrm{H}}(\varepsilon) \leqslant t_{\mathrm{H}}(1/2).$$

The following proposition offers an upper bound on $t_{\mathrm{H}}(\varepsilon) - t_{\mathrm{H}}(1 - \varepsilon)$, which in some cases is considerably better than the bound in Theorem B (in particular, this is the case under reversibility when the product condition holds).

Proposition 1.13. *There exists an absolute constant $C > 0$ such that for every finite irreducible chain (Ω, P, π),*

$$t_{\mathrm{H},\mu}(\varepsilon) - t_{\mathrm{H},\mu}(1 - \varepsilon) \leqslant C\varepsilon^{-1} t_{\mathrm{rel}}, \text{ for every } 0 < \varepsilon < 1/2 \text{ and } \mu \in \mathscr{P}(\Omega). \quad (1.4)$$

The same holds when μ is omitted from the l.h.s..

Remark 1.14. In the non-reversible setup it is possible that $t_{\mathrm{H}}(\varepsilon) \leqslant C\sqrt{t_{\mathrm{rel}}}$ for all $\varepsilon \in (0, 1)$. To see this, consider a walk on the n-cycle with a fixed clockwise bias. Here $\max_{x,y} \mathbb{E}_x[T_y] \leqslant Cn$, while the additive symmetrization is simple random walk on the n-cycle and hence $t_{\mathrm{rel}} = \Theta(n^2)$ (e.g. Levin et al., 2017, Section 12.3). Conversely, under reversibility we have that $t_{\mathrm{H}}(1/2) \geqslant t_{\mathrm{rel}}$ (e.g. Aldous and Fill, 2002, Lemma 4.39) and thus (1.4) implies that $t_{\mathrm{H}}(\varepsilon) - t_{\mathrm{H}}(1 - \varepsilon) \leqslant C\varepsilon^{-1} t_{\mathrm{H}}(1/2)$, recovering Theorem B, up to a constant.

2. Starr's Maximal inequality

In this section we present the machinery that will be utilized in the proof of the main results. The most important tool we shall utilize is Starr's L^2 maximal inequality (Theorem 2.3). We start with a few basic definitions and facts. We denote $\mathbb{Z}_+ := \{n \in \mathbb{Z} : n \geqslant 0\}$ and $\mathbb{R}_+ := \{t \in \mathbb{R} : t \geqslant 0\}$.

Definition 2.1. Let (Ω, P, π) be a finite chain. For $f \in \mathbb{R}^\Omega$, let

$$\mathbb{E}_\pi[f] := \sum_{x \in \Omega} \pi(x)f(x) \quad \text{and} \quad \mathbf{Var}_\pi f := \mathbb{E}_\pi[(f - \mathbb{E}_\pi f)^2].$$

The inner-product $\langle \cdot, \cdot \rangle_\pi$ and L^p norms ($1 \leqslant p < \infty$) are

$$\langle f, g \rangle_\pi := \mathbb{E}_\pi[fg] \quad \text{and} \quad \|f\|_p := (\mathbb{E}_\pi[|f|^p])^{1/p}.$$

We identify P^k and H_t with the operators $P^k, H_t : L^2(\mathbb{R}^\Omega, \pi) \to L^2(\mathbb{R}^\Omega, \pi)$, defined by $P^k f(x) := \sum_{y \in \Omega} P^k(x, y) f(y)$ and $H_t f(x) := \sum_{y \in \Omega} H_t(x, y) f(y) = \mathbb{E}_x[f(X_t)]$.

Recall that if P is reversible, then also H_t is self-adjoint w.r.t. $\langle \cdot, \cdot \rangle_\pi$. Recall that the spectral gap can be characterized as $\lambda = \inf\{\frac{\langle (I-P)f, f \rangle_\pi}{\mathbf{Var}_\pi f} : f \in \mathbb{R}^\Omega \text{ non-constant}\}$ and that $\frac{d}{dt}\mathbf{Var}_\pi H_t f = -\langle (I - P)H_t f, H_t f \rangle_\pi \leqslant -2\lambda \mathbf{Var}_\pi H_t f$, from which the following lemma, known as the Poincaré inequality or the L^2-contraction Lemma, follows.

Lemma 2.2. *Let (Ω, P, π) be a finite irreducible chain. Let $f \in \mathbb{R}^\Omega$. Then*

$$\forall t \geqslant 0, \quad \mathbf{Var}_\pi H_t f \leqslant e^{-2\lambda t} \mathbf{Var}_\pi f. \tag{2.1}$$

We now state a particular case of Starr's maximal inequality (Starr, 1966 Proposition 3). The proof in the discrete time setup could be found in Basu et al. (2017). Let $f \in \mathbb{R}^\Omega$. Define its **maximal function** by $f^*(x) := \sup_{t \geqslant 0} |H_t f(x)|$.

Theorem 2.3 (Maximal inequality, Starr, 1966)**.** *Let (Ω, P, π) be a finite reversible irreducible Markov chain. Then for every $f \in \mathbb{R}^\Omega$*

$$\|f^*\|_2 \leqslant 2\|f\|_2. \tag{2.2}$$

For any $B \subseteq \Omega$ and $s \in \mathbb{R}_+$, set $\rho(B) := \sqrt{\pi(B)(1 - \pi(B))} = \sqrt{\mathbf{Var}_\pi 1_B}$ and $\sigma_s := \rho(B)e^{-\lambda s}$. Note that by Lemma 2.2, $\sigma_s \geqslant \sqrt{\mathbf{Var}_\pi H_s 1_B}$. We define the *good set for B from time s within m standard-deviations* to be

$$G_s(B, m) := \left\{ y : |\mathrm{H}_y^t(B) - \pi(B)| < m\sigma_s, \text{ for all } t \geqslant s \right\}. \tag{2.3}$$

The following corollary follows by combining Lemma 2.2 with Theorem 2.3.

Corollary 2.4. *Let (Ω, P, π) be a finite reversible irreducible chain. Then*

$$\pi(G_s(B, m)) \geqslant 1 - \frac{4}{m^2}, \text{ for all } B \subseteq \Omega, s \geqslant 0 \text{ and } m > 0. \tag{2.4}$$

Proof: For any $s \geqslant 0$, let $f_s(x) := H_s(1_B(x) - \pi(B)) = \mathrm{H}_x^s(B) - \pi(B)$. Then by Lemma 2.2 and Theorem 2.3

$$\|f_s^*\|_2 \leqslant 2\|f_s\|_2 \leqslant 2e^{-\lambda s}\|1_B(x) - \pi(B)\|_2 = 2\sigma_s. \tag{2.5}$$

For any $t \geqslant 0$, $H_t f_s(x) = f_{t+s}(x) = \mathrm{H}_x^{t+s}(B) - \pi(B)$. Then, in the notation of Theorem 2.3,

$$f_s^*(x) := \sup_{t \geqslant 0} |H_t f_s(x)| = \sup_{t \geqslant s} |\mathrm{H}_x^t(B) - \pi(B)|.$$

Hence $D := \{x \in \Omega : f_s^*(x) \geqslant m\sigma_s\}$ is the complements of $G_s(B, m)$. Thus by Markov inequality and (2.5)

$$1 - \pi(G_s(B, m)) = \pi(D) = \pi\left\{ (f_s^*)^2 \geqslant (m\sigma_s)^2 \right\} \leqslant 4m^{-2}. \quad \square$$

Conjecture 2.5. *Let (Ω, P, π) be a finite irreducible Markov chain (not necessarily reversible). Define $t_* := \inf\{t : \mathbf{Var}_\pi H_t f \leqslant e^{-2}\mathbf{Var}_\pi f \text{ for all } f \in \mathbb{R}^\Omega\}$ as in Conjecture 1.10. Let $f_s^*(x) := \sup_{t \geqslant s} |H_t f(x)|$. Then for every $\varepsilon \in (0, 1)$ there exists a constant C_ε (independent of (Ω, P, π)) such that for every $f \in \mathbb{R}^\Omega$ of mean zero (i.e. $\mathbb{E}_\pi f = 0$) we have that*

$$\mathbb{E}_\pi f_{C_\varepsilon t_*}^* = \|f_{C_\varepsilon t_*}^*\|_1 \leqslant \varepsilon\|f\|_\infty. \tag{2.6}$$

Remark 2.6. If (2.6) holds then one could prove the upper bound on $t_{\mathrm{mix}}(\varepsilon)$ from Conjecture 1.10 by imitating the proof of the upper bound on $t_{\mathrm{mix}}(\varepsilon)$ from Proposition 1.9, with (2.6) replacing (2.2).

3. Inequalities relating $t_{\mathrm{mix}}(\cdot)$ and $\mathrm{hit}.(\cdot)$

Our aim in this section is to obtain inequalities relating $t_{\mathrm{mix}}(\varepsilon)$ and $\mathrm{hit}_\beta(\delta)$ for suitable values of β, ε and δ using Corollary 2.4. As was shown in Basu et al. (2017), these two notions of mixing are intimately connected to each other. In this section we refine the analysis from Basu et al. (2017). Corollary 3.2 below contains the more difficult half of Proposition 1.9.

Lemma 3.1. *Let (Ω, P, π) be a finite irreducible reversible chain. Let $\mu \in \mathscr{P}(\Omega)$, $\alpha, p \in (0,1)$, $s \geqslant 0$ and $B \subseteq \Omega$. Then*

$$\pi(B) - \mathrm{H}_\mu^{\mathrm{hit}_{\alpha,\mu}(p)+s}(B) \leqslant p\pi(B) + 2(1-p)e^{-\lambda s}\sqrt{(1-\alpha)^{-1}\pi(B)\pi(B^c)}. \quad (3.1)$$

Proof: Denote $\tau := \mathrm{hit}_{\alpha,\mu}(p)$, $\rho(B) := \pi(B)\pi(B^c)$ and $\ell := 2e^{-\lambda s}\sqrt{(1-\alpha)^{-1}}$. Consider

$$H := \left\{ y : |\mathrm{H}_y^t(B) - \pi(B)| < \ell\sqrt{\rho(B)} \text{ for all } t \geqslant s \right\}.$$

By Corollary 2.4, $\pi(H) \geqslant \alpha$. By the Markov property and the definition of H,

$$\mathrm{H}_\mu[X_{\tau+s} \in B \mid T_H \leqslant \tau] \geqslant \pi(B) - \ell\sqrt{\rho(B)}.$$

Since $\tau = \mathrm{hit}_{\alpha,\mu}(p)$ and $\pi(H) \geqslant \alpha$, we get that $\mathrm{H}_\mu[T_H \leqslant \tau] \geqslant 1 - p$. Thus

$$\pi(B) - \mathrm{H}_\mu^{\tau+s}(B) \leqslant \pi(B) - \mathrm{H}_\mu[X_{\tau+s} \in B, T_H \leqslant \tau]$$
$$= \mathrm{H}_\mu[T_H > \tau]\pi(B) + \mathrm{H}_\mu[T_H \leqslant \tau]\left(\pi(B) - \mathrm{H}_\mu[X_{\tau+s} \in B \mid T_H \leqslant \tau]\right)$$
$$\leqslant p\pi(B) + (1-p)\left(\pi(B) - \mathrm{H}_\mu[X_{\tau+s} \in B \mid T_H \leqslant \tau]\right) \leqslant p\pi(B) + (1-p)\ell\sqrt{\rho(B)}.$$

This concludes the proof of (3.1). $\qquad\square$

Corollary 3.2. *Let (Ω, P, π) be a reversible irreducible finite chain. Let $\mu \in \mathscr{P}(\Omega)$. Let $0 < \varepsilon, \alpha < 1$ and let $\delta \in (0, \varepsilon)$. Denote $s := 0 \vee t_{\mathrm{rel}} \log\left(\frac{1-\varepsilon+\delta}{\sqrt{(1-\alpha)\varepsilon\delta}}\right)$. Then*

$$t_{\mathrm{mix},\mu}(\varepsilon) \leqslant \mathrm{hit}_{\alpha,\mu}(\varepsilon - \delta) + s. \quad (3.2)$$

Proof: Denote $t := \mathrm{hit}_{\alpha,\mu}(\varepsilon - \delta)$. If $1 - \varepsilon + \delta < \sqrt{(1-\alpha)\varepsilon\delta}$, then we can prove (3.2) for $\delta' < \delta$ such that $1 - \varepsilon + \delta' = \sqrt{(1-\alpha)\varepsilon\delta'}$, so we may assume that $s = t_{\mathrm{rel}} \log\left(\frac{1-\varepsilon+\delta}{\sqrt{(1-\alpha)\varepsilon\delta}}\right) \geqslant 0$. Let $B \subseteq \Omega$. Denote $\pi(B) = z$. By (3.1) and the fact that $h(z) := (\varepsilon - \delta)z + 2\sqrt{\varepsilon\delta z(1-z)}$ attains its maximum in $[0,1]$ at $z_* = \frac{\varepsilon}{\varepsilon+\delta}$ and $h(z_*) = \varepsilon$, we get that

$$\pi(B) - \mathrm{H}_\mu^{t+s}(B) \leqslant (\varepsilon - \delta)z + 2\sqrt{\varepsilon\delta z(1-z)} \leqslant \varepsilon,$$

The claim now follows by maximizing over B. $\qquad\square$

4. Proofs of Theorem 1 and Propositions 1.9 and 1.13

Let $A \subsetneq \Omega$. Let Q_A (resp. P_A) be the restriction of Q (resp. P) to A. Note that Q_A and P_A are substochastic. The spectral gap of P_A, denoted by $\lambda(A)$, is defined as the minimal eigenvalue of $I - Q_A$ (it is often referred to as the Dirichlet eigenvalue of $I - Q$ on A). Denote $\Lambda(c) := \min_{A:\pi(A) \leqslant c} \lambda(A)$. Let π_A denote π conditioned on A (i.e. $\pi_A(y) = 1_{y\in A}\pi(y)/\pi(A)$).

Lemma 4.1. *Let (Ω, P, π) be a finite irreducible Markov chain. Let $A \subsetneq \Omega$ be non-empty. Let $\alpha > 0$ and $w \geqslant 0$. Let $B(A, t, \alpha) := \{y : \mathrm{H}_y [T_{A^c} > t] \geqslant \alpha\}$. Then*

$$\lambda(A) \geqslant \pi(A^c)\lambda \quad \text{and so} \quad \forall c \in (0, 1), \ \Lambda(c) \geqslant (1 - c)\lambda. \tag{4.1}$$

$$\frac{\mathrm{H}_\pi[T_{A^c} > t]}{\pi(A)} = \mathrm{H}_{\pi_A}[T_{A^c} > t] \leqslant e^{-\lambda(A)t} \leqslant e^{-\Lambda(\pi(A))t} \leqslant e^{-\lambda\pi(A^c)t}, \text{ for all } t. \tag{4.2}$$

$$\pi(B(A, t, \alpha)) \leqslant \pi(A)e^{-2\lambda(A)t}\alpha^{-2} \leqslant \pi(A)e^{-2\pi(A^c)\lambda t}\alpha^{-2}. \tag{4.3}$$

Remark 4.2. The inequality (4.1) is well-known (cf. Goel et al., 2006, Eq. (1.4)). We include its proof for the sake of completeness.

Proof of Lemma 4.1: Consider the inner-product $\langle \cdot, \cdot \rangle_{\pi_A}$ on $\mathbb{R}^A$ given by $\langle f, g \rangle_{\pi_A} := \sum_{a \in A} \pi_A(a)f(a)g(a)$. We identify every $g \in \mathbb{R}^A$ with its extension to Ω obtained by setting $g \equiv 0$ on $\Omega \setminus A$. Using the fact that for all $g \in \mathbb{R}^A$ we have $\langle (I - Q)g, g \rangle_\pi = \langle (I - P)g, g \rangle_\pi$, together with the Perron-Frobenius (for the non-negativity) and the Courant-Fischer variational characterization of eigenvalues we have

$$\lambda(A) = \inf \left\{ \langle (I - P_A)g, g \rangle_{\pi_A} / \langle g, g \rangle_{\pi_A} : g \in \mathbb{R}_+^A, \ g \text{ non-constant} \right\}. \tag{4.4}$$

We may identify each $g \in \mathbb{R}^A$ with $g \in \mathbb{R}^\Omega$ s.t. $g \equiv 0$ on A^c and vice-versa. Hence

$$\lambda(A) = \inf \left\{ \langle (I - P)g, g \rangle_\pi / \langle g, g \rangle_\pi : g \in \mathbb{R}_+^\Omega \text{ non-constant and } g = 0 \text{ on } A^c \right\}.$$

Also observe that by the Cauchy-Schwarz inequality, for all $g \geqslant 0$ such that $g = 0$ on A^c, we have that

$$\mathbf{Var}_\pi g = \mathbb{E}_\pi g^2 - (\mathbb{E}_\pi g)^2 = \mathbb{E}_\pi g^2 - (\pi(A)\mathbb{E}_{\pi_A} g)^2$$
$$\geqslant \mathbb{E}_\pi g^2 - [\pi(A)]^2 \mathbb{E}_{\pi_A} g^2 = \pi(A^c)\langle g, g \rangle_\pi.$$

Thus by the extremal characterization of the spectral gap

$$\lambda = \inf \left\{ \langle (I - P)g, g \rangle_\pi / \mathbf{Var}_\pi g : g \in \mathbb{R}^\Omega \text{ non-constant} \right\} \leqslant \lambda(A)/\pi(A^c). \tag{4.5}$$

Denote the heat kernel of the chain killed upon escaping A by H_t^A. Then $H_t^A = e^{-t} \sum_{k=0}^\infty (kP_A)^t/k!$. Let $g \in \mathbb{R}^A$. Denote $g_t := H_t^A g$ and $\|g\|_{A,p}^p := \sum_{a \in A} \pi_A(a)|g(a)|^p$, for $1 \leqslant p < \infty$. Then by (4.4)

$$\frac{d}{dt}\|g_t\|_{A,2}^2 = \frac{d}{dt}\langle H_t^A g, H_t^A g \rangle_{\pi_A} = -2\langle (I - P_A)H_t^A g, H_t^A g \rangle_{\pi_A} \leqslant -2\lambda(A)\|g_t\|_{A,2}^2. \tag{4.6}$$

Hence $\|g_t\|_{A,1}^2 \leqslant \|g_t\|_{A,2}^2 \leqslant \|g\|_{A,2}^2 e^{-2\lambda(A)t}$. Taking $g = 1_A$, by (4.5) we get that

$$\mathrm{H}_{\pi_A}[T_{A^c} > t] = \|g_t\|_{A,1} \leqslant \|g_t\|_{A,2} \leqslant \|1_A\|_{A,2}e^{-\lambda(A)t} = e^{-\lambda(A)t} \leqslant e^{-\pi(A^c)\lambda t}. \tag{4.7}$$

Write $B = B(A, t, \alpha)$. Then $B \subseteq \{a \in A : g_t^2(a) > \alpha^2\}$. By (4.7)

$$\pi(B)/\pi(A) = \pi_A(B) \leqslant \alpha^{-2}\|g_t\|_{A,2}^2 \leqslant \alpha^{-2}e^{-2\pi(A^c)\lambda t}. \quad \square$$

Corollary 4.3. *Let (Ω, P, π) be a finite irreducible Markov chain. Let $\mu \in \mathscr{P}(\Omega)$, $0 < \varepsilon, \alpha < 1$ and $0 < \delta \leqslant 1 - \varepsilon$. Denote $s := \frac{\log(\frac{1-\alpha}{\delta})}{\Lambda(1-\alpha)} \leqslant \frac{\log(\frac{1-\alpha}{\delta})}{\alpha\lambda}$. Then*

$$\mathrm{hit}_{\alpha,\mu}(\varepsilon + \delta) \leqslant t_{\mathrm{mix},\mu}(\varepsilon) + s \vee 0.$$

Proof: By decreasing δ if necessary, we may assume that $s \geqslant 0$. Take an arbitrary set A with $\pi(A) \geqslant \alpha$ and $\mu \in \mathscr{P}(\Omega)$. It follows by coupling of the chain with initial distribution H_μ^t with the stationary chain that for all $t \geqslant 0$

$$\mathrm{H}_\mu[T_A > t + s] \leqslant d_\mu(t) + \mathrm{H}_\pi[T_A > s] \leqslant d_\mu(t) + (1-\alpha)e^{-\lambda(A^c)s} \leqslant d_\mu(t) + \delta. \quad (4.8)$$

where the penultimate inequality follows from (4.2) and the last from the choice of s. Plugging $t = t_{\mathrm{mix},\mu}(\varepsilon)$ in (4.8) and maximizing over A with $\pi(A) \geqslant \alpha$ concludes the proof. $\qquad\square$

Proof of Proposition 1.9: By combining Corollaries 3.2 and 4.3 it remains only to show that $t_{\mathrm{mix},\mu}(\varepsilon) \leqslant \mathrm{hit}_{1-\delta,\mu}(\varepsilon-\delta)+\log(2/\delta^3)t_{\mathrm{rel}}\log(t_{\mathrm{rel}}\vee e)$, for $0 < \delta \leqslant \varepsilon < 1$. Set $t := \log(2/\delta^3)t_{\mathrm{rel}}\log(t_{\mathrm{rel}} \vee e)$. Let $A \subseteq \Omega$. For all $x \in \Omega$ and $s \geqslant 0$ we have that $|\frac{d}{ds}H_s(x,A)| = |\delta_x H_s(P-I)1_A| \leqslant 1$ and so by (2.1) (used to deduce that $(\sum_x \pi(x)|H_s(x,A) - \pi(A)|)^2 \leqslant \mathbf{Var}_\pi H_s 1_A \leqslant e^{-2\lambda s}\mathbf{Var}_\pi 1_A \leqslant \frac{e^{-2\lambda s}}{4})$

$$\sum_x \pi(x)|\frac{d}{ds}[H_s(x,A) - \pi(A)]^2| \leqslant 2\sum_x \pi(x)|H_s(x,A) - \pi(A)| \leqslant e^{-\lambda s}.$$

Denote $g(x) := \int_t^\infty |\frac{d}{ds}[H_s(x,A) - \pi(A)]^2|ds$. Then

$$\|g\|_1 \leqslant \int_t^\infty e^{-\lambda s}ds = e^{-\lambda t}/\lambda \leqslant \delta^3/2$$

and $\mathbf{Var}_\pi H_t 1_A \leqslant \frac{1}{4}e^{-2\lambda t} \leqslant \delta^6/8$. Let

$$B := \{b \in \Omega : g(b) + |H_t(b,A) - \pi(A)|^2 > \delta^2\}.$$

Then

$$\pi(B) \leqslant \delta^{-2}(\|g\|_1 + \mathbf{Var}_\pi H_t 1_A) \leqslant \delta.$$

From the definition of B it follows that if $x \notin B$ then $\sup_{s \geqslant t}|H_s(x,A) - \pi(A)| \leqslant \delta$. Thus

$$|\mathrm{H}_\mu^{\mathrm{hit}_{1-\delta,\mu}(\varepsilon-\delta)+t}(A) - \pi(A)| \leqslant \mathrm{H}_\mu[T_{B^c} > \mathrm{hit}_{1-\delta,\mu}(\varepsilon - \delta)] + \delta \leqslant \varepsilon. \quad \square$$

Corollary 4.4. *Let (Ω, P, π) be an irreducible finite Markov chain. Let $\mu \in \mathscr{P}(\Omega)$. Let $0 < \varepsilon < \delta < 1$ and $0 < \beta \leqslant \gamma < 1$. Denote $s := \frac{\log(\frac{1-\beta}{(1-\gamma)\varepsilon^2})}{2\Lambda(1-\beta)} \leqslant \frac{1}{2\beta}t_{\mathrm{rel}}\log\left(\frac{1-\beta}{(1-\gamma)\varepsilon^2}\right)$. Then*

$$\mathrm{hit}_{\gamma,\mu}(\delta) \leqslant \mathrm{hit}_{\beta,\mu}(\delta) \leqslant \mathrm{hit}_{\gamma,\mu}(\delta - \varepsilon) + s. \quad (4.9)$$

Proof: The first inequality in (4.9) is trivial. We now prove the second inequality in (4.9). Let A be an arbitrary set with $\pi(A) \geqslant \beta$. Using the notation from Lemma 4.1, let $B := B(A^c, s, \varepsilon) := \{y : \mathrm{H}_y[T_A > s] > \varepsilon\}$. Then by (4.3)

$$\pi(B) \leqslant \varepsilon^{-2}\pi(A^c)e^{-\log\left(\frac{1-\beta}{(1-\gamma)\varepsilon^2}\right)} \leqslant 1 - \gamma.$$

Set $t_1 := \mathrm{hit}_{\gamma,\mu}(\delta - \varepsilon)$. Then $\mathrm{H}_\mu[T_{B^c} > t_1] \leqslant \delta - \varepsilon$. By the definition of B and the Markov property,

$$\mathrm{H}_\mu[T_A > t_1 + s \mid T_{B^c} \leqslant t_1] \leqslant \max_{d \notin B}\mathrm{H}_d[T_A > s] \leqslant \varepsilon.$$

Whence

$$\mathrm{H}_\mu[T_A > t_1 + s] \leqslant \mathrm{H}_\mu[T_{B^c} > t_1] + \mathrm{H}_\mu[T_A > t_1 + s \mid T_{B^c} \leqslant t_1] \leqslant (\delta - \varepsilon) + \varepsilon = \delta.$$

Since A was arbitrary, this concludes the proof of (4.9). $\qquad\square$

Proposition 4.5. *Let (Ω_n, P_n, π_n) be a sequence of finite irreducible reversible chains. Let $\mu_n \in \mathscr{P}(\Omega_n)$. Assume that $t_{\text{rel}}^{(n)} = o(t_{\text{mix},\mu_n}^{(n)}(\delta))$ for some $0 < \delta < 1$. Then for all $\beta \in (0,1)$,*

$$(1 - o(1))t_{\text{mix},\mu_n}^{(n)}(\delta) \leqslant \text{hit}_{\beta,\mu_n}^{(n)}(\delta/2) \leqslant (1 + o(1))t_{\text{mix},\mu_n}^{(n)}(\delta/8). \tag{4.10}$$

$$t_{\text{rel}}^{(n)} = o(\text{hit}_{\beta,\mu_n}^{(n)}(\delta/2)) \tag{4.11}$$

Moreover, (1) below implies (2):

(1) *There exists some $\alpha \in (0,1)$ such that the sequence exhibits a $\text{hit}_{\alpha,\mu_n}$-cutoff.*
(2) *For every $\alpha \in (0,1)$ the sequence exhibits a $\text{hit}_{\alpha,\mu_n}$-cutoff.*

Moreover, if (1) holds for some α then

$$\lim_{n \to \infty} \text{hit}_{\beta,\mu_n}^{(n)}(1/4)/\text{hit}_{\alpha,\mu_n}^{(n)}(1/4) = 1, \quad for\ all \quad \beta \in (0,1). \tag{4.12}$$

Proof: First note that (4.10) follows from Proposition 1.9 together with the assumption that $t_{\text{rel}}^{(n)} = o(t_{\text{mix},\mu_n}^{(n)}(\delta))$, while (4.11) follows from the (4.10). Assume (1) holds for some α. This in conjunction with Corollary 4.4 and (4.11) implies (4.12), which combined with another application of Corollary 4.4 yield (2) (cf. Basu et al., 2017, Proposition 3.6).   □

We now present two lemmas regarding expected hitting times inequalities. Proposition 1.13 follows by combining these two lemmas. The first of which is simpler and gives better bounds for some poruses. The second one gives better asymptotic in the sense that it follows from it that $t_{\text{H}}(\varepsilon) - t_{\text{H}}(1 - \varepsilon) \leqslant Ct_{\text{rel}}\varepsilon^{-1}$, for some absolute constant C, whereas the first lemma only implies that $t_{\text{H}}(\varepsilon) - t_{\text{H}}(1 - \varepsilon) \leqslant Ct_{\text{rel}}\varepsilon^{-1} \log(1/\varepsilon)$.

Lemma 4.6. *Let (Ω, P, π) be a finite irreducible Markov chain. Let $\varepsilon \in (0,1)$ and $k \geqslant 0$. Let $A \subseteq \Omega$ be such $\pi(A^c) \leqslant \varepsilon$. There exists $I = I(A,k) \subseteq \Omega$ with $\pi(I^c) \leqslant \frac{\varepsilon}{2 \cdot 3^k}$, such that*

$$\mathbb{E}_z[T_A] \leqslant (9/4 + k/2)(\log 3)/\lambda(A^c) \leqslant (9/4 + k/2)(1 - \varepsilon)^{-1}t_{\text{rel}} \log 3, \ for\ all\ z \in I.$$

Proof: Fix $k \geqslant 0$. Denote $a := (\log 3)/\lambda(A^c) \leqslant \Lambda^{-1}(\varepsilon) \log 3 \leqslant (1 - \varepsilon)^{-1}t_{\text{rel}} \log 3$. Consider

$$I_i = I_i(A,k) := \left\{ y : \text{H}_y\left[2T_A > (k + 3i)a\right] \leqslant 3^{-i} \right\}.$$

Then by (4.3) we have that

$$\pi(I_i^c) \leqslant (3^{-i})^{-2}\pi(A^c)e^{-(k+3i)\log 3} \leqslant \varepsilon(1/3)^{i+k}.$$

Let $I := \bigcap_{i \in \mathbb{N}} I_i$. Then $\pi(I^c) \leqslant \sum_{i \in \mathbb{N}} \pi(I_i^c) \leqslant \varepsilon \sum_{i \in \mathbb{N}} 3^{-(i+k)} = \frac{\varepsilon}{2 \cdot 3^k}$. By construction,

$$\text{H}_z\left[T_A - \frac{ka}{2} > \frac{3ia}{2}\right] \leqslant 3^{-i}, \ \text{for all } z \in I \text{ and } i \in \mathbb{N}.$$

Hence $\mathbb{E}_z[T_A] \leqslant \frac{3a/2}{1 - \frac{1}{3}} + ka = (9/4 + k/2)a$ for all $z \in I$, as desired.   □

Lemma 4.7. *Let (Ω, P, π) be a finite irreducible chain. Let $A \subseteq \Omega$ be non-empty. Denote $\varepsilon := \pi(A)$. Let $t \geqslant 1$. There exists $J = J(A,t) \subseteq \Omega$ such that*

(i) *$\pi(J) \geqslant 1 - 3e^{-2t}/2$.*
(ii) *For any $z \in J$ we have that $\mathbb{E}_z[T_A] \leqslant t_{\text{rel}}[t + \log 2(1 + 9\varepsilon^{-1}) + |\log \varepsilon|]$.*

(iii) *For any $z \in J$ and $i \in \mathbb{N}$ we have that*

$$H_z[\lambda T_A \leqslant t + i \log 2(1 + \frac{3}{2\varepsilon}) + |\log \varepsilon|] \geqslant (1 - 2^{-i})^2.$$

Proof: Denote $r = r(\varepsilon, t) := (t + |\log \varepsilon|)t_{\text{rel}}$, $\ell := t_{\text{rel}} \log 2$ and $s = s(\varepsilon) := (3/2)\varepsilon^{-1} t_{\text{rel}} \log 2$. For any $i \in \mathbb{N}$, let $C_i := \{y : H_y[T_A \leqslant is] \geqslant 1 - 2^{-i}\}$. Then by (4.3) we have that

$$\pi(C_i^c) \leqslant 2^{2i}\pi(A^c)e^{-3i\log 2} = (1 - \varepsilon)2^{-i}. \tag{4.13}$$

Define

$$J_i := \left\{ y : H_y^{r+i\ell}[C_i] \geqslant 1 - 2^{-i} = 1 - (1 - \varepsilon)2^{-i} - \varepsilon 2^{-i} \right\}.$$

Denote $g_i := H_{r+i\ell}1_{C_i}$. By stationarity $\mathbb{E}_\pi[g_i] = \mathbb{E}_\pi[1_{C_i}] = \pi(C_i) \geqslant 1 - (1 - \varepsilon)2^{-i}$. Hence

$$J_i^c \subseteq \{x : g_i(x) \geqslant \mathbb{E}_\pi[g_i] + \varepsilon 2^{-i}\}. \tag{4.14}$$

By (2.1) and the fact that $z_1(1 - z_1) < z_2(1 - z_2)$ if $z_1, z_2 \in [0,1]$ and $|z_1 - 1/2| > |z_2 - 1/2|$,

$$\mathbf{Var}_\pi g_i \leqslant e^{-2\lambda(r+i\ell)}\mathbf{Var}_\pi 1_{C_i} \leqslant \varepsilon^2 2^{-2i}e^{-2t}\pi(C_i)\pi(C_i^c) \leqslant \varepsilon^2 2^{-3i}(1 - 2^{-i})e^{-2t}.$$

By the Chebyshev's inequality and (4.14)

$$\pi(J_i^c) \leqslant \varepsilon^{-2}2^{2i}\mathbf{Var}_\pi g_i \leqslant 2^{-i}e^{-2t}(1 - 2^{-i}).$$

Denote $J := \bigcap_{i=1}^\infty J_i$. By a union bound

$$\pi(J^c) \leqslant \sum_{i \in \mathbb{N}} \pi(J_i^c) \leqslant e^{-2t}(1 + 3/16 + \sum_{i \geqslant 3} 2^{-i}) < 3e^{-2t}/2.$$

Fix some $z \in J$ and $i \geqslant 1$. Note that because $J \subseteq J_i$ we get from the definitions of J_i and C_i together with the Markov property that

$$H_z[T_A \leqslant r + i(\ell + s)] \geqslant H_z[X_{r+i\ell} \in C_i] \min_{y \in C_i} H_y[T_A \leqslant is] \geqslant (1 - 2^{-i})^2.$$

From this it is easy to verify that indeed $\mathbb{E}_z[T_A] \leqslant r + 6(s + \ell)$. $\qquad\square$

The following lemma asserts that, for a fixed starting distribution μ such that t_{rel} is much smaller than $t_{\text{mix},\mu}$, a set A which is "worst" in expectation (i.e. $\mathbb{E}_\mu[T_A] = t_{\text{H},\mu}(\pi(A))$) is almost the "*worst in probability*" (in the sense of Definition 1.2) for all times. By this we mean that this is the case up to a small size and time shifts and up to a small difference in the chance of not being hit by any given time.

Lemma 4.8. *Let (Ω, P, π) be a finite irreducible reversible chain. Let $\mu \in \mathscr{P}(\Omega)$ and $0 < \varepsilon < 1$. Let $A \subseteq \Omega$ such that $\mathbb{E}_\mu[T_A] = t_{\text{H},\mu}(1 - \varepsilon)$ and $\pi(A) \geqslant 1 - \varepsilon$. Denote $\rho = \rho(\varepsilon) := 3(1 - \varepsilon)^{-1}t_{\text{rel}} \log 3$. Then for all $r \geqslant 1$*

$$\max_{B \subseteq \Omega : \pi(B) \geqslant 1 - \varepsilon/2} H_\mu[T_B - T_A \geqslant r\rho] \leqslant r^{-1}. \tag{4.15}$$

In particular, for all $t \geqslant 0$, $r \geqslant 1$, $q \in (0,1)$ and $B \subseteq \Omega$ with $\pi(B^c) \leqslant \varepsilon/2$ we have that

$$\begin{aligned}
&H_\mu[T_A \leqslant t] - H_\mu[T_B \leqslant t + r\rho] \leqslant H_\mu[T_B \geqslant t + r\rho, T_A \leqslant t] \leqslant r^{-1} \quad \textit{and} \\
&H_\mu[T_A \leqslant \text{hit}_{1-\varepsilon/2,\mu}(1 - q) - r\rho] \leqslant q + r^{-1}.
\end{aligned} \tag{4.16}$$

Proof. We first note that the first row in (4.16) follows from (4.15) trivially. The second row in (4.16) follows from the first by taking $t = \mathrm{hit}_{1-\varepsilon/2,\mu}(1-q) - r\rho$ and picking some $B \subseteq \Omega$ such that $\pi(B^c) \leqslant \varepsilon/2$ and $\mathrm{H}_\mu[T_B \leqslant \mathrm{hit}_{1-\varepsilon/2,\mu}(1-q)] \leqslant q$.

We now prove (4.15). Let I be as in Lemma 4.6 w.r.t. A with the choice of $k = 0$. Then $\pi(I^c) \leqslant \varepsilon/2$ and for any $z \in I$ we have $\mathbb{E}_z[T_A] \leqslant \rho$. Denote $D := I \cap B$. Then by the assumption that $\pi(B^c) \leqslant \varepsilon/2$, we have that $\pi(D) \geqslant 1 - \pi(B^c) - \pi(I^c) \geqslant 1 - \varepsilon$. Hence

$$\mathbb{E}_\mu[T_D] \leqslant t_{\mathrm{H},\mu}(1-\varepsilon) = \mathbb{E}_\mu[T_A]. \tag{4.17}$$

For any $\ell \in \mathbb{R}$ denote $\ell^+ := \max\{\ell, 0\}$. Since $D \subseteq I$, by the Markov property we have that $\mathbb{E}_\mu[(T_A - T_D)^+] \leqslant \max_{z \in I} \mathbb{E}_z[T_A] \leqslant \rho$. Thus by (4.17)

$$\mathbb{E}_\mu[(T_D - T_A)^+] = \mathbb{E}_\mu[T_D - T_A] + \mathbb{E}_\mu[(T_A - T_D)^+] \leqslant 0 + \rho = \rho.$$

By Markov inequality and the fact that $D \subseteq B$ we get that for all $r \geqslant 1$

$$\mathrm{H}_\mu[T_B - T_A \geqslant r\rho] \leqslant \mathrm{H}_\mu[T_D - T_A \geqslant r\rho] \leqslant r^{-1}. \quad \square$$

We are now ready to prove Theorem 1.

Proof of Theorem 1. The equivalence between (i) and (ii) follows from Proposition 1.9 together with (4.11). We now show that (ii)$\Longrightarrow$(iii). Let $\alpha \in (0,1)$. Assume that

$$\mathrm{hit}^{(n)}_{1-\alpha,\mu_n}(\varepsilon) - \mathrm{hit}^{(n)}_{1-\alpha,\mu_n}(1-\varepsilon) = o\left(\mathrm{hit}^{(n)}_{1-\alpha,\mu}(1/4)\right), \quad \text{for all } 0 < \varepsilon < 1/4.$$

Then by Proposition 4.5,

$$\mathrm{hit}^{(n)}_{\beta,\mu_n}(\varepsilon) - \mathrm{hit}^{(n)}_{\beta,\mu_n}(1-\varepsilon) = o\left(\mathrm{hit}^{(n)}_{\beta,\mu_n}(1/4)\right), \quad \text{for all } 0 < \varepsilon < 1/4 \text{ and } 0 < \beta < 1.$$

Let $A_n \subseteq \Omega_n$ be an arbitrary sequence of sets such that $\mathbb{E}_{\mu_n}[T_{A_n}] = t^{(n)}_{\mathrm{H},\mu_n}(1-\alpha)$ and $\pi_n(A_n) \geqslant 1 - \alpha$. By the equivalence between (i) and (ii) in Theorem 1, we have that

$$t^{(n)}_{\mathrm{mix},\mu_n}(\varepsilon) - t^{(n)}_{\mathrm{mix},\mu_n}(1-\varepsilon) = o(t^{(n)}_{\mathrm{mix},\mu_n}), \quad \text{for all } 0 < \varepsilon \leqslant 1/4.$$

Fix some $0 < \varepsilon \leqslant 1/8$. Using Proposition 4.5 and similar reasoning as in the proof of the equivalence between (i) and (ii) in Theorem 1, we have that for all $\beta \in (0,1)$

$$(1 - o(1))t^{(n)}_{\mathrm{mix},\mu_n} \leqslant \mathrm{hit}^{(n)}_{\beta,\mu_n}(1-\varepsilon) \leqslant \mathrm{hit}^{(n)}_{\beta,\mu_n}(\varepsilon) \leqslant (1+o(1))t^{(n)}_{\mathrm{mix},\mu_n}, \tag{4.18}$$

Let $k_n(p) := \inf\{t : \mathrm{H}_{\mu_n}[T_{A_n} > t] \leqslant p\}$. Then by the definition of $\mathrm{hit}^{(n)}_{1-\alpha,\mu_n}(\varepsilon)$ and the fact that $\pi_n(A_n) \geqslant 1 - \alpha$ (first inequality), together with (4.18) we get that

$$k_n(\varepsilon) \leqslant \mathrm{hit}^{(n)}_{1-\alpha,\mu_n}(\varepsilon) \leqslant (1+o(1))t^{(n)}_{\mathrm{mix},\mu_n}. \tag{4.19}$$

Conversely, let $\ell_n = \ell_n(\varepsilon) := 3\varepsilon^{-1}(1-\alpha)^{-1}t^{(n)}_{\mathrm{rel}}\log 3 = o(t^{(n)}_{\mathrm{mix},\mu_n})$. Then by (4.16) (first inequality) and (4.18) we get that

$$k_n(1-2\varepsilon) \geqslant \mathrm{hit}^{(n)}_{1-\alpha/2,\mu_n}(1-\varepsilon) - \ell_n \geqslant (1-o(1))t^{(n)}_{\mathrm{mix},\mu_n}. \tag{4.20}$$

Whence (4.19) implies that $k_n(\varepsilon) - k_n(1-2\varepsilon) = o(t^{(n)}_{\mathrm{mix},\mu_n})$. By (4.20) we get that $t^{(n)}_{\mathrm{mix},\mu_n} = O(\mathbb{E}_{\mu_n}[T_{A_n}])$ and thus we also have that $k_n(\varepsilon) - k_n(1-2\varepsilon) = o(\mathbb{E}_{\mu_n}[T_{A_n}])$, for all $0 < \varepsilon < 1/8$. This concludes the proof of (ii)$\Longrightarrow$(iii). We now show that (iii)$\Longrightarrow$(ii).

Let $\alpha \in (0,1)$ and $A_n \subseteq \Omega_n$ be an arbitrary sequence of sets such that $\mathbb{E}_{\mu_n}[T_{A_n}] = t_{\mathrm{H},\mu_n}^{(n)}(1-\alpha)$ and $\pi_n(A_n) \geqslant 1 - \alpha$. Assume that

$$\lim_{n \to \infty} \mathrm{H}_{\mu_n}[|T_{A_n} - t_{\mathrm{mix},\mu_n}^{(n)}| < \varepsilon \mathbb{E}_{\mu_n}[T_{A_n}]] = 1, \text{ for all } \varepsilon > 0. \qquad (4.21)$$

As before, denote $k_n(p) := \inf\{t : \mathrm{H}_{\mu_n}[T_{A_n} > t] \leqslant p\}$. Then by (4.21),

$$k_n(\varepsilon) - k_n(1 - \varepsilon/2) = o(k_n(1-\varepsilon)), \text{ for all } 0 < \varepsilon < 1/4. \qquad (4.22)$$

Recall that by assumption, there exists some $0 < \delta < 1$ such that $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix},\mu_n}^{(n)}(\delta))$. Fix some $0 < \varepsilon < \delta/4$. By (4.9) we have that

$$\mathrm{hit}_{1-\alpha,\mu_n}^{(n)}(1 - \varepsilon/2) \leqslant \mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(1-\varepsilon) + o(t_{\mathrm{mix},\mu_n}^{(n)}(\delta)). \qquad (4.23)$$

As in (4.19),

$$k_n(1-\varepsilon/2) \leqslant \mathrm{hit}_{1-\alpha,\mu_n}^{(n)}(1-\varepsilon/2). \qquad (4.24)$$

By (4.10), we have that

$$(1 - o(1))t_{\mathrm{mix},\mu_n}^{(n)}(\delta) \leqslant \mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(\delta/2).$$

Hence by (4.23)-(4.24) we get that

$$k_n(1-\varepsilon/2) \leqslant \mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(1-\varepsilon) + o(\mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(\delta/2)). \qquad (4.25)$$

This, in conjunction with (4.22), yields that for all $0 < \varepsilon < \delta/4$,

$$k_n(\varepsilon) - k_n(1-\varepsilon/2) = o(\mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(\delta/2)) \qquad (4.26)$$

Conversely, let $\ell_n(\varepsilon)$ be as before. Fix some $0 < \varepsilon < \delta/4$. Then $\ell_n(\varepsilon/2) = o(t_{\mathrm{mix},\mu_n}^{(n)}(\delta))$ and by (4.11) $\ell_n(\varepsilon/2) = o(\mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(\delta/2))$. Similarly to the derivation of (4.20), by (4.16)

$$k_n(\varepsilon) \geqslant \mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(\varepsilon/2) - \ell_n(\varepsilon/2) = \mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(\varepsilon/2) - o(\mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(\delta/2)). \qquad (4.27)$$

This, in conjunction with (4.25)-(4.26), implies that

$$\mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(\varepsilon/2) - \mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(1-\varepsilon) = o(\mathrm{hit}_{1-\alpha/2,\mu_n}^{(n)}(\delta/2)), \text{ for all } 0 < \varepsilon < \delta/4. \quad \square$$

5. Aldous' Example

We now present a version of Aldous' example (see Figure 2) for a sequence of reversible Markov chains (Ω_n, P_n, π_n) which satisfies the product condition but do not exhibit cutoff and analyze it. Our version of Aldous' demonstrates the behavior described in Proposition 1.7. Namely, we show that the sequence does not exhibit cutoff although there exist $A_n \subseteq \Omega_n$ with $\pi_n(A_n) \geqslant 1/2$ and $x_n \in \Omega_n$ satisfying $t_{\mathrm{H}}(1/2) = \mathbb{E}_{x_n}[T_{A_n}]$ such that the hitting times of A_n started from x_n are concentrated under the initial starting positions $x_n \in \Omega_n$.

Example 5.1. Consider the sequence of chains (Ω_n, P_n, π_n), where $\Omega_n := A \cup B \cup C \cup \{z\}$, where $A = A_n := \{a_{2n+1}, a_{2n}, a_{2n-1}, \ldots, a_{n+1}\}$, $B = B_n := \{b_n, b_{n-1}, \ldots, b_1\}$ and $C = C_n := \{c_n, c_{n-1}, \ldots, c_1\}$. For notational convenience we write $a := a_{2n+1}$, $v := a_{n+1} = b_{n+1} = c_{n+1}$ and $b_0 = z = c_0$.
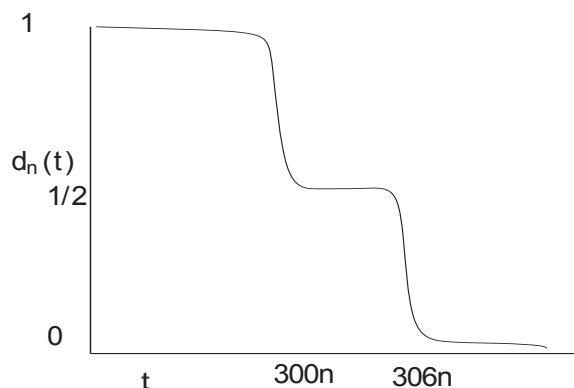Define the transition matrix P_n by

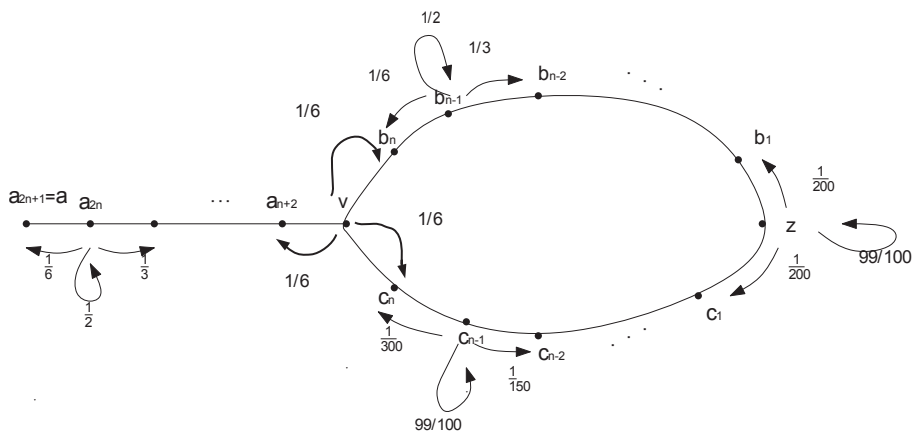FIGURE 5.1. Decay in total variation distance for Aldous' example: it does not have cutoff.

FIGURE 5.2. We consider a Markov chain with the transition probabilities specified above.

· Holding probabilities:

$$P_n(x,x) = \begin{cases} 1/2 & x \in A_n \cup B_n, \\ 99/100 & x \in C_n \cup \{z\}. \end{cases}$$

· Values at the special three states $a = a_{2n+1}, v = a_{n+1}, z = b_0 = c_0$: $P_n(a, a_{2n}) = 1/2$,

$$P_n(v, a_{n+2}) = P_n(v, b_n) = P_n(v, c_n) = \frac{1}{6},$$

$$P_n(z, b_1) = \frac{1}{200} = P_n(z, c_1) = \frac{1}{200}.$$

· $P_n(a_{n+k}, a_{n+k-1}) = P_n(b_k, b_{k-1}) = 1/3 = 2P_n(a_{n+k}, a_{n+k+1}) = 2P_n(b_k, b_{k+1})$ for all $k \in [n]$.
· $P_n(c_k, c_{k-1}) = \frac{1}{150} = 2P_n(c_k, c_{k+1})$ for all $k \in [n]$.

Think of Ω_n as a nearest neighbor walk (on an interval of length $2n + 1$), biased towards state z, which in the middle of the interval (at state v) splits into two parallel paths, B and C (which we refer to as "branches") of length n leading to z (see Figure 5.2). The difference between the two branches is that on branch C the holding probability is much larger (i.e. $P_n(c, c) = 99/100$ for all $c \in C$, while $P_n(b, b) = 1/2$ for all $b \in B$).

Conditionally on not making a lazy step, the chain moves with a fixed bias towards z. More precisely, let (Ω_n, Q_n, π'_n) be the non-lazy version of (Ω_n, P_n, π_n). That is, $Q_n(x, x) = 0$ for all x and $Q_n(x, y) = \frac{P_n(x,y)}{1 - P_n(x,x)}$ for all $x \neq y$. Let $f : \Omega_n \to \{0, 1, \ldots, 2n + 1\}$ be $f(b_i) = i = f(c_i)$ and $f(a_{n+1+i}) = n + 1 + i$ for all $0 \leqslant i \leqslant n$. Let (Y_t) be a realization of (Ω_n, Q_n, π'_n). It is easy to see that the projection $Z_t = f(Y_t)$ is a nearest neighbor biased random walk on the interval $\{0, 1, \ldots, 2n + 1\}$ (with reflecting boundary conditions) with a fixed bias of $2/3$ of making a step towards 0. In particular, T_0 under H_k (w.r.t. the chain (Z_t)) is concentrated around $\mathbb{E}_k[T_0] = 3k - O(1)$ within a time window of size $O(\sqrt{n})$ (where the equality $\mathbb{E}_k[T_0] = 3k - O(1)$ follows from (5.11) below).

It is easy to check that the chains (Ω_n, P_n, π_n) are indeed reversible. One way to see this is to note that Kolmogorov's cycle condition holds. Alternatively, the corresponding (symmetric) edge weights are $w_n(a_{n+m}, a_{n+m+1}) = 2^{-(n+m)}$, $w_n(b_m, b_{m+1}) = 2^{-m} = w_n(c_m, c_{m+1})$ and

$$w_n(x, x) = \begin{cases} \sum_{y:y\neq x} w_n(x, y) & x \in A \cup B, \\ 99 \sum_{y:y\neq x} w_n(x, y) & \text{otherwise.} \end{cases}$$

By the well-known discrete analog of Cheeger inequality (e.g. Levin et al., 2017 Theorem 13.14) we have that $t_{\text{rel}}^{(n)} = O(1)$, as the bottleneck-ratio is bounded from below (which can readily be seen from the above edge weights). In particular, the product condition holds.

For $0 < \varepsilon < 1$ let $k_n(\varepsilon) := \inf\{t : \max_{x \in \Omega_n} P_x[T_z > t] \leqslant \varepsilon\}$. As $\pi_n(z) > 1/2$, we get that for all $0 < \varepsilon < 1$ we have $\text{hit}_{1/2}^{(n)}(\varepsilon) = k_n(\varepsilon)$. To see this, observe that if A is such that $\pi(A) \geqslant 1/2$ then it must be the case that $z \in A$ and hence for all $x \in \Omega$ and $t \geqslant 0$ we have that $\max_{A \subseteq \Omega : \pi(A) \geqslant 1/2} P_x[T_A > t] \leqslant P_x[T_z > t]$. Conversely, $\pi(\{z\}) \geqslant 1/2$ and so the opposite inequality holds as well.

We define CB (a shorthand for "chosen branch") to equal B (resp. C) if the first visit to z was made by crossing the edge (b_1, z) (resp. (c_1, z)).

Note that for all $x \in A$ we have that $H_x[CB = B] = 1/2 = H_x[CB = C]$. Let $S \in \{B, C\}$. It is easy to see for every $\ell \in [n]$, conditioned on $CB = S$, the conditional distribution of T_z under $H_{a_{n+1+\ell}}[\cdot \mid CB = S]$, is concentrated around $6\ell + 6n1_{S=B} + 300n1_{S=C}$.

Using the aforementioned projection (Z_t) together with elementary results about hitting probabilities for a nearest neighbor biased walk on a segment (e.g. Levin et al., 2017 Exam. 9.9) we get that

$$H_{b_\ell}[T_v < T_z] = \frac{2^\ell - 1}{2^{n+1} - 1} = H_{c_\ell}[T_v < T_z], \text{ for all } \ell \in [n]. \tag{5.1}$$

Consequently,

$$H_{c_\ell}[CB = B] = \frac{1}{2} \cdot \frac{2^\ell - 1}{2^{n+1} - 1} = H_{b_\ell}[CB = C], \text{ for all } \ell \in [n]. \tag{5.2}$$

In particular, we get that for all $\ell > \lceil \log_2 n \rceil$ the law of T_z under $\mathrm{H}_{c_{n+1-\ell}}$ (respectively $\mathrm{H}_{b_{n+1-\ell}}$) is concentrated around $300(n - \ell)$ (resp. $6(n - \ell)$), within a time window of size $O(\sqrt{n})$.

Let $S \in \{B, C\}$. It follows from (5.10) below that $\mathbb{E}_{c_{n+1-r}}[T_v \mid \mathrm{CB} = B] \leqslant 300r + O(1)$, for all $0 \leqslant r \leqslant n$. Using Markov inequality, and the analysis of the case $x \in A$, with $x = a_{n+1} = v$, it is easy to verify that for all $0 \leqslant \ell \leqslant \lceil \log_2 n \rceil$, conditioned on $\mathrm{CB} = S$, the conditional distribution of T_z under $\mathrm{H}_{c_{n+1-\ell}}[\cdot \mid \mathrm{CB} = S]$ is concentrated around $6n 1_{S=B} + 300n 1_{S=C}$. The same holds for $b_{n+1-\ell}$.

By the analysis above,

$$k_n(1/2 - o(1)) \geqslant 306n - o(n) \text{ and } k_n(1/2 + o(1)) \leqslant 300n + o(n). \tag{5.3}$$

In particular, there is no $\mathrm{hit}_{1/2}$-cutoff. By Theorem A, the sequence does not exhibit a cutoff.

Let $x_n \in \Omega_n$ be such that $t_{\mathrm{H}}^{(n)}(1/2) = \max_{y \in \Omega_n} \mathbb{E}_y[T_z] = \mathbb{E}_{x_n}[T_z]$. We now argue that the $x_n = c_{j_n}$ for some $j_n \in [n]$ such that $\min(n - j_n, j_n) \to \infty$ (in fact, we shall show that $n - j_n = \Theta(\log n)$). Note that starting from such x_n, the hitting time of z is concentrated, although the sequence of chain does not exhibit a cutoff.

Most readers should be satisfied by the following explanation. It is clear that either $x_n \in C$ or $x_n = a$. If $\ell_n = o(n)$ and $\ell_n \to \infty$, then the distribution of T_z under $\mathrm{H}_{c_{n-\ell_n}}$ is concentrated around $300n - o(n)$ and $\mathbb{E}_{c_{n-\ell_n}}[T_z] = 300n - o(n)$. On the other hand, $\mathbb{E}_a[T_z] \leqslant 159n$. Lastly, if $\ell_n = O(1)$, then $\mathrm{H}_{c_{n+1-\ell_n}}[\mathrm{CB} = B]$ is bounded from below, and so $\limsup_{n \to \infty} \mathbb{E}_{c_{n+1-\ell_n}}[T_z]/n < 300$.

We now present a more detailed proof for the fact that $x_n = c_{j_n}$ for some j_n such that $n - j_n = \Theta(\log n)$. First write

$$\mathbb{E}_{c_r}[T_z] = \mathbb{E}_{c_r}[T_z \mid T_z < T_v]\frac{2^{n+1} - 2^r}{2^{n+1} - 1} + (\mathbb{E}_{c_r}[T_v \mid T_v < T_z] + \mathbb{E}_v[T_z])\frac{2^r - 1}{2^{n+1} - 1}. \tag{5.4}$$

We shall show that there exist absolute constants $K_1, K_2, K_3 > 0$ such that for all $r \in [n]$

$$300 - K_1 2^{-r} \leqslant \mathbb{E}_{c_{r+1}}[T_{c_r}] \leqslant 300 \text{ and } 6 - K_1 2^{-r} \leqslant \mathbb{E}_{a_{n+r+1}}[T_{a_{n+r}}], \mathbb{E}_{b_{r+1}}[T_{b_r}] \leqslant 6. \tag{5.5}$$

$$\frac{300n}{2} + \frac{6n}{2} - K_2 \leqslant \mathbb{E}_v[T_z] \leqslant \frac{300(n+1)}{2} + \frac{6(n+1)}{2} = 153(n+1). \tag{5.6}$$

$$\mathbb{E}_a[T_z] = \mathbb{E}_a[T_v] + \mathbb{E}_v[T_z] \leqslant 159(n+1). \tag{5.7}$$

$$\mathbb{E}_{c_{n+1-r}}[T_{c_{n-r}} \mid T_z < T_v] = 300 \pm K_3 2^{-r} = \mathbb{E}_{c_r}[T_{c_{r+1}} \mid T_v < T_z]. \tag{5.8}$$

$$\mathbb{E}_{c_{n+1-r}}[T_z \mid T_z < T_v] = 300(n+1-r) \pm 2K_3 2^{-r}. \tag{5.9}$$

$$\mathbb{E}_{c_{n+1-r}}[T_v \mid T_v < T_z] = 300r \pm 2K_3 2^{-(n-r)}. \tag{5.10}$$

Combining (5.6)-(5.10) with (5.4) it is easy to verify that indeed $x_n = c_{j_n}$ for some $j_n \in [n]$ such that $n - j_n = \Theta(\log n)$.

We first note that (5.6) and (5.7) follow easily from (5.5). We now prove (5.5).

It is a standard result (e.g. Aldous and Fill, 2002 Lemma 1 Chapter 5, or Basu et al., 2017 Lemma 5.2) that for a birth and death chain on $\{0, 1, \ldots, 2n + 1\}$ with symmetric edge weights $(w_{i,j})_{i,j:|i-j| \leqslant 1}$,

$$\mathbb{E}_{r+1}[T_r] = \frac{\sum_{i,j:i \geqslant r, j \geqslant r+1, |i-j| \leqslant 1} w_{i,j}}{w_{r,r+1}} - 1. \tag{5.11}$$

It follows from (5.11) that the projected chain (Z_t) satisfies that

$$\forall 0 \leqslant k \leqslant 2n, \quad \in 3 - 2^{-(2n-k)} \leqslant \mathbb{E}_{k+1}[T_k] \leqslant 3,$$

which together with (5.1) imply (5.5). As (5.9) and (5.10) follow from (5.8), we conclude the proof by verifying (5.8).

Using the Doob's transform (see e.g. Levin et al., 2017 Section 17.6) we have that the law of (X_t) up to time T_v (resp. T_z) conditioned on $T_v < T_z$ (resp. $T_z < T_v$) is a Markov chain whose transition matrix is given by $P_v(x,y) := \frac{P_n(x,y)\mathrm{H}_y[T_v<T_z]}{\mathrm{H}_x[T_v<T_z]}$ (resp. $P_z(x,y) = \frac{P_n(x,y)\mathrm{H}_y[T_z<T_v]}{\mathrm{H}_x[T_z<T_v]}$).

By (5.1) we get that for all $r \in [n]$ there exists an absolute constant $K_4 > 0$ such that $P_z(c_{n+1-r},y) = P_n(c_{n+1-r},y) \pm K_4 2^{-r}$ for all y, while $P_v(c_r,c_{r+i}) = P_n(c_r,c_{r-i}) \pm K_4 2^{-r}$ for $i \in \{0,\pm 1\}$ (i.e., up to negligible terms, P_v restricted to C is a nearest neighbor walk with an opposite bias compared to the original chain). This, in conjunction with (5.5), implies (5.8).

Acknowledgements

References

D. Aldous and J. Fill. Reversible Markov chains and random walks on graphs (2002). Unfinished manuscript.

D. J. Aldous. Some inequalities for reversible Markov chains. *J. London Math. Soc. (2)* **25** (3), 564–576 (1982). MR657512.

R. Basu, J. Hermon and Y. Peres. Characterization of cutoff for reversible Markov chains. *Ann. Probab.* **45** (3), 1448–1487 (2017). MR3650406.

G.-Y. Chen. *The cutoff phenomenon for finite Markov chains.* ProQuest LLC, Ann Arbor, MI (2006). ISBN 978-0542-71243-2. Thesis (Ph.D.)–Cornell University. MR2708835.

G.-Y. Chen and L. Saloff-Coste. Comparison of cutoffs between lazy walks and Markovian semigroups. *J. Appl. Probab.* **50** (4), 943–959 (2013). MR3161366.

S. Goel, R. Montenegro and P. Tetali. Mixing time bounds via the spectral profile. *Electron. J. Probab.* **11**, no. 1, 1–26 (2006). MR2199053.

S. Griffiths, R. J. Kang, R. I. Oliveira and V. Patel. Tight inequalities among set hitting times in Markov chains. *Proc. Amer. Math. Soc.* **142** (9), 3285–3298 (2014). MR3223383.

D. A. Levin, Y. Peres and E. L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI (2017). ISBN 978-1-4704-2962-1. MR3726904.

L. Lovász and P. Winkler. Mixing times. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 85–133. Amer. Math. Soc., Providence, RI (1998). MR1630411.

R. Montenegro and P. Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.* **1** (3), x+121 (2006). MR2341319.

R. I. Oliveira. Mixing and hitting times for finite Markov chains. *Electron. J. Probab.* **17**, no. 70, 12 (2012). MR2968677.

Y. Peres and P. Sousi. Mixing times are hitting times of large sets. *J. Theoret. Probab.* **28** (2), 488–519 (2015). MR3370663.

N. Starr. Operator limit theorems. *Trans. Amer. Math. Soc.* **121**, 90–115 (1966). MR0190757.