THE CUTOFF PHENOMENON FOR FINITE MARKOV CHAINS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Guan-Yu Chen

August 2006

THE CUTOFF PHENOMENON FOR FINITE MARKOV CHAINS

Guan-Yu Chen, Ph.D.

Cornell University 2006

A card player may ask the following question: how many shuffles are needed to mix up a deck of cards? Mathematically, this question falls in the realm of the quantitative study of the convergence of finite Markov chains. Similar convergence rate questions for finite Markov chains are important in many fields including statistical physics, computer science, biology and more. In this dissertation, we discuss a behavior —the cutoff phenomenon— that is known to appear in many models. For these models, after a waiting period, the chain abruptly converges to its stationary distribution.

Our aim is to develop a theory of this phenomenon and to illustrate this theory with interesting examples. We focus on the case when the convergence is measured at the ℓ^p-distance for $1 \leq p \leq \infty$. For $p = 1$, one recovers the classical total variation distance.

One of the main result of the thesis is that for families of reversible Markov chains and $1 < p \leq \infty$, the existence of an ℓ^p-cutoff can be characterized using two parameters: the spectral gap and the mixing time. This fails when $p = 1$.

The notion of cutoff for a family of Markov chains indexed by n involves a cutoff time sequence $(t_n)_1^\infty$ and window size sequence $(b_n)_1^\infty$. Ideally, when a cutoff exists, we would like to determine precisely t_n and b_n. When $p = 2$, spectral theory allows for a deeper analysis of the cutoff phenomenon producing in some cases the

asymptotic behavior of the sequences $(t_n)_1^\infty$ and $(b_n)_1^\infty$.

Throughout the thesis, examples are provided to illustrate the theoretical results. In particular, the last chapter is devoted to the study of the cutoff for the randomized riffle shuffle.

BIOGRAPHICAL SKETCH

Guan-Yu Chen was born in Hsinchu, Taiwan in 1975. He studied mathematics in National Chiao Tung University. After finishing his undergraduate study in 1997, he performed a two year obligatory military service. In fall 2000, Guan-Yu enrolled in a master program in National Chiao Tung University and, after one year, he transferred to a Ph.D. program. In 2003, he visited Professor Saloff-Coste and registered at Cornell University as a non-degree graduate student. In fall 2004, he became an official graduate student in Cornell University and started pursuing the Ph.D. there. Under the supervision of Prof. Saloff-Coste, he graduated from Cornell University in summer 2006. After leaving Ithaca, Guan-Yu will go back to Taiwan and work in the National Center for Theoretical Science in Hsinchu as a visiting assistant professor for two years.

To Min-Yu

ACKNOWLEDGEMENTS

I am sincerely grateful to my advisor and mentor Laurent Saloff-Coste for enthusiastically and patiently sharing his knowledge with me. I thank him for his guidance in exploring the world of mathematics and his help with composing this dissertation. I thank Richard Durrett and Leonard Gross for reading through my thesis and generously directing me to interesting problems as well as precious comments during the Ph.D. defense.

I thank my friends and colleagues in the Mathematics department for brightening my life at Cornell and making Ithaca an enjoyable place to live. I want to share my immense appreciation for the energetic support and encouragement from my family and friends in Taiwan. At last, I thank my wife, Min-Yu, for her kindness and generosity and her priceless company.

TABLE OF CONTENTS

Chapter 1

Introduction

A card player may ask the following question: how many shuffles are needed to mix up a deck of cards? Mathematically, this question falls in the realm of the quantitative study of the convergence of finite Markov chains. Similar convergence rate questions for finite Markov chains are important in many fields including statistical physics, computer science, biology and more. In statistical physics, it is common to have to estimate the average entropy of a dynamical mechanism. In biology, the question could concern the position of the last common ancestor of two related species in the history of evolution or the expected spatial structure of a protein. The common problem posed by these questions is to estimate the average of a function f defined on Ω with respect to a probability measure π on Ω. From the perspective of the Markov Chain Monte Carlo method, this is achieved by simulating a Markov process with limiting distribution π and choosing the state at a certain time T as a random sample. However, knowing the qualitative behavior of convergence is not sufficient to determine the sampling time T. A quantitative understanding of the mixing time is essential for theoretical results. In practice, various heuristics are used to choose T.

Diverse techniques have been introduced to estimate the mixing time. Coupling and strong uniform time are discussed by Aldous and Diaconis in [1, 2]. Jerrum and Sinclair use conductance to bound mixing time in [27]. Application of representation theory appears in [21] and Diaconis and Saloff-Coste used comparison techniques in [15, 16]. For lower bound, important techniques are described in [9] and in more recent work of Wilson [36].

In this dissertation, we define the mixing time and the cutoff phenomenon for a family of finite Markov chains and discuss their relationships. Based on classical results in spectral theory and real analysis, we give bounds for the mixing time and derive equivalent conditions for the cutoff phenomenon.

1.1 Preliminaries

Let $\mathcal{X}$ be a finite set. A discrete time Markov chain is a sequence of $\mathcal{X}$-valued random variables $(X_n)_0^\infty$ satisfying

$$\mathbb{P}\{X_{n+1} = x_{n+1} | X_i = x_i, \forall 0 \le i \le n\} = \mathbb{P}\{X_{n+1} = x_{n+1} | X_n = x_n\}$$

for all $x_i \in \mathcal{X}$ with $0 \le i \le n$ and $n \ge 0$. A Markov chain is *time homogeneous* if the quantity in the right hand side above is independent of n. In this case, such a Markov chains is specified by the initial distribution (the distribution of X_0) and the one-step transition kernel $K : \mathcal{X} \times \mathcal{X} \to [0, 1]$(also called the Markov kernel) which is defined by

$$\forall x, y \in \mathcal{X}, \quad K(x, y) = \mathbb{P}\{X_{n+1} = y | X_n = x\}.$$

An immediate observation on the Markov kernel K is that $\sum_{y \in \mathcal{X}} K(x, y) = 1$ for all $x \in \mathcal{X}$. Throughout this thesis, all Markov chains are assumed to be time homogeneous. For any Markov chain $(X_n)_0^\infty$ with transition matrix K and initial distribution μ, that is, $\mathbb{P}\{X_0 = x\} = \mu(x)$ for all $x \in \mathcal{X}$, the distribution of X_n is given by

$$\forall x \in \mathcal{X}, \quad \mathbb{P}\{X_n = x\} = (\mu K^n)(x) = \sum_{y \in \mathcal{X}} \mu(y) K^n(y, x),$$

where K^n is a matrix defined iteratively by

$$\forall x, y \in \mathcal{X}, \quad K^n(x, y) = \sum_{z \in \mathcal{X}} K^{n-1}(x, z) K(z, y).$$

Similarly, one can also consider a continuous-time Markov process. Here we consider only the following specific type. For any Markov kernel K, let $(X_t)_{t\geq 0}$ be a Markov process with infinitesimal generator $K-I$(the Q-matrix defined in [28]). One way to realize this process is to stay in a state for an exponential(1) time and then move to another state according to the Markov kernel K. In other words, the law of X_t is determined by the initial distribution μ and the continuous-time semigroup $H_t = e^{-t(I-K)}$(a matrix defined formally by $H_t(x,y) = e^{-t}\sum_{n=0}^{\infty}\frac{t^n K^n(x,y)}{n!}$ for $x, y \in \mathcal{X}$ and $t \geq 0$, where $K^0 = I$) through the formula

$$\forall x \in \mathcal{X}, t \geq 0 \quad \mathbb{P}\{X_t = x\} = \sum_{y\in\mathcal{X}} \mu(y)H_t(y,x).$$

Note that if $(Y_n)_0^\infty$ is a Markov chain with transition kernel K and N_t is a Poisson process with rate 1 and independent of $(Y_n)_0^\infty$, then the above Markov process $(X_t)_{t\geq 0}$ satisfies that $X_t \overset{d}{=} Y_{N_t}$(in distribution) for $t \geq 0$, since

$$\forall x, y \in \mathcal{X}, \quad H_t(x,y) = \mathrm{E}[K^{N_t}(x,y)] = \mathbb{P}\{Y_{N_t} = y | Y_0 = x\}.$$

Another view point on the continuous-time semigroup H_t is the following. For any Markov kernel K, let $\mathcal{L} = \mathcal{L}_K$ be a linear operator on $\mathbb{R}^{|\mathcal{X}|}$ defined by

$$\forall x \in \mathcal{X}, \quad \mathcal{L}f(x) = (K-I)f(x) = \sum_{y\in\mathcal{X}} K(x,y)f(y) - f(x). \tag{1.1}$$

The operator $\mathcal{L}$ can be viewed intuitively as a Laplacian operator on $\mathcal{X}$. A direct computation then shows that, for any real-valued function f on $\mathcal{X}$, the function $u(t,x) = H_t f(x)$ is a solution for the initial value problem of the discrete-version heat equation, i.e.,

$$\begin{cases} (\partial_t + \mathcal{L})u = 0 \quad u : \mathbb{R}_+ \times \mathcal{X} \to \mathbb{R} \\ u(0,x) = f(x) \quad \forall x \in \mathcal{X}. \end{cases}$$

For any Markov kernel K, a measure π on $\mathcal{X}$ is called *invariant*(with respect to K) if $\pi K = \pi$ or equivalently

$$\forall x \in \mathcal{X}, \quad \sum_{y \in \mathcal{X}} \pi(y)K(y,x) = \pi(x).$$

A measure π on $\mathcal{X}$ is called *reversible* if the following identity holds

$$\forall x, y \in \mathcal{X}, \quad \pi(x)K(x,y) = \pi(y)K(y,x).$$

In this case, K is said to be *reversible* with respect to π. From these definitions, it is obvious that a reversible measure is an invariant measure. Besides, if π is invariant(resp. reversible) with respect to K, then, for all $t \geq 0$, $\pi H_t = \pi$ or equivalently $\sum_{y \in \mathcal{X}} \pi(y)H_t(y,x) = \pi(x)$ for all $x \in \mathcal{X}$(resp. $\pi(x)H_t(x,y) = \pi(y)H_t(y,x)$ for all $x, y \in \mathcal{X}$).

Note that, for any Markov kernel K on $\mathcal{X}$, a constant vector on $\mathcal{X}$ is a right eigenvector of K associated to eigenvalue 1. This implies the existence of a real-valued function f on $\mathcal{X}$ satisfying $f = fK$, that is, $f(x) = \sum_y f(y)K(y,x)$ for all $x \in \mathcal{X}$. Then, by the following computation,

$$\sum_{x \in \mathcal{X}} |f(x)| = \sum_{x \in \mathcal{X}} \left| \sum_{y \in \mathcal{X}} f(y)K(y,x) \right| \leq \sum_{x,y \in \mathcal{X}} |f(y)|K(y,x) = \sum_{y \in \mathcal{X}} |f(y)|,$$

one finds that $|f|$ is also a left eigenvector of K with eigenvalue 1. Hence, one can always find a probability measure π, which is invariant with respect to K. In that case, π is called a *stationary* distribution for K.

A Markov kernel K is called *irreducible* if, for any $x, y \in \mathcal{X}$, there exists $n = n(x,y)$ such that $K^n(x,y) > 0$. A state $x \in \mathcal{X}$ is called *aperiodic* if $K^n(x,x) > 0$ for sufficiently large n,and K is called aperiodic if all states are aperiodic. It is known that under the assumption of irreducibility of K, there exists a unique stationary distribution π. In particular, the distribution π is positive everywhere.

In addition, if K is irreducible, then K is aperiodic if and only if some state in $\mathcal{X}$ is aperiodic.

Proposition 1.1. *Let K be an irreducible Markov kernel on a finite set $\mathcal{X}$ with the stationary distribution π. Then*

$$\forall x, y \in \mathcal{X}, \quad \lim_{t\to\infty} H_t(x,y) = \pi(y).$$

If K is irreducible and aperiodic, then

$$\forall x, y \in \mathcal{X}, \quad \lim_{n\to\infty} K^n(x,y) = \pi(y).$$

Under mild assumptions —irreducibility for continuous-time Markov processes and irreducibility and aperiodicity for discrete-time Markov chains— Proposition 1.1 shows the qualitative result that Markov chains converge to their stationarity as time tends to infinity. If such a convergence happens, the Markov kernel is called *ergodic*.

Proposition 1.2. *Let K be a Markov kernel on a finite set $\mathcal{X}$ and π is a positive probability measure on $\mathcal{X}$. If, for all $x, y \in \mathcal{X}$,*

$$\lim_{t\to\infty} H_t(x,y) = \pi(y),$$

then K is irreducible. If the following holds

$$\lim_{n\to\infty} K^n(x,y) = \pi(y), \quad \forall x, y \in \mathcal{X},$$

then K is irreducible and aperiodic.

Note that the positiveness of π in Proposition 1.2 is sufficient but not necessary for the ergodicity. A simple example is to consider a two point space $\{0,1\}$ and the Markov kernel

$$K(0,0) = 1,\ K(0,1) = 0,\ K(1,0) = 1-p,\ K(1,1) = p,$$

where $p \in (0,1)$. It is clear that K is not irreducible because $K^n(0,1) = 0$ for all $n \geq 0$. A simple computation shows that for $n \geq 1$ and $t > 0$,

$$K^n = \begin{pmatrix} 1 & 0 \\ 1 - p^n & p^n \end{pmatrix}, \quad H_t = \begin{pmatrix} 1 & 0 \\ 1 - e^{(p-1)t} & e^{(p-1)t} \end{pmatrix}.$$

Thus, the limiting distribution exists and equals to $(1,0)$ whatever the starting state is. However, by Proposition 1.1 and 1.2, if the limiting distribution is assumed positive, then, in continuous-time cases, K is ergodic if and only if K is irreducible, whereas, in discrete-time cases, ergodicity is equivalent to irreducibility and aperiodicity.

Before we can make a quantitative analysis, the function used to measure the distance should be specified. The following distances are frequently used to study this convergence.

Definition 1.1. Let μ and ν be two measures on $\mathcal{X}$. The *total variation* distance(or briefly the *variation norm*) between μ and ν is denoted and defined by

$$d_{\mathrm{TV}}(\mu,\nu) = \|\mu - \nu\|_{\mathrm{TV}} = \max_{A \subset \mathcal{X}} |\mu(A) - \nu(A)|.$$

Let π be a nowhere vanishing finite measure on $\mathcal{X}$. For $1 \leq p \leq \infty$ and any (complex-valued) function f on $\mathcal{X}$, the $\ell^p(\pi)$-*norm*(or briefly the ℓ^p-*norm*, if there is no confusion) of f is defined by

$$\|f\|_p = \|f\|_{\ell^p(\pi)} = \begin{cases} \left(\sum_{x \in \mathcal{X}} |f(x)|^p \pi(x) \right)^{1/p} & \text{if } 1 \leq p < \infty \\ \max_{x \in \mathcal{X}} |f(x)| & \text{if } p = \infty \end{cases}.$$

Definition 1.2. Let μ, ν and π be finite measures on $\mathcal{X}$ and assume that π is positive everywhere. The $\ell^p(\pi)$-*distance*(or briefly the ℓ^p-*distance*) between μ and ν is defined to be

$$d_{\pi,p}(\mu,\nu) = \|f - g\|_{\ell^p(\pi)},$$

where f and g are corresponding densities of μ and ν with respect to π, which means $\mu = f\pi$ and $\nu = g\pi$.

According to the above definitions, we have $\|\mu - \nu\|_{\mathrm{TV}} \le d_{\pi,1}(\mu,\nu)$. In particular, if $\mu(\mathcal{X}) = \nu(\mathcal{X})$, then

$$\forall \pi > 0, \quad d_{\pi,1}(\mu,\nu) = 2\|\mu - \nu\|_{\mathrm{TV}} = 2\max_{A\subset\mathcal{X}}\{\mu(A) - \nu(A)\}.$$

Remark 1.1. In general, if $(\mathcal{X},\pi)$ is a measure space and f has a finite $\ell^p(\pi)$-norm, then

$$\|f\|_{\ell^p(\pi)} = \sup_{\|g\|_{\ell^q(\pi)}\le 1} \int_{\mathcal{X}} f(x)g(x)d\pi(x), \tag{1.2}$$

where $p^{-1} + q^{-1} = 1$.

By the above remark, one can rewrite the $\ell^p(\pi)$-distance as follows.

Proposition 1.3. *Let π, μ, ν, f, g be the same as in Definition 1.2. Then, for $1 \le p \le \infty$,*

$$d_{\pi,p}(\mu,\nu) = \max_{\|h\|_q\le 1} \|(f-g)h\|_1,$$

where $p^{-1} + q^{-1} = 1$. In particular, if $p = 1$, one has

$$d_{\pi,1}(\mu,\nu) = \max_{\|h\|_\infty\le 1} \{\mu(h) - \nu(h)\}.$$

By Jensen's inequality, if π is a positive probability measure, then

$$\|f\|_p \le \|f\|_q, \quad \forall 1 \le p < q \le \infty.$$

With this fact, one can easily obtain a relation between the distances mentioned above.

Proposition 1.4. *Let π be a positive probability measure on $\mathcal{X}$. Then, for any two finite measures μ, ν on $\mathcal{X}$,*

$$d_{\pi,p}(\mu,\nu) \le d_{\pi,q}(\mu,\nu), \quad \forall 1 \le p \le q \le \infty.$$

The following fact shows that, for fixed $1 \le p \le \infty$, the ℓ^p-distance of Markov chains to their stationarity decays exponentially.

Proposition 1.5. *Let K be an irreducible Markov kernel and π be the stationary distribution of K. Then, for $1 \le p \le \infty$, the maps*

$$n \mapsto \max_{x\in\mathcal{X}} d_{\pi,p}(K^n(x,\cdot),\pi) \quad \text{and} \quad t \mapsto \max_{x\in\mathcal{X}} d_{\pi,p}(H_t(x,\cdot),\pi)$$

are non-increasing and submultiplicative. In particular, if there exists $\beta > 0$ such that

$$\max_{x\in\mathcal{X}} d_{\pi,p}(K^m(x,\cdot),\pi) \le \beta \text{ (resp. } \max_{x\in\mathcal{X}} d_{\pi,p}(H_s(x,\cdot),\pi) \le \beta),$$

then for $n \ge m$(resp. *$t \ge s$*),

$$\max_{x\in\mathcal{X}} d_{\pi,p}(K^n(x,\cdot),\pi) \le \beta^{\lfloor n/m \rfloor} \text{ (resp. } \max_{x\in\mathcal{X}} d_{\pi,p}(H_t(x,\cdot),\pi) \le \beta^{\lfloor t/s \rfloor}).$$

Remark 1.2. By Proposition 1.5, if $\beta \in (0,1)$, then the exponential convergence of ℓ^p-distance has rate at least $m^{-1}\log(1/\beta)$ in discrete-time cases and $s^{-1}\log(1/\beta)$ in continuous-time cases.

To any Markov kernel K, we associate a linear operator denoted by K and defined by $Kf(x) = \sum_{y\in\mathcal{X}} K(x,y)f(y)$ for $f \in \mathbb{R}^{|\mathcal{X}|}$. Similarly, we can view H_t, π as linear operators on $\mathbb{R}^{|\mathcal{X}|}$ by setting $H_tf(x) = \sum_{y\in\mathcal{X}} H_t(x,y)f(y)$ and interpreting $\pi(f)$ as a constant vector on $\mathcal{X}$ with value $\sum_{x\in\mathcal{X}} f(x)\pi(x)$. We let L^* be the adjoint operator of L. The following proposition follows easily from this view point.

Proposition 1.6. *Let K be an irreducible Markov operator with stationary distribution π. Then for $1 \le p \le \infty$,*

$$\max_{x\in\mathcal{X}} d_{\pi,p}(K^n(x,\cdot),\pi) = \|K^n - \pi\|_{q\to\infty} \quad \text{for } n \ge 0$$

and

$$\max_{x\in\mathcal{X}} d_{\pi,p}(H_t(x,\cdot),\pi) = \|H_t - \pi\|_{q\to\infty} \quad \text{for } t\ge 0$$

where $p^{-1}+q^{-1}=1$ *and for any linear operator* $L:\ell^r(\pi)\to\ell^s(\pi)$,

$$\|L\|_{r\to s} = \sup_{\|f\|_{\ell^r(\pi)}\le 1} \|Lf\|_{\ell^s(\pi)}. \tag{1.3}$$

Proof. Please confer to Lemma A.2. □

In this work, we will mostly consider the ℓ^p-distance. However, many other distances are considered in this literature. We end this section by introducing three other quantities which are not distances in mathematical sense. For any probability measure μ and any positive probability measure π on $\mathcal{X}$, let h be the density of μ with respect to π, that is, $\mu = h\pi$. The *separation* of μ respect to π is defined by

$$d_{\text{sep}}(\mu,\pi) = \max_{x\in\mathcal{X}}\{1-h(x)\}.$$

The *Kullback-Leibler separation* or the (relative) *entropy* of μ w.r.t. π is defined by

$$d_{\text{ent}}(\mu,\pi) = \text{Ent}_\pi(\mu) = \sum_{x\in\mathcal{X}}[h(x)\log h(x)]\pi(x).$$

(Generally, the entropy of any nonnegative function f with respect to any measure π is defined by $\text{Ent}_\pi(f) = \text{E}_\pi[f\log(f/\text{E}_\pi f)]$ where $\text{E}_\pi(f) = \sum_{x\in\mathcal{X}} f(x)\pi(x)$.) The *Hellinger* distance between μ and π is given by

$$\begin{aligned} d_{\text{H}}(\mu,\pi) &= \sum_{x\in\mathcal{X}}\left|\sqrt{h(x)}-1\right|^2\pi(x) = \sum_{x\in\mathcal{X}}\left|\sqrt{\mu(x)}-\sqrt{\pi(x)}\right|^2 \\ &= 2\left(1-\sum_{x\in\mathcal{X}}\sqrt{h(x)}\pi(x)\right). \end{aligned}$$

Let K_1 and K_2 be two Markov kernels on the same state space $\mathcal{X}$ and let $d(\cdot,\cdot)$ be any one of the above three functions. Applying Jensen's inequality, one has

$$\max_{x\in\mathcal{X}} d(K_1K_2(x,\cdot),\pi) \le \max_{x\in\mathcal{X}} d(K_1(x,\cdot),\pi) \quad \text{if } \pi K_2 = \pi,$$

This implies that the maps

$$\forall n \geq 0,\ n \mapsto \max_{x\in\mathcal{X}} d(K^n(x,\cdot),\pi), \quad \forall t \geq 0,\ t \mapsto \max_{x\in\mathcal{X}} d(H_t(x,\cdot),\pi)$$

are non-increasing.

As mentioned in Proposition 1.1 and Proposition 1.5, an ergodic Markov chain has distributions tending to its stationarity and the ℓ^p-distance decays exponentially for $1 \leq p \leq \infty$. Thus, it is interesting to ask *what is the first time that the total variation distance or the ℓ^p-distance is less than* $1/2$*, if a chain starts at some specific state.* By Proposition 1.5, one obtains an upper bound on this quantity and even gets an asymptotic rate of the exponential convergence. However, this is not enough to answer that question since the obtained quantity is sufficient for the chain to get close to the stationary distribution but usually not of the same order as the exact one. In the next section, we give a short discussion on functions used to measure the distance of a Markov chain to its stationarity. In section 1.3, we give definitions on the quantity stated in the above question and the cut-off phenomenon.

1.2 Distances

Let $\mathcal{X}$ be a finite set, $\mathcal{D}$ be the set of all probability measures on $\mathcal{X}$ and M be the set of all $|\mathcal{X}| \times |\mathcal{X}|$ stochastic matrices, where a $|\mathcal{X}| \times |\mathcal{X}|$ matrix A is stochastic if $A(x,\cdot) \in \mathcal{D}$ for all $x \in \mathcal{D}$ or equivalently

$$A(x,y) \geq 0, \forall x,y \in \mathcal{X} \quad \text{and} \quad \sum_{y\in\mathcal{X}} A(x,y) = 1, \forall x \in \mathcal{X}.$$

For any positive probability measure $\pi \in \mathcal{D}$, we consider the subset M_π of M given by

$$M_\pi = \{A \in M : \pi A = \pi\}.$$

Note that M_π contains the identity matrix and is closed under the matrix multiplication and convex combination. Our purpose now is to define a rich collection of nonnegative functions ρ_π on M_π such that, for $K \in M_\pi$, the following maps

$$\forall n \geq 0,\ n \mapsto \rho_\pi(K^n) \quad \text{and} \quad \forall t \geq 0,\ t \mapsto \rho_\pi(H_t) \tag{1.4}$$

are non-increasing.

Let $\rho_\pi^{(1)}$ and $\rho_\pi^{(2)}$ be nonnegative functions defined respectively on $\mathcal{D}$ and $\mathbb{R}_+^{|\mathcal{X}|}$. Define a function $\rho_\pi : M_\pi \to [0, \infty)$ by letting

$$\forall A \in M_\pi, \quad \rho_\pi(A) = \rho_\pi^{(2)}(v) \quad \text{where} \quad v(x) = \rho_\pi^{(1)}(A(x, \cdot)), \forall x \in \mathcal{X}. \tag{1.5}$$

Note that both $\rho_\pi^{(1)}$ and $\rho_\pi^{(2)}$ need not depend on π. Because of the convergence of ergodic chains, $\rho_\pi^{(1)}$ is usually related to π. However, there is no reason to put any restriction on $\rho_\pi^{(2)}$ a priori. The following lemma gives a simple sufficiency for the monotonicity of functions in (1.4).

Lemma 1.1. *Let ρ_π be a function on M_π defined by* (1.5)*. Assume that either*

$$\rho_\pi(AB) \leq \rho_\pi(A) \quad \forall A, B \in M_\pi, \tag{1.6}$$

or

$$\rho_\pi(AB) \leq \rho_\pi(B) \quad \forall A, B \in M_\pi. \tag{1.7}$$

Then, for $K \in M_\pi$, both maps in (1.4) *are non-increasing.*

Remark 1.3. Note that the operator norm $\|\cdot\|_{q\to\infty}$ mentioned in Proposition 1.6 is of the form in (1.5) with

$$\rho_\pi^{(1)}(v) = d_{\pi,p}(v, \pi) = \|v/\pi - 1\|_p, \quad \rho_\pi^{(2)}(v) = \|v\|_\infty,$$

where $p^{-1}+q^{-1}=1$. For other concrete examples, one may choose $\rho_\pi^{(1)}$ to be the separation , the entropy and the Hellinger distance defined in section 1.1. That is, $\rho_\pi^{(1)}(v)$ is defined respectively by

$$\max_{y\in\mathcal{X}}\{1-v(y)/\pi(y)\},\quad \sum_{y\in\mathcal{X}} v(y)\log(v(y)/\pi(y)),\quad \sum_{y\in\mathcal{X}}\left|\sqrt{v(y)/\pi(y)}-1\right|^2.$$

Besides the $\ell^\infty(\pi)$-norm, if one is particularly interested in a Markov chain with specific starting states, we may choose $\rho_\pi^{(2)}(v)$ to be $\|v\delta_S\|_\infty$ for some set $S\subset\mathcal{X}$, where $\delta_S(x)=1$ if $x\in S$ and $\delta_S(x)=0$ if $x\notin S$. Then the special case $S=\mathcal{X}$ is the ℓ^∞-norm mentioned above.

The following proposition provides some classes of functions(ρ_π) satisfying the assumption in Lemma 1.1, which include those quantities in Remark 1.3.

Proposition 1.7. *Let $\mathcal{X}$ be a finite set and $\mathcal{D}$ be the collection of all probability measures on $\mathcal{X}$. Fix a positive measure $\pi\in\mathcal{D}$ and let, for $\mu\in\mathcal{D}$, f_μ be the density of μ with respect to π. Consider a function ρ_π defined in* (1.5). *Assume that $\rho_\pi^{(2)}:\mathbb{R}_+^{|\mathcal{X}|}\to\mathbb{R}_+$ is non-decreasing in the following sense:*

$$\rho_\pi^{(2)}(u)\le\rho_\pi^{(2)}(v),\quad \text{for } u,v\in\mathbb{R}_+^{|\mathcal{X}|} \text{ satisfying } u(x)\le v(x),\forall x\in\mathcal{X}. \tag{1.8}$$

If one of the following conditions holds, then the function ρ_π is nonnegative and satisfies one of (1.6) *and* (1.7).

(1) $\rho_\pi^{(1)}(\mu)=\sum_{x\in\mathcal{X}}F(f_\mu(x))\pi(x)$ for all $\mu\in\mathcal{D}$, where F is a convex function on $\mathbb{R}_+$ with $F(1)\ge 0$.

(2) $\rho_\pi^{(1)}(\mu)=\sum_{x\in\mathcal{X}}F(|1-f_\mu(x)|)\pi(x)$ for all $\mu\in\mathcal{D}$, where F is a non-decreasing convex function on $\mathbb{R}_+$ with $F(0)\ge 0$.

(3) $\rho_\pi^{(1)}(\mu)=G(1-f_\mu)$ for all $\mu\in\mathcal{D}$, where G is a nonnegative function on $\mathbb{R}^{|\mathcal{X}|}$ satisfying $\|v\|_\infty\le G(v)$ and $G(u)\le G(v)$ if $u(x)\le v(x)$ for all $x\in\mathcal{X}$.

(4) $\rho_\pi^{(1)}(\mu) = G(1 - f_\mu)$ *for all* $\mu \in \mathcal{D}$*, where* G *is a nonnegative convex function on* $\mathbb{R}^{|\mathcal{X}|}$ *and* $\rho_\pi^{(2)}$ *is convex on* $\mathbb{R}_+^{|\mathcal{X}|}$*.*

Proof. We will prove that (1), (2) and (3) imply (1.6) and (4) implies (1.7). Note first that (2) is a special case of (1) since $F(|1-t|)$ is a convex function for $t \geq 0$ while F is a nonnegative non-decreasing convex function on $\mathbb{R}_+$.

For (1), note that

$$\frac{AB(x,y)}{\pi(y)} = \sum_{z \in \mathcal{X}} \left(\frac{A(x,z)}{\pi(z)} \right) \frac{\pi(z)B(z,y)}{\pi(y)}.$$

Since, for fixed y, $\{\pi(z)B(z,y)/\pi(y)\}_{z \in \mathcal{X}}$ is a probability measure, by Jensen's inequality, one has

$$F\left(\frac{AB(x,y)}{\pi(y)} \right) \leq \sum_{z \in \mathcal{X}} F\left(\frac{A(x,z)}{\pi(z)} \right) \frac{\pi(z)B(z,y)}{\pi(y)}.$$

This implies $\rho_\pi^{(1)}(AB(x,\cdot)) \leq \rho_\pi^{(1)}(A(x,\cdot))$ and then proves (1.6) by the monotonicity of $\rho_\pi^{(2)}$.

For (3), observe that, for $y \in \mathcal{X}$,

$$1 - \frac{AB(x,y)}{\pi(y)} = \sum_{z \in \mathcal{X}} \left(1 - \frac{A(x,z)}{\pi(z)} \right) \frac{\pi(z)B(z,y)}{\pi(y)} \leq \rho_\pi^{(1)}(A(x,\cdot)).$$

The monotonicity of G then implies $\rho_\pi^{(1)}(AB(x,\cdot)) \leq \rho_\pi^{(1)}(A(x,\cdot))$ and then, as in (1), $\rho_\pi(AB) \leq \rho_\pi(A)$.

For (4), we rewrite $1 - AB(x,y)/\pi(y)$ as follows.

$$1 - \frac{AB(x,y)}{\pi(y)} = \sum_{z \in \mathcal{X}} A(x,z) \left(1 - \frac{B(z,y)}{\pi(y)} \right)$$

The convexity of G implies

$$\rho_\pi^{(1)}(AB(x,\cdot)) \leq \sum_{z \in \mathcal{X}} A(x,z) \rho_\pi^{(1)}(B(z,\cdot)).$$

Hence, the desired identity is proved by the convexity and monotonicity of $\rho_\pi^{(2)}$. □

Remark 1.4. Note that Proposition 1.7(2) remains true if the term $|1 - f_\mu(x)|$ is replaced by $H(f_\mu(x))$, where H is any nonnegative convex function on $\mathbb{R}_+$.

Corollary 1.1. *Let $\mathcal{D}$ be the set of all probability measures on the finite set $\mathcal{X}$ and, for fixed positive probability measure π, let $\rho_\pi^{(1)}$ be any one of the following function on $\mathcal{D}$.*

$$\|f_\mu - 1\|_p, \quad \mathrm{Ent}_\pi(\mu), \quad d_{\mathrm{sep}}(\mu, \pi), \quad \|\mu - \pi\|_{\mathrm{H}}$$

where $1 \le p \le \infty$ and $\mu = f_\mu \pi$ for all $\mu \in \mathcal{D}$. Assume that $\rho_\pi^{(2)}$ is a nonnegative function on $\mathbb{R}_+^{|\mathcal{X}|}$ satisfying (1.8)*. Then the function ρ_π defined in* (1.5) *is nonnegative and, for $K \in M_\pi$, those maps in* (1.4) *are non-increasing.*

Proof. It can be easily checked that in Proposition 1.7, case (1) is satisfied by the entropy and the Hellinger distance, case (2) holds for the $\ell^p(\pi)$-norm and the separation fits case (3). □

Because the intended use of the functions ρ_π is to measure convergence to π, it is natural to request

$$\rho_\pi^{(1)}(\mu) = 0 \Leftrightarrow \mu = \pi, \quad \rho_\pi^{(2)}(0) = 0. \tag{1.9}$$

This implies $\rho_\pi(\Pi) = 0$, where $\Pi \in M_\pi$ is a matrix with rows π. To achieve such a requirement, one needs only to assume further in Proposition 1.7 that $F(1) = 0$ in (1), $F(0) = 0$ in (2), $G(0) = 0$ in (3) and (4), and $\rho_\pi^{(2)}(0) = 0$.

The following are some other interesting possibilities for $\rho_\pi^{(2)}$. For any positive measure ν on $\mathcal{X}$, define $\rho_\pi^{(2)}$ to be

$$\forall u \in \mathbb{R}_+^{|\mathcal{X}|},\ \rho_\pi^{(2)}(u) = \|u\|_{\ell^q(\nu)} \quad \text{for some } 1 \le q \le \infty.$$

When ν is a counting measure on $\mathcal{X}$, one has $\|u\|_{\ell^\infty(\nu)} = \max_{x \in \mathcal{X}}\{u(x)\}$, which is the one used in section 1.1. If g is a nonnegative function on $\mathcal{X}$, then the function

$\rho_\pi^{(2)}(u) = \|gu\|_{\ell^q(\nu)}$ still fits the requirement for Proposition 1.7(1), (2) and (3) and for Corollary 1.1. Particularly, if $g = \delta_x$ for some $x \in \mathcal{X}$ and $q = \infty$, then the maps

$$\forall n \geq 0, \quad d_{\pi,p}(K^n(x,\cdot),\pi),\ d_{\mathrm{sep}}(K^n(x,\cdot),\pi),\ \mathrm{Ent}_\pi(K^n(x,\cdot)),\ \|K^n(x,\cdot)-\pi\|_{\mathrm{H}}$$

and

$$\forall t \geq 0, \quad d_{\pi,p}(H_t(x,\cdot),\pi),\ d_{\mathrm{sep}}(H_t(x,\cdot),\pi),\ \mathrm{Ent}_\pi(H_t(x,\cdot)),\ \|H_t(x,\cdot)-\pi\|_{\mathrm{H}}$$

are non-increasing.

Specifically, consider that $\rho_\pi^{(1)}(u)$ is the $\ell^p(\pi)$-norm of $u/\pi - 1$ and $\rho_\pi^{(2)}(v)$ is the $\ell^r(\pi)$-norm of v. Then, by Lemma A.2, one has

$$\rho_\pi(A) \geq \|A-\pi\|_{q\to r},$$

where $p^{-1}+q^{-1}=1$. In particular, if $r=\infty$, then

$$\rho_\pi(A) = \|A-\pi\|_{q\to\infty}.$$

1.3 Mixing time and cutoff phenomenon

In this section, we will define the quantity reflecting the distance between the distribution of a Markov chain and its stationarity. First, recall that, for any positive probability measure π on a finite set $\mathcal{X}$, M_π is the set containing all $|\mathcal{X}|\times|\mathcal{X}|$ stochastic matrices with stationary distribution π.

Definition 1.3. Let π be a positive probability measure on the finite set $\mathcal{X}$ and ρ_π be a nonnegative function on M_π. For $\epsilon > 0$ and $K \in M_\pi$, the ρ_π*-mixing time* is defined by

$$T^c_{\rho_\pi}(K,\epsilon) := \inf\{t \geq 0 : \rho_\pi(H_t) \leq \epsilon\}$$

and

$$T^d_{\rho_\pi}(K,\epsilon) := \inf\{n \geq 0 : \rho_\pi(K^n) \leq \epsilon\},$$

where $T^c_{\rho_\pi}(K,\epsilon)$ or respectively $T^d_{\rho_\pi}(K,\epsilon)$ is infinity if the infimum is taken on an empty set. For convenience, we use $T_{\rho_\pi}(K,\epsilon)$ to denote both $T^c_{\rho_\pi}(K,\epsilon)$ and $T^d_{\rho_\pi}(K,\epsilon)$.

Remark 1.5. (1) Note that if ρ_π is a nonnegative function on M_π satisfying (1.6) or (1.7), then, for any $K \in M_\pi$, the mixing time $T_{\rho_\pi}(K,\epsilon)$ is non-increasing for $\epsilon \in (0,\infty)$.

(2) The definition of ρ_π-mixing time does not imply the finiteness of $T_{\rho_\pi}(K,\epsilon)$ for small $\epsilon > 0$, even though $K \in M_\pi$ is irreducible and $\rho_\pi(\Pi) = 0$, where $\Pi \in M_\pi$ is a matrix having rows π. Consider M_π as a subset of the metric space $\mathbb{R}^{|\mathcal{X}|^2}$ whose metric is given by the Euclidean norm. Then the continuity of ρ_π at Π is sufficient for the finiteness of T_{ρ_π}. That is, if K is ergodic in discrete-time(resp. continuous-time) cases, then for $\epsilon > 0$,

$$T^d_{\rho_\pi}(K,\epsilon) < \infty. \quad (\text{resp. } T^c_{\rho_\pi}(K,\epsilon) < \infty.)$$

The mixing time reflects the finite-time behavior of Markov chains we are interested in. For $\epsilon > 0$, if $T_{\rho_\pi}(K,\epsilon) \in (0,\infty)$, then

$$\rho_\pi(H_{t+s}) < \epsilon,\ \forall s > 0, \quad \rho_\pi(H_{t-s}) > \epsilon,\ \forall s \in (0,t),$$

and

$$\rho_\pi(K^m) \leq \epsilon, \quad \rho_\pi(K^{m-1}) > \epsilon,$$

if $t = T^c_{\rho_\pi}(K,\epsilon) > 0$ and $m = T^d_{\rho_\pi}(K,\epsilon) > 0$.

For any family of finite Markov chains $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, ...\}$, where $\mathcal{X}_n$ is the state space and K_n is the Markov kernel with stationary distribution

π_n, we denote $H_{n,t} = e^{-t(I-K_n)}$ and use $\mathcal{F}_c$ and $\mathcal{F}_d$ to distinguish the family of continuous-time Markov processes and the family of discrete-time Markov chains.

Definition 1.4. For $n \geq 1$, let π_n be a positive probability measure on a finite set $\mathcal{X}_n$ and M_n be the set of all $|\mathcal{X}_n| \times |\mathcal{X}_n|$ stochastic matrices with stationary distribution π_n. Consider a sequence of pairs $\mathcal{M} = \{(M_n, \rho_n)|n = 1, 2, ...\}$ where ρ_n is a nonnegative function on M_n satisfying one of the monotonicity conditions (1.6) or (1.7). Assume that ρ_n is continuous at Π_n, the matrix in M_n with rows π_n, and satisfies

$$\lim_{n\to\infty} \rho_n(\Pi_n) = 0, \quad \lim_{n\to\infty} \rho_n(I_n) = U \in (0, \infty],$$

where I_n is the $|\mathcal{X}_n| \times |\mathcal{X}_n|$ identity matrix. For any family $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ of finite Markov chains, we say that $\mathcal{F}_c$ presents:

(1) A $\mathcal{M}$-*pre-cutoff* if there exist $0 < a < b$ and a sequence of positive numbers $(t_n)_1^\infty$ such that

$$\liminf_{n\to\infty} \rho_n(H_{n,at_n}) > 0, \quad \lim_{n\to\infty} \rho_n(H_{n,bt_n}) = 0.$$

(2) A $\mathcal{M}$-*cutoff* if there exists a sequence of positive numbers $(t_n)_1^\infty$ such that

$$\forall \epsilon > 0, \quad \lim_{n\to\infty} \rho_n(H_{n,(1+\epsilon)t_n}) = 0$$

and

$$\forall \epsilon \in (0,1), \quad \lim_{n\to\infty} \rho_n(H_{n,(1-\epsilon)t_n}) = U.$$

(3) A (t_n, b_n) $\mathcal{M}$-*cutoff* if $t_n > 0$, $b_n > 0$ satisfy $b_n = o(t_n)$ and

$$\lim_{c\to\infty} \overline{f}(c) = 0, \quad \lim_{c\to-\infty} \underline{f}(c) = U,$$

where

$$\overline{f}(c) = \limsup_{n\to\infty} \rho_n(H_{n,t_n+cb_n}), \quad \underline{f}(c) = \liminf_{n\to\infty} \rho_n(H_{n,t_n+cb_n}). \tag{1.10}$$

In the case of Definition 1.4(2) and (3), we refer to t_n as the $\mathcal{M}$-cutoff critical time and in the case of Definition 1.4(3), we refer to b_n as the window of the $\mathcal{M}$-cutoff.

Definition 1.5. The definition of cutoffs(the $\mathcal{M}$-pre-cutoff, the $\mathcal{M}$-cutoff and the (t_n, b_n) $\mathcal{M}$-cutoff) for $\mathcal{F}_d$ are given by replacing the following terms

$$at_n,\ bt_n,\ (1-\epsilon)t_n,\ (1+\epsilon)t_n,\ t_n + cb_n$$

with

$$\lfloor at_n \rfloor,\ \lceil bt_n \rceil,\ \lfloor (1-\epsilon)t_n \rfloor,\ \lceil (1+\epsilon)t_n \rceil,\ \begin{cases} \lceil t_n + cb_n \rceil & \text{if } c > 0 \\ \lfloor t_n + cb_n \rfloor & \text{if } c < 0 \end{cases}$$

in Definition 1.4 and requiring only $b_n \geq 0$ in (3), where $b_n > 0$ is necessary in continuous-time cases.

In the following, we introduce four well-known models as examples for the above definitions. In these examples, we will discuss ℓ^p-cutoffs for $1 \leq p \leq \infty$. This means that

$$\rho_n(A) = \max_{x \in \mathcal{X}_n} \|A(x,\cdot)/\pi_n - 1\|_p = \|A - \pi_n\|_{q \to \infty}$$

with $p^{-1} + q^{-1} = 1$. In this case, $U = 2$ if $p = 1$ and $U = \infty$ otherwise. Note that the ℓ^1-cutoff is the same as the total variation cutoff whose distance is defined by

$$\rho_n(A) = \max_{x \in \mathcal{X}_n} \|A(x,\cdot) - \pi_n\|_{\mathrm{TV}}.$$

In the above setting, we will call the $\mathcal{M}$-mixing time as the ℓ^p-mixing time, for $1 \leq p \leq \infty$, and the total variation mixing time respectively. For a further discussion on the ℓ^p-cutoff, please refer to Chapter 2, 3 and 4.

Example 1.1. **(Simple random walk on a cycle.)** For $n \geq 1$, we consider a Markov chain on the n-cycle $\mathbb{Z}/n\mathbb{Z}$ whose transitions are from x to either $x-1$ or $x+1$ with the same probability. To avoid the parity problem occurring in the discrete-time case, we assume that n is odd. It has been shown by many authors with different techniques that none of the families $\mathcal{F}_c$ and $\mathcal{F}_d$ presents a total variation or a ℓ^2 pre-cutoff, but the mixing time(for both discrete-time and continuous-time cases) is of order n^2. For a proof on this fact, see [9] and [30].

Example 1.2. **(Simple random walk on the hypercube.)** This model is in fact a nearest neighbor random walk on the hypercube, which is essentially the same process as Ehrenfest model of diffusion. For $n \geq 1$, the state space $\mathcal{X}_n$ consists of n-vectors whose entries are either 0 or 1, and the transition is done by uniformly selecting a coordinate i from $\{0, 1, ..., n\}$ and then changing the value of the ith entry. If $i = 0$, then the transition does nothing.

In [9], Diaconis proved that, for $p = 1, 2$, the family $\mathcal{F}_d$ has a $(\frac{n \log n}{4}, n)$ ℓ^p-cutoff. Latter in [14], Diaconis, Graham and Morrison proved that both $\mathcal{F}_d$ and $\mathcal{F}_c$ have a $(\frac{n \log n}{4}, n)$ total variation cutoff.

Example 1.3. **(Top to random shuffle.)** This model is first studied by Aldous and Diaconis in [2]. For a deck of n cards, a top to random shuffle is made by removing the top card from the deck and inserting it uniformly back to the deck. Another interpretation of this model is to identify the state space(all deck arrangements of n cards) with the symmetric group S_n of n elements. For the transition kernel, the present permutation σ is moved to $\sigma\tau$, where τ is uniformly selected from the set $\{(1, ..., i) : i = 1, ..., n\}$. Aldous and Diaconis proved that $\mathcal{F}_d$ presents a $(n \log n, b_n)$ total variation cutoff, where $(b_n)_1^\infty$ is any sequence satisfying $b_n = o(n \log n)$ and $n = o(b_n)$.

Later in 1992, Diaconis, Fill and Pitman in [13] improved the result by studying a generalized model of top to random. In their work, the shuffling is called top m to random and cards are shuffled by removing the top m cards from the deck and then randomly inserting them back. In this setting, the top to random card shuffling is just the special case $m = 1$. In that paper, for fixed m, they give a formula on the functions $\overline{f}$ and $\underline{f}$ in Definition 1.4 which suffices to show that $\mathcal{F}_d$ has a $(\frac{n}{m}\log n, n)$ total variation cutoff.

Example 1.4. **(Standard riffle shuffle.)** The standard riffle shuffle models how a card player shuffles a deck of cards. First, a deck of n cards is cut into two piles according to a binomial $(n, 1/2)$ random variable. Then forming a deck by dropping cards one by one from the bottom of each pile with probability proportional to respective sizes. There are many equivalent ways to defined such a model. For a detailed description and discussion, please refer to Chapter 5 and references given there.

Aldous proved in [1] that $\mathcal{F}_d$ presents a total variation cutoff with critical time $\frac{3}{2}\log_2 n(\log_a b = \log b/\log a)$. In [6], Bayer and Diaconis obtained an exact formula on the distribution after k riffle shuffles. Based on this observation, they determined the functions $\overline{f}$ and $\underline{f}$ in Definition 1.4 and proved that the family $\mathcal{F}_d$ presents a $(\frac{3}{2}\log_2 n, 1)$ total variation cutoff.

Later, we shall prove in Chapter 5 that $\mathcal{F}_c$ presents a $(\frac{3}{2}\log_2 n, \sqrt{\log n})$ total variation cutoff. In section 2.3, we show that $\mathcal{F}_d$ presents a $(\frac{3}{2}\log_2 n, 1)$ ℓ^p-cutoff for $1 < p < \infty$ and has a $(2\log_2 n, 1)$ ℓ^∞-cutoff. For continuous-time cases, the family $\mathcal{F}_c$ has a $\left(\frac{p-1}{p}(n\log n - n), (\log n)^2\right)$ ℓ^p-cutoff for $1 < p \leq \infty$, where $(p-1)/p = 1$ if $p = \infty$.

Remark 1.6. Considering cutoffs for discrete-time chains, one might think that, in

Definition 1.5, it is possible to exchange $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ at any time without changing the critical time t_n and the window size b_n at all. However this can fail in some cases. Under the assumption of $t_n \to \infty$ and $\inf_{n\geq 1} b_n > 0$, all cutoffs in Definition 1.5 are preserved with the same t_n and b_n whatever $\lfloor\cdot\rfloor$ or $\lceil\cdot\rceil$ is used. This is because $\lfloor t\rfloor \leq \lceil t\rceil$ for $t \in \mathbb{R}$ and we can choose, for each $\epsilon \in (0, 1/2)$, a constant $N(\epsilon) > 0$ such that

$$\begin{cases} \lceil(1+\epsilon)t_n\rceil \leq \lfloor(1+2\epsilon)t_n\rfloor \\ \lceil(1-\epsilon)t_n\rceil \leq \lfloor(1-\epsilon/2)t_n\rfloor \end{cases} \quad \forall n \geq N(\epsilon), \epsilon \in (0,1).$$

Similarly, by the assumption $\inf_{n\geq 1} b_n > 0$, one may choose $c_1 > 0$ such that for all $n \geq 1$,

$$\lceil t_n + cb_n\rceil \leq \begin{cases} \lfloor t_n + 2cb_n\rfloor & \text{if } c > c_1 \\ \lfloor t_n + cb_n/2\rfloor & \text{if } c < -c_1 \end{cases}.$$

Remark 1.7. When the mixing time $(t_n)_1^\infty$ is bounded or the window size b_n tends to 0, one cannot exchange the floor and the ceiling arbitrarily because the cutoff happens in one or two steps and, if any, the window size tends to 0. In such cases, instead of looking at the critical time and the type of cutoff, it is natural to ask for the actual step at which a cutoff occurs and for the limiting distances at that step. See examples in Chapter 5. Whatever the cutoff is, we use Definition 1.5 throughout this work unless another one is specified.

Remark 1.8. (1) By the monotonicity of $\rho_n(H_{n,t})$ and $\rho_n(K_n^m)$(respectively in t and m), both functions $\overline{f}$ and $\underline{f}$ in Definition 1.4(3) are non-increasing on $\mathbb{R}$ and, as far as the $\mathcal{M}$-cutoff is concerned, one needs only to prove each identity in Definition 1.4(2) with sufficiently small ϵ.

(2) Clearly, (3)$\Rightarrow$(2)$\Rightarrow$(1). For instance, if $\mathcal{F}$ has (t_n, b_n) $\mathcal{M}$-cutoff, then it presents a $\mathcal{M}$-cutoff with critical time t_n.

1.4 The optimality of the window

The cutoffs given in Definition 1.4(3) specifies the asymptotic behavior of the mixing time(please refer to Proposition 1.10 and Proposition 1.11) but say nothing about the distance at time t_n+cb_n, or $\overline{f}(c)$ and $\underline{f}(c)$. In fact, one may construct an example that presents a (t_n,b_n) $\mathcal{M}$-cutoff with $\underline{f}(c_1)=\infty$ and $\overline{f}(c_2)<\infty$ for some $-\infty<c_1<c_2<\infty$. This means that asymptotically the nth Markov chain is far from its stationarity at time $t_n+c_1b_n$. To distinguish the difference of window, we define the optimality of the window size(the difference between the critical time and the mixing time) as follows.

Definition 1.6. Let $\mathcal{F}$, $\mathcal{M}$ and U be the same as in Definition 1.4. A (t_n,b_n) $\mathcal{M}$-cutoff for $\mathcal{F}$ is

(1) weakly optimal if, given any (t_n,c_n) $\mathcal{M}$-cutoff for $\mathcal{F}$, one has $b_n=O(c_n)$.

(2) optimal if, given any (s_n,c_n) $\mathcal{M}$-cutoff for $\mathcal{F}$, one has $b_n=O(c_n)$. In this case, b_n is called an optimal window size of the $\mathcal{M}$-cutoff.

(3) strongly optimal if the functions $\overline{f}$ and $\underline{f}$ given in Definition 1.4(3) satisfy $\overline{f}(-c)<U$ and $\underline{f}(c)>0$ for all $c>0$.

Remark 1.9. Note that the strong optimality implies that one may choose $0<c_1<c_2<U$ such that the sequence $\rho_{\pi_n}(H_{n,t_n})$ is bounded from above by c_2 and from below by c_1. However, if the family has only an optimal cutoff, nothing can be said about the sequence $\rho_{\pi_n}(H_{n,t_n})$, but see Corollary 1.6.

Remark 1.10. In the discrete-time cases, as the window size b_n converges to 0, it makes no sense to discuss the optimality of a cutoff and it is worthwhile to determine the following limiting values,

$$\limsup_{n\to\infty} \rho_n(K_n^{[t_n]+k}),\ \liminf_{n\to\infty} \rho_n(K_n^{[t_n]+k}) \quad \text{for } k = -1, 0, 1.$$

As the above remark says, if a discrete-time family presents a cutoff with critical time t_n and the window size converges to 0, then the cutoff phenomenon ranges over these steps, $[t_n]-1$, $[t_n]$ and $[t_n]+1$. This is sufficient to show that no strongly optimal cutoff exists since the functions $\overline{f}(c)$ and $\underline{f}(c)$ in Definition 1.5 take values on a finite set and, hence, mush equal to 0 or U for some $c \in \mathbb{R}$. The following lemma remark this fact.

Lemma 1.2. *Let $\mathcal{F}$ and $\mathcal{M}$ be as in Definition 1.5. If $\mathcal{F}_d$ presents a strongly optimal (t_n, b_n) $\mathcal{M}$-cutoff, then* $\inf_n b_n > 0$.

These definitions show that there are more than one way to discuss the optimality of a cutoff and the difference is somewhat subtle. Please refer to Corollary 1.6 for a relation between the optimality and the weak one. In the following, we give a comparison of the optimal window size when two families present $\mathcal{M}$-cutoffs with the same critical time.

Lemma 1.3. *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be families of finite Markov chains. Assume that both of them present $\mathcal{M}$-cutoffs with the same critical time. Then the following are equivalent.*

(1) $\mathcal{F}_1$ and $\mathcal{F}_2$ have the same optimal window size(in the sense of order), if any.

(2) $\mathcal{F}_1$ presents a (t_n, b_n) $\mathcal{M}$-cutoff if and only if $\mathcal{F}_2$ has a (t_n, b_n) $\mathcal{M}$-cutoff.

Proof. Immediate from the definition of the optimality for a cutoff. □

The following are examples whose optimality had been or shall be proved in the reference or the oncoming chapters.

Example 1.5. **(Simple random walks on the hypercube.)** In [14], the families $\mathcal{F}_d$ and $\mathcal{F}_c$ are proved to have an optimal $(\frac{n\log n}{4}, n)$ total variation cutoff. In Chapter 3, we will show that, for $1 \leq p \leq 2$, the family $\mathcal{F}_c$ has an optimal $(\frac{n\log n}{4}, n)$ ℓ^p-cutoff and presents an optimal $(\frac{n\log n}{2}, n)$ ℓ^∞-cutoff. In particular, if $p = 1, 2, \infty$, the ℓ^p-cutoff is strongly optimal.

Example 1.6. **(Top m to random shuffle.)** Diaconis, Fill and Pitman proved in [13] that, for fixed $m \geq 1$, the family $\mathcal{F}_d$ has a strongly optimal $(\frac{n}{m}\log n, 1)$ total variation cutoff.

Example 1.7. **(Standard riffle shuffle.)** In [6], Bayer and Diaconis shows that $\mathcal{F}_d$ has a strongly optimal $(\frac{3}{2}\log_2 n, 1)$ total variation cutoff. In Chapter 5, we will show that $\mathcal{F}_c$ presents a strongly optimal $(\frac{3}{2}\log_2 n, \sqrt{\log n})$ total variation cutoff. In section 2.3, the family $\mathcal{F}_d$ is proved to have a strongly optimal $(\frac{3}{2}\log_2 n, 1)$ ℓ^p-cutoff for $1 < p < \infty$ and has a strongly optimal $(2\log_2 n, 1)$ ℓ^∞-cutoff.

Note that the optimality in Definition 1.6 is not the only way to discuss the window size of a cutoff. Consider the following simple example(a general setting will be given in section 2.1.3). For $n \geq 1$, let $\mathcal{X}_n = (\mathbb{Z}_2)^n$, $\pi_n \equiv 2^{-n}$ and K_n be a Markov kernel defined by

$$K_n(x,y) = \begin{cases} \frac{1}{2} & \text{if } y = s(x) + (0,...,0,i) \text{ for } i \in \{0,1\} \\ 0 & \text{otherwise} \end{cases},$$

where $s(x) = (x_2, x_3, ..., x_n, x_1)$ for all $x = (x_1, ..., x_n) \in \mathcal{X}_n$. In the total variation distance, one can easily compute that for $m \geq 1$,

$$\max_{x\in\mathcal{X}_n} \|K_n^m(x,\cdot) - \pi_n\|_{\mathrm{TV}} = \begin{cases} 1 - 2^{m-n} & \text{for } 0 \leq m \leq n-1 \\ 0 & \text{for } m \geq n \end{cases}.$$

From this fact, it is clear that the family has an optimal $(n, 1)$ total variation cutoff. However, one can find that the "left" window is strongly optimal but the "right" window is not optimal at all.

For a further categorization of the optimality of window sizes, it is natural to consider each side individually. For instance, in discrete-time cases, one can reset the functions $\overline{f}$ and $\underline{f}$ in Definition 1.4 by considering two window sizes for each of the cases $c > 0$ and $c < 0$. That is, we set

$$\forall c < 0, \quad \overline{f}(c) = \limsup_{n\to\infty} \rho_n(K_n^{\lceil t_n + cb_n \rceil}), \quad \underline{f}(c) = \liminf_{n\to\infty} \rho_n(K_n^{\lfloor t_n + cb_n \rfloor}),$$

and

$$\forall c > 0, \quad \overline{f}(c) = \limsup_{n\to\infty} \rho_n(K_n^{\lceil t_n + cc_n \rceil}), \quad \underline{f}(c) = \liminf_{n\to\infty} \rho_n(K_n^{\lfloor t_n + cb_n \rfloor}).$$

Then the left window $(b_n)_1^\infty$ is called optimal if any cutoff with left window $(b'_n)_1^\infty$, one has $b_n = O(b'_n)$. Similarly, one can define the optimality for the right window. In this setting, it can be easily seen that the cutoff in the above example has an optimal left window 1 and an optimal right window 0. Though this dissertation, we treat only the simplest classification of the optimality given in Definition 1.6.

1.5 The weak cutoff

As one can see from Definition 1.4 and Definition 1.5, the cutoff can be defined in many different ways. Here, we introduce another cutoff which is first introduced by Saloff-Coste in his survey [29].

Definition 1.7. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ and $\mathcal{M}$ be as in Definition 1.4 and $H_{n,t}$ be the continuous-time semigroup associated to K_n. A family $\mathcal{F}_c$(resp. $\mathcal{F}_d$) is said to present a *weak* $\mathcal{M}$*-cutoff* if there exists a sequence of positive numbers $(t_n)_1^\infty$

such that

$$\liminf_{n\to\infty} \rho_n(H_{n,t_n}) > 0, \quad \lim_{n\to\infty} \rho_n(H_{n,(1+\epsilon)t_n}) = 0, \ \forall \epsilon > 0.$$

$$\left(\text{resp. } \liminf_{n\to\infty} \rho_n(K_n^{\lfloor t_n \rfloor}) > 0, \quad \lim_{n\to\infty} \rho_n(K_n^{\lceil (1+\epsilon)t_n \rceil}) = 0 \ \forall \epsilon > 0\right)$$

We refer to t_n as the critical time for the weak $\mathcal{M}$-cutoff.

Remark 1.11. (1) By definition, it is clear that the weak $\mathcal{M}$-cutoff is weaker than the $\mathcal{M}$-cutoff but stronger that the $\mathcal{M}$-pre-cutoff.

(2) Note that if $t_n \to \infty$, the ceiling of the term $\lceil (1+\epsilon)t_n \rceil$ in discrete-time case can be changed into the floor without changing the critical time. However, this does not necessarily hold for the term $\lfloor t_n \rfloor$.

(3) To show the weak $\mathcal{M}$-cutoff for a family, it suffices to prove the second requirement in Definition 1.7 for sufficiently small ϵ.

Note that it is easy to see from Definition 1.4 and Definition 1.7 the difference between the weak $\mathcal{M}$-cutoff and the $\mathcal{M}$-cutoff. But how distinct between the weak $\mathcal{M}$-cutoff and the $\mathcal{M}$-pre-cutoff is not so obvious. The following proposition clearly specifies the unlikeness.

Proposition 1.8. *Let $\mathcal{F}$ and $\mathcal{M}$ be the families in Definition 1.4. Assume that $\mathcal{F}_c$ has a $\mathcal{M}$-pre-cutoff at time t_n and set*

$$c_1 = \inf\{c > 0 : \underline{g}(c) = 0\}, \quad c_2 = \inf\{c > 0 : \overline{g}(c) = 0\},$$

where

$$\underline{g}(c) = \liminf_{n\to\infty} \rho_n(H_{n,ct_n}), \quad \overline{g}(c) = \limsup_{n\to\infty} \rho_n(H_{n,ct_n}), \quad \forall c > 0.$$

Then $\mathcal{F}_c$ has a weak $\mathcal{M}$-cutoff if and only if $c_1 = c_2$ and $\underline{g}$ is discontinuous at c_1. Furthermore, the critical time of a weak $\mathcal{M}$-cutoff for $\mathcal{F}_c$ can be taken to be $T^c_{\rho_n}(K_n, \epsilon)$, where $\epsilon \in (0, U_1)$ and $U_1 = \lim_{c\uparrow c_1} \underline{g}(c)$.

The following proposition says that Proposition 1.8 also holds in discrete-time cases if the mixing time tends to infinity.

Proposition 1.9. *The result in Proposition 1.8 also holds for $\mathcal{F}_d$ if one assumes $t_n \to \infty$, replaces $T^c_{\rho_n}(K_n, \epsilon)$ with $T^d_{\rho_n}(K_n, \epsilon) - 1$ and resets $\underline{g}$ and $\overline{g}$ as follows.*

$$\underline{g}(c) = \liminf_{n\to\infty} \rho_n(K_n^{\lfloor ct_n \rfloor}), \quad \overline{g}(c) = \limsup_{n\to\infty} \rho_n(K_n^{\lfloor ct_n \rfloor}), \quad \forall c > 0.$$

Proof of Proposition 1.8 and Proposition 1.9. Let a, b be constants for the $\mathcal{M}$-precutoff in Definition 1.4. Then, by assumption, $a \le c_1 \le c_2 \le b$. Assume first that $\mathcal{F}_c$ has a weak $\mathcal{M}$-cutoff at time s_n. If $c_1 < c_2$, one may choose $\delta \in (0,1)$ such that $c_1(1+\delta)^2 < c_2$ and, by the monotonicity of $\underline{g}$, we have $\underline{g}(1+\delta) = 0$. Since $\liminf_{n\to\infty} \rho_n(H_{n,s_n}) > 0$, one may select a subsequence $(n_k)_1^\infty$ such that

$$s_{n_k} \le c_1(1+\delta)t_{n_k} \quad \forall k \ge 1.$$

This implies $c_1(1+\delta)^2 t_{n_k} > (1+\delta)s_{n_k}$ for all $k \ge 1$, and then $\overline{g}(c_1(1+\delta)^2) = 0$, which contradicts the definition of c_2. Hence $c_1 = c_2$.

For the discontinuity of $\underline{g}$ at c_1, note that the fact $c_1 = c_2$ implies the existence of an integer $N(\delta)$, for each $\delta \in (0,1)$, such that

$$(1-\delta)c_1 t_n \le (1+\delta)s_n \quad \forall n \ge N(\delta).$$

The discontinuity of $\underline{g}$ at c_1 is then proved by the following.

$$\lim_{c\uparrow c_1} \underline{g}(c) = \lim_{\delta\downarrow 0} \underline{g}\left(\frac{1-\delta}{1+\delta} c_1\right) \ge \liminf_{n\to\infty} \rho_n(H_{n,s_n}) > 0.$$

We prove the inverse direction and the second part at the same time. Set, as in the assumption,

$$U_1 = \lim_{c\uparrow c_1} \underline{g}(c) > 0, \quad \epsilon \in (0, U_1), \quad s_n = T^c_{\rho_n}(K_n, \epsilon).$$

By the monotonicity of $\underline{g}$, one can choose, for each $\delta \in (0,1)$, an integer $N(\delta)$ such that

$$(1-\delta/2)c_1 t_n \leq s_n \quad \forall n \geq N(\delta).$$

Since $c_1 = c_2$ and $(1+\delta)(1-\delta/2) > 1$ for $\delta \in (0,1)$, the above inequality implies $\lim_{n\to\infty} \rho_n(H_{n,(1+\delta)s_n}) = 0$. Hence $\mathcal{F}_c$ has a weak $\mathcal{M}$-cutoff with critical time $T^c_{\rho_n}(K_n, \epsilon)$.

For discrete-time cases, since t_n tends to infinity, one may replace $\lceil\cdot\rceil$ with $\lfloor\cdot\rfloor$ in the definition of a weak $\mathcal{M}$-cutoff. The proof goes word for word as above. $\square$

According to Definition 1.7, without choosing t_n to be the mixing time, it is not easy to prove the weak $\mathcal{M}$-cutoff for a family. The following corollary provides an easier criterion to inspect a weak $\mathcal{M}$-cutoff for a family.

Corollary 1.2. *Let $\mathcal{F}$, $\mathcal{M}$ and U be the same as in Definition 1.4. The family $\mathcal{F}_c$ has a weak $\mathcal{M}$-cutoff if and only if there exists a sequence $(t_n)_1^\infty$ of positive numbers such that*

$$\lim_{n\to\infty} \rho_n(H_{n,(1+\delta)t_n}) = 0 \quad \forall \delta > 0, \tag{1.11}$$

and

$$\lim_{\delta\downarrow 0} \liminf_{n\to\infty} \rho_n(H_{n,(1-\delta)t_n}) > 0. \tag{1.12}$$

Similarly, the family $\mathcal{F}_d$ presents a weak $\mathcal{M}$-cutoff with critical time tending to infinity if and only if there exists a sequence $(t_n)_1^\infty$ tending to infinity such that

$$\lim_{n\to\infty} \rho_n(K_n^{\lceil(1+\delta)t_n\rceil}) = 0 \quad \forall \delta > 0$$

and

$$\lim_{\delta\downarrow 0} \liminf_{n\to\infty} \rho_n(K_n^{\lfloor(1-\delta)t_n\rfloor}) > 0.$$

Furthermore, one can exchange $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ at any time.

Remark 1.12. Corollary 1.2 is also a simple corollary of Proposition 1.10 and 1.11.

1.6 Relations between cutoff and mixing time

As one can see from the definition of mixing time and cutoff, the critical time and the mixing time are closely related. We will make a connection between them in this section. The following is a result on continuous-time cases.

Proposition 1.10. *Let $\mathcal{M}$, $\mathcal{F}$ and U be the same as in Definition 1.4.*

(1) $\mathcal{F}_c$ has a $\mathcal{M}$-pre-cutoff if and only if there exist a sequence $(t_n)_1^\infty$ and constants $0 < c_1 < c_2 < \infty, \epsilon_0 > 0$ such that for all $0 < \epsilon < \epsilon_0$,

$$c_1 t_n < T^c_{\rho_n}(K_n, \epsilon) \le c_2 t_n \quad \forall n \ge N(\epsilon),$$

where $N(\epsilon)$ is a positive integer depending on ϵ.

(2) $\mathcal{F}_c$ presents a weak $\mathcal{M}$-cutoff if and only if there exists $0 < U_1 \le U$ such that for all $0 < \epsilon < \eta < U_1$,

$$T^c_{\rho_n}(K_n, \epsilon) \sim T^c_{\rho_n}(K_n, \eta). \tag{1.13}$$

(3) $\mathcal{F}_c$ presents a $\mathcal{M}$-cutoff if and only if (1.13) *holds for all $0 < \epsilon < \eta < U$.*

(4) $\mathcal{F}_c$ has a (t_n, b_n) $\mathcal{M}$-cutoff if and only if $t_n > 0$, $b_n > 0$, $b_n = o(t_n)$ and, for all $0 < \epsilon < U$,

$$\left| t_n - T^c_{\rho_n}(K_n, \epsilon) \right| = O_\epsilon(b_n). \tag{1.14}$$

Proof. For (1), assume first that $\mathcal{F}_c$ presents a $\mathcal{M}$-pre-cutoff. Let a, b, t_n be constants in Definition 1.4 and set

$$L = \liminf_{n\to\infty} \rho_n(H_{n,at_n}).$$

Then for $\epsilon \in (0, L)$, we may choose $N(\epsilon) > 0$ such that

$$at_n < T^c_{\rho_n}(K_n, \epsilon) \le bt_n \quad \forall n \ge N(\epsilon). \tag{1.15}$$

For the other direction, we assume that (1.15) holds for $0 < \epsilon < L$. This implies that, for $\epsilon \in (0, L)$,

$$\limsup_{n\to\infty} \rho_n(H_{n,bt_n}) \le \epsilon, \quad \liminf_{n\to\infty} \rho_n(H_{n,at_n}) \ge \epsilon.$$

Taking $\epsilon \to 0$ in the first inequality proves the $\mathcal{M}$-pre-cutoff.

For (2), assume that $\mathcal{F}_c$ has a weak $\mathcal{M}$- cutoff at time t_n and set

$$U_1 = \liminf_{n\to\infty} \rho_n(H_{n,t_n}).$$

By definition, $U_1 > 0$ and we choose, for each $\epsilon \in (0, U_1)$ and $\delta \in (0, 1)$, an integer $N(\delta, \epsilon)$ such that

$$t_n < T^c_{\rho_n}(K_n, \epsilon) \le (1+\delta)t_n \quad \forall n \ge N(\delta, \epsilon), \tag{1.16}$$

which implies (1.13). For the inverse, we prove it by applying the equivalence of the weak $\mathcal{M}$-cutoff given by Corollary 1.2. Fix $0 < \eta < U_1$ and set $t_n = T^c_{\rho_n}(K_n, \eta)$. By assumption, we may choose, for each $\epsilon \in (0, U_1)$ and $\delta \in (0, 1)$, an integer $N(\delta, \epsilon)$ such that

$$(1-\delta)t_n < T^c_{\rho_n}(K_n, \epsilon) \le (1+\delta)t_n \quad \forall n \ge N(\delta, \epsilon). \tag{1.17}$$

This implies

$$\limsup_{n\to\infty} \rho_n(H_{n,(1+\delta)t_n}) \le \epsilon, \quad \liminf_{n\to\infty} \rho_n(H_{n,(1-\delta)t_n}) \ge \epsilon > 0.$$

Letting $\epsilon \to 0$ in the first inequality and letting $\delta \downarrow 0$ in the second identity proves the weak cutoff of $\mathcal{F}_c$.

For (3), we first assume that $\mathcal{F}_c$ presents a $\mathcal{M}$-cutoff with critical time $(t_n)_1^\infty$. Note that it suffices to prove $T^c_{\rho_n}(K_n,\epsilon)\sim t_n$ for all $\epsilon\in(0,U)$. By the monotonicity of $\rho_n(H_{n,t})$(as a function of t), we may choose, for each $\delta\in(0,1)$ and $\epsilon\in(0,U)$, an integer $N(\delta,\epsilon)$ such that

$$(1-\delta)t_n < T^c_{\rho_n}(K_n,\epsilon)\le(1+\delta)t_n \quad \forall n\ge N(\delta,\epsilon). \tag{1.18}$$

This is equivalent to $T^c_{\rho_n}(K_n,\epsilon)\sim t_n$.

For the inverse direction, choose $\eta\in(0,U)$ and let $t_n=T^c_{\rho_n}(K_n,\eta)$ for $n\ge 1$. By assumption, for $\epsilon\in(0,U)$ and $\delta\in(0,1)$, the inequality (1.18) holds for some integer $N(\delta,\epsilon)$. This implies

$$\limsup_{n\to\infty}\rho_n(H_{n,(1+\delta)t_n})\le\epsilon,\quad \liminf_{n\to\infty}\rho_n(H_{n,(1-\delta)t_n})\ge\epsilon.$$

Letting $\epsilon\to 0$ in the former and $\epsilon\to U$ in the latter derives the $\mathcal{M}$-cutoff.

For (4), assume that $\mathcal{F}_c$ presents a (t_n,b_n) $\mathcal{M}$-cutoff. By definition, for $\epsilon\in(0,U)$, there exist $C(\epsilon)>0$ and $N(\epsilon)\in\mathbb{N}$ such that

$$\sup_{n\ge N(\epsilon)}\rho_n(H_{n,t_n+C(\epsilon)b_n})<\epsilon,\quad \inf_{n\ge N(\epsilon)}\rho_n(H_{n,t_n-C(\epsilon)b_n})>\epsilon.$$

By the above fact, one can easily prove

$$t_n-C(\epsilon)b_n<T^c_{\rho_n}(K_n,\epsilon)\le t_n+C(\epsilon)b_n \quad \forall n\ge N(\epsilon), \tag{1.19}$$

which is equivalent to (1.14).

For the other direction, assume that (1.19) holds for $\epsilon\in(0,U)$. Then those functions $\overline{f}$ and $\underline{f}$ in Definition 1.4 satisfy $\overline{f}(C(\epsilon))\le\epsilon$ and $\underline{f}(-C(\epsilon))\ge\epsilon$. Since $\overline{f}$ and $\underline{f}$ are non-increasing on $\mathbb{R}^+\cup\mathbb{R}^-$, we have

$$\limsup_{c\to\infty}\overline{f}(c)\le\epsilon,\quad \liminf_{c\to-\infty}\underline{f}(c)\ge\epsilon.$$

Letting $\epsilon\to 0$ and $\epsilon\to U_1$ respectively in the above derives the desired cutoff. $\square$

One can imagine a similar proof for the discrete-time cases, but will find that the similar statements are not true when the ρ_n-mixing time sequence $(T^d_{\rho_n}(K_n,\epsilon))_1^\infty$ is bounded. However, as mentioned in section 1.3, we should treat independently the case where the critical time t_n is bounded or the window size b_n tends to 0. The following proposition deals with the case $t_n \to \infty$ whose results are the same as Proposition 1.10.

Proposition 1.11. *Let $\mathcal{M}$, $\mathcal{F}$ and U be the same as in Definition 1.4. Assume that* $t_n \to \infty$, $\inf_{n\ge 1} b_n > 0$ *and $T^d_{\rho_n}(K_n,\epsilon) \to \infty$ for some $\epsilon \in (0,U)$. Then Proposition 1.10 remains true if one replaces $\mathcal{F}_c$ and $T^c_{\rho_n}(K_n,\epsilon)$ with $\mathcal{F}_d$ and $T^d_{\rho_n}(K_n,\epsilon)$.*

Proof. According to the discussion in the paragraph after Definition 1.5, we may replace $\lceil\cdot\rceil$ with $\lfloor\cdot\rfloor$ in the definition of cutoffs for $\mathcal{F}_d$. The proof is almost stated word for word by following the proof of Proposition 1.11 and correlating the inequalities in (1.15), (1.16), (1.17), (1.18) and (1.19) with

$$\lfloor at_n \rfloor < T^d_{\rho_n}(K_n,\epsilon) \le \lfloor bt_n \rfloor, \quad \lfloor t_n \rfloor < T^d_{\rho_n}(K_n,\epsilon) \le \lfloor (1+\epsilon)t_n \rfloor,$$

$$\lfloor (1-\delta)t_n \rfloor < T^d_{\rho_n}(K_n,\epsilon) \le \lfloor (1+\delta)t_n \rfloor,$$

$$\lfloor t_n - C(\epsilon)b_n \rfloor < T^d_{\rho_n}(K_n,\epsilon) \le \lfloor t_n + C(\epsilon)b_n \rfloor,$$

through the following fact

$$\lfloor a \rfloor < c \le \lfloor b \rfloor \Leftrightarrow a < c \le b \quad \forall a,b \in \mathbb{R}, c \in \mathbb{Z}.$$

□

By Proposition 1.10 and Proposition 1.11, one can find that the established relationship between the cutoff and the mixing time implies that the critical time is asymptotically the same as the mixing time if there is a cutoff.

Corollary 1.3. *Let $\mathcal{F}$, $\mathcal{M}$ and U be the same as in Definition 1.4.*

(1) If $\mathcal{F}_c$ presents a weak $\mathcal{M}$-cutoff with critical time $(t_n)_1^\infty$, then $t_n \sim T_{\rho_n}^c(K_n,\epsilon)$ for all $\epsilon \in (0, U_1)$, where

$$U_1 = \liminf_{n\to\infty} \rho_n(H_{n,t_n}).$$

(2) If $\mathcal{F}_c$ presents a $\mathcal{M}$-cutoff with critical time $(t_n)_1^\infty$, then $t_n \sim T_{\rho_n}^c(K_n,\epsilon)$ for all $\epsilon \in (0, U)$.

The above facts also hold for discrete-time cases if t_n tends to infinity.

The following question arises. Suppose two critical time sequences $(t_n)_1^\infty$ and $(s_n)_1^\infty$ have been found for a given family $\mathcal{F}$. What can we say about these sequences? The following corollary provides some answer.

Corollary 1.4. *Let $\mathcal{F}$, $\mathcal{M}$ and U be the same as in Definition 1.4.*

(1) Assume that $\mathcal{F}_c$ presents a weak $\mathcal{M}$-cutoff with critical time t_n and s_n. Then $t_n \sim s_n$.

(2) Assume that $\mathcal{F}_c$ presents a $\mathcal{M}$-cutoff with critical time $(t_n)_1^\infty$. Then $\mathcal{F}_c$ presents a $\mathcal{M}$-cutoff with critical time $(s_n)_1^\infty$ if and only if $s_n \sim t_n$. In particular, for $\epsilon \in (0, U)$, the critical time for the $\mathcal{M}$-cutoff can be taken to be $T_{\rho_n}^c(K_n,\epsilon)$.

The above statements remain true for $\mathcal{F}_d$ if we assume further $t_n \to \infty$.

Remark 1.13. Note that the inverse direction of Corollary 1.4(1) is not necessarily true since the definition of a weak cutoff requires a critical time "not too large"(please refer to Definition 1.4(2)). In any case, one can always choose, for small enough ϵ, $T_{\rho_n}^c(K_n,\epsilon)$ as a critical time for a weak cutoff.

For cutoffs with windows, a similar question arises. Suppose a family has a (t_n, b_n) and a (s_n, d_n) cutoff. What can we say about those quantities t_n, s_n, b_n, d_n? It has been known from Corollary 1.4 that $t_n \sim s_n$. For the window sizes b_n and d_n, the following corollary gives some answer.

Corollary 1.5. *Let $\mathcal{F}$, $\mathcal{M}$ and U be the same as in Definition 1.4.*

(1) If $\mathcal{F}_c$ has both (t_n, b_n) and (s_n, d_n) $\mathcal{M}$-cutoff, then $|t_n - s_n| = O(b_n + d_n)$.

(2) Suppose that $\mathcal{F}_c$ presents a (t_n, b_n) $\mathcal{M}$-cutoff and $(s_n)_1^\infty$ and $(d_n)_1^\infty$ are sequences satisfying $b_n = O(d_n)$ and $d_n = o(s_n)$. Then $\mathcal{F}_c$ has a (s_n, d_n) $\mathcal{M}$-cutoff if and only if $|t_n - s_n| = O(d_n)$.

The above statements remain true for $\mathcal{F}_d$ if we assume further

$$\lim_{n\to\infty} t_n = \infty, \quad \inf_{n\geq 1} b_n > 0.$$

The following is a useful consequence of Proposition 1.10 and Proposition 1.11 which gives a necessary and sufficient condition for the critical time to possesses an optimal window size.

Corollary 1.6. *Let $\mathcal{M}$, $\mathcal{F}$ and U be the same as in Definition 1.4. Assume that $\mathcal{F}_c$ presents a $\mathcal{M}$-cutoff. Then the following are equivalent.*

(1) The $\mathcal{M}$-cutoff for $\mathcal{F}_c$ has an optimal window size b_n.

(2) $\mathcal{F}_c$ presents a weakly optimal (t_n, b_n) $\mathcal{M}$-cutoff, where t_n is a sequence satisfying

$$0 < \liminf_{n\to\infty} \rho_n(H_{n,t_n}) \leq \limsup_{n\to\infty} \rho_n(H_{n,t_n}) < \infty.$$

In particular, if $\mathcal{F}_c$ presents a weakly optimal $(T^c_{\rho_n}(K_n, \epsilon), b_n)$ $\mathcal{M}$-cutoff, then it is optimal.

For discrete-time cases, the above remains true if one assumes further

$$\inf_{n\ge 1} b_n > 0.$$

Proof. By Corollary 1.4, it is clear that (1)⇒(2). For (2)⇒(1), by the assumptions in (2) and Corollary 1.4, the family $\mathcal{F}_c$ has a weakly optimal (t_n, b_n) $\mathcal{M}$-cutoff, where $t_n = T^c_{\rho_n}(K_n, \epsilon)$. Assume that $\mathcal{F}_c$ also presents a (s_n, c_n) $\mathcal{M}$-cutoff. Then, by Proposition 1.10, $\mathcal{F}_c$ has a (t_n, c_n) $\mathcal{M}$-cutoff. Hence, the weak optimality implies $b_n = O(c_n)$. □

Frequently, one uses a different distance ρ'_π to bound the original one ρ_π from above or below. The following proposition says that if a family presents a cutoff in both distances with the same critical time and one of them has a strongly optimal window, then the window size of the other can not be too small.

Proposition 1.12. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of irreducible Markov chains and $\mathcal{M} = \{(M_n, \rho_n)\}_1^\infty$ and $\mathcal{M}' = \{(M_n, \rho'_n)\}_1^\infty$ be families satisfying* (1.6) *or* (1.7)*. Set*

$$U = \lim_{n\to\infty} \rho_n(I_n), \quad U' = \lim_{n\to\infty} \rho'_n(I_n),$$

where I_n is a $|\mathcal{X}_n| \times |\mathcal{X}_n|$ identity matrix.

(1) Assume that $\rho_n \le \rho'_n$ for all $n \ge 1$. If $\mathcal{F}$ presents a strongly optimal (t_n, b_n) $\mathcal{M}$-cutoff and a (s_n, c_n) $\mathcal{M}'$-cutoff with $|t_n - s_n| = O(b_n)$, then $b_n = O(c_n)$.

(2) Assume that $U = U'$ and, for $n \ge 1$, either $\rho_n \le \rho'_n$ or $\rho_n \ge \rho'_n$. If $\mathcal{F}$ presents a strongly optimal (t_n, b_n) $\mathcal{M}$-cutoff and a (s_n, c_n) $\mathcal{M}'$-cutoff with $|t_n - s_n| = O(b_n)$, then $b_n = O(c_n)$.

Proof. We prove by contradiction and first deal with the continuous-time cases. Assume that $b_n \ne O(c_n)$, that is, one may find a subsequence $(k_n)_1^\infty$ such that

$c_{k_n} = o(b_{k_n})$. Let $C > 0$ be such that $|t_n - s_n| \le Cb_n$. Then, for $c > 0$, there exists $N = N(c)$ such that

$$t_{k_n} + 2Cb_{k_n} \ge s_{k_n} + cc_{k_n}, \quad \forall n \ge N. \tag{1.20}$$

This implies that for $c > 0$,

$$0 < \limsup_{n\to\infty} \rho_{k_n}(H_{n,t_{k_n}+2Cb_{k_n}}) \le \limsup_{n\to\infty} \rho'_{k_n}(H_{n,s_{k_n}+cc_{k_n}}). \tag{1.21}$$

This contracts the assumption of the (s_n, c_n) $\mathcal{M}'$-cutoff of $\mathcal{F}_c$.

For (2), it suffices to prove the desired property by assuming further that $\rho_n \ge \rho'_n$ for all $n \ge 1$. This is because one can separate $\mathbb{N}$ into two subsequences $(k_n)_1^\infty$ and $(k'_n)_1^\infty$, where $\rho_{k_n} \le \rho'_{k_n}$ and $\rho_{k'_n} \ge \rho'_{k'_n}$ for all $n \ge 1$, and apply part (1) to conclude $b_{k'_n} = O(c_{k'_n})$. For the case $\rho_n \ge \rho'_n$ for all $n \ge 1$, let C be the constant as before. Assume that $b_n \ne O(c_n)$ and $(m_n)_1^\infty$ is a subsequence of $\mathbb{N}$ such that $c_{m_n} = o(b_{m_n})$. Then, for $c > 0$, one can choose an integer $N = N(c)$ such that

$$t_{m_n} - 2Cb_{m_n} \le s_{m_n} - cc_{m_n}, \quad \forall n \ge N, \tag{1.22}$$

which implies

$$U > \liminf_{n\to\infty} \rho_{m_n}(H_{n,t_{m_n}-2Cb_{m_n}}) \ge \liminf_{n\to\infty} \rho'_{m_n}(H_{n,s_{m_n}-cc_{m_n}}), \quad \forall c > 0. \tag{1.23}$$

This contradicts the assumption of a (s_n, c_n) $\mathcal{M}'$-cutoff.

For discrete-time cases, note that Lemma 1.2 implies that $b = \inf_n b_n > 0$. Since $|t_n - s_n| = O(b_n)$, we may choose $C > b^{-1}$, such that $|t_n - s_n| \le Cb_n$ for all $n \ge 1$. Then, by replacing (1.20) and (1.22) with

$$t_{k_n} + 3Cb_{k_n} \ge s_{k_n} + cc_{k_n} + 1, \quad \forall n \ge N$$

and

$$t_{m_n} - 3Cb_{m_n} \le s_{m_n} - cc_{m_n} - 1, \quad \forall n \ge N,$$

we get

$$0 < \limsup_{n\to\infty} \rho_{k_n}(K_n^{\lfloor t_{k_n}+3Cb_{k_n}\rfloor}) \le \limsup_{n\to\infty} \rho'_{k_n}(K_n^{\lceil s_{k_n}+cc_{k_n}\rceil})$$

and

$$U > \liminf_{n\to\infty} \rho_{m_n}(K_n^{\lceil t_{m_n}-3Cb_{m_n}\rceil}) \ge \liminf_{n\to\infty} \rho'_{m_n}(K_n^{\lfloor s_{m_n}-cc_{m_n}\rfloor}), \quad \forall c > 0.$$

Hence, $b_n = O(c_n)$. □

1.7 A short history of cutoff phenomenon

Chains presenting a cutoff show a sharp phase transition in their behavior: The distance $\|K^m - \pi\|_{\mathrm{TV}}$ holds at almost its maximum for a while, then goes down in a relatively short time to a small value and converges to 0 exponentially fast. One of the most striking observation in the quantitative study of Markov chains is that many models presenting such a phase transition. The first example presenting such a phenomenon is the random transposition model studied by Diaconis and Shahshahani in [21] using the group representation theory.

After this first example, diverse techniques were invented and classical tools were developed to bound the total variation mixing time. In [2], Aldous and Diaconis implemented a stopping time argument to derive the cutoff for the top to random shuffle. This is the first time that the phrase, "cutoff phenomenon", appears. Aldous and Diaconis introduced systemically in [3] the coupling, the strong uniform time and the method of discrete Fourier analysis. A rigorous definition of cutoff phenomenon had never been given until Diaconis' article [11]. In that paper, the definition is the same as that given in Definition 1.4(3). In [29, 30], Saloff-Coste clarified different cutoffs according the shape(the functions $\overline{f}$ and $\underline{f}$

in Definition 1.4(4)) and the window size(b_n). A detailed introduction of various techniques and a list of existing results are given in his survey [30].

In this dissertation, we focus on the ℓ^p-distance for $1 \leq p \leq \infty$. In chapter 2, various equivalent conditions for the ℓ^p-cutoff are established and comparisons between the ℓ^p and ℓ^q cutoffs are made. In chapter 3, we restrict ourselves to normal Markov chains, and, based on an observation in chapter 2, the study of the ℓ^p-cutoff for $1 < p < \infty$ reduces to that of the ℓ^2-cutoff. Under the circumstances, the ℓ^2-cutoff is determined by the spectrum and eigenvectors of the transition matrices and a method to test the ℓ^2-cutoff is introduced. It is remarkable that the method not only determines the ℓ^2-cutoff but also gives a critical time and a window. In chapter 4, we discuss the ℓ^1-cutoff and compare the discrete-time and continuous-time cases. With the developed techniques, a counterexample to Peres' conjecture in ARCC workshop is built based on Aldous' idea. In the last chapter, we illustrate the notion of cutoff by introducing a specific card shuffling modified from the riffle shuffle.

Chapter 2

The ℓ^p-cutoff phenomenon

From the examples presented in Chapter 1, one can see that most of the known results are given for the total variation mixing time and the total variation cutoff(equivalent to the ℓ^1-mixing time and the ℓ^1-cutoff). One reason to study the ℓ^2-mixing time is the application of classical techniques, e.g. the operator theory and the group representation theory. As one can see from the definition of the ℓ^∞-norm, the ℓ^∞-distance is an upper bound for all other distances.

In this chapter, we will concentrate mostly on the ℓ^p-cutoff for $1 < p \leq \infty$. In section 2.1, we have a short discussion on the comparison of the ℓ^p and ℓ^q mixing time. Based on Riesz-Thorin interpolation theorem, we establish an equivalence relation for the ℓ^p-cutoff with $1 < p < \infty$. In section 2.2, we discuss how the ℓ^p-mixing time for discrete-time chains affects the ℓ^p-mixing time for continuous-time processes. They can be very different.

2.1 The ℓ^p-mixing time and the ℓ^p-cutoff

In this section, we restrict the function ρ_n to be the following type

$$\rho_n(A) = \max_{x \in \mathcal{X}_n} d_{\pi_n,p}(A(x,\cdot),\pi_n) = \|A - \pi_n\|_{q\to\infty},$$

where $1 < p \leq \infty$, $p^{-1} + q^{-1} = 1$ and A is any $|\mathcal{X}_n| \times |\mathcal{X}_n|$ matrix. To distinguish the difference of ρ_n as p ranges over the set $[1,\infty]$, we denote, for $1 \leq p \leq \infty$,

$$\rho_{n,p}(A) = \|A - \pi_n\|_{q\to\infty}, \quad \mathcal{M}_p = \{(M_n, \rho_{n,p})\}_{n=1}^\infty,$$

where $p^{-1}+q^{-1}=1$. From the view point of continuous state spaces, it is natural to assume that $U=2$ for $p=1$ and

$$U=\lim_{n\to\infty}\rho_{n,p}(I_n)=\lim_{n\to\infty}\left((1-\pi_{n,*})^p\pi_{n,*}^{1-p}+1-\pi_{n,*}\right)^{1/p}=\infty,\quad \forall 1<p\le\infty,$$

where $\pi_{n,*}=\min_{x\in\mathcal{X}}\pi_n(x)$. Obviously, this is equivalent to $\pi_{n,*}\to 0$ and is a very weak assumption in discrete-time cases.

In the above setting, if a family $\mathcal{F}$ presents a $\mathcal{M}_p$-pre-cutoff, (weak) $\mathcal{M}_p$-cutoff or (t_n,b_n) $\mathcal{M}_p$-cutoff, we briefly say that $\mathcal{F}$ has a ℓ^p-pre-cutoff, (weak) ℓ^p-cutoff or (t_n,b_n) ℓ^p-cutoff. Similarly, we let $T_p(K_n,\epsilon)$ denote the mixing time $T_{\rho_{n,p}}(K_n,\epsilon)$.

For convenience, for $x,y\in\mathcal{X}_n$, we set

$$h_{n,t}^x(y)=\frac{H_{n,t}(x,y)}{\pi_n(y)},\quad k_{n,x}^m(y)=\frac{K_n^m(x,y)}{\pi_n(y)}$$

and

$$h_{n,t}^{*,x}(y)=\frac{H_{n,t}^*(x,y)}{\pi_n(y)}=\frac{H_{n,t}(y,x)}{\pi_n(x)},\quad k_{n,x}^{*,m}(y)=\frac{(K_n^*)^m(x,y)}{\pi_n(y)}=\frac{K_n^m(y,x)}{\pi_n(x)}.$$

The following is a simple observation from the definition of ℓ^p-norm and Jensen's inequality.

Lemma 2.1. *Let $(\mathcal{X},K,\pi)$ be an irreducible Markov chain. Then for $1\le p\le\infty$, the following mappings*

$$m\mapsto\max_{x\in\mathcal{X}}\|k_x^m-1\|_p,\quad t\mapsto\max_{x\in\mathcal{X}}\|h_t^x-1\|_p$$

are non-increasing and submultiplicative.

In particular, if $T_p(K,\epsilon)>0$ for some $\epsilon\in(0,1)$, then

$$T_p(K,\epsilon)\le T_p(K,\delta)\le\left\lceil\frac{\log\delta}{\log\epsilon}\right\rceil T_p(K,\epsilon)\quad\forall\delta\in(0,\epsilon).$$

In many cases, the underlying Markov chain $(\mathcal{X}, K, \pi)$ are assumed to be reversible. In addition to the diagonalizability of K with orthogonal matrix, one may prove by applying Lemma A.1 that for $1 \le p \le \infty$ and $t > 0, m > 0$,

$$\max_{x\in\mathcal{X}} \|k_x^m - 1\|_p = \max_{x\in\mathcal{X}} \|k_x^{*,m} - 1\|_p, \quad \max_{x\in\mathcal{X}} \|h_t^x - 1\|_p = \max_{x\in\mathcal{X}} \|h_t^x - 1\|_p. \tag{2.1}$$

This means that if K is reversible, there is no difference in the ℓ^p-distance whatever K or $K*$ is studied. However, there are still many other cases whose Markov kernels are not reversible. The following lemma gives another class of stochastic matrices satisfying (2.1).

Lemma 2.2. *Let $\mathcal{X}$ be a finite set and K be an irreducible Markov kernel on $\mathcal{X}$ with stationary distribution π. Assume that there exists a finite group G acting transitively on $\mathcal{X}$ such that*

$$K(gx, gy) = K(x, y) \quad \forall x, y \in \mathcal{X}, g \in G.$$

Then $\pi \equiv 1/|\mathcal{X}|$ and, for $1 \le p \le \infty$ and $m, t \ge 0$, the following quantities

$$\|k_x^m - 1\|_p \quad \text{and} \quad \|h_t^x - 1\|_p$$

are independent of x and satisfy

$$\|k_x^m - 1\|_p = \|(k^*)_x^m - 1\|_p \quad \text{and} \quad \|h_t^x - 1\|_p = \|(h^*)_t^x - 1\|_p.$$

Proof. For the first part, fix $g \in G$ and let μ be a probability measure on $\mathcal{X}$ defined by $\mu(x) = \pi(g^{-1}x)$ for all $x \in \mathcal{X}$. A simple computation then shows, for $y \in \mathcal{X}$,

$$\sum_{x\in\mathcal{X}} \mu(x)K(x, y) = \sum_{x\in\mathcal{X}} \pi(g^{-1}gx)K(gx, y) = \sum_{x\in\mathcal{X}} \pi(x)K(x, g^{-1}y) = \mu(y).$$

This implies that μ is also a stationary distribution for K. By the uniqueness of π, one has $\pi(gx) = \pi(x)$. Since G acts transitively on $\mathcal{X}$, π has to be uniform on $\mathcal{X}$.

For the second part, choose $a \in \mathcal{X}$ and set G_a be the stabilizer of a. Since G acts transitively on $\mathcal{X}$, we have $|G| = |G_a| \times |\mathcal{X}|$. Let $|\mathcal{X}| = n$ and $g_1 G_a, ..., g_n G_a$ be all left cosets of G_a in G, where $g_1, ..., g_n \in G$ are representatives. Then one has $G = g_1 G_a \cup \cdots \cup g_n G_a$ and

$$g_i a = g_j a \quad \Leftrightarrow \quad i = j.$$

By this fact, the ℓ^p-distances for K^m and $(K^*)^m$ are given by

$$\begin{aligned} \|k_x^m - 1\|_p^p &= n^{p-1} \sum_{y \in \mathcal{X}} |K^m(x,y) - n^{-1}|^p \\ &= \frac{n^{p-1}}{|G_a|} \sum_{g \in G} |K^m(x, ga) - n^{-1}|^p \\ &= \frac{n^{p-1}}{|G_a|} \sum_{g \in G} |K^m(a, ga) - n^{-1}|^p \end{aligned}$$

and then

$$\begin{aligned} \|(k^*)_x^m - 1\|_p^p &= \frac{n^{p-1}}{|G_a|} \sum_{g \in G} |(K^*)^m(a, ga) - n^{-1}|^p \\ &= \frac{n^{p-1}}{|G_a|} \sum_{g \in G} |K^m(ga, a) - n^{-1}|^p \\ &= \frac{n^{p-1}}{|G_a|} \sum_{g \in G} |K^m(a, ga) - n^{-1}|^p. \end{aligned}$$

For the continuous-time cases, since $H(gx, gy) = H(x, y)$ for all $x, y \in \mathcal{X}$ and $g \in G$, one can prove this lemma by the same method as above. □

Remark 2.1. It can be easily checked that the requirements in Lemma 2.2 are satisfied if $\mathcal{X}$ is a group and $K(x, y) = P(x^{-1}y)$ for all $x, y \in \mathcal{X}$, where P is probability measure on $\mathcal{X}$.

2.1.1 Comparison of ℓ^p and ℓ^q mixing time

In this subsection, we will establish relations between the ℓ^p- and ℓ^q-mixing time for $1 < p, q \leq \infty$. The following lemma says that the ℓ^p and ℓ^q distances are not too different if the adjoint operator is considered.

Lemma 2.3. *Let K be a finite irreducible Markov kernel with stationary distribution π. Assume that $1 \leq q, r, s \leq \infty$ satisfy $1 + q^{-1} = r^{-1} + s^{-1}$. Then, for all positive numbers ϵ, η, δ,*

$$T_q(K, \epsilon^{s/q}\eta^{1-s/q}\delta) \leq \max\{\mathbf{1}_{[1,\infty)}(q)T_s(K,\epsilon), \mathbf{1}_{(1,\infty]}(q)T_s(K^*,\eta)\} + T_r(K,\delta).$$

Proof. By Lemma A.2 and Lemma A.3, one has

$$\max_{x\in\mathcal{X}} \|h^x_{u+v} - 1\|_q \leq \max_{x\in\mathcal{X}} \|h^x_u - 1\|_s^{s/q} \max_{x\in\mathcal{X}} \|h^{*,x}_u - 1\|_s^{1-s/q} \max_{x\in\mathcal{X}} \|h^x_v - 1\|_r$$

In the case $1 < q < \infty$, replacing u, v with $\max\{T^c_s(K,\epsilon), T^c_s(K^*,\eta)\}$ and $T^c_r(K,\delta)$ implies the desired identity. For the case $q = 1$ and $q = \infty$, one can find that the second term of the right hand side in the above inequality has the power 0 if $q = 1$, and so does the first term if $q = \infty$.

For discrete-time Markov chains, one can prove the lemma in the same way as above. □

The following propositions are useful facts in comparing different mixing times.

Proposition 2.1. *Let K be an irreducible Markov kernel on the finite set $\mathcal{X}$ with stationary distribution π. Then one has, for all $\epsilon > 0$,*

$$T_p(K,\epsilon) \leq T_q(K,\epsilon) \quad \text{if } 1 \leq p < q \leq \infty,$$

and

$$T_\infty(K,\epsilon^2) \leq T_p(K,\epsilon) + T_{p'}(K^*,\epsilon), \tag{2.2}$$

for any $1 \leq p \leq \infty$*, where* $p^{-1} + (p')^{-1} = 1$*. In particular,* $T_\infty(K, \epsilon^2) \leq T_2(K, \epsilon) + T_2(K^*, \epsilon)$.

If K *is reversible, then for* $\epsilon > 0$,

$$T_\infty^c(K, \epsilon^2) = 2T_2^c(K, \epsilon)$$

and

$$2T_2^d(K, \epsilon) - 1 \leq T_\infty^d(K, \epsilon^2) \leq 2T_2^d(K, \epsilon).$$

Proof. The first inequality is implied by Proposition 1.4 and the second one is implied by Lemma 2.3 with $q = \infty$, $s = p$, $r = p'$ and $\eta = \delta = \epsilon$.

In the case of reversible Markov chains, one has that, for $t \geq 0$,

$$h_{2t}(y, y) - 1 = \sum_{z \in \mathcal{X}} (h_t(y, z) - 1)(h_t(y) - 1)\pi(z) = \|h_t^y - 1\|_2^2,$$

and

$$|h_{2t}(x, y) - 1| = \left| \sum_{z \in \mathcal{X}} (h_t(x, z) - 1)(h_t(y) - 1)\pi(z) \right| \leq \|h_t^x - 1\|_2 \|h_t^y - 1\|_2,$$

where the last inequality is obtained by applying Cauchy-Schwartz inequality. This implies $\|H_{2t} - \pi\|_{1 \to \infty} = \|H_t - \pi\|_{2 \to \infty}^2$ and then $2T_2^c(K, \epsilon) = T_\infty^c(K, \epsilon^2)$. For discrete-time cases, the inequality $T_\infty^d(K, \epsilon^2) \leq 2T_2^d(K, \epsilon)$ can be derived from (2.2) with $p = p' = 2$. For the other part, note that the same computation as above implies

$$\max_{x \in \mathcal{X}} \|k_x^{2m} - 1\|_\infty = \max_{x \in \mathcal{X}} \|k_x^m - 1\|_2^2.$$

If $T_\infty^d(K, \epsilon^2)$ is even, then $T_2^d(K, \epsilon) \leq \frac{1}{2} T_\infty^d(K, \epsilon^2)$; if $T_\infty^d(K, \epsilon^2)$ is odd, then $T_2^d(K, \epsilon) \leq \frac{1}{2}[T_\infty^d(K, \epsilon^2) + 1]$. This proves the last inequality. □

Remark 2.2. Recall Example 2.3. By Proposition 2.3, one has

$$T_p^c(K_n, \epsilon) \sim t_{n,p} \quad \forall 1 < p \leq \infty,$$

where $t_{n,p} = \frac{(1-p^{-1})n \log a_n}{1-a_n^{p^{-1}-1}}$. Letting $a_n \to \infty$, we get

$$T_p^c(K_n, \epsilon) + T_{p'}^c(K_n, \epsilon) \sim t_{n,p} + t_{n,p'} \sim t_{n,\infty} \sim T_\infty^c(K_n, \epsilon^2).$$

This implies that (2.2) is sharp in continuous-time cases in the sense that, for any $1 < p < \infty$, one can't find a constant $0 < C < 1$ universal for any Markov kernel such that

$$T_\infty(K_n, \epsilon^2) \leq C[T_p^c(K_n, \epsilon) + T_{p'}^c(K_n, \epsilon)],$$

where $p^{-1} + (p')^{-1} = 1$.

Remark 2.3. (1) Note that, if the Markov kernel is reversible(that is, the linear operator K is self adjoint), then by Lemma 2.1,

$$T_p(K, \epsilon) = T_p(K^*, \epsilon) \quad \forall 1 \leq p \leq \infty. \tag{2.3}$$

(2) Assume that $\mathcal{X}$ is equipped with a group structure and P is a probability measure on $\mathcal{X}$. If the Markov kernel K is given by $K(x, y) = P(x^{-1}y)$ for $x, y \in \mathcal{X}$, then, by Lemma 2.2, the identity in (2.3) holds.

The following proposition is a complementary of Proposition 2.1, which allows one to bound the ℓ^q-mixing time from above with the ℓ^p-mixing time, where $1 < p < q < \infty$.

Proposition 2.2. *Let K be an irreducible Markov kernel on the finite set $\mathcal{X}$. Then one has, for $1 < p < q < \infty$ and $\epsilon > 0$,*

$$T_q(K, \epsilon) \leq m_{p,q} \max\{T_p(K, \epsilon^{1/m_{p,q}}), T_p(K^*, \epsilon^{1/m_{p,q}})\},$$

where $m_{p,q} = \left\lceil \frac{p(q-1)}{q(p-1)} \right\rceil$, and

$$T_\infty(K, \epsilon) \leq (1 + m_{p,p'}) \max\{T_p(K, \epsilon_0), T_p(K^*, \epsilon_0)\} \quad \textit{for } 1 < p < 2,$$

where $p^{-1} + (p')^{-1} = 1$ and $\epsilon_0 = \min\{\epsilon^{1/2}, \epsilon^{1/2m_{p,p'}}\}$.

Proof. By Lemma 2.3, one has, for $1 \le q, r, s \le \infty$ with $1 + q^{-1} = r^{-1} + s^{-1}$ and $1 < q < \infty$,

$$T_q(K, \epsilon\delta) \le \max\{T_s(K, \epsilon), T_s(K^*, \epsilon)\} + T_r(K, \delta). \tag{2.4}$$

Let $(p_n)_0^\infty$ be a sequence of positive numbers satisfying

$$p_0 = q, \quad 1 + \frac{1}{p_i} = \frac{1}{p_{i+1}} + \frac{1}{p} \quad \forall i \ge 0.$$

Note that $p_j^{-1} = (1 - p^{-1})j + q^{-1}$ for all $j \ge 0$. This implies $p_{j-1} \le p$ if and only if $j \ge \frac{p(q-1)}{q(p-1)}$. Let $m_{p,q}$ be the quantity given in the assumption, then by iterating the inequality (2.4) for $m_{p,q}$ times, we get the desired inequality.

For $q = \infty$, the inequality is a combination of the above result and Proposition 2.1. □

Note that in Proposition 2.1 and Proposition 2.2, to relate different ℓ^p-mixing time, one always needs to consider the adjoint operator. The following corollary restricts Markov kernels to some specific type, which allows one to use inequalities in Proposition 2.1 and Proposition 2.2 without using the adjoint of Markov kernels.

Corollary 2.1. *Let K be an irreducible Markov kernel on a finite set $\mathcal{X}$ with the stationary distribution π. Assume that either K is reversible or there exists a finite group G acting transitively on $\mathcal{X}$ such that*

$$K(gx, gy) = K(x, y), \quad \forall x, y \in \mathcal{X}, g \in G. \tag{2.5}$$

Then

$$T_\infty(K, \epsilon) \le (1 + m_{p,p'})T_p(K, \epsilon_0) \quad \text{for } 1 < p \le 2,$$

where $p^{-1} + (p')^{-1} = 1$, $\epsilon_0 = \min\{\epsilon^{1/2}, \epsilon^{1/2m_{p,p'}}\}$ and $m_{p,q} = \left\lceil \frac{p(q-1)}{q(p-1)} \right\rceil$.

Proof. Using Lemma 2.2 and Proposition 2.2. □

Remark 2.4. Note that Corollary 2.1 needs only the assumption of $T_p^c(K,\epsilon) = T_p^c(K^*,\epsilon)$ and $T_p^d(K,\epsilon) = T_p^d(K^*,\epsilon)$ for all $1 < p < \infty$ and $\epsilon > 0$, while the reversibility of K and the existence of a transitive group action satisfying (2.5) are sufficient for that.

2.1.2 The ℓ^p-cutoff for general Markov chains

In this section, we will establish some equivalence for the ℓ^p-cutoff defined in Definition 1.4(4).

Theorem 2.1. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of irreducible Markov chains and λ_n be the spectral gap of K_n(the smallest nonzero eigenvalue of $I-\frac{1}{2}(K_n+K_n^*)$). Assume that*

$$\lim_{n\to\infty} \pi_{n,*} = 0, \tag{2.6}$$

where $\pi_{n,*} = \min_{x\in\mathcal{X}_n} \pi_n(x)$. Then, for fixed $1 < p < \infty$, the following are equivalent.

(1) For all $\epsilon > 0$, $\mathcal{F}_c$ presents a $(T_p^c(K_n,\epsilon), \lambda_n^{-1})$ ℓ^p-cutoff.

(2) For some $\epsilon > 0$, $\lambda_n^{-1} = o(T_p^c(K_n,\epsilon))$.

Proof. (1)⇒(2) is obvious from Definition 1.4. For (2)⇒(1), we prove it by modifying the proof of Theorem 2.4.7 in [29].

Denote $t_n = T_p^c(K_n,\epsilon)$. Observe that the second identity of (2.6) implies that $t_n > 0$ for all but finitely many n. Since the distance $\rho_{n,p}(H_{n,t})$ is continuous in t, one has $\rho_{n,p}(H_{n,t_n}) = \epsilon > 0$ for n large enough. Recall that: For $s \geq 0$,

$$\left\|H_{n,s}^* - \pi_n\right\|_{2\to 2} \leq e^{-s\lambda_n}, \ \left\|H_{n,s}^* - \pi_n\right\|_{1\to 1} \leq 2, \ \left\|H_{n,s}^* - \pi_n\right\|_{\infty\to\infty} \leq 2. \tag{2.7}$$

For $1 < p < \infty$, set $\theta_p = \left|1 - \frac{2}{p}\right|$. A simple computation shows that $1 - \theta_p > 0$ and

$$\frac{1}{p} = \begin{cases} \frac{1-\theta_p}{2} + \frac{\theta_p}{\infty} & \text{for } 2 < p < \infty \\ \frac{\theta_p}{1} + \frac{1-\theta_p}{2} & \text{for } 1 < p \leq 2 \end{cases}$$

By Theorem A.1, we have

$$\left\|H_{n,s}^* - \pi_n\right\|_{p\to p} \leq 2^{\theta_p} e^{-\lambda_n s(1-\theta_p)}. \tag{2.8}$$

The above fact then implies

$$\begin{aligned} \left\|h_{n,t_n+s}^x - 1\right\|_p &= \left\|(H_{n,s}^* - \pi_n)(h_{n,t_n}^x - 1)\right\|_p \\ &\leq \left\|H_{n,s}^* - \pi_n\right\|_{p\to p} \left\|h_{n,t_n}^x - 1\right\|_p \leq \epsilon 2^{\theta_p} e^{-\lambda_n s(1-\theta_p)}. \end{aligned}$$

Similarly, one has

$$\epsilon = \rho_{n,p}(H_{n,t_n}) \leq \left\|H_{n,s}^* - \pi_n\right\|_{p\to p} \rho_{n,p}(H_{n,t_n-s}),$$

which implies

$$\rho_{n,p}(H_{n,t_n-s}) \geq \epsilon 2^{-\theta_p} e^{\lambda_n s(1-\theta_p)}.$$

By replacing s with $c\lambda_n^{-1}$, the functions $\overline{f}$ and $\underline{f}$ in Definition 1.4 are bounded as follows.

$$\forall c > 0, \quad \overline{f}(c) \leq \epsilon 2^{\theta_p} e^{-c(1-\theta_p)}, \quad \underline{f}(-c) \geq \epsilon 2^{-\theta_p} e^{c(1-\theta_p)}. \tag{2.9}$$

This proves the desired cutoff. □

Remark 2.5. The second inequality in (2.9) says that if the quantity $U < \infty$(or equivalently $\pi_{n,*}$ is bounded) in Definition 1.4, then, for $\epsilon > 0$, the mixing time sequence $(T_p^c(K_n, \epsilon))_1^\infty$ is bounded from above by $c(p,\epsilon)\lambda_n^{-1}$ for some $c(p,\epsilon) > 0$.

Example 2.1. **(The ℓ^p-cutoff for the top to random shuffle in continuous-time cases with $1 < p < \infty$).** Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be the family of top to random shuffles, where n denotes the number of cards in a deck. Lemma 2.7(proved later in section 2.2) shows that the ℓ^p-mixing time for the family $\mathcal{F}_c$ can be bounded from below by

$$T_p^c(K_n, \epsilon) \geq \frac{p-1}{2p} n \log n, \quad \text{for } n \text{ large enough.}$$

To apply Theorem 2.1, we need to get a bound on the spectral gap. Let λ_n be the spectral gap of the top to random shuffle for a deck on n cards. By the comparison technique introduced in [16], comparing the top to random shuffle with the random transposition(a card shuffling made by randomly choosing one card respectively and independently from each hand and then exchanging them) implies that there exists a constant c such that

$$\lambda_n \geq \frac{c}{n}, \quad \forall n \geq 2.$$

For a proof of the above inequality, one can use the comparison by rewriting a transposition $(i, j) \in S_n$ as

$$(i, j) = (j, ..., 1)(1, ..., i + 1)(i, ..., 1)(1, ..., j), \quad \forall 1 \leq i < j \leq n,$$

and apply the fact that the spectral gap of the random transposition obtained in [21] is equal to $2/n$.

Combining all the above, we get $\lambda_n^{-1} = o(T_p^c(K_n, \epsilon))$. Then, by Theorem 2.1, the family $\mathcal{F}_c$ presents a $(T_p^c(K_n, \epsilon), \lambda_n^{-1})$ ℓ^p-cutoff. It is an open problem to find what the critical time is in, say, the ℓ^2-cutoff.

One can observe that the proof of Theorem 2.1 is based on the inequality (2.8), which is provided by Riesz-Thorin interpolation and (2.7). Hence, for discrete-time cases, one has to find the rate of exponential decay of $\|K_n^m - \pi_n\|_{2\to 2}$. This comes from the operator theory, which says

$$\|K_n - \pi_n\|_{2\to 2} = \mu_n,$$

where μ_n is the second largest singular value of K_n and

$$\|K_n^m - \pi_n\|_{2\to 2} \le \mu_n^m \quad \forall m \ge 1. \tag{2.10}$$

Note that the ergodicity of K_n is not sufficient for the positiveness of $1-\mu_n$, that is, $K_nK_n^*$ is not necessarily irreducible. Example 2.3 illustrates this fact.

Theorem 2.2. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)|n = 1, 2, ...\}$ be a family of finite ergodic Markov chains whose second largest singular values $(\mu_n)_1^\infty$ are contained in $(0,1)$ and set $b_n = \min\{-\log \mu_n, 1\}$. Assume that*

$$\lim_{n\to\infty} \pi_{n,*} = 0.$$

Then, for fixed $1 < p < \infty$, the following are equivalent.

(1) For any $\epsilon > 0$, $\mathcal{F}_d$ has a $(T_p^d(K_n,\epsilon), b_n^{-1})$ ℓ^p-cutoff.

(2) For some $\epsilon > 0$, one has $b_n^{-1} = o(T_p^d(K_n,\epsilon))$.

Proof. (1)⇒(2) is the definition of cutoff phenomena. For (2)⇒(1), the proof is similar to that of Theorem 2.1.

By (2.10), one has

$$\|K_n^m - \pi_n\|_{2\to 2} \le e^{-mb_n},$$

and by Theorem A.1, we get

$$\|(K_n^m)^* - \pi_n\|_{p\to p} \le 2^{\theta_p} e^{-mb_n(1-\theta_p)} \quad \forall m \ge 1, \tag{2.11}$$

where $\theta_p = \left|1 - \frac{2}{p}\right|$. Set $t_n = T_p^d(K_n, \epsilon)$. Then a similar argument as in the proof of Theorem 2.1 implies that for $m \geq 1$,

$$\max_{x \in \mathcal{X}_n} \|k_{n,x}^{t_n+m} - 1\|_p \leq \epsilon 2^{\theta_p} e^{-mb_n(1-\theta_p)},$$

and for $m \geq 2$,

$$\max_{x \in \mathcal{X}_n} \|k_{n,x}^{t_n-m} - 1\|_p \geq \epsilon 2^{-\theta_p} e^{(m-1)b_n(1-\theta_p)}.$$

Note that for $c > 0$,

$$\lceil cb_n^{-1} \rceil \geq cb_n^{-1} \geq (c-1)b_n^{-1} + 1.$$

By the above inequality, replacing m with $\lfloor cb_n^{-1} \rfloor$ in the previous computations implies

$$\max_{x \in \mathcal{X}_n} \|k_{n,x}^{\lceil t_n + cb_n^{-1} \rceil} - 1\|_p \leq \epsilon 2^{\theta_p} e^{-c(1-\theta_p)}$$

and

$$\max_{x \in \mathcal{X}_n} \|k_{n,x}^{\lfloor t_n - cb_n \rfloor} - 1\|_p \geq \epsilon 2^{-\theta_p} e^{(1-c)(1-\theta_p)}.$$

Then both functions $\overline{f}$ and $\underline{f}$ defined in Definition 1.4 satisfy

$$\overline{f}(c) \leq \epsilon 2^{\theta_p} e^{-c(1-\theta_p)}, \quad \underline{f}(-c) \geq \epsilon 2^{-\theta_p} e^{(1-c)(1-\theta_p)}, \ \forall c > 2.$$

and hence $\mathcal{F}_d$ presents a (t_n, b_n^{-1}) ℓ^p-cutoff. □

Example 2.2. **(The ℓ^p-cutoff for the top to random shuffle in discrete-time cases with $1 < p < \infty$).** Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be the family of top to random shuffles, where n denotes the number of cards in a deck. Note that the family $\mathcal{F}_d$ is proved in [13] to present a total variation cutoff with critical time $n \log n$. Then the monotonicity of the ℓ^p-norm(in p) implies that $T_p^d(K_n, \epsilon) \geq \frac{1}{2} n \log n$ for n large enough.

As in the continuous-time case, we need to bound the quantity b_n defined in Theorem 2.2. By definition, the square of the second largest singular value μ_n of

K_n is the second largest eigenvalue of $K_n^*K_n$. Note that the Markov kernel $K_n^*K_n$ describes the random insertion, which is a card shuffling modelled by randomly drawing out a card from a deck and then randomly inserting it back. Again, by the comparison technique in [16] with the following identity

$$(i,j) = (j,...,i+1)(i,...,j), \quad \forall 1 \le i < j \le n,$$

one may choose a constant c such that $1-\mu_n^2 \ge c/n$. This implies that

$$-\log \mu_n \ge \frac{c}{2n}, \quad \forall n > c.$$

Combining those results in the above, we get $b_n^{-1} = O(n) = o(T_p^d(K_n,\epsilon))$ and then, by Theorem 2.2, the family $\mathcal{F}_d$ has a $(T_p^d(K_n,\epsilon), b_n^{-1})$ ℓ^p-cutoff for $1<p<\infty$. The ℓ^p-critical time for the top to random shuffle is an open problem.

The following lemma says that the window sizes of $\mathcal{F}_c$ given by Theorem 2.1 is smaller(in the sense of order) than that of $\mathcal{F}_d$ given by Theorem 2.2.

Lemma 2.4. *Let λ_n and b_n be quantities defined in Theorem 2.1 and Theorem 2.2. Then $b_n \le 2\lambda_n$. Moreover, if $\lambda_n = O(1-\mu_n)$, then $\lambda_n = O(b_n)$.*

Proof. One can easily obtain the relation $1-\lambda_n \le \mu_n$ by the characterizations of both constants λ_n and μ_n. A proof can also be found in [26].

Assume first that $\mu_n \in (e^{-1},1)$. In this case, $b_n = -\log \mu_n$. Note that

$$\log t \ge \frac{t-1}{1-e^{-1}} \quad \forall t \in (e^{-1},1).$$

This implies $-\log \mu_n \le \frac{1-\mu_n}{1-e^{-1}} \le 2\lambda_n$. For $\mu_n \in (0,e^{-1})$, it is obvious that $b_n = 1$ and then $\lambda_n \ge 1-\mu_n \ge b_n/2$.

For the second part, let $c>0$ such that $\lambda_n \le c(1-\mu_n)$ for all $n \ge 1$. Note that

$$-\log \mu_n \ge 1-\mu_n \ge c^{-1}\lambda_n.$$

Since $\lambda_n \le 2$, we have

$$b_n = \min\{-\log \mu_n, 1\} \ge \min\{c^{-1}, 2^{-1}\}\lambda_n.$$

□

2.1.3 An example

By Remark 1.8, the consequence of Theorem 2.1 and Theorem 2.2 implies that, for $1 < p \le \infty$ and $\epsilon > 0$,

$$\lambda_n^{-1} = o(T_p^c(K_n, \epsilon)) \Rightarrow \mathcal{F}_c \text{ presents a } \ell^p\text{-cutoff} \tag{2.12}$$

and

$$b_n^{-1} = o(T_p^d(K_n, \epsilon)) \Rightarrow \mathcal{F}_d \text{ presents a } \ell^p\text{-cutoff.}$$

As in those two theorems, it is natural to consider the inverse direction of the above implications. Generally, this is not true unless the normality of Markov kernels is assumed. For an explicit description of the equivalence and a detailed proof, please see Theorem 2.4 and 2.5.

In the following, we will construction a counterexample in continuous-time cases for the inverse implication of (2.12). This means that there exists a family of irreducible Markov chains presenting a ℓ^p-cutoff but the window size can not be λ_n^{-1}.

Example 2.3. Let $(a_n)_1^\infty$ be a sequence of positive integers greater than 1 and set $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of Markov chains where $\mathcal{X}_n = (\mathbb{Z}_{a_n})^n$, $\pi_n \equiv a_n^{-n}$ and the Markov kernel K_n is given by

$$K_n(x,y) = \begin{cases} \frac{1}{a_n} & \text{if } y = s(x) + (0, ..., 0, i) \text{ for some } i \in \mathbb{Z}_{a_n} \\ 0 & \text{otherwise} \end{cases}, \tag{2.13}$$

where $s(x) = (x_2, x_3, ..., x_n, x_1)$ for all $x = (x_1, ..., x_n) \in \mathcal{X}_n$.

We will show that, in the case $a_n \equiv 2$, this family has a ℓ^p-cutoff for $1 < p \le \infty$ but the mixing time $T_p(K_n, \epsilon)$ is of order n and the spectral gap λ_n of K_n satisfies $\lambda_n = O(1/n)$.

Proposition 2.3. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be the family of irreducible Markov chains, where $\mathcal{X}_n = (\mathbb{Z}_{a_n})^n$, $\pi_n \equiv a_n^{-n}$ and K_n satisfies (2.13). Assume that $1 < p \le \infty$ and $a_n > 1$ for $n \ge 1$.*

(1) The family $\mathcal{F}_c$ presents a $(t_{n,p}, b_n)$ ℓ^p-cutoff, where

$$t_{n,p} = \frac{(1-p^{-1})n \log a_n}{1 - a_n^{p^{-1}-1}}, \quad b_n = \begin{cases} \log n & \text{for } 1 < p < \infty \\ 1 & \text{for } p = \infty \end{cases}.$$

Moreover, the family $\mathcal{F}_c$ presents a strongly optimal $(t_{n,\infty}, 1)$ ℓ^∞-cutoff.

(2) Set $\widetilde{K}_n = \frac{1}{2}(K_n + K_n^*)$, then, for any $\epsilon > 0$ and for $2 < p \le \infty$, $T_p^c(\widetilde{K}_n, \epsilon) > n^2/20$ for n large enough.

We are now ready to construct a counterexample for the following implication

$$\mathcal{F}_c \text{ presents a } \ell^p\text{-cutoff} \Rightarrow \lambda_n^{-1} = o(T_p^c(K_n, \epsilon)).$$

Let $a_n = 2$ for $n \ge 1$. By Proposition 2.3(1), for $1 < p \le \infty$, the family $\mathcal{F}_c$ presents a ℓ^p-cutoff and the mixing time $T_p^c(K_n, \epsilon)$ is of order n. Recalling (2.7) and (2.8), a simple computation shows

$$\|\widetilde{H}_{n,t} - \pi_n\|_{q\to\infty} \le \|\widetilde{H}_{n,t} - \pi_n\|_{q\to q} \|I_n\|_{q\to\infty} \le 2^{\theta_q} e^{-\lambda_n(1-\theta_q)t} 2^{n(1-1/p)},$$

where λ_n is the spectral gap of K_n and $\widetilde{K}_n$, $p^{-1} + q^{-1} = 1$ and $\theta_q \in (0,1)$ for $1 < q < \infty$. By Proposition 2.3(2), one has

$$\text{for } 2 < p < \infty, \quad \lim_{n\to\infty} e^{-\lambda_n(1-\theta_q)n^2/20} 2^{n(1-1/p)} = \infty.$$

This implies that $\lambda_n n^2 \le Cn$ for some $C > 0$, or equivalently $\lambda_n = O(1/n)$. Thus, for all $\epsilon > 0$ and $1 < p \le \infty$,

$$\lambda_n T_p^c(K_n, \epsilon) = O(1) \quad \text{as } n \to \infty.$$

In discrete-time cases, it is obvious that the family $\mathcal{F}_d$ presents a ℓ^p-cutoff with critical time n for $1 \le p \le \infty$. Furthermore, $\mathcal{F}_d$ presents an optimal(in the sense of Definition 1.6(2)) (t_n, b_n) ℓ^p-cutoff, where $(t_n, b_n) = (n, 1)$ if a_n is bounded and $(t_n, b_n) = (n - \frac{1}{2}, 0)$ if $a_n \to \infty$. In details, the distance $d(p, c) = \max_{x \in \mathcal{X}_n} \|k_{n,x}^{n-c} - 1\|_p$ satisfies $d(p, 0) = 0$ for $1 \le p \le \infty$,

$$\forall 1 \le c \le n, \quad d(1, c) = 2(1 - a_n^{-c}), \quad f(\infty, c) = a_n^c - 1,$$

and for $1 < p < \infty$,

$$d(p, c) = a_n^{c(1-1/p)} \left\{(1 - a_n^{-c})^p + a_n^{c(1-p)} - a_n^{-cp}\right\}^{1/p} \begin{cases} \le 2a_n^{c(1-1/p)} \\ \ge \frac{1}{2}a_n^{c(1-1/p)} \end{cases}$$

However, $K_n K_n^*$ is not irreducible and hence $\mu_n = 1$.

Remark 2.6. Proposition 2.3 illustrates a possibility that, comparing with K_n, the reversibility of $\frac{1}{2}(K_n + K_n^*)$ slows down the convergence to its stationarity.

To prove Proposition 2.3, we need the following lemma.

Lemma 2.5. *For $n > 0$, let $a_n \in \mathbb{R}^+$, $b_n \in \mathbb{Z}^+$, $c_n = \frac{b_n - a_n}{\sqrt{a_n}}$ and $d_n = e^{-a_n} \sum_{i=0}^{b_n} \frac{a_n^i}{i!}$. Assume that $a_n + b_n \to \infty$. Then*

$$\limsup_{n\to\infty} d_n = \Phi\left(\limsup_{n\to\infty} c_n\right), \quad \liminf_{n\to\infty} d_n = \Phi\left(\liminf_{n\to\infty} c_n\right), \tag{2.14}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$.

In particular, if c_n converges(the limit can be $+\infty$ and $-\infty$), then $\lim_{n\to\infty} d_n = \Phi\left(\lim_{n\to\infty} c_n\right)$.

Proof. Here we give a proof for the first identity of (2.14) while the second one is done in a similar way. Note that if (2.14) fails, one can always find a subsequence of $(a_n)_1^\infty$ which is either bounded or tending to infinity such that

$$\limsup_{n\to\infty} d_n < \Phi\left(\limsup_{n\to\infty} c_n\right).$$

Hence it suffices to prove Lemma 2.5 by assuming the sequence $(a_n)_1^\infty$ is either bounded or tending to infinity. In the former case, one can easily prove it by Taylor expansion of an exponential function and the boundedness of a_n.

Now assume that a_n tends to infinity. We first deal with the case $a_n \in \mathbb{Z}^+$ for all $n \geq 1$. Let $Y_1, Y_2, ...$ be i.i.d. Poisson(1) random variables and F_n the distribution function of $a_n^{-1/2}(Y_1 + Y_2 + ... + Y_{a_n} - a_n)$. Then $d_n = F_n(c_n)$ and, by the central limit theorem, F_n converges uniformly to the distribution function Φ of the standard normal random variable.

Set $L = \limsup_{n\to\infty} c_n$. We first assume that $|L| < \infty$. For all $\epsilon > 0$, if k is large enough, one has

$$\sup_{n\geq k} F_n(L-\epsilon) \leq \sup_{n\geq k} F_n(c_n) \leq \sup_{n\geq k} F_n(L+\epsilon).$$

Letting $k \to \infty$ and then $\epsilon \to 0$ implies the desired identity.

In the case $|L| = \infty$, observe that, for $l \in \mathbb{R}$, if k is large enough, one has

$$\sup_{n\geq k} F_n(c_n) \begin{cases} \geq \sup_{n\geq k} F_n(l) & \text{if } L = \infty \\ \\ \leq \sup_{n\geq k} F_n(l) & \text{if } L = -\infty \end{cases}.$$

Then the first identity with integer a_n is proved by letting $k \to \infty$ and $l \to \pm\infty$.

For $a_n \in \mathbb{R}^+$, we consider these two sequences, $(\lfloor a_n \rfloor)_{n=1}^\infty$ and $(\lceil a_n \rceil)_{n=1}^\infty$. Note that, for fixed $k, l > 0$, both $\frac{l-t}{\sqrt{t}}$ and $e^{-t}\sum_{i=0}^k \frac{t^i}{i!}$ are strictly decreasing for $t \in \mathbb{R}^+$, which implies

$$\frac{b_n - \lceil a_n \rceil}{\sqrt{\lceil a_n \rceil}} \leq c_n \leq \frac{b_n - \lfloor a_n \rfloor}{\sqrt{\lfloor a_n \rfloor}},$$

and

$$e^{-\lceil a_n \rceil} \sum_{i=0}^{b_n} \frac{\lceil a_n \rceil^i}{i!} \leq d_n \leq e^{-\lfloor a_n \rfloor} \sum_{i=0}^{b_n} \frac{\lfloor a_n \rfloor^i}{i!}. \tag{2.15}$$

Note also that for $[\cdot] \in \{\lfloor \cdot \rfloor, \lceil \cdot \rceil\}$,

$$\frac{b_n - [a_n]}{\sqrt{[a_n]}} = \frac{b_n - a_n}{\sqrt{a_n}} \times \sqrt{\frac{a_n}{[a_n]}} + \frac{a_n - [a_n]}{\sqrt{[a_n]}}.$$

One then has $\limsup_{n\to\infty} \frac{b_n - [a_n]}{\sqrt{[a_n]}} = \limsup_{n\to\infty} c_n$. Hence, the first identity for nonnegative real-valued a_n is proved by applying (2.15) and the result in the case $a_n \in \mathbb{Z}_+$ for $n \geq 1$. □

Proof of Proposition 2.3(1). For (1), note that for K_n and $\widetilde{K}_n$, the ℓ^p-distance(for both discrete-time and continuous-time cases) is independent of the initial state. It is obvious that, for $i \geq n$, $K_n^i(x, \cdot) \equiv \pi_n$ for all $x \in \mathcal{X}_n$, which means that the discrete-time Markov chain with transition matrix K_n perfectly mixes after the nth step. This implies

$$h_{n,t}(0,0) - 1 = e^{-t} \sum_{j=0}^{n} \frac{t^j}{j!} (a_n^{n-j} - 1)$$

and

$$h_{n,t}(0,y) - 1 = e^{-t} \sum_{j=i}^{n} \frac{t^j}{j!} (a_n^{n-j} - 1) - e^{-t} \sum_{j=0}^{i-1} \frac{t^j}{j!},$$

if $y = (y_1, ..., y_n)$ satisfies $y_1 = y_2 = \cdots = y_{n-i} = 0$ and $y_{n-i+1} \neq 0$ for some $1 \leq i \leq n$.

For $1 < p < \infty$, the ℓ^p-distance is given by

$$\begin{aligned} \|h_{n,t}^0 - 1\|_p^p = &\sum_{i=1}^{n} \left| e^{-t} \sum_{j=i}^{n} \frac{t^j}{j!} (a_n^{n-j} - 1) - e^{-t} \sum_{j=0}^{i-1} \frac{t^j}{j!} \right|^p a_n^{-n} a_n^{i-1} (a_n - 1) \\ &+ \left(e^{-t} \sum_{j=0}^{n} \frac{t^j}{j!} (a_n^{n-j} - 1) \right)^p a_n^{-n}. \end{aligned}$$

By the triangle inequality and the following fact

$$n^{1-p} (c_1 + \cdots + c_n)^p \leq c_1^p + \cdots + c_n^p \leq (c_1 + \cdots + c_n)^p$$

for $c_i \geq 0$ and $1 \leq p < \infty$, one has that for $t > 0$,

$$2^{\frac{-1}{p}}(n+1)^{\frac{1-p}{p}}[f_{p,1}(n,t) - f_{p,2}(n,t)] \leq \|h^0_{n,t} - 1\|_p \leq [f_{p,1}(n,t) + f_{p,2}(n,t)] \quad (2.16)$$

where

$$\begin{aligned} f_{p,1}(n,t) &= e^{-t}a_n^{-n/p}\left(\sum_{i=1}^n \sum_{j=i}^n \frac{t^j}{j!}(a_n^{n-j} - 1)a_n^{i/p} + \sum_{j=0}^n \frac{t^j}{j!}(a_n^{n-j} - 1)\right) \\ &= e^{-t}a_n^{n(1-1/p)} \sum_{j=0}^{n-1} \frac{1}{j!}(ta_n^{(1-p)/p})^j \frac{(1 - a_n^{j-n})(1 - a_n^{-(j+1)/p})}{1 - a_n^{-1/p}} \end{aligned}$$

and

$$f_{p,2}(n,t) = e^{-t}a_n^{-n/p}\left(\sum_{i=1}^n \sum_{j=0}^{i-1} \frac{t^j}{j!}a_n^{i/p}\right) \leq 2e^{-t} \sum_{j=0}^n \frac{t^j}{j}.$$

Let $t_{n,p}$ and b_n be quantities defined in Proposition 2.3 and $t_n = t_{n,p}$. Note that for $s > 1$, the function $s \mapsto \frac{\log s}{1-s^{-1}}$ is increasing and has limit 1 as $s \downarrow 1$. This implies $t_n(1-\delta) > n$ for some $\delta > 0$ and hence, by Lemma 2.5, one has

$$\lim_{n\to\infty} f_{p,2}(n, t_n + cb_n) \leq \lim_{n\to\infty} f_{p,2}(n, t_n(1 - \delta/2)) = 0 \quad \forall c \in \mathbb{R}. \quad (2.17)$$

By this fact, it suffices to consider only the function $f_{p,1}$. Moreover, by the following inequality

$$1 - 2^{-1/p} \leq (1 - a_n^{j-n})(1 - a_n^{-(j+1)/p})(1 - a_n^{-1/p})^{-1} \leq (1 - 2^{-1/p})^{-1},$$

for $0 \leq j \leq n-1$, it is equivalent to concern the following function

$$g_p(n,t) = e^{-t}a_n^{n(1-1/p)} \sum_{j=0}^{n-1} \frac{1}{j!}(ta_n^{(1-p)/p})^j.$$

A simple computation shows

$$g_p(n, t_n + cb_n) = \exp\{-cb_n(1 - a_n^{(1-p)/p})\}e^{-s_n} \sum_{j=0}^{n-1} \frac{s_n^j}{j!} \quad (2.18)$$

where $s_n = (t_n + cb_n)a_n^{(1-p)/p}$, and for fixed $c \in \mathbb{R}$,

$$n - s_n = n\left(1 - (1 + o(1))\frac{\log a_n^{1-1/p}}{a_n^{1-1/p} - 1}\right) \quad \text{as } n \to \infty.$$

Since the mapping $s \mapsto \frac{\log s}{s-1}$ for $s \geq 1$ is strictly decreasing and has limit 1 as $s \downarrow 1$, one may choose $\delta \in (0,1)$ and $N = N(\delta, p, c) \in \mathbb{N}$ such that

$$n - s_n \geq \delta n \quad \forall n \geq N,$$

which implies, by Lemma 2.5,

$$\lim_{n\to\infty} e^{-s_n} \sum_{j=0}^{n-1} \frac{s_n^j}{j!} = 1 \quad \forall c \in \mathbb{R}. \tag{2.19}$$

Now combining (2.16), (2.17), (2.18) and (2.19), we get

$$\limsup_{n\to\infty} \|h^0_{n,t_n+cb_n} - 1\|_p \leq \limsup_{n\to\infty} n^{-c\left(1-2^{1/p-1}\right)}, \quad \forall c > 0,$$

and

$$\liminf_{n\to\infty} \|h^0_{n,t_n+cb_n} - 1\|_p \geq \liminf_{n\to\infty} 2^{-1/p} n^{-c\left(1-2^{1/p-1}\right)+1/p-1}, \quad \forall c < 0.$$

Hence both functions $\overline{f}$ and $\underline{f}$ defined in Definition 1.4 satisfy

$$\overline{f}(c) = 0 \ \forall c > 0, \quad \underline{f}(c) = \infty \ \forall c < \frac{p^{-1}-1}{1-2^{1/p-1}},$$

which proves the desired ℓ^p-cutoff for $1 < p < \infty$.

For $p = \infty$, the ℓ^∞-distance is given by

$$\|h^0_{n,t} - 1\|_\infty = h_{n,t}(0,0) - 1 \begin{cases} \leq a_n^n e^{-t} \sum_{j=0}^{n-1} \frac{(t/a_n)^j}{j!} \\ \geq \frac{1}{2} a_n^n e^{-t} \sum_{j=0}^{n-1} \frac{(t/a_n)^j}{j!} \end{cases}.$$

For $c \in \mathbb{R}$, let $t = t_{n,\infty} + c$. Then one has

$$t - n\log a_n = \frac{n \log a_n}{a_n - 1} + c = \frac{t}{a_n} + (1 - a_n^{-1})c.$$

which implies

$$\frac{1}{2} e^{-c} c_n \leq \|h^0_{n,t} - 1\|_\infty \leq e^{-c/2} c_n,$$

where $c_n = e^{-t/a_n} \sum_{j=0}^{n-1} \frac{(t/a_n)^j}{j!}$. Since $\frac{\log a_n}{a_n - 1} \le \log 2 < 1$ for $n \ge 1$, by Lemma 2.5, one has $c_n \to 1$ as $n \to \infty$. Hence those functions $\overline{f}$ and $\underline{f}$ defined in Definition 1.4(3) are bounded by

$$\frac{e^{-c}}{2} \le \underline{f}(c) \le \overline{f}(c) \le e^{-c/2} \quad \forall c \in \mathbb{R}.$$

This proves the desired ℓ^∞-cutoff. □

Proof of Proposition 2.3(2). One can see that the mixing time of $\widetilde{K}_n$ depends strongly on how many digits are randomized via the mapping s in Example 2.3 and hence is related to the ruin problem of the simple random walk on $\mathbb{Z}$.

We first consider the following realization of the Markov kernel $\widetilde{K}_n$. For $n \ge 1$, let $X_n^1, X_n^2, ...$ be a sequence of i.i.d. random variables on the n-cycle $\mathbb{Z}_n$ satisfying

$$\mathbb{P}\{X_n^1 = 1\} = \mathbb{P}\{X_n^1 = -1\} = \frac{1}{2},$$

and let $U_n^1, U_n^2, ...$ be a sequence of i.i.d. random variables which is independent of $(X_n^i)_{i=1}^\infty$ and U_n^1 is uniformly distributed on $\{1, ..., a_n\}$. Set $S_n^0 = 0$, $S_n^k = X_n^1 + \cdots + X_n^k$ for $k \ge 1$ and, for $1 \le i \le n$, let $e_{n,i}$ be the element in $(\mathbb{Z}_{a_n})^n$ with entry 0 in each coordinate except the ith one which is equal to 1. For $n \ge 1$, let $(Y_n^k)_{k=0}^\infty$ and $(Z_n^k)_{k=0}^\infty$ be random variables on $(\mathbb{Z}_{a_n})^n$ satisfying

$$Y_n^{k+1} = Y_n^k + U_n^{k+1} e_{n, S_n^{k+1}}, \quad Z_n^k = s^{S_n^k}(Y_n^k), \quad \forall k \ge 0.$$

where, s is the function defined in Example 2.3 with the inverse s^{-1} and s^i denotes the composition of s with itself for i times if $i > 0$. For $i < 0$, $s^i = (s^{-1})^{-i}$ and s^0 stands for the identity map on $(\mathbb{Z}_{a_n})^n$. Then the transition matrix $\widetilde{K}_n$ can be specified by

$$\widetilde{K}_n(x, y) = \mathbb{P}\left\{Z_n^{k+1} = y | Z_n^k = x\right\}.$$

Note that, for $k \geq 1$, $\mathbb{P}\{Z_n^k \in \{e_{n,1}, e_{n,n}\} | Z_n^0 = 0\} > 0$.

To finish the proof, we need the following fact. Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables satisfying $\mathbb{P}\{X_1 = 1\} = \mathbb{P}\{X_1 = -1\} = 1/2$ and $S_n = X_1 + \cdots + X_n$. For $n \geq 1$, let $A_n = \{\max_{1 \leq k \leq n^2/20} |S_k| \leq n/4\}$. By Kolmogorov's inequality, one has

$$\mathbb{P}\{A_n\} \geq \frac{1}{5}, \quad \forall n \geq 8.$$

For $n \geq 1$, let $m = \lfloor n^2/20 \rfloor$ and for $j \geq 1$, let $W_{n,j} = \max\{|S_n^i| : 1 \leq i \leq j\}$ and $B_n = \{W_{n,m} \leq n/4\}$. It is clear that $\mathbb{P}(A_n) = \mathbb{P}(B_n)$. Without loss of generality, one may assume that $(X_n^i)_{i=1}^{\infty}$ are independent of Y_n^0. In the above setting, we have that, for $1 \leq j \leq m$,

$$\begin{aligned}
&\widetilde{K}_n^j(0, e_{n,1}) + \widetilde{K}_n^j(0, e_{n,n}) = \mathbb{P}\{Z_n^j \in \{e_{n,1}, e_{n,n}\} | Z_n^0 = 0\} \\
\geq &\sum_{l=1}^{\lfloor n/4 \rfloor} \mathbb{P}\{Z_n^j \in \{e_{n,1}, e_{n,n}\} | W_{n,j} = l, Z_n^0 = 0\} \mathbb{P}\{W_{n,j} = l\} \\
\geq &a_n^{-n/2} \mathbb{P}\{B_n\} \geq \frac{a_n^{-n/2}}{5}
\end{aligned}$$

This implies that for $n \geq 8$,

$$\begin{aligned}
\widetilde{h}_{n,m}^0(e_{n,1}) + \widetilde{h}_{n,m}^0(e_{n,n}) &\geq a_n^n e^{-m} \sum_{j=1}^{m} \frac{m^j}{j!} \left[\widetilde{K}_{n,0}^j(e_{n,1}) + \widetilde{K}_{n,0}^j(e_{n,n}) \right] \\
&\geq \frac{a_n^{n/2} e^{-m}}{5} \sum_{j=1}^{m} \frac{m^j}{j!}.
\end{aligned}$$

By the above computation, we have

$$\begin{aligned}
\|\widetilde{h}_{n,m}^0 - 1\|_p \geq \|\widetilde{h}_{n,m}^0\|_p - 1 &\geq a_n^{-n/p} \left(\widetilde{h}_{n,m}^0(e_{n,1})^p + \widetilde{h}_{n,m}^0(e_{n,1})^p \right)^{1/p} - 1 \\
&\geq a_n^{-n/p} 2^{1/p-1} \left(\widetilde{h}_{n,m}^0(e_{n,1}) + \widetilde{h}_{n,m}^0(e_{n,n}) \right) - 1 \\
&\geq a_n^{n(1/2-1/p)} 2^{1/p-1} 5^{-1} \left(e^{-m} \sum_{j=1}^{m} \frac{m^j}{j!} \right).
\end{aligned}$$

Hence, by Lemma 2.5, one may choose, for $\epsilon > 0$ and $2 < p \le \infty$, $N = N(p,\epsilon)$ such that

$$T_p^c(\widetilde{K}_n, \epsilon) > n^2/20, \quad \forall n \ge N.$$

□

2.1.4 Comparing the ℓ^p and ℓ^q cutoffs

By the monotonicity(in p) of the ℓ^p-norm, one may relate the ℓ^p-cutoff and ℓ^q-cutoff for $1 < p, q < \infty$ in the following way.

Theorem 2.3. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of irreducible Markov chains and λ_n and μ_n be the spectral gap and the second largest singular value of K_n. Assume that*

$$\lim_{n\to\infty} \pi_{n,*} = 0.$$

If $\mathcal{F}_c$ presents a $(T_p^c(K_n,\epsilon), \lambda_n^{-1})$ ℓ^p-cutoff for some $1 \le p \le \infty$, then

(1) for $p < q < \infty$, the family $\mathcal{F}_c$ presents a $(T_q^c(K_n,\epsilon), \lambda_n^{-1})$ ℓ^q-cutoff.

(2) for $1 < q < p$, there exist a sequence $(i_n)_1^\infty$ tending to infinity such that, by setting

$$\mathcal{F}^{(1)} = \{(\mathcal{X}_{i_n}, K_{i_n}, \pi_{i_n})\}_{n=1}^\infty, \quad \mathcal{F}^{(2)} = \{(\mathcal{X}_{i_n}, K_{i_n}^*, \pi_{i_n})\}_{n=1}^\infty,$$

we have either $\mathcal{F}_c^{(1)}$ presents a $(T_q^c(K_{i_n},\epsilon), \lambda_{i_n}^{-1})$ ℓ^q-cutoff or $\mathcal{F}_c^{(2)}$ presents a $(T_q^c(K_{i_n}^*,\epsilon), \lambda_{i_n}^{-1})$ ℓ^q-cutoff.

In discrete-time cases, assume that, for $n \ge 1$, K_n is aperiodic and $\mu_n \in (0,1)$. Set $b_n = \min\{-\log \mu_n, 1\}$. If $\mathcal{F}_d$ has a $(T_p^d(K_n,\epsilon), b_n^{-1})$ ℓ^p-cutoff for some $1 \le p \le \infty$, then

(3) *for* $p < q < \infty$, *the family* $\mathcal{F}_d$ *presents a* $(T_q^d(K_n, \epsilon), b_n^{-1})$ ℓ^q*-cutoff.*

(4) *for* $1 < q < p$, *there exist an increasing sequence* $(j_n)_1^\infty$ *such that, by setting*

$$\mathcal{F}^{(3)} = \{(\mathcal{X}_{j_n}, K_{j_n}, \pi_{j_n})\}_{n=1}^\infty, \quad \mathcal{F}^{(4)} = \{(\mathcal{X}_{j_n}, K_{j_n}^*, \pi_{j_n})\}_{n=1}^\infty,$$

we have either $\mathcal{F}_d^{(3)}$ *presents a* $(T_q^d(K_{j_n}, \epsilon), b_{i_n}^{-1})$ ℓ^q*-cutoff or* $\mathcal{F}_d^{(4)}$ *presents a* $(T_q^d(K_{j_n}^*, \epsilon), b_{i_n}^{-1})$ ℓ^q*-cutoff.*

Proof. By applying Proposition 2.2, Theorem 2.1 and Theorem 2.2. □

The following corollary improves the above results for some specific Markov kernel K with $T_p(K, \epsilon) = T_p(K^*, \epsilon)$.

Corollary 2.2. *Let* $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ *be a family of irreducible Markov chains with* $\pi_{n,*} \to 0$ *and* λ_n *and* μ_n *be the spectral gap and the second largest singular value of* K_n. *Assume that, for* $n \geq 1$, *there exists a finite group* G_n *acting transitively on* $\mathcal{X}_n$ *such that*

$$K_n(gx, gy) = K_n(x, y), \quad \forall x, y \in \mathcal{X}_n, g \in G_n.$$

Let $\epsilon > 0$, *then*

(1) *for* $1 < p < q < \infty$,

$$\mathcal{F}_c \text{ has a } (T_p^c(K_n, \epsilon), \lambda_n^{-1}) \ \ell^p\text{-cutoff} \Leftrightarrow \mathcal{F}_c \text{ has a } (T_q^c(K_n, \epsilon), \lambda_n^{-1}) \ \ell^q\text{-cutoff.}$$

(2) *If* K_n *is aperiodic and there exist* $1 < r < \infty$ *and* $\eta > 0$ *such that* $T_r^d(K_n, \eta)$ *tends to infinity, then for* $1 < p < q < \infty$,

$$\mathcal{F}_d \text{ has a } (T_p^d(K_n, \epsilon), b_n^{-1}) \ \ell^p\text{-cutoff} \Leftrightarrow \mathcal{F}_d \text{ has a } (T_q^d(K_n, \epsilon), b_n^{-1}) \ \ell^q\text{-cutoff,}$$

where $b_n = \min\{-\log \mu_n, 1\}$.

In particular, if $\mathcal{F}_c$ presents a ℓ^p-cutoff for some $p > 1$, then, for $1 < q < \infty$, the ℓ^q-critical time and the ℓ^p-critical time are of the same order. This also holds for discrete-time cases if one assumes further that $T_r^d(K_n, \epsilon) \to \infty$ for some $1 < r < \infty$ and $\epsilon > 0$.

Proof. By Theorem 2.1, Theorem 2.2 and Corollary 2.1. □

Example 2.4. **(The ℓ^p-cutoff for the random transposition in discrete-time cases with $p \in (1, \infty)$.)** Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be the family of random transpositions, where n denotes the number of cards in a deck. In [21], the ℓ^2-mixing time of K_n was proved to satisfy $T_2^d(K_n, \epsilon) \sim \frac{1}{2} n \log n$ for $\epsilon \in (0, 1)$ and the spectral gap of K_n is equal to $2/n$. Recall a simple but useful fact modified from [16, Lemma 1]: If K is a reversible and irreducible Markov kernel on $\mathcal{X}$ and β is an eigenvalue of K, then

$$\beta \geq -1 + 2 \max_{x \in \mathcal{X}} \{K(x, x)\}. \tag{2.20}$$

In addition to this fact, we have $b_n = -\log(1 - \frac{2}{n}) \sim \frac{2}{n}$. This implies that, by Theorem 2.2, the family $\mathcal{F}_d$ presents a $(T_2^d(K_n \epsilon), n)$ ℓ^2-cutoff and hence, by Corollary 2.2, it has a $(T_p^d(K_n, \epsilon), n)$ ℓ^p-cutoff for $1 < p < \infty$. It has been proved by Diaconis and Shahshahani in [21] that $T_p^d(K_n, \epsilon) \sim \frac{1}{2} n \log n$ for $1 \leq p \leq 2$. For $2 < p < \infty$, the ℓ^p-critical time is open and a conjecture is $T_p^d(K_n, \epsilon) \sim \frac{1}{2} n \log n$, for $2 < p \leq \infty$.

The following is a corollary of Proposition 1.12 which gives an upper bound on the window size of a ℓ^p-cutoff if a family presents a strongly optimal ℓ^q-cutoff with the same critical time.

Proposition 2.4. *Let $\mathcal{F}$ be a family of irreducible Markov chains. Assume that, for some $1 \leq p \leq \infty$ and $1 < q \leq \infty$, $\mathcal{F}$ presents a strongly optimal (t_n, b_n) ℓ^p-cutoff and a (s_n, c_n) ℓ^q-cutoff with $|t_n - s_n| = O(b_n)$. Then $b_n = O(c_n)$.*

When a family presents a ℓ^p and a ℓ^q cutoff with the same critical time, a question arises from the monotonicity of the ℓ^p-norm in p : Does the family present a ℓ^r-cutoff for $p < r < q$? The following gives part of the answer.

Proposition 2.5. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of irreducible Markov chains. Assume that, for $1 < p < q \leq \infty$, $\mathcal{F}_c$ presents a (t_n, b_n) ℓ^p-cutoff and a (s_n, c_n) ℓ^q-cutoff with $t_n \sim s_n$ and $|t_n - s_n| = o(t_n)$. Then, for $p < r < q$, the family $\mathcal{F}_c$ has a (t_n, d_n) ℓ^r-cutoff, where*

$$d_n = \max\{b_n, c_n, |t_n - s_n|\}.$$

The above remains true for discrete-time cases if one assumes further that t_n tends to infinity and sets

$$d_n = \max\{1, b_n, c_n, |t_n - s_n|\}$$

Proof. By definition, we may choose $C > 0$ and $N > 0$ such that, for $c \geq C$ and $n \geq N$,

$$\max_{x \in \mathcal{X}_n} d_{\pi_n,p}(H^x_{n,t_n - cb_n}, \pi_n) \geq 1/2 \geq \max_{x \in \mathcal{X}_n} d_{\pi_n,q}(H^x_{n,s_n + cc_n}, \pi_n).$$

This implies that

$$t_n - Cb_n \leq T^c_r(K_n, 1/2) \leq s_n + Cc_n, \quad \forall n \geq N.$$

Then one can prove the desired cutoff by following the definition.

For discrete-time cases, a similar statement as above can only show that, for some $C > 0$ and $N > 0$,

$$\lfloor t_n - Cb_n \rfloor \leq T^d_r(K_n, 1/2) \leq \lceil s_n + Cc_n \rceil, \quad \forall n \geq N.$$

To take care of the floor and ceiling, we need assume further that the window size is bounded from below by a positive number. □

2.1.5 The ℓ^p-cutoff for normal and reversible Markov chains

In Example 2.3, one can check that the family $\mathcal{F}$ contains no normal matrices. The following theorem gives a positive answer on the reverse statement of (2.12) with the assumption of normal Markov kernels.

Theorem 2.4. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of irreducible Markov chains and λ_n be the spectral gap of K_n. Assume that K_n is normal for all $n \geq 1$ and $\pi_{n,*} \to 0$. Then, for $1 < p < \infty$, the following are equivalent.*

(1) For all $\epsilon > 0$, $\mathcal{F}_c$ has a $(T_p^c(K_n, \epsilon), \lambda_n^{-1})$ ℓ^p-cutoff.

(2) $\mathcal{F}_c$ presents a ℓ^p-cutoff.

(3) $\mathcal{F}_c$ presents a weak ℓ^p-cutoff.

(4) $\lambda_n^{-1} = o(T_p^c(K_n, \epsilon))$ for some $\epsilon > 0$.

Proof. By Theorem 2.1 and Definition 1.4, it remains to prove (3)⇒(4). By the normality of K_n, one has

$$\max_{x \in \mathcal{X}_n} \|h_{n,t}^x - 1\|_p = \|H_{n,t} - \pi_n\|_{p' \to \infty} \geq \|H_{n,t} - \pi_n\|_{p' \to p'} \geq e^{-\lambda_n t}$$

for $1 < p < \infty$ and $p^{-1} + (p')^{-1} = 1$. Assume that $\mathcal{F}_d$ presents a weak ℓ^p-cutoff with critical time t_n. By Corollary 1.3, there exists $\epsilon > 0$ such that $t_n \sim T_p^c(K_n, \epsilon)$. Putting $t = 2T_p^c(K_n, \epsilon)$ in the above inequality implies

$$\begin{aligned} 0 &= \lim_{n \to \infty} \max_{x \in \mathcal{X}_n} \|h_{n,3t_n/2}^x - 1\|_p \\ &\geq \lim_{n \to \infty} \max_{x \in \mathcal{X}_n} \|h_{n,2T_p^c(K_n,\epsilon)}^x - 1\|_p \geq \limsup_{n \to \infty} e^{-2\lambda_n T_p^c(K_n,\epsilon)}. \end{aligned}$$

This proves the desired limit $\lambda_n T_p^c(K_n, \epsilon) \to \infty$ as $n \to \infty$. $\square$

Example 2.5. **(The ℓ^p-cutoff for the random insertion in continuous-time cases with $1 < p \leq \infty$)** The random insertion is a card shuffling done by randomly drawing out a card from a deck and randomly inserting it back. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of random insertions, where n denotes that number of cards in a deck. It has been proved in Example 2.2 that the spectral gap λ_n of K_n is bounded from below by c/n, where c is a constant independent of n. By applying Lemma 2.7, the ℓ^p-mixing time is bounded from below by $\frac{p-1}{2p} n \log n$ for n large enough. Then, Theorem 2.4 gives the ℓ^p-cutoff for the family $\mathcal{F}_c$ for all $1 < p \leq \infty$. The ℓ^p-critical time is an open problem.

Similarly, one can obtain a discrete version of the above theorem. Note that the assumption of $\mu_n \neq 1$ is not needed in this case since the normality and ergodicity of K_n implies the irreducibility of $K_n K_n^*$.

Theorem 2.5. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of ergodic Markov chains and $b_n = \min\{-\log \mu_n, 1\}$, where μ_n is the second largest singular value of K_n. Fix $1 < p < \infty$ and assume that K_n is normal and*

$$\lim_{n\to\infty} \pi_{n,*} = 0, \quad \lim_{n\to\infty} T_p^d(K_n, \eta) = \infty$$

for some $\eta > 0$. Then the following are equivalent.

(1) For all $\epsilon > 0$, $\mathcal{F}_d$ has a $(T_p^d(K_n, \epsilon), b_n^{-1})$ ℓ^p-cutoff.

(2) $\mathcal{F}_d$ presents a ℓ^p-cutoff.

(3) $\mathcal{F}_d$ presents a weak ℓ^p-cutoff.

(4) There exists $\epsilon > 0$ such that $b_n^{-1} = o(T_p^d(K_n, \epsilon))$.

Proof. As in the proof of Theorem 2.4, it remains to show (3)⇒(4). By the normality of K_n, we have

$$\max_{x\in\mathcal{X}_n} \|k^m_{n,x}-1\|_p = \|K_n^m-\pi_n\|_{q\to\infty} \geq \|K_n^m-\pi_n\|_{q\to q} \geq e^{-m(-\log\mu_n)}.$$

Assume that $\mathcal{F}_d$ presents a weak ℓ^p-cutoff with critical time t_n. By Corollary 1.3, one has $t_n \sim T_p^d(K_n,\epsilon)$ for some $\epsilon > 0$. Then a similar argument as in the proof of Theorem 2.4 implies $(-\log\mu_n)T_p^d(K_n,\epsilon)\to\infty$ and hence $b_nT_p^d(K_n,\epsilon)\to\infty$. □

Based on the above theorems, we may relate the ℓ^p-cutoff and ℓ^q-cutoff as follows.

Corollary 2.3. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of normal and irreducible Markov chains. Assume that $\lim_{n\to\infty}\pi_{n,*} = 0$. If $\mathcal{F}_c$ presents a ℓ^p-cutoff for some $1 < p \leq \infty$, then*

(1) for $p < q < \infty$, $\mathcal{F}_c$ presents a ℓ^q-cutoff.

(2) for all $1 < q < p$, there exist a sequence $(i_n)_1^\infty$ tending to infinity such that, setting

$$\mathcal{F}^{(1)} = \{(\mathcal{X}_{i_n}, K_{i_n}, \pi_{i_n})\}_{n=1}^\infty, \quad \mathcal{F}^{(2)} = \{(\mathcal{X}_{i_n}, K^*_{i_n}, \pi_{i_n})\}_{n=1}^\infty,$$

we have that either $\mathcal{F}_c^{(1)}$ or $\mathcal{F}_c^{(2)}$ presents a ℓ^q-cutoff.

In discrete-time cases, assume that K_n is aperiodic. If, for some $1 < p \leq \infty$, $\mathcal{F}_d$ presents a ℓ^p-cutoff with critical time tending to infinity, then

(3) for $p < q < \infty$, the family $\mathcal{F}_d$ presents a ℓ^q-cutoff.

(4) for $1 < q < p$, there exist a sequence $(j_n)_1^\infty$ tending to infinity such that, by setting

$$\mathcal{F}^{(3)} = \{(\mathcal{X}_{j_n}, K_{j_n}, \pi_{j_n})\}_{n=1}^\infty, \quad \mathcal{F}^{(4)} = \{(\mathcal{X}_{j_n}, K^*_{j_n}, \pi_{j_n})\}_{n=1}^\infty,$$

we have that either $\mathcal{F}_d^{(3)}$ *or* $\mathcal{F}_d^{(4)}$ *presents a* ℓ^q*-cutoff.*

Proof. By Proposition 2.2, Theorem 2.4 and Theorem 2.5. □

As the group structure of $\mathcal{X}_n$ and a specific Markov kernel K_n equates $T_p(K_n, \epsilon)$ and $T_p(K_n^*, \epsilon)$, one may derive a stronger version of Corollary 2.3 as follows.

Corollary 2.4. *Let* $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ *be a family of normal and irreducible Markov chains. Assume that* $\pi_{n,*} \to 0$ *and, for* $n \geq 1$*, there exists a finite group* G_n *acting transitively on* $\mathcal{X}_n$ *such that*

$$K_n(gx, gy) = K_n(x, y), \quad \forall x, y \in \mathcal{X}_n, g \in G_n.$$

Then:

(1) For $1 < p < q < \infty$*,*

$$\mathcal{F}_c \text{ presents a } \ell^p\text{-cutoff} \quad \Leftrightarrow \quad \mathcal{F}_c \text{ presents a } \ell^q\text{-cutoff.}$$

(2) If K_n *is aperiodic and there exist* $1 < r < \infty$ *and* $\epsilon > 0$ *such that* $T_r^d(K_n, \epsilon) \to \infty$*, then for* $1 < p < q < \infty$*,*

$$\mathcal{F}_d \text{ presents a } \ell^p\text{-cutoff} \quad \Leftrightarrow \quad \mathcal{F}_d \text{ presents a } \ell^q\text{-cutoff.}$$

In particular, if $\mathcal{F}_c$ *presents a* ℓ^p*-cutoff for some* $p > 1$*, then, for* $1 < q < \infty$*, the* ℓ^q*-critical time and the* ℓ^p*-critical time are of the same order. This also holds for discrete-time cases if one assumes further that* $T_r^d(K_n, \epsilon) \to \infty$ *for some* $1 < r < \infty$ *and* $\epsilon > 0$*.*

Proof. By Corollary 2.1, Theorem 2.4 and Theorem 2.5. □

As a consequence of the above corollary, one always has the ℓ^p-cutoff for all $1 < p < \infty$, if it is proved for some specific p. In this case, it is natural to consider the case $p = 2$. Then, by Lemma 3.1, we may obtain a sufficient condition for the ℓ^p-cutoff by considering only the multiplicity of the spectral gap, for continuous-time cases, or the second largest singular value, for the discrete-time cases.

Corollary 2.5. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of normal and irreducible Markov chains. For $n \geq 1$, let λ_n and μ_n be the spectral gap and the second largest singular value of K_n whose multiplicities are m_n and m'_n respectively. Assume that, for $n \geq 1$, there exists a finite group G_n acting transitively on $\mathcal{X}_n$ such that*

$$K_n(gx, gy) = K_n(x, y), \quad \forall x, y \in \mathcal{X}_n, \; g \in G_n.$$

Then:

(1) If $m_n \to \infty$, then $\mathcal{F}_c$ presents a ℓ^p-cutoff for all $1 < p < \infty$.

(2) If $0 < \mu_n < 1$ for all $n \geq 1$ and

$$\lim_{n\to\infty} m'_n = \infty, \quad \lim_{n\to\infty} \frac{\log m'_n}{\log \mu_n^{-1}} = \infty, \tag{2.21}$$

then $\mathcal{F}_d$ presents a ℓ^p-cutoff for all $1 < p < \infty$.

Proof. By Lemma 3.1, one has

$$\max_{x\in\mathcal{X}_n} \|h_{n,t}(x, \cdot) - 1\|_2^2 \geq m_n e^{-2t\lambda_n}, \quad \max_{x\in\mathcal{X}_n} \|k_n^m(x, \cdot) - 1\|_2^2 \geq m'_n \mu_n^{2m}.$$

This implies that

$$T_2^c(K_n, 1) \geq \frac{\log m_n}{2} \lambda_n^{-1}, \quad T_2^d(K_n, 1) \geq \frac{\log m'_n}{2} (-\log \mu_n)^{-1}.$$

Then, for $1 < p < \infty$, the ℓ^p-cutoff for $\mathcal{F}_c$ is proved by Theorem 2.4 and Corollary 2.4. For the discrete-time cases, the assumption in (2.21) implies that $T_2^d(K_n, 1)$ tends to infinity and hence $(\min\{-\log \mu_n, 1\})^{-1} = o(T_2^d(K_n, 1))$. Then, by Theorem 2.5 and Corollary 2.4, the family $\mathcal{F}_d$ presents a ℓ^p-cutoff for all $1 < p < \infty$. □

By Proposition 2.1, one can find a connection between the ℓ^2-mixing time and the ℓ^∞-mixing time when a Markov kernel is assumed to be reversible. This is sufficient to show the ℓ^∞-cutoff.

Theorem 2.6. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of reversible and irreducible Markov chains. Then Theorem 2.4 and Theorem 2.5 also hold for $p = \infty$.*

In particular, if $\mathcal{F}_c$(resp. $\mathcal{F}_d$) presents a ℓ^∞-cutoff, then, for any $\epsilon > 0$, $\mathcal{F}_c$(resp. $\mathcal{F}_d$) presents a $(2T_2^c(K_n,\epsilon), \lambda_n^{-1})$(resp. $(2T_2^d(K_n,\epsilon), b_n^{-1})$) ℓ^∞-cutoff, where λ_n is the spectral gap of K_n, μ_n is the second largest singular value of K_n and $b_n = \min\{-\log\mu_n, 1\}$.

Proof. By Theorem 2.4 and Theorem 2.5, we only need to deal with the case $p = \infty$. According to Definition 1.4, it remains to prove (3)⇒(4)⇒(1).

For (3)⇒(4), assume that $\mathcal{F}_c$ presents a weak ℓ^∞-cutoff with critical time t_n. Then there exists $\epsilon > 0$ such that $t_n \sim T_\infty^c(K_n,\epsilon)$. By Proposition 2.1, one has $t_n \sim 2T_2^c(K_n,\epsilon^{1/2})$ and by Theorem 2.4, we get the desired property.

To prove (4)⇒(1), we assume that $\lambda_n T_\infty^c(K_n,\epsilon) \to \infty$. By Proposition 2.1, one has $\lambda_n T_2(K_n,\epsilon^{1/2}) \to \infty$ and then, by Theorem 2.4, the family $\mathcal{F}_c$ has a $(T_2^c(K_n,\epsilon^{1/2}), \lambda_n^{-1})$ ℓ^2-cutoff, or equivalently(by Proposition 1.10(4)),

$$|T_2^c(K_n,\epsilon^{1/2}) - T_2^c(K_n,\eta)| = O_\eta(\lambda_n^{-1}), \quad \forall \eta > 0.$$

Again, by Proposition 2.1, the above identity is equivalent to

$$|T_\infty^c(K_n,\epsilon) - T_\infty^c(K_n,\eta^2)| = O_\eta(\lambda_n^{-1}), \quad \forall \eta > 0.$$

Hence $\mathcal{F}_c$ has a $(T_\infty^c(K_n,\epsilon), \lambda_n^{-1})$ ℓ^∞-cutoff.

The discrete-time case can be proved in a similar way with almost the same statements as above. □

Since the reversibility of K also makes $T_p(K,\epsilon)$ and $T_p(K^*,\epsilon)$ equal, we have an implication similar to Corollary 2.4.

Corollary 2.6. *Let $\mathcal{F}=\{(\mathcal{X}_n,K_n,\pi_n)\}_1^\infty$ be a family of reversible and irreducible Markov chains and assume that $\pi_{n,*}\to 0$.*

(1) For $1<p<q\le\infty$,

$$\mathcal{F}_c \text{ presents a } \ell^p\text{-cutoff} \quad \Leftrightarrow \quad \mathcal{F}_c \text{ presents a } \ell^q\text{-cutoff}.$$

(2) If K_n is aperiodic and there exist $1<r<\infty$ and $\epsilon>0$ such that $T_r^d(K_n,\epsilon)\to\infty$, then for $1<p<q\le\infty$,

$$\mathcal{F}_d \text{ presents a } \ell^p\text{-cutoff} \quad \Leftrightarrow \quad \mathcal{F}_d \text{ presents a } \ell^q\text{-cutoff}.$$

In particular, if $\mathcal{F}_c$ presents a ℓ^p-cutoff for some $p>1$, then, for $1<q\le\infty$, the ℓ^q-critical time and the ℓ^p-critical time are of the same order. This also holds for discrete-time cases if one assumes further that $T_r^d(K_n,\epsilon)\to\infty$ for some $1<r<\infty$ and $\epsilon>0$.

Proof. Using Corollary 2.1 and Theorem 2.6. □

By the above fact, one may obtain a similar result as in Corollary 2.5.

Corollary 2.7. *Let $\mathcal{F}=\{(\mathcal{X}_n,K_n,\pi_n)\}_1^\infty$ be a family of reversible and irreducible Markov chains. For $n\ge 1$, let λ_n and μ_n be the spectral gap and the second largest singular value of K_n with multiplicities m_n and m_n' respectively. Then the conclusion in Corollary 2.5 holds for all $1<p\le\infty$.*

Example 2.6. **(The ℓ^∞-cutoff for the random transposition in discrete-time cases.)** Let $\mathcal{F}=\{(\mathcal{X}_n,K_n,\pi_n)\}_1^\infty$ be the family of random transpositions. Recall that we have previously proved that $\mathcal{F}_d$ presents a ℓ^p-cutoff with critical time tending to infinity. By Corollary 2.6 and Theorem 2.6, the family $\mathcal{F}_d$ also has a $(2T_2^d(K_n,\epsilon),n)$ ℓ^∞-cutoff. It is known in [21] that $|T_2^d(K_n,\epsilon)-\frac{1}{2}n\log n|=O(n)$. Hence, $\mathcal{F}_d$ has a $(n\log n,n)$ ℓ^∞-cutoff.

Note that the equivalence given in Corollary 2.6 is not necessary true for $p=1$. In fact, one direction must hold and is given in the following theorem, and a counterexample for the other direction is presented in section 4.2.

Theorem 2.7. *Let $\mathcal{F}=\{(\mathcal{X}_n,K_n,\pi_n)\}_1^\infty$ be a family of normal and irreducible Markov chains.*

(1) If $\mathcal{F}_c$ presents a weak total variation cutoff, then, for $1<p<\infty$, the family $\mathcal{F}_c$ presents a ℓ^p-cutoff.

(2) If $\mathcal{F}$ contains ergodic Markov kernels and $\mathcal{F}_d$ presents a weak total variation cutoff with critical time tending to infinity, then, for $1<p<\infty$, $\mathcal{F}_d$ presents a ℓ^p-cutoff.

Furthermore, if K_n is reversible for $n\geq 1$, then the above also holds for $p=\infty$.

Proof. The proof is done by applying Theorem 2.4, Theorem 2.5, Theorem 2.6, Proposition 2.1 and the following inequalities.

$$\max_{x\in\mathcal{X}_n}\|H_{n,t}(x,\cdot)-\pi_n\|_{\mathrm{TV}}=\frac{1}{2}\|H_{n,t}-\pi_n\|_{\infty\to\infty}\geq\frac{1}{2}e^{-\lambda_n t}$$

and

$$\max_{x\in\mathcal{X}_n}\|K_n^m(x,\cdot)-\pi_n\|_{\mathrm{TV}}=\frac{1}{2}\|K_n^m-\pi_n\|_{\infty\to\infty}\geq\frac{1}{2}e^{-m(-\log\mu_n)}$$

where λ_n and μ_n are the spectral gap and the second largest singular value of K_n. □

Remark 2.7. Assume that $\mathcal{F}$ is a family containing normal and irreducible Markov kernels. If $\mathcal{F}_c$ presents a weak ℓ^1-cutoff, then one has

$$\lambda_n^{-1}=o(T_1^c(K_n,\epsilon)),\quad \text{for } \epsilon \text{ small enough,}$$

and if $\mathcal{F}_d$ presents a weak ℓ^1-cutoff with critical time tending to infinity, then

$$b_n^{-1} = o(T_1^d(K_n, \epsilon)), \quad \text{for } \epsilon \text{ small enough.}$$

We end this subsection by giving a complementary result for Proposition 2.5 if a family consists of normal Markov chains.

Proposition 2.6. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of normal and irreducible Markov chains and λ_n be the spectral gap of K_n. Assume that, for $1 < p \le \infty$, $\mathcal{F}_c$ has a (t_n, b_n) ℓ^p-cutoff and a (s_n, c_n) ℓ^1-cutoff with $t_n \sim s_n$ and $|t_n - s_n| = o(t_n)$. Then, for $1 < r < p$, $\mathcal{F}_c$ has a (t_n, d_n) ℓ^r-cutoff, where*

$$d_n = \max\{b_n, c_n, |t_n - s_n|, \lambda_n^{-1}\}.$$

The above also holds for discrete-time cases if one assume further that $t_n \ge \infty$ and set

$$d_n = \max\{b_n, c_n, |t_n - s_n|, b_n^{-1}\},$$

where $b_n = \min\{1, -\log \mu_n\}$ and μ_n is the second largest singular value of K_n.

2.1.6 The ℓ^p-cutoff for normal random walks on symmetric groups

For an illustration of the theorems in the previous subsection, we consider a family $\mathcal{F} = \{(S_n, K_n, \pi_n)\}_1^\infty$ of irreducible Markov chains, where S_n is the symmetric group of degree n and $\pi \equiv 1/n!$. Suppose that, for $n \ge 1$, the Markov kernel K_n is given by

$$K_n(x, y) = p_n(x^{-1}y), \quad \forall x, y \in S_n,$$

where p_n is a probability measure on S_n. In this setting, by Lemma 2.2, the ℓ^p-distances $\|k_n^m(x, \cdot) - 1\|_p$ and $\|h_{n,t}(x, \cdot) - 1\|_p$ are independent of the initial state x.

If the reversibility of K_n is assumed further, then, by Corollary 2.6, for $1 < p \leq \infty$, the existence of the ℓ^p-cutoff is equivalent to that of the ℓ^2-cutoff.

For any representation ρ of S_n, we define the Fourier transform of p_n at ρ as follows.

$$\widehat{p_n}(\rho) = \sum_{\sigma \in S_n} p_n(\sigma)\rho(\sigma).$$

Let e be the identity of S_n and $\{\rho_{n,0}, \rho_{n,1}, ...\}$ be the set of all irreducible representations of S_n, where $\rho_{n,0}$ is the trivial representation, that is, $\rho_{n,0} \equiv 1$. Then, the ℓ^2-distance can be expressed by

$$\|k_n^m(e, \cdot) - 1\|_2^2 = \sum_{i \geq 1} d_{\rho_{n,i}} \mathrm{tr} \left(\widehat{p_n^{*m}}(\rho_{n,i}) \widehat{p_n^{*m}}(\rho_{n,i})^* \right) \tag{2.22}$$

and

$$\|h_{n,t}(e, \cdot) - 1\|_2^2 = \sum_{i \geq 1} d_{\rho_{n,i}} \mathrm{tr} \left(\widehat{H_{n,t}^e}(\rho_{n,i}) \widehat{H_{n,t}^e}(\rho_{n,i})^* \right), \tag{2.23}$$

where p_n^{*m} denotes the convolution $p_n * p_n * \cdots * p_n$ of p_n for m times, tr is the trace of matrices and d_{ρ_i} is the dimension of ρ_i. A proof of the above two identities can be found in [9].

We now assume that K_n is normal, that is,

$$\sum_z p_n(x^{-1}z)p_n(y^{-1}z) = \sum_z p_n(z^{-1}x)p_n(z^{-1}y), \quad \forall x, y \in S_n.$$

By the normality of K_n, one may easily check that for any irreducible representation ρ of S_n,

$$\widehat{p_n}(\rho)\widehat{p_n}(\rho)^* = \widehat{p_n}(\rho)^*\widehat{p_n}(\rho).$$

This means that $\widehat{p_n}(\rho)$ is a normal matrix. Let $D_\rho = \{\beta_{\rho,1}, ..., \beta_{\rho,d_\rho}\}$ be the spectrum of $\widehat{p_n}(\rho)$. In addition to the following fact,

$$\widehat{p_n^{*m}}(\rho) = \left(\widehat{p_n}(\rho)\right)^m,$$

the matrices $\widehat{p_n^{*m}}(\rho)\widehat{p_n^{*m}}(\rho)^*$ and $\widehat{H_{n,t}^e}(\rho_{n,i})\widehat{H_{n,t}^e}(\rho_{n,i})^*$ can be diagonalized with respective diagonals, regardless of the order,

$$[|\beta_{\rho,1}|^{2m}, ..., |\beta_{\rho,d_\rho}|^{2m}] \quad \text{and} \quad [e^{-2t(1-\text{Re}\beta_{\rho,1})}, ..., e^{-2t(1-\text{Re}\beta_{\rho,d_\rho})}].$$

For convenience, we let $\beta_{n,i,j}$ denote $\beta_{\rho_{n,i},j}$ for $1 \leq j \leq d_{\rho_{n,i}}$, $i \geq 0$, and set $R(S_n)$ be the set of all irreducible representations of S_n. By the discussion in the previous paragraph, (2.22) and (2.23) can be rewritten in the following way.

$$\|k_n^m(e,\cdot) - 1\|_2^2 = \sum_{i\geq 1} \sum_{j=1}^{d_{\rho_{n,i}}} d_{\rho_{n,i}} |\beta_{n,i,j}|^{2m}, \quad \forall m \geq 0 \tag{2.24}$$

and

$$\|h_{n,t}(e,\cdot) - 1\|_2^2 = \sum_{i\geq 1} \sum_{j=1}^{d_{\rho_{n,i}}} d_{\rho_{n,i}} e^{-2t(1-\text{Re}\beta_{n,i,j})}, \quad \forall t \geq 0. \tag{2.25}$$

Let us consider the alternating representation of S_n. Recall that in algebra a permutation can be decomposed into a product of either an even number or an odd number of transpositions, and it is called respectively even or odd. Then the alternating representation of S_n is a one-dimensional representation $(\text{sgn}, \mathbb{C})$, where

$$\forall z \in \mathbb{C}, \quad \text{sgn}(\sigma)(z) = \begin{cases} z & \text{if } \sigma \text{ is even} \\ -z & \text{if } \sigma \text{ is odd} \end{cases}.$$

Let p be a probability measure on S_n, K be a Markov kernel on S_n defined by

$$K(x,y) = p(x^{-1}y), \quad \forall x, y \in S_n,$$

and λ and μ be the spectral gap and the second largest singular value of K. Consider the following sets.

$$R(\lambda) = \{\rho \in R(S_n) : 1 - \lambda \in D_\rho^{(1)}, \}$$

and

$$R(\mu) = \{\rho \in R(S_n) : \mu_n \in D_\rho^{(2)}\},$$

where $D_\rho = \{\beta_{\rho,1}, ..., \beta_{\rho,d_\rho}\}$ is the spectrum of $\widehat{p}(\rho)$ and $D_\rho^{(1)}$ and $D_\rho^{(2)}$ contain respectively the real parts and the absolute values of elements in D_ρ. In this setting, we have the following theorem.

Theorem 2.8. *Let $\mathcal{F} = \{(S_n, K_n, \pi_n)\}_1^\infty$ be a family of normal and irreducible Markov chains, where S_n is the symmetric group of degree n and $\pi_n \equiv 1/n!$. For $n \geq 1$, let p_n be a probability measure on S_n and K_n be given by*

$$K_n(x, y) = p_n(x^{-1}y), \quad \forall x, y \in S_n.$$

Let λ_n and μ_n be the spectral gap and the second largest singular value of K_n. Then:

(1) Assume that, for $n \geq 1$, $R(\lambda_n)$ contains irreducible representations other than the alternating one. Then the family $\mathcal{F}_c$ presents a ℓ^p-cutoff for all $1 < p < \infty$.

(2) Assume that $\inf_n \mu_n > 0$ and, for $n \geq 1$, K_n is aperiodic and $R(\mu_n)$ contains irreducible representations other than the alternating one. Then the family $\mathcal{F}_d$ presents a ℓ^p-cutoff for $1 < p < \infty$.

In particular, if K_n is assumed further reversible, then the above conclusions also holds for $p = \infty$.

Remark 2.8. A well-known result stated in [30, Proposition 2.3] is the following: Let p be a probability measure on a finite group $\mathcal{X}$ with support E and K be a Markov kernel given by

$$K(x, y) = p(x^{-1}y), \quad \forall x, y \in \mathcal{X}.$$

Assume that K is irreducible. Then K is aperiodic if and only if E is not contained in any coset of a proper normal subgroup of $\mathcal{X}$. Note also that, for $n \geq 5$, the alternating group A_n is the only proper normal subgroup of S_n. Hence, if $\mathcal{X} = S_n$ with $n \geq 5$, then the Markov kernel K is aperiodic if and only if

$$\sum_{\sigma \in S_n} p(\sigma)\mathrm{sgn}(\sigma) \in (-1, 1).$$

Moreover, as A_n is simple, if $\mathcal{X} = A_n$ and K is irreducible, then K is aperiodic.

The following corollary is an immediate consequence from Theorem 2.8 and Remark 2.8.

Corollary 2.8. *Let $\mathcal{F} = \{(S_n, K_n, \pi_n)\}_1^\infty$ be the family of normal Markov chains in Theorem 2.8 and λ_n and μ_n be the spectral gap and the second largest singular value of K_n. Set*

$$\beta_n = \sum_{\sigma \in S_n} p_n(\sigma)\mathrm{sgn}(\sigma).$$

Then:

(1) *Assume that*

$$\mathrm{Re}\beta_n < 1 - \lambda_n, \quad \forall n \geq 1. \tag{2.26}$$

Then $\mathcal{F}_c$ presents a ℓ^p-cutoff for all $1 < p \leq \infty$.

(2) *Assume that $\inf_n \mu_n > 0$ and*

$$|\beta_n| < \mu_n, \quad \forall n \geq 1. \tag{2.27}$$

Then $\mathcal{F}_d$ presents a ℓ^p-cutoff for all $1 < p \leq \infty$.

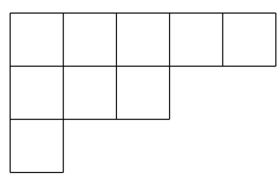
To prove Theorem 2.8, we need the following lemma.

Lemma 2.6. *Let S_n be the symmetric group of degree n and ρ be an irreducible representation of S_n with degree d_ρ. Assume that ρ is neither a trivial representation nor an alternating representation. Then $d_\rho \geq n-1$ for $n \geq 5$.*

Remark 2.9. (1) By Lemma 2.6, one can easily see that, in Theorem 2.8, the assumption that $R(\lambda_n)$(resp. $R(\mu_n)$) contains irreducible representations other than the alternating one is equivalent to the requirement that the multiplicity of λ_n(resp. μ_n) is at least 2.

(2) The fact in (1) implies that the inequalities (2.26) and (2.27) in Corollary 2.8 are stronger than the assumptions given in Theorem 2.8.

Proof of Lemma 2.6. It is well-known that there is a one to one correspondence between the set of all irreducible representation and the Young diagrams. The following is a Young diagram of S_9



For a short hand, we write $(5, 3, 1)$ for the above diagram, which indicates the numbers of boxes in each row from the top. Note that the trivial representation and the alternating representation of S_n are associated respectively to (n) and $(1, 1, ..., 1)$.

In the above setting, the dimension of ρ is the number of ways putting the integers $1, 2, ..., 9$ into each box such that the numbers in each row and column is increasing. For example, one can fill in the tableau in the following way.

1	2	4	5	8
3	6	7		
9				

One can see that flipping a diagram by changing the rows into columns does not change the dimension. For example, irreducible representations associated to $(5,3,1)$ and $(3,2,2,1,1)$ have the same dimension.

We first prove this lemma for rectangular Young diagrams. Note that the irreducible representation of S_5 which has a rectangular Young diagram is either the trivial one or the alternating one. For $n \geq 6$, let $(m, m, ..., m)$, where m repeats for l times, be a rectangular Young diagram with $m \geq 2$ and $l \geq 2$. Then, by Hook formula, the dimension of the irreducible representation is

$$\frac{(ml)!}{\prod_{k=1}^{l}[k \times (k+1) \times \cdots \times (k+m-1)]}. \tag{2.28}$$

Without loss of generality, one may assume further that $m \geq l$. In this case, the assumption $n \geq 6$ implies that $m \geq 3$. Consider first the case $l = 2$. A simple computation show that $ml - m + 1 \geq l + 3$ and, by (2.28, the dimension bounded from below by

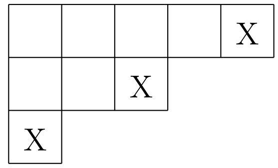
$$\frac{(ml-m+1)(ml-m+2)\cdots(ml)}{(l+1)(l+2)\cdots(l+m)} \geq \frac{ml(ml-1)}{l(l+1)} \geq ml - 1.$$

For $l \geq 3$, the dimension is bounded from below by

$$\frac{2(ml-m+1)(ml-m+2)\cdots(ml)}{(l+1)(l+2)\cdots(l+m)} \geq \frac{2ml(ml-1)}{l(l+1)} \geq ml - 1.$$

This proves the lemma for rectangular Young diagrams.

For the general case, we prove this lemma by induction. Consider first the case $n = 5$. The diagrams to be checked are $(4,1)$, $(3,2)$ and $(3,1,1)$, and they have respective dimensions 4, 5 and 6. Now assume that this lemma is true for $n \geq 5$. Let $D = (d_1, ..., d_k)$ be a diagram, where $d_i \geq d_{i+1}$ for $1 \leq i < k$ and $\sum_i d_i = n+1$. The case that D is a rectangle has be proved in the previous paragraph. We assume further that $d_i > d_j$ for some $i < j$. Since D is not rectangular, there are at least two boxes for n to fill in. For example, in the following diagram,



one can put the integer $n+1$ in any one the three boxes marked by X. Let D' be the diagram(of n boxes) obtained by removing one marked box from D. Then the inductive assumption implies that the irreducible representation associated to D' has at least dimension $n-1$. Hence, the irreducible representation associated to D has dimension at least $2(n-1) > n$. □

Remark 2.10. Note that Lemma 2.6 is not true if $n = 4$, since the irreducible representation associated to the diagram $(2,2)$ has dimension 2.

Proof of Theorem 2.8. By comparing the identities in (2.24) and (2.25) with those in Lemma 3.1, one can see that the eigenvalue $\beta_{n,i,j}$ has multiplicity at least $d_{\rho_{n,i}}$ for $1 \le j \le d_{\rho_{n,i}}$. Let m_n and m'_n be the multiplicities of the spectral gap and the second largest singular value of K_n. Then, by Lemma 2.6, we have $m_n \ge n-1$ and $m'_n \ge n-1$ for $n \ge 5$. With these facts, one can easily prove this theorem by applying Corollary 2.5 and 2.7. □

Example 2.7. **(The ℓ^p-cutoff for the random inversion with $1 < p \le \infty$.)** This model is first introduced by Durrett in [23] for the study of chromosome rearrangements and called an n reversal chain in that paper. For a description of the random inversion, let n be a positive integer and S_n be the symmetric group of degree n. The transition kernel is driven by the probability measure p_n on S_n defined by

$$p_n(id) = \frac{2}{n+1}, \quad p_n(c_{i,j}) = \frac{2}{n(n+1)}, \quad \forall 1 \le i < j \le n,$$

where

$$c_{i,j} = (i,j)(i+1,j-1)(i+2,j-2)\cdots(\tfrac{\lfloor i+j\rfloor}{2},\tfrac{\lceil i+j\rceil}{2}).$$

From the view point of biology, the arrangements of numbers $1,...,n$ denote the arrangements of genomes in a chromosome of length n. In that setting, the inversion $c_{i,j}$ represents a rearrangement of a chromosome which reverses the order of genomes in the segment ranging from the ith position to the jth one.

From the above definition, one can see that $c_{i,i+4j}$ and $c_{i,i+4j+3}$ are even permutations and $c_{i,i+4j+1}$ and $c_{i,i+4j+2}$ are odd permutations. This implies that, for $0 \le j \le \lfloor n/4\rfloor - 1$,

$$\sum_{1\le i\le n-4j} p_n(c_{i,i+4j}) - \sum_{1\le i\le n-4j-1} p_n(c_{i,i+4j+1}) - \sum_{1\le i\le n-4j-2} p_n(c_{i,i+4j+2}) + \sum_{1\le i\le 4j-3} p_n(c_{i,i+4j+3}) = 0.$$

A simple computation then shows

$$\beta_n = \sum_{\sigma\in S_n} p_n(\sigma)\mathrm{sgn}(\sigma) = \frac{2}{n+1} + \frac{2}{n(n+1)}\delta_{\{1,2\}}(n \bmod 4).$$

To apply Corollary 2.8, we need to determine the spectral gap λ_n and the second largest singular value μ_n of the n reversal chain. For $n \ge 1$ and $1 \le i < n$, let ϕ_i be a function on S_n defined by

$$\phi_{n,i}(\sigma) = \begin{cases} n-2 & \text{if } |\sigma(i)-\sigma(i+1)| = 1 \\ -2 & \text{otherwise} \end{cases},$$

where $\sigma(i)$ denotes the position of the card whose face value is i. Let K_n be the transition kernel of the n reversal chain. By the above definition, it has been proved in [23] that

$$\forall 1 \le i \le n-1, \quad K_n\phi_{n,i} = \frac{n-1}{n+1}\phi_{n,i}.$$

Hence, we have for $n \geq 4$,

$$\lambda_n \leq \frac{2}{n+1} < 1 - \beta_n, \quad \mu_n \geq 1 - \frac{2}{n+1} > \frac{1}{2} \geq \beta_n.$$

Let $\mathcal{F} = \{(S_n, K_n, \pi_n)\}_1^\infty$ be the family of the random inversion. By Corollary 2.8, both families $\mathcal{F}_c$ and $\mathcal{F}_d$ present a ℓ^p-cutoff for $1 < p \leq \infty$. It has been proved by Durrett in [23] using the comparison technique that the ℓ^1 and ℓ^2 mixing times are both of order $n \log n$. By Corollary 2.6, the ℓ^p-critical time is of order $n \log n$ for $1 < p \leq \infty$.

Example 2.8. **(The ℓ^p-cutoff for the random insertion with $1 < p \leq \infty$)** Let $\mathcal{F}$ be the family of the random insertion introduced in Example 2.5. By Example 2.2 and (2.20), one may choose $c_2 > c_1 > 0$ such that

$$\frac{c_1}{n} \leq 1 - \mu_n \leq \lambda_n \leq \frac{c_2}{n}, \quad \forall n \geq 1.$$

To compute β defined in Theorem 2.8, we need to explicitly describe the probability measure p_n. In details, let

$$c_{i,j} = (j, j-1, ..., j+i, i), \quad \forall i \leq j,$$

and $c_{i,j} = c_{j,i}^{-1}$ if $i > j$. Then one has

$$p_n(c_{i,j}) = n^{-2}, \quad \forall 1 \leq i, j \leq n,$$

and a simple computation shows

$$\beta = \sum_{1 \leq i,j \leq n} p_n(c_{i,j}) \mathrm{sgn}(c_{i,j}) = \sum_{1 \leq i,j \leq n} \frac{(-1)^{i+j}}{n^2} = \frac{1 + (-1)^{n+1}}{2n^2}.$$

This implies that (2.26) and (2.27) are satisfied for large n. Hence, for $1 < p \leq \infty$, the families $\mathcal{F}_c$ and $\mathcal{F}_d$ present ℓ^p-cutoffs.

As one can see from Theorem 2.8, if the alternating representation is not the only irreducible representation contributing to the multiplicity of the spectral gap or, in discrete-time cases, the second largest singular value of the Markov kernel, then the ℓ^p-cutoff exists for $1 < p \leq \infty$. Hence, if one can get rid of the alternating representation from the summation in the ℓ^2-distance(see (2.22) and (2.23)), then there is almost no requirement for the ℓ^p-cutoff for all $1 < p \leq \infty$.

In the following, we consider a random walk on the alternating group A_n of degree n, which is the normal subgroup containing all even permutations in S_n. Let p be a probability measure on A_n, K be a Markov kernel defined by $K(x, y) = p(x^{-1}y)$ for $x, y \in A_n$, and $\pi \equiv 1/|A_n|$. By definition, one has

$$\|k(e, \cdot) - 1\|_2^2 = \|p/\pi\|_2^2 - 1 = |A_n| \sum_{x \in A_n} p(x)^2 - 1. \tag{2.29}$$

Note that one may also consider p as a probability measure on S_n whose support generates A_n. Let $\{\rho_{n,i} : i \geq 0\}$ be the set of all irreducible representations of S_n, where $\rho_{n,0}$ is the trivial representation and $\rho_{n,1}$ is the alternating representation. Since $\rho_{n,1}(\sigma) = 1$ if σ is even and $\rho_{n,1}(\sigma) = -1$ if σ is odd, we have $\widehat{p}(\rho_{n,1}) = 1$. Then, by the representation theory and (2.29), one has

$$\begin{aligned} \|k(e, \cdot) - 1\|_2^2 &= \frac{|A_n|}{|S_n|} \left(|S_n| \sum_{x \in S_n} p(x)^2 \right) - 1 \\ &= \frac{1}{2} \sum_{i \geq 2} d_{\rho_{n,i}} \mathrm{tr} \left(\widehat{p}(\rho_{n,i}) \widehat{p}(\rho_{n,i})^* \right). \end{aligned}$$

As before, for any irreducible representation ρ, let $\{\beta_{\rho,1}, ..., \beta_{\rho,d_\rho}\}$ be all the eigenvalues of the matrix $\widehat{p}(\rho)$. If p is assumed further normal on S_n, then the above identity can be rewritten as

$$\|k(e, \cdot) - 1\|_2^2 = \frac{1}{2} \sum_{i \geq 2} \sum_{j=1}^{d_{\rho_{n,i}}} d_{\rho_{n,i}} |\beta_{n,i,j}|^2, \tag{2.30}$$

where $\beta_{n,i,j} = \beta_{\rho_{n,i},j}$. By the above identity, we have the following theorem for random walks on alternating groups.

Theorem 2.9. *Let $\mathcal{F} = \{(A_n, K_n, \pi_n)\}_1^\infty$ be a family of irreducible Markov chains, where A_n is the alternating group of degree n and $\pi_n = 2/n!$. For $n \geq 1$, let p_n be a probability measure on A_n and $K_n(x,y) = p_n(x^{-1}y)$ for $x, y \in A_n$. Assume that p_n, as a function defined on S_n, is normal for $n \geq 1$. That is,*

$$\sum_{z \in S_n} p_n(xz)p_n(yz) = \sum_{z \in S_n} p_n(zx)p_n(zy), \quad \forall x, y \in S_n.$$

Then:

(1) For $1 < p < \infty$, the family $\mathcal{F}_c$ presents a ℓ^p-cutoff.

(2) Let μ_n be the second largest singular value of K_n. If $\inf_n \mu_n > 0$, then, for $1 < p < \infty$, the family $\mathcal{F}_d$ presents a ℓ^p-cutoff.

In particular, if K_n is assumed further reversible, then the above also holds for $p = \infty$.

Proof. Let $\{\rho_{n,i} : i \geq 0\}$ be the set of all irreducible representation of S_n, where $\rho_{n,0}$ is the trivial representation and $\rho_{n,1}$ is the alternating representation. By (2.30), we have

$$\|k_n^m(e,\cdot) - 1\|_2^2 = \sum_{i \geq 2} \sum_{j=1}^{d_{\rho_{n,i}}} d_{\rho_{n,i}} |\beta_{n,i,j}|^{2m}, \quad \text{for } m \geq 0,$$

and

$$\|h_{n,t}(e,\cdot) - 1\|_2^2 = \sum_{i \geq 2} \sum_{j=1}^{d_{\rho_{n,i}}} d_{\rho_{n,i}} e^{-2t(1-\beta_{n,i,j})}, \quad \text{for } t \geq 0.$$

In discrete-time cases, Remark 2.8 implies that the Markov kernel K_n is aperiodic and then $\beta_{n,i,j} \in (-1, 1)$ for $1 \leq j \leq d_{\rho_{n,i}}$ and $i \geq 2$. This is equivalent to saying that the second largest singular value of K_n is less than 1. By the above facts, the theorem is then proved by Corollary 2.5, 2.7 and Lemma 2.6. □

2.2 Comparison between continuous-time and discrete-time ℓ^p-cutoffs

In this section, we discuss how the mixing time $T_p^d(K,\epsilon)$ affects $T_p^c(K,\epsilon)$. To illustrate the difference between them, we first consider the following example.

Example 2.9. Let $\mathcal{F}=\{(\mathcal{X}_n,K_n,\pi_n)\}_1^\infty$ be a family of Markov chains, where $|\mathcal{X}_n|\ge 2$, π_n is a positive probability measure on $\mathcal{X}_n$ and $K_n(x,y)=\pi_n(y)$ for all $x,y\in\mathcal{X}_n$. It is obvious that any Markov chain in $\mathcal{F}$ are perfectly mixed once the transition starts, and hence $T_p^d(K_n,\epsilon)\le 1$ for all $n\ge 1$.

For continuous-time cases, a simple computation shows that

$$H_{n,t}(x,y)=1+e^{-t}(\delta_x(y)-1),$$

where $\delta_x(\cdot)$ is a function taking value 1 at x and 0 otherwise. This implies

$$\max_{x\in\mathcal{X}_n}\|h_{n,t}^x-1\|_p=\begin{cases} e^{-t}\left(\frac{\left(\pi_{n,*}^{-1}-1\right)^p+\left(\pi_{n,*}^{-1}-1\right)}{\pi_{n,*}^{-1}}\right)^{1/p} & \text{for } 1\le p<\infty \\ e^{-t}\left(\pi_{n,*}^{-1}-1\right) & \text{for } p=\infty \end{cases},$$

where $\pi_{n,*}=\min_{x\in\mathcal{X}_n}\pi_n(x)$. Hence the ℓ^p-mixing time is given by

$$T_p^c(K_n,\epsilon)=\max\{f_p(\pi_{n,*}^{-1})-\log\epsilon,0\},$$

where $f_p(s)=\frac{1}{p}\log\left(\frac{(s-1)^p+s-1}{s}\right)$ for $1\le p<\infty$ and $f_\infty(s)=\log(s-1)$.

Assume that $\pi_{n,*}\to 0$, then for $\epsilon>0$,

$$T_p^c(K_n,\epsilon)=\frac{p-1}{p}\log\left(\pi_{n,*}^{-1}\right)-\log\epsilon+o_p(1)\quad\forall 1\le p\le\infty,$$

which implies

$$|T_p^c(K_n,\epsilon)-T_p^c(K_n,\delta)|=|\log\epsilon-\log\delta|+o_p(1)=o\left(\log\left(\pi_{n,*}^{-1}\right)\right).$$

Hence, by Proposition 1.10, the family $\mathcal{F}_c$ presents an optimal $\left(\frac{p-1}{p}\log\left(\pi_{n,*}^{-1}\right),1\right)$ ℓ^p-cutoff for $1<p\le\infty$. However, one can prove that $\mathcal{F}_c$ does not have a ℓ^1-cutoff.

By this example, one can find that if the discrete-time Markov chain mixes fast enough, then the mixing time of the continuous Markov process depends mostly on the stationary distribution, not the transition matrix. The same idea can also be applied to Markov chains which mix slow enough. In the following subsections, we discuss the relations between $T_p^c(K,\epsilon)$ and $T_p^d(K,\epsilon)$.

2.2.1 Discrete-time Markov chains with small ℓ^p-mixing time for $p > 1$

In this subsection, we are concerned with the case where the family $\mathcal{F}_d$ consists Markov chains mixing fast(compared with the quantity $\pi_{n,*} = \min_{x\in\mathcal{X}_n}\pi_n(x)$). Before stating the theorems, we first make some observations on the continuous-time semigroup H_t.

Let $(\mathcal{X},K,\pi)$ be an irreducible Markov chain with stationary distribution π and $H_t = e^{-t(I-K)}$ be the continuous-time semigroup associated to K. Denote h_t^x to be the density of $H_t(x,\cdot)$ with respect to π. By the triangle inequality, one has, for $1 < p \le 1$ and $m \in \mathbb{Z}$,

$$\max_{x\in\mathcal{X}} \|h_t^x - 1\|_p \ge (\pi_*)^{\frac{1-p}{p}} e^{-t} - 1 \tag{2.31}$$

and

$$\begin{aligned}\max_{x\in\mathcal{X}} \|h_t^x - 1\|_p &\le e^{-t}\sum_{j=0}^{\infty}\frac{t^j}{j!}\max_{x\in\mathcal{X}}\|k_x^j - 1\|_p \\ &\le 2(\pi_*)^{\frac{1-p}{p}}\left(e^{-t}\sum_{j=0}^{m-1}\frac{t^j}{j!}\right) + \max_{x\in\mathcal{X}}\|k_x^m - 1\|_p\end{aligned} \tag{2.32}$$

where $\pi_* = \min_{x\in\mathcal{X}}\pi(x)$ and $\frac{p-1}{p} = 1$ if $p = \infty$. Note that the last inequality in the above is implied by the fact that for any $|\mathcal{X}| \times |\mathcal{X}|$ stochastic matrix A,

$$\begin{aligned}
\|A(x,\cdot)/\pi(\cdot) - 1\|_p &= \left(\sum_{x\in\mathcal{X}} |A(x,y) - \pi(y)|^p (\pi(y))^{1-p}\right)^{1/p} \\
&\le \left(\sum_{x\in\mathcal{X}} |A(x,y) - \pi(y)| (\pi(y))^{1-p}\right)^{1/p} \\
&\le 2(\pi_*)^{\frac{1-p}{p}}
\end{aligned}$$

It seems that if one can control both terms $\sum_{j=0}^{m} \frac{t^j}{j!}$ and $\max_{x\in\mathcal{X}} \|k_x^m - 1\|_p$, then the ℓ^p-distance $\|h_t^x - 1\|_p$ depends only on the stationary distribution. This derives the following lemma.

Lemma 2.7. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chains, $\pi_* = \min_{x\in\mathcal{X}} \pi(x)$ and set $t_p = \frac{p-1}{p}\log(\pi_*^{-1})$ for $1 < p \le \infty$. Then, for all $\epsilon > 0$,*

$$T_p^c(K,\epsilon) \ge t_p - \log(\epsilon + 1).$$

Furthermore:

(1) If $T_p^d(K,\epsilon) < t_p$, then for $\delta \in (0,1)$,

$$T_p^c(K, 2\delta + \epsilon) \le \frac{t_p}{t_p - T_p^d(K,\epsilon)} \left(t_p + T_p^d(K,\epsilon)\log\frac{t_p}{T_p^d(K,\epsilon)} + \log\frac{1}{\delta}\right).$$

(2) It $T_p^d(K,\epsilon) > t_p$, then for $\delta \in (0, e^{t_p})$

$$T_p^c(K, \delta + \epsilon) \le \left[1 + f^{-1}\left(\frac{t_p - \log\delta}{T_p^d(K,\epsilon)}\right)\right] T_p^d(K,\epsilon),$$

where $f(r) = r - \log(1+r)$ for $r > 0$.

Proof. The first inequality can be obtained by (2.31). For the second one, let $X_1, X_2, ...$ be a sequence of i.i.d. exponential(1) random variable and $Y_n = X_1 + \cdots + X_n$. Then for $m, t > 0$,

$$\mathbb{P}\{Y_m > t\} = e^{-t} \sum_{j=0}^{m-1} \frac{t^j}{j!}.$$

By the large deviation estimate, one has

$$\mathbb{P}\{Y_m > t\} \le \exp\left\{m\gamma\left(\frac{t}{m}\right)\right\}, \tag{2.33}$$

where $\gamma(a) = 1 - a + \log a$ for $a > 1$. For a proof on this useful fact, please confer [22, Section 1.9]. With this inequality and (2.32), one has

$$\begin{aligned}\max_{x\in\mathcal{X}} \|h_t^x - 1\|_p \le& 2\exp\left\{\frac{p-1}{p}\log(\pi_*^{-1}) - t + m\left(1 + \log\frac{t}{m}\right)\right\}\\ &+ \max \|k_x^m - 1\|_p\end{aligned}$$

Let $t_p = \frac{p-1}{p}\log(\pi_*^{-1})$ and $t = t_p + cm\log(t_p/m)$ with $c > 0$. A simple computation shows

$$\begin{aligned}&t_p - t + m\left(1 + \log\frac{t}{m}\right)\\ =&m\left\{1 + (c-1)\log\frac{m}{t_p} + \log\left(1 - c\frac{m}{t_p}\log\frac{m}{t_p}\right)\right\}\\ \le&m\left\{1 - \log\frac{m}{t_p} + c\left(1 - \frac{m}{t_p}\right)\log\frac{m}{t_p}\right\}\end{aligned}$$

Letting $m = T_p^d(K,\epsilon)$ and $c = [(1 - m/t_p)\log(m/t_p)]^{-1}[\log(m/t_p) + \frac{1}{m}\log\delta - 1]$ in the above inequality implies

$$\max_{x\in\mathcal{X}} \|h_t^x - 1\|_p \le 2\delta + \epsilon.$$

Hence, for $\delta \in (0,1)$,

$$T_p^c(K, 2\delta+\epsilon) \le \frac{t_p}{t_p - m}\left(m\log\frac{t_p}{m} + t_p + \log\frac{1}{\delta}\right)$$

For the last identity, letting $r > 0$, $\epsilon > 0$ and replacing t and m with $(1+r)T_p^d(K,\epsilon)$ and $T_p^d(K,\epsilon)$ into the first term of (2.33), we get

$$t_p - t + m\left(1 + \log\frac{t}{m}\right) = t_p - m(r - \log(1+r))$$

and

$$t_p - m(r - \log(1+r) < \log\delta \quad \Leftrightarrow \quad r > f^{-1}\left(\frac{t_p - \log\delta}{m}\right).$$

□

Example 2.10. **(The ℓ^p-cutoff for the random transposition in continuous-time cases with $1 < p \le \infty$)** Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be the family of random transposition introduced in Example 2.4. It has been shown in [21] that the spectral gap λ_n is equal to $2/n$. Note that Lemma 2.7 implies that

$$\liminf_{n\to\infty} \frac{T_p^c(K_n,\epsilon)}{n\log n} \ge \frac{p-1}{p}, \quad \forall 1 < p \le \infty.$$

Thus, by Theorem 2.1, the family $\mathcal{F}_c$ presents a ℓ^p-cutoff for all $1 < p \le \infty$.

By Lemma 2.7, one can easily obtain the following theorem.

Theorem 2.10. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of ergodic Markov chains. Set $\pi_{n,*} = \min_{x\in\mathcal{X}_n} \pi_n(x)$. Assume that $\pi_{n,*} \to 0$ and, for some $1 < p \le \infty$ and $\epsilon \in (0,1)$,*

$$T_p^d(K_n,\epsilon) = o\left(\log(\pi_{n,*}^{-1})\right). \tag{2.34}$$

Then, for $1 < q \le p$, the family $\mathcal{F}_c$ has $(t_q(n), b_q(n))$ ℓ^q-cutoff, where

$$t_q(n) = \frac{q-1}{q}\log(\pi_{n,*}^{-1}), \quad b_q(n) = T_q^d(K_n,\epsilon)\log\frac{t_q(n)}{T_q^d(K_n,\epsilon)}.$$

Proof. Note that it suffices to prove only the ℓ^p-cutoff. Obviously, one has $b_p(n) = o(t_p(n))$. By Lemma 2.1, we have

$$T_p^d(K_n,\delta) = o(\log(\pi_{n,*}^{-1})) \quad \forall\delta > 0$$

and then, by Lemma 2.7,

$$|T_p^c(K_n,3\delta) - t_p(n)| \le \frac{t_p(n)}{t_p(n) - T_p^d(K_n,\delta)} \times \left\{T_p^d(K_n,\delta)\left|\log\frac{t_p(n)}{T_p^d(K_n,\delta)}\right|\right.$$
$$\left. + T_p^d(K_n,\delta) + |\log\delta| + |\log(3\delta+1)|\right\}$$

This implies that for all $\delta > 0$, there exists an integer $N = N(\delta) > 0$ such that

$$|T_p^c(K_n, 3\delta) - t_p(n)| \leq 2T_p^d(K_n, \delta) \left| \log \frac{t_p(n)}{T_p^d(K_n, \delta)} \right| \leq 2C(\delta)b_p(n),$$

where $C(\delta) = 1$ if $\delta \geq \epsilon$, and $C(\delta) = \left\lceil \frac{\log \delta}{\log \epsilon} \right\rceil$ if $0 < \delta < \epsilon$. □

Example 2.11. Recall Example 2.3. For $n \geq 1$, let $a_n > 1$, $\mathcal{X}_n = \mathbb{Z}_{a_n}^n$ and K_n be the Markov kernel on $\mathcal{X}_n$ given by (2.13). Note that the stationary distribution is given by $\pi_n \equiv a_n^{-n}$, which implies $\pi_{n,*} = a_n^{-n}$. Assume that $a_n \to \infty$. Since $T_p^d(K_n, \epsilon) \sim n = o(n \log a_n)$ for all $\epsilon > 0$ and $1 < p \leq \infty$, the family $\mathcal{F}_c$, by Theorem 2.10, presents a $\left(\frac{p-1}{p} n \log a_n, n \log\log a_n\right)$ ℓ^p-cutoff for $1 < p \leq \infty$.

Example 2.12. Recall Example 2.9. For $n \geq 1$, let π_n be a positive probability measure on $\mathcal{X}_n$ and $K_n(x,y) = \pi_n(y)$ for all $x, y \in \mathcal{X}_n$. Assume that $\pi_{n,*} \to 0$. Since $T_\infty^d(K_n, \epsilon) \sim 1 = o\left(\log(\pi_{n,*}^{-1})\right)$ for all $\epsilon > 0$, the family $\mathcal{F}_c$ presents a $\left(\frac{p-1}{p} \log(\pi_{n,*}^{-1}), \log\log(\pi_{n,*}^{-1})\right)$ ℓ^p-cutoff for $1 < p \leq \infty$.

Remark 2.11. Let $(\mathcal{X}, K, \pi)$ be an ergodic Markov chain and m be a positive integer such that

$$\zeta = \max_{x \in \mathcal{X}} K^m(x,x) > 0.$$

Note that the irreducibility of K implies $\zeta < 1$. Then for $1 < p \leq \infty$ and $\epsilon > 0$, we have

$$\max_{x \in \mathcal{X}} \|k_x^n - 1\| \geq \zeta^{\lceil n/m \rceil} (\pi^*)^{(1-p)/p} - 1,$$

where $\pi^* = \max_{x \in \mathcal{X}} \pi(x)$. This implies

$$T_p^d(K, \epsilon) \geq m \left\{ \frac{(1 - 1/p)\log(1/\pi^*) - \log(1+\epsilon)}{\log(\zeta^{-1})} - 1 + \delta_1(m) \right\},$$

where $\delta_1(x)$ is a function taking value 1 if $x = 1$ and 0 otherwise.

By the above remark, the following lemma provides a necessary condition of (2.34) for some specific Markov chains, which include random walks on finite groups and the simple random walks on finite graphs..

Lemma 2.8. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of ergodic Markov chains with*

$$\pi_{n,*} \to 0, \quad \pi_n^* = \max_{x \in \mathcal{X}_n} \pi_n(x) = O(\pi_{n,*}).$$

If (2.34) holds, then one has, for any sequence $(m_n)_1^\infty$ such that $m_n = o(\pi_{n,*}^{-1})$ and $\zeta_n = \max_{x \in \mathcal{X}_n} K_n^{m_n}(x,x) > 0$,

$$\lim_{n\to\infty} \frac{\log(\zeta_n^{-1})}{m_n} = \infty.$$

By the above lemma, the family of lazy walks $\{(\mathcal{X}_n, \frac{1}{2}(I + K_n), \pi_n)\}_1^\infty$ does not fit the requirement (2.34). In fact, the ℓ^p-mixing time is bounded by

$$T_p^d(K_n, \epsilon) \geq \frac{(\log 2)(p-1)}{p} \left(\log(\pi_{n,*}^{-1}) - \log(1+\epsilon)\right).$$

The next corollaries are applications of Theorem 2.10 with further assumptions on the transition matrices.

Corollary 2.9. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of ergodic Markov chains with $\pi_{n,*} = \min_{x \in \mathcal{X}_n} \pi_n(x) \to 0$. Assume that either $\mathcal{F}$ contains reversible chains or, for $n \geq 1$, there exists a finite group G_n acting transitively on $\mathcal{X}_n$ such that*

$$K_n(gx, gy) = K_n(x,y), \quad \forall x, y \in \mathcal{X}_n, g \in G_n.$$

If there are $1 < p \leq \infty$ and $\epsilon \in (0,1)$ such that

$$T_p^d(K_n, \epsilon) = o\left(\log(\pi_{n,*}^{-1})\right),$$

then, for $1 < q \leq \infty$, the family $\mathcal{F}_c$ has a $(t_q(n), b_q(n))$ ℓ^q-cutoff, where

$$t_q(n) = \frac{q-1}{q} \log(\pi_{n,*}^{-1}), \quad b_q(n) = T_q^d(K_n, \epsilon) \log \frac{t_q(n)}{T_q^d(K_n, \epsilon)}.$$

Proof. Proved by Theorem 2.10 and Corollary 2.1. □

Recall (2.11) in the proof of Theorem 2.2: For $1 < p < \infty$,

$$\|k_x^m - 1\|_p \leq 2^{\theta_p} \mu^{m(1-\theta_p)} \pi(x)^{(1-p)/p},$$

where $\theta_p = |1 - 2/p|$ and μ is the second largest singular value of K. Set $\pi_* = \min_{x \in \mathcal{X}} \pi(x)$. If $\mu < 1$, then the ℓ^p-mixing time is given by

$$T_p^d(K, \epsilon) \leq \left\lceil \frac{(1 - 1/p)\log(\pi_*^{-1}) + \theta_p \log 2 + \log(\epsilon^{-1})}{(1-\theta_p)\log(\mu^{-1})} \right\rceil. \tag{2.35}$$

This implies that if μ is close to 0, then the ℓ^p-mixing time is much smaller than $\log(\pi_*^{-1})$.

Corollary 2.10. *Let $\mathcal{X}_n = \{(\mathcal{X}_n, K_n, \pi_)\}_1^\infty$ be a family of ergodic Markov chains with $\pi_{n,*} \to 0$. Assume that the second largest singular value μ_n of K_n converges to 0. Then, for $1 < p \leq \infty$, $\mathcal{F}_c$ presents a $(t_p(n), b_n)$ ℓ^p-cutoff, where*

$$t_p(n) = \frac{p-1}{p} \log(\pi_{n,*}^{-1}), \quad b_n = \frac{\log(\pi_{n,*}^{-1})}{\log(\mu_n^{-1})} \log\log(\mu_n^{-1}).$$

Proof. By (2.35), one has $T_p^d(K_n, 1/2) = o\left(\log(\pi_{n,*}^{-1})\right)$ for all $1 < p < \infty$ and by Theorem 2.10, $\mathcal{F}_c$ has a $(t_p(n), b_p(n))$ ℓ^p-cutoff with

$$t_p(n) = \frac{p-1}{p} \log(\pi_{n,*}^{-1}), \quad b_p(n) = T_p^d(K_n, 1/2) \log \frac{t_p(n)}{T_p^d(K_n, 1/2)}.$$

Note that

$$T_p^d(K_n, 1/2) \sim \frac{(1-1/p)\log(\pi_{n,*}^{-1})}{(1-\theta_p)\log(\mu_n^{-1})}$$

and then

$$\log \frac{t_p(n)}{T_p^d(K_n, 1/2)} \sim \log\log(\mu_n^{-1}).$$

Hence $b_p(n) \sim \frac{1-1/p}{1-\theta_p} b_n$.

For $p = \infty$, note that (2.35) also holds for the adjoint Markov kernel. Then, by Proposition 2.1, we have $T_2^d(K_n, 1/2) \leq T_\infty^d(K_n, 1/4) \leq 2T_2^d(K_n, 1/2)$, which implies $b_\infty(n) = O(b_n)$. □

2.2.2 Discrete-time Markov chains with large ℓ^p-mixing time for $p > 1$

In this section, we deal with the case where the ℓ^p-mixing time $T_p^d(K,\epsilon)$ is large, that is, the Markov chain mixes slowly. First, we start from the following observation which says that $T_2^c(K,\epsilon)$ can't be of smaller order than $T_2^d(K,\epsilon)$ for some specific K.

Lemma 2.9. *Let $(\mathcal{X},K,\pi)$ be a reversible Markov chain and $c > 0$ be the following constant*

$$c = \frac{1}{2}\left(\min_{m\ge 0}\left\{e^{-m}\sum_{j=0}^{m}\frac{m^j}{j!}\right\}\right)^2.$$

Then for all $\epsilon > c^{-1/2}$ such that $T_p^d(K,\epsilon) \ge 1$,

$$T_2^c(K,(c\epsilon^2-1)^{1/2}) \ge T_2^d(K,\epsilon) - 1.$$

Proof. By a simple computation, one has

$$\|k_x^m - 1\|_2^2 = \|k_x^m\|_2^2 - 1 = k^{2m}(x,x) - 1,$$

and

$$\|h_t^x - 1\|_2^2 = \|h_t^x\|_2^2 - 1 = h_{2t}(x,x) - 1,$$

for all $m, t \ge 0$. Note that

$$h_{2t}(x,x) = e^{-2t}\sum_{j=0}^{\infty}\frac{(2t)^j}{j!}k^j(x,x) \ge e^{-2t}\sum_{j=0}^{m}\frac{(2t)^{2j}}{(2j)!}k^{2j}(x,x).$$

Let X, Y be independent Poisson random variables with intensity t. Then $X+Y$ is of Poisson distribution with intensity $2t$ and, for $t \ge m$,

$$2e^{-2t}\sum_{j=0}^{m}\frac{(2t)^{2j}}{(2j)!} \ge \mathbb{P}\{X+Y \le 2m\} \ge \mathbb{P}\{X \le m\}^2.$$

Hence we have, for $m \geq 0$,

$$\max_{x \in \mathcal{X}} \|h_m^x - 1\|_2^2 \geq c \left(\max_{x \in \mathcal{X}} \|k_x^m - 1\|_2^2 \right) - 1,$$

where c is the constant defined in the lemma. The desired identity is then proved by taking $m = T_p^d(K, \epsilon) - 1$ for $\epsilon > c^{-1/2}$. □

Remark 2.12. A simple observation from Lemma 2.9 is: Let $\{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of reversible and ergodic Markov chains with $\pi_{n,*} \to 0$. If $(s_n)_1^\infty$ is a sequence of positive integers such that

$$\lim_{n \to \infty} \max_{x \in \mathcal{X}_n} \|k_x^{s_n} - 1\|_2 = \infty,$$

then

$$\lim_{n \to \infty} \max_{x \in \mathcal{X}_n} \|h_{s_n - 1}^x - 1\|_2 = \infty.$$

Theorem 2.11. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of reversible and ergodic Markov chains with $\pi_{n,*} = \min_{x \in \mathcal{X}_n} \pi_n(x) \to 0$. For $n \geq 1$, let μ_n and λ_n be the second largest singular value and the spectral gap of K_n. Assume that there exist $1 \leq p \leq \infty$ and $\epsilon_0 > 0$ such that $\log(\pi_{n,*}^{-1}) = o(T_p^d(K_n, \epsilon_0))$. Then:*

(1) If $\mathcal{F}_d$ presents a ℓ^q-cutoff for some $1 < q \leq \infty$, then both $\mathcal{F}_d$ and $\mathcal{F}_c$ present a ℓ^q-cutoff for all $1 < q \leq \infty$.

(2) If $\mathcal{F}_c$ presents a ℓ^q-cutoff for some $1 < q \leq \infty$ and $\lambda_n = O(1 - \mu_n)$, then both $\mathcal{F}_d$ and $\mathcal{F}_c$ present a ℓ^q-cutoff for all $1 < q \leq \infty$.

In particular, if any one of the above two conditions holds, then for all $\delta > 0, \epsilon > 0$,

$$|T_2^c(K_n, \epsilon) - T_2^d(K_n, \delta)| = O_{\delta,\epsilon}(c_n),$$

and

$$|T^c_\infty(K_n,\epsilon)-T^d_\infty(K_n,\delta)|=O_{\delta,\epsilon}(c_n),$$

where η is any fixed positive number and

$$c_n=\max\left\{(-\log(\mu_n))^{-1},\sqrt{(\log\pi_{n,*}^{-1})T^d_2(K_n,\eta)}\right\}=o(T^d_2(K_n,\eta)).$$

Proof. By Corollary 2.6, it suffices to prove the ℓ^2-cutoff of $\mathcal{F}_c$ in case (1) and of $\mathcal{F}_d$ in case (2). Note that by Proposition 2.2, we have $\log(\pi_{n,*}^{-1})=o\left(T^d_2(K_n,\epsilon_0^{1/m_p})\right)$, where $m_p=\left\lceil\frac{2(p-1)}{p}\right\rceil$. Set $t_n=T^d_2(K_n,\epsilon_0^{1/m_p})$.

In case (1), since $\mathcal{F}_d$ presents a ℓ^2-cutoff, we have, by Proposition 1.11(3), $T^d_2(K_n,\delta)\sim t_n$ for all $\delta>0$ and, by Theorem 2.6,

$$b_n^{-1}=o(T^d_2(K_n,\epsilon))\quad\forall\delta>0.$$

Recall the fact $b_n\le 2\lambda_n$ in Lemma 2.4. Then, by Lemma 2.9, we obtain

$$\lambda_n^{-1}=o(T^c_2(K_n,\delta))\quad\forall\delta>0,$$

and hence, by Theorem 2.6, $\mathcal{F}_c$ presents a ℓ^2-cutoff.

In case (2), since $\mathcal{F}_c$ presents a ℓ^2-cutoff, one has $\lambda_n^{-1}=o(T^c_2(K_n,\delta))$ for all $\delta>0$. By Lemma 2.7(2), the fact $\log(\pi_{n,*}^{-1})=o\left(T^d_2(K_n,\epsilon_0^{1/m_p})\right)$ implies that for n large enough,

$$T^c_2(K_n,2\epsilon_0^{1/m_p})\le 2T^d_2(K_n,\epsilon_0^{1/m_p}).$$

Hence we have, by Lemma 2.4,

$$b_n^{-1}=O(\lambda_n^{-1})=o\left(T^d_2(K_n,\epsilon_0^{1/m_p})\right),$$

and, by Theorem 2.6, $\mathcal{F}_d$ presents a ℓ^2-cutoff.

For the last part, by Theorem 2.6, it suffices to prove the desired identity with any specified ϵ,δ. Let $f(r)=r-\log(1+r)$ for $r\ge 0$ and denote g as the inverse function of f. A simple computation shows

$$\lim_{s\downarrow 0}\frac{g(s)}{\sqrt{s}}=\lim_{r\downarrow 0}\frac{r}{\sqrt{f(r)}}=\sqrt{2}.$$

By Lemma 2.7, we may choose $N_1(\epsilon)>0$ such that

$$T_2^c(K_n,2\epsilon)\le\left(1+2\sqrt{\frac{\log(\pi_{n,*}^{-1})}{T_2^d(K_n,\epsilon)}}\right)T_2^d(K_n,\epsilon)\quad\forall n\ge N_1(\epsilon).$$

For the other direction, let $c>0$ be the constant defined in Lemma 2.9. Then one has, for $\epsilon>c^{-1/2}$,

$$T_2^c(K_n,2\epsilon)\ge T_2^d(K_n,\epsilon_1),$$

where $\epsilon_1=\sqrt{(4\epsilon^2+1)/c}$. This implies that we may choose $N_2(\epsilon)>0$ and $C(\epsilon)>0$ such that

$$T_2^c(K_n,2\epsilon)\ge T_2^d(K_n,\epsilon)-C(\epsilon)b_n\quad\forall n\ge N_2(\epsilon).$$

Combining both inequalities derives

$$|T_2^c(K_n,2\epsilon)-T_2^d(K_n,\epsilon)|\le C(\epsilon)b_n+2\sqrt{\log(\pi_{n,*}^{-1})T_2^d(K_n,\epsilon)}.$$

This proves the case of ℓ^2-mixing time. For the ℓ^∞-mixing time, one needs only the proved result and Proposition 2.1. □

Example 2.13. **(The ℓ^p-cutoff for adjacent transpositions with $p\in(1,\infty]$.)** An adjacent transposition is a card shuffling made by choosing two contiguous cards from a deck of n cards(there are entirely $n-1$ choices) uniformly with probability $1/n$ and doing nothing with the remaining probability, which is $1/n$. This model has been studied by many authors. In [16], Diaconis and Saloff-Coste derived an upper bound on the total variation mixing time by their comparison technique and the result was improved later by Wilson in [36] in the way that the obtained upper bound on the total variation mixing time has the correct order. It is worth noting that in [16], comparing the adjacent transposition with the random transposition though the following path

$$(i,j) = (i,i+1)\cdots(j-2,j-1)(j,j-1)\cdots(i+2,i+1)(i+1,i),$$

one obtained that the spectral gap of the adjacent transposition on a deck of n cards is bounded from below by c/n^3, where c is a universal constant. In [5], Bacher proved that the bound is of the correct order.

Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be the family of adjacent transpositions, where the index indicates the number of cards. In [5], Bacher showed that the spectral gap λ_n of K_n is equal to $\frac{2(1-\cos(\pi/n))}{n}$ with multiplicity at least $n-1$. By Lemma 2.2, the ℓ^2-distance is bounded by

$$d_{\pi_n,2}^2(K_n^m, \pi_n) \geq (n-1)\left(1 - \frac{2(1-\cos(\pi/n))}{n}\right)^{2m}, \quad \forall n, m \geq 1,$$

which is sufficient to find a constant $c > 0$ such that $T_2^d(K_n, \pi_n) \geq cn^3 \log n$ for $n \geq 1$. Since λ_n^{-1} is of order $1/n^3$, by Theorem 2.2, the family $\mathcal{F}_d$ presents a ℓ^2-cutoff. Note that $\log \pi_{n,*}^{-1} \sim n \log n = o(T_2^d(K_n, \epsilon))$. Hence, by Theorem 2.11, both $\mathcal{F}_d$ and $\mathcal{F}_c$ present a ℓ^p-cutoff for $1 < p \leq \infty$, and furthermore

$$T_2^d(K_n, \epsilon) \sim T_2^c(K_n, \epsilon), \quad \forall \epsilon > 0.$$

The ℓ^p-critical time is an open problem for $1 < p \leq \infty$.

Note that the ℓ^1-cutoff for the adjacent transposition is still open. In discrete-time cases, the exact order of the total variation mixing time is determined by Wilson in [36]. In that paper, he gave a bound on the mixing time whose coefficients are the tightest so far.

The following corollary is a simple implication of the proof of Theorem 2.11.

Corollary 2.11. *Let* $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ *be a family of reversible and ergodic Markov chains with* $\pi_{n,*} = \min_{x\in\mathcal{X}_n} \pi_n(x) \to 0$. *Assume that* $\mathcal{F}_d$ *present a* ℓ^2*-cutoff and*

$$\log(\pi_{n,*}^{-1}) = o(T_2^d(K_n, \epsilon)).$$

Then the family $\mathcal{F}_c$ *presents a* ℓ^2*-cutoff with critical time* $T_2^d(K_n, \epsilon)$.

At last, we consider the case where $T_p^d(K, \epsilon)$ is comparable to $\log(\pi_*^{-1})$. By Lemma 2.7, one may easily conclude that $T_p^c(K, \epsilon)$ is also comparable to $\log(\pi_*^{-1})$. However, the cutoff phenomenon is hard to obtain because knowing the order of mixing-time is not enough. Please refer to Proposition 1.10 and Proposition 1.11. Here, we give a result on the pre-cutoff.

Proposition 2.7. *Let* $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ *be a family of ergodic Markov chains with* $\pi_{n,*} \to 0$. *Assume that, for some* $1 < p \le \infty$, $\mathcal{F}_d$ *presents a* ℓ^p*-pre-cutoff in the following way: There exist* $\epsilon_0 > 0$ *and* $c_2 > c_1 > 0$ *satisfying either* $c_1 > 1$ *or* $c_2 < 1$ *such that*

$$c_1 \le \liminf_{n\to\infty} \frac{T_p^d(K_n, \epsilon)}{t_p(n)} \le \limsup_{n\to\infty} \frac{T_p^d(K_n, \epsilon)}{t_p(n)} \le c_2 \quad \forall \epsilon \in (0, \epsilon_0),$$

where $t_p(n) = \frac{p-1}{p} \log(\pi_{n,*}^{-1})$. *Then* $\mathcal{F}_c$ *presents a* ℓ^p*-pre-cutoff.*

Proof. We prove the case $c_1 > 1$ while the case $c_2 < 1$ is done in a similar way. In this case, for $\epsilon \in (0, \epsilon_0)$, one may choose an integer $N(\epsilon) > 0$ such that

$$c_1 t_p(n) \le T_p^d(K_n, \epsilon) \le c_2 t_p(n).$$

Then, by Lemma 2.7, we have

$$\frac{1}{2} t_p(n) \le T_p^c(K_n, 2\epsilon) \le c_3 t_p(n) \quad \forall n \ge N(\epsilon),$$

where $c_3 = \left(1 + f^{-1}(c_1^{-1})\right) c_2$. This implies that $\mathcal{F}_c$ presents a ℓ^p-pre-cutoff. □

2.3 The ℓ^p-cutoff for standard riffle shuffle with $1 < p \leq \infty$

The first two subsections provide methods to examine the ℓ^p-cutoff and some models introduced in chapter 1 are shown to have such a phenomenon. One can see from those theorems that the exact order of the ℓ^p-mixing time is sufficient to determine the ℓ^p-cutoff but, to obtain a critical time, further tricks need to be developed. Whatever the transition kernels are, if a cutoff is proved to possess by a family of Markov chains, then we should, theoretically, be given a critical time. However, it is never an easy work even the explicit distribution of a chain at any time is well understood.

In this section, we consider the well-known card shuffling, the riffle shuffle, which models the way a good card-player shuffles cards. For a detailed introduction on this model, please go to chapter 5 and the references there. A generalization of the standard riffle shuffle is given as follows. For any integer $a \geq 2$ and $n \geq 1$, an a-shuffle is a card shuffling done by cutting a deck of n cards into a piles according to a multinomial random variable. That is, the probability of cutting a deck of n cards into a piles with sizes from top to bottom $n_1, .., n_a$ is $a^{-n}\binom{n!}{n_1! \cdots n_a!}$. Then forming a deck by dropping cards one by one from the bottoms of each pile with probability proposition to its size. For example, if these a piles have sizes $n_1, ..., n_a$, then the bottom card in pack i is dropped with probability $\frac{n_i}{n_1 + \cdots + n_a}$.

For $n \geq 1$, let $\mathcal{X}_n$ be the set of all deck arrangements of n cards(which can be identified with the symmetric group S_n of n elements) and π_n be the uniform distribution on S_n. Fix an integer $a \geq 2$ and let, for $n \geq 1$, $Q_{n,a}$ be the distribution of card arrangements after one a-shuffles starting from a deck in order. This distribution is explicitly determined by Bayer and Diaconis in [6] in the following way.

Lemma 2.10. *Let $n, a \geq 1$. Then, starting for a deck of n cards in order, the probability distribution $Q_{n,a}$ of the deck arrangements after one a-shuffle is equal to*

$$Q_{n,a}(\sigma) = a^{-n}\binom{n+a-r}{n},$$

where r is the number of rising sequence in σ.

For an example on the rising sequence, suppose that $n = 9$ and

$$\sigma = 7, 1, 3, 4, 2, 6, 5, 9, 8.$$

Then the rising sequences of σ are $\{1,2\}, \{3,4,5\}, \{6\}, \{7,8\}, \{9\}$ and $r = 5$. Since $Q_{n,a}(\sigma)$ depends only the number of rising sequences of σ, we also denote $Q_{n,a}(r)$ as the quantity $a^{-n}\binom{n+a-r}{n}$.

The following is another important observation in [6].

Lemma 2.11. *In distribution, an a-shuffle followed by a b-shuffle is equivalent to an ab-shuffle.*

Combining the above two lemmas, the distribution of S_n after m a-shuffles starting from a deck in order is equal to

$$Q^m_{n,a}(\sigma) = Q_{n,a^m}(\sigma) = a^{-mn}\binom{n+a^m-r}{n},$$

where r is the number of rising sequence in σ.

To compute the ℓ^p-distance, we need to know how many permutations of r rising sequences in S_n. Tanny gave in [35] the following estimation of that quantity.

Lemma 2.12. *For $n \geq 1$ and $1 \leq r = \frac{n}{2} + h \leq n$, let R_{nh} be the number of permutations in S_n with r rising sequences. Then*

$$\frac{R_{nh}}{n!} = \frac{e^{-6h^2/n}}{\sqrt{\pi n/6}}\left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right) \quad \textit{uniformly in } h.$$

Based on the above results, we may prove the ℓ^p-cutoff for the riffle shuffle.

Theorem 2.12. *For fixed integer $a \geq 2$, let $\mathcal{F} = \{(S_n, Q_{n,a}, \pi_n)\}_1^\infty$ be a family of a-shuffles. Then, for $1 < p \leq \infty$, $\mathcal{F}_d$ presents a strongly optimal $(t_p(n), 1)$ ℓ^p-cutoff, where*

$$t_p(n) = \begin{cases} \frac{3}{2}\log_a n & \text{for } 1 < p < \infty \\ 2\log_a n & \text{for } p = \infty \end{cases},$$

and $\mathcal{F}_c$ presents a $(s_p(n), b_n)$ ℓ^p-cutoff, where

$$s_p(n) = \frac{p-1}{p}(n\log n - n), \quad b_n = (\log n)^2.$$

The following lemma is needed for the proof of Theorem 2.12.

Lemma 2.13. *For $a \geq 2$, let $Q_{n,a}$ and π_n be as in Theorem 2.12 and denote $q_{n,a}$ to be the density of $Q_{n,a}$ with respect to π_n. Then, for fixed $c > 0$, if $p \geq 2$ is an even integer, one has*

$$\forall k > c, \quad \|q_{n,n^{\frac{3}{2}}k} - 1\|_p^p = \sum_{l=2}^{p}\binom{p}{l}(-1)^{p-l}e^{\frac{l(l-1)}{24k^2}} + O_{c,p}(n^{-1/2}), \tag{2.36}$$

where $O_{c,p}$ is uniform for $k > c$. For $1 < p < 2$, one has that for $0 < k < c$,

$$\begin{aligned} \|q_{n,n^{\frac{3}{2}}k} - 1\|_p^p \geq \left(\exp\left\{\frac{1}{24k^2} + O_c\left(n^{-\frac{1}{4}}\right)\right\} - 1\right)^{p-1} \\ \times\left[\Phi\left(\frac{1}{2\sqrt{3}k}\right) - \frac{1}{2} + O\left(n^{-\frac{1}{4}}\right)\right], \end{aligned} \tag{2.37}$$

where $O_c(\cdot)$ and $O(\cdot)$ are uniform for $0 < k < c$.

Proof of Theorem 2.12. We first prove the ℓ^p-cutoff for the family $\mathcal{F}_d$ with $1 < p < \infty$. Let $p_1 \in (1, 2)$ and p_2 be a positive even number such that $p_1 < p < p_2$. Note that for $m \geq 0$ $Q_{n,a}^m = Q_{n,a^m}$. Fix $\theta \in \mathbb{R}$ and let $m = \lfloor \frac{3}{2}\log_a n + \theta \rfloor = \log_a\left(n^{3/2}c_n\right)$. Then $a^{\theta-1} \leq c_n \leq a^\theta$. Let $q_{n,a}^m$ be the density of $Q_{n,a}^m$ with respect to π_n. By

Lemma 2.13, one has

$$\|q_{n,a}^m - 1\|_p \le \|q_{n,a}^m - 1\|_{p_2} = \left\{ \sum_{l=2}^{p_2} \binom{p_2}{l} (-1)^{p_2-l} e^{\frac{l(l-1)}{24a^{\theta-1}}} + O_{\theta,p_2}\left(n^{-1/2}\right) \right\}^{1/p_2}$$

and

$$\|q_{n,a}^m - 1\|_p \ge \|q_{n,a}^m - 1\|_{p_1} \ge \left(\exp\left\{ \frac{1}{24a^{2\theta}} + O_\theta\left(n^{-\frac{1}{4}}\right) \right\} - 1 \right)^{1-1/p_1}$$
$$\times \left[\Phi\left(\frac{1}{2\sqrt{3}a^\theta} \right) - \frac{1}{2} + O\left(n^{-\frac{1}{4}}\right) \right]^{1/p_1}.$$

Then the functions $\overline{f}$ and $underlinef$ defined in Definition 1.4 satisfy

$$\overline{f}(\theta) \le \left\{ \sum_{l=2}^{p_2} \binom{p_2}{l} (-1)^{p_2-l} e^{\frac{l(l-1)}{24a^{\theta-1}}} \right\}^{1/p_2} < \infty, \quad \forall \theta \in \mathbb{R},$$

and

$$\underline{f}(\theta) \ge \left(e^{\frac{1}{24a^{2\theta}}} - 1 \right)^{1-1/p_1} \times \left\{ \Phi\left(\frac{1}{2\sqrt{3}a^\theta} \right) - \frac{1}{2} \right\}^{1/p_1} > 0, \quad \forall \theta \in \mathbb{R}.$$

This proves the strongly optimal ℓ^p-cutoff for $\mathcal{F}_d$ with $1 < p < \infty$.

For $p = \infty$, note that, by Lemma 2.10, $Q_{n,a}(r)$ is monotone decreasing in r. The following fact

$$\prod_{i=1}^k (1 + a_i) - 1 \ge 1 - \prod_{i=1}^k (1 - a_i), \quad \forall a_1, ..., a_k \ge 0,$$

then implies that

$$\|q_n^m - 1\|_\infty = Q_{n,a^m}(1) n! - 1, \quad \forall m \ge 0.$$

Hence we have, for $m \ge 0$,

$$\|q_n^m - 1\|_\infty = \prod_{i=1}^n \frac{a^m + i - 1}{a^m} - 1 = \exp\left\{ \sum_{i=1}^{n-1} \log\left(1 + \frac{i}{a^m}\right) \right\} - 1.$$

Observe that $x - \frac{x^2}{2} < \log(1+x) < x$, for $0 < x < 1$. Standard summation formulas then give

$$\sum_{i=1}^{n-1} \frac{i}{a^m} = \frac{n(n-1)}{2a^m}, \quad \sum_{i=1}^{n-1} \frac{i^2}{a^{2m}} = \frac{n(n-1)(2n-1)}{6a^{2m}}.$$

For $\theta \in \mathbb{R}$, set $m = \lfloor 2\log_a n + \theta \rfloor = 2\log_a n + \theta_n$ and let $N > 0$ be such that $\theta_n + \log_a n > 0$ for $n \geq N$. In the above setting, it is clear that $\theta - 1 \leq \theta_n \leq \theta$. By the previous computation, we get

$$\|q_n^m - 1\|_\infty \leq \exp\left\{\frac{n(n-1)}{2n^2 a^{\theta_n}}\right\} - 1 \leq \exp\left\{a^{1-\theta}\right\} - 1,$$

and

$$\|q_n^m - 1\|_\infty \geq \exp\left\{\frac{n(n-1)}{2n^2 a^{\theta_n}} + O_\theta\left(\frac{1}{n}\right)\right\} - 1 \geq \exp\left\{a^{-\theta-2} + O_\theta\left(\frac{1}{n}\right)\right\} - 1.$$

This implies that, for the ℓ^∞-distance, the functions $\underline{f}$ and $\overline{f}$ in Definition 1.4 are bounded as follows.

$$\overline{f}(\theta) \leq e^{a^{1-\theta}} - 1 < \infty, \quad \underline{f}(\theta) \geq e^{a^{-\theta-2}} - 1 > 0, \quad \forall \theta \in \mathbb{R}.$$

This proves the desired ℓ^∞-cutoff.

The ℓ^p-cutoff can be easily obtained by Theorem 2.10. □

Proof of Lemma 2.13. Note that if p is an even number, then the ℓ^p-distance can be rewritten as follows.

$$\|q_{n,a} - 1\|_p^p = \sum_{l=2}^{p} \binom{p}{l} (-1)^{p-l} d_{n,l}(a),$$

where R_{nh} is the quantity in Lemma 2.12 and

$$d_{n,l}(a) = \sum_{h=-\frac{n}{2}+1}^{\frac{n}{2}} \left[n! Q_{n,a}\left(\tfrac{n}{2} + h\right)\right]^l \frac{R_{nh}}{n!}.$$

To bound $d_{n,l}$, we need the following two identities which are modified from the proof of Proposition 1 in [6]. Let $a = n^{\frac{3}{2}} k \in \mathbb{Z}$.

(1) Assume that $k \geq c$ for some $c > 0$. Then

$$Q_{n,a}\left(\tfrac{n}{2} + h\right) = \frac{1}{n!} \exp\left\{ - \frac{h}{k\sqrt{n}} - \frac{1}{24k^2} - \frac{1}{2}\left(\frac{h}{kn}\right)^2 + O_c\left(\frac{1}{k\sqrt{n}}\right)\right\}, \tag{2.38}$$

where $O_c(\cdot)$ is uniform for all h and $k \geq c$.

(2) Assume that $k \geq n^{\frac{-1}{16}}$ and $|h| \leq n^{\frac{3}{4}}$. Then one has

$$Q_{n,a}\left(\tfrac{n}{2}+h\right) = \frac{1}{n!}\exp\left\{ -\frac{h}{k\sqrt{n}} - \frac{1}{24k^2} - \frac{1}{2}\left(\frac{h}{kn}\right)^2 + O\left(\frac{1}{k\sqrt{n}}\right)\right\}, \tag{2.39}$$

where $O(\cdot)$ is uniform for all $|h| \leq n^{\frac{3}{4}}$ and $k \geq n^{\frac{-1}{16}}$.

Now letting $a = n^{\frac{3}{2}}k \in \mathbb{Z}$ and considering the following two regions,

$$I_1 = \left\{h : |h| \leq n^{3/4}, h + \frac{n}{2} \in \mathbb{Z}\right\}$$

and

$$I_2 = \left\{h : -\frac{n}{2} + 1 \leq h \leq \frac{n}{2}, h + \frac{n}{2} \in \mathbb{Z}\right\} \setminus I_1.$$

By Lemma 2.12 and (2.38), one has, for $k \geq c > 0$,

$$\begin{aligned}
&\sum_{h\in I_1}\left[n!Q_{n,a}\left(\tfrac{n}{2}+h\right)\right]^l \frac{R_{nh}}{n!}\\
=&\sqrt{\frac{1}{2\pi(n/12)}}\sum_{h\in I_1}\exp\left\{-\frac{1}{2}\left(\frac{h}{\sqrt{n/12}}+\frac{l}{\sqrt{12}k}\right)^2 + \frac{l(l-1)}{24k^2} + O_c\left(n^{-\frac{1}{2}}\right)\right\}\\
=&\Phi\left(\sqrt{3n}+\frac{l}{\sqrt{12}k}\right)\exp\left\{\frac{l(l-1)}{24k^2} + O_c\left(n^{-\frac{1}{2}}\right)\right\}
\end{aligned}$$

and

$$\sum_{h\in I_2}\left[n!Q_{n,a}\left(\tfrac{n}{2}+h\right)\right]^l \frac{R_{nh}}{n!} = o_{c,p}\left(e^{-6\sqrt{n}}\right),$$

where $O_c(\cdot)$ is uniform for $k \geq c$. Combining both bounds implies that for $k > c$ and $a = n^{3/2}k \in \mathbb{Z}$,

$$d_{n,l}(a) = e^{\frac{l(l-1)}{24k^2}} + O_{c,p}(n^{-1/2}),$$

and then

$$\|q_{n,a}-1\|_p^p=\sum_{l=2}^{p}\binom{p}{l}(-1)^{p-l}e^{\frac{l(l-1)}{24k^2}}+O_{c,p}(n^{-1/2}).$$

For $p\in(1,2)$, (2.39) implies that one may choose $N>0$ such that, for $n^{\frac{-1}{16}}<k<n^{\frac{1}{9}}$ and $-n^{\frac{3}{4}}\le h\le -\frac{\sqrt{n}}{12k}$,

$$Q_{n,n^{\frac{3}{2}}k}\left(\frac{n}{2}+h\right)\ge\frac{1}{n!}\exp\left\{\frac{1}{24k^2}+O\left(n^{-\frac{1}{4}}\right)\right\}>\frac{1}{n!},\quad \forall n\ge N.$$

Then a similar computation as before derives, for $0<k<c$,

$$\begin{aligned}\|q_{n,n^{\frac{3}{2}}k}-1\|_p^p\ge&\sum_{h=-n^{\frac{3}{4}}}^{-\frac{\sqrt{n}}{12k}}\left[n!Q_{n,n^{\frac{3}{2}}k}\left(\tfrac{n}{2}+h\right)-1\right]^p\frac{R_{nh}}{n!}\\ \ge\left(\exp\left\{\frac{1}{24k^2}+O\left(n^{-\frac{1}{4}}\right)\right\}-1\right)^{p-1}&\sum_{h=-n^{\frac{3}{4}}}^{-\frac{\sqrt{n}}{12k}}\left[n!Q_{n,n^{\frac{3}{2}}k}\left(\tfrac{n}{2}+h\right)-1\right]\frac{R_{nh}}{n!}\end{aligned}$$

and

$$\begin{aligned}&\sum_{h=-n^{\frac{3}{4}}}^{-\frac{\sqrt{n}}{12k}}\left[n!Q_{n,n^{\frac{3}{2}}k}\left(\tfrac{n}{2}+h\right)-1\right]\frac{R_{nh}}{n!}\\ =&\frac{1}{\sqrt{2\pi(n/12)}}\sum_{h=-n^{\frac{3}{4}}}^{-\frac{\sqrt{n}}{12k}}\exp\left\{-\frac{1}{2}\left(\frac{h}{\sqrt{n/12}}+\frac{1}{\sqrt{12}k}\right)^2+o\left(n^{-\frac{1}{4}}\right)\right\}\\ &\qquad-\frac{1}{\sqrt{2\pi(n/12)}}\sum_{h=-n^{\frac{3}{4}}}^{-\frac{\sqrt{n}}{12k}}\exp\left\{\frac{1}{2}\left(\frac{h}{\sqrt{n/12}}\right)^2+o\left(n^{-\frac{1}{2}}\right)\right\}\\ =&\Phi\left(\frac{1}{2\sqrt{3}k}\right)-\frac{1}{2}+O\left(n^{-\frac{1}{4}}\right)\end{aligned}$$

Combining both estimation implies the desired bound. □

Chapter 3

The ℓ^2-cutoff for random walks on finite groups

As Theorem 2.4 and Theorem 2.5 say, the ℓ^p-cutoff for a family of normal Markov chains can be determined by looking at the ℓ^p-mixing time and the spectral gap or, in discrete-time cases, the second largest singular value. However, the order of magnitude of these quantities is not easy to obtain. Under the assumption of reversibility, Corollary 2.1 implies that the ℓ^p-mixing time is of the same order as the ℓ^2-mixing time and, for $1 < p \leq \infty$, a cutoff in ℓ^p occurs if and only if a cutoff in ℓ^2 occurs. This section focuses on the ℓ^2-cutoff.

Spectral theory is a standard tool to study the ℓ^2-convergence of Markov chains to their stationarity. Under the assumption of reversibility, the ℓ^2-distance can be expressed as a function of eigenvalues and eigenvectors of a transition matrix. In particular, by Lemma 2.2, if the state space $\mathcal{X}$ possesses a group structure and the Markov kernel K is driven by a probability measure p on $\mathcal{X}$ satisfying $K(x,y) = p(x^{-1}y)$, then the ℓ^2-distance is independent of the initial starting state and involves only the spectrum of the transition matrix. To determine the ℓ^2-cutoff, one needs to analyze the distribution of these eigenvalues.

In this chapter, we develop methods for a more general class of Markov kernels which contains random walks on finite groups. In section 3.1, we introduce a class of transition matrices, whose ℓ^2-distances depend only on their spectrum, and derive some basic results for them. In section 3.3 and 3.4, by partitioning the spectrum into small subsets, we obtain a criterion for testing the ℓ^2-cutoff, which also gives a formula for the ℓ^2-mixing time. In section 3.5, we use this method to study the ℓ^2-cutoff for specific Markov chains, e.g. direct products of Markov chains.

3.1 Basic results and settings

Let K be an irreducible Markov kernel on a finite set $\mathcal{X}$ with stationary distribution π. Recall that K is normal if $KK^*(x,y) = K^*K(x,y)$ for all $x, y \in \mathcal{X}$, where AB denotes the multiplication of matrices A and B. A classical result in matrix analysis shows that if K is normal, then it is unitarily diagonalizable. Let $\beta_0 = 1, \beta_1, ..., \beta_{|\mathcal{X}|-1}$ be eigenvalues of K and $\psi_0 \equiv 1, \psi_1, ..., \psi_{|\mathcal{X}|-1}$ be the corresponding orthonormal eigenvectors. Then the ℓ^2-distance is given by

$$\|h_t^x - 1\|_2^2 = \sum_{i=1}^{|\mathcal{X}|-1} e^{-2t(1-\mathrm{Re}\beta_i)} |\psi_i(x)|^2$$

and

$$\|k_x^m - 1\|_2^2 = \sum_{i=1}^{|\mathcal{X}|-1} |\beta_i|^2 |\psi_i(x)|^2.$$

For a proof on the above facts, please refer to [29].

By Lemma 2.2, one has

Lemma 3.1. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain. Assume that there is a finite group G acting transitively on $\mathcal{X}$ such that*

$$K(gx, gy) = K(x, y), \quad \forall x, y \in \mathcal{X}, g \in G,$$

and K is normal with eigenvalues $\beta_0 = 1, \beta_1, ..., \beta_{|\mathcal{X}|-1}$. Then for $t, m \geq 0$,

$$\|h_t^x - 1\|_2^2 = \sum_{i=1}^{|\mathcal{X}|-1} e^{-2t(1-\mathrm{Re}\beta_i)}, \quad \|k_x^m - 1\|_2^2 = \sum_{i=1}^{|\mathcal{X}|-1} |\beta_i|^2 \quad \forall x \in \mathcal{X}.$$

Remark 3.1. Note that the invariance requirement in the above lemma is satisfied if $\mathcal{X}$ is a group and $K(x, y) = P(x^{-1}y)$ for all $x, y \in \mathcal{X}$, where P is a probability measure on $\mathcal{X}$ whose support generates $\mathcal{X}$(with positive power).

With the above observation, we may rewrite, for $p = 2$, Theorem 2.4 and Theorem 2.5 as follows.

Theorem 3.1. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of irreducible Markov chains with $|\mathcal{X}_n| \to \infty$. Assume that, for $n \geq 1$, there exists a finite group G_n acting transitively on $\mathcal{X}_n$ such that*

$$K_n(gx, gy) = K_n(x, y), \quad \forall x, y \in \mathcal{X}_n, g \in G_n, \tag{3.1}$$

Suppose that K_n is normal and $\lambda_{n,0} = 0 < \lambda_{n,1} \leq \cdots \leq \lambda_{n,|\mathcal{X}_n|-1}$ are the real parts of eigenvalues of $I - K_n$. Then the spectral gap λ_n of K_n is equal to $\lambda_{n,1}$ and the following are equivalent.

(1) $\mathcal{F}_c$ presents a ℓ^2-cutoff.

(2) For some $\epsilon > 0$, $\lambda_{n,1}^{-1} = o(T_2^c(K_n, \epsilon))$.

Furthermore, if any of these two conditions holds, then, for any fixed $k \geq 1$, the family $\mathcal{F}_c$ has a $\left(T_2^c(K_n, \epsilon), \lambda_{n,k\wedge(|\mathcal{X}_n|-1)}^{-1}\right)$ ℓ^2-cutoff.

Proof. The first part is the special case, $p = 2$, of Theorem 2.4. For the second part, let $t_n = T_2^c(K_n, \epsilon)$ and assume that $\lambda_n^{-1} = o(t_n)$. This implies $\lambda_{n,i}^{-1} = o(t_n)$ for all $1 \leq i < |\mathcal{X}_n|$. By Lemma 3.1, we have

$$\max_{x \in \mathcal{X}_n} \|h_{n,t}^x - 1\|_2^2 = \sum_{i=1}^{|\mathcal{X}_n|-1} e^{-2\lambda_{n,i}t}.$$

Then replacing t with $t_n + c\lambda_{n,k}^{-1}$ implies

$$\max_{x \in \mathcal{X}_n} \|h_{n,t}^x - 1\|_2^2 \begin{cases} \leq \epsilon^2 e^{-2c} + b_n & \text{if } c > 0 \\ \geq e^{-2c}(\epsilon^2 - b_n) & \text{if } c < 0 \end{cases}$$

where $b_n = e^{-2t_n\lambda_{n,1}} + \cdots + e^{-2t_n\lambda_{n,k}}$. Then the functions $\overline{f}$ and $\underline{f}$ defined in Definition 1.4 satisfy

$$\overline{f}(c) \le \epsilon e^{-c},\ \forall c > 0, \quad \underline{f}(c) \ge \epsilon e^{-c},\ \forall c < 0.$$

This proves the desired cutoff. □

The same line of reasoning gives the almost word for word proof derives the following theorem for discrete-time Markov chains.

Theorem 3.2. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of normal and ergodic Markov chains with $|\mathcal{X}_n| \to \infty$, and, for $n \ge 1$, there exists a finite group G_n acting transitively on $\mathcal{X}_n$ such that (3.1) holds. Let $\beta_{n,0} = 1 > \beta_{n,1} \ge \cdots \ge \beta_{n,|\mathcal{X}_n|-1}$ be the absolute values of eigenvalues of K_n. Assume that $T_2^d(K_n, \epsilon) \to \infty$ for some $\epsilon > 0$. Then the following are equivalent.*

(1) $\mathcal{F}_d$ presents a ℓ^2-cutoff.

(2) For some $\epsilon > 0$, $b_n^{-1} = o(T_2^d(K_n, \epsilon))$, where $b_n = \min\{-\log \beta_{n,1}, 1\}$.

Moreover, if any of the above two conditions holds, then, for $k \ge 1$, $\mathcal{F}_d$ has a $(T_2^d(K_n, \epsilon), b_{n,k}^{-1})$ ℓ^2-cutoff, where $b_{n,k} = \min\{-\log \beta_{n,k\wedge(|\mathcal{X}_n|-1)}, 1\}$

For an equivalence between ℓ^p-cutoffs, please refer to Theorem 2.6, Corollary 2.4 and Corollary 2.6.

3.2 Triangular arrays of positive numbers

Under the assumption of Lemma 3.1, one can find that the ℓ^2-cutoff for $\mathcal{F}$ is determined by a sequence of positive numbers(or strictly, of numbers in $(0, 2)$). However, even knowing the spectrum of K_n, it is not an easy job to determine the

ℓ^2-mixing time. We will give an idea of how to compute such a quantity in the next section. Here, we introduce a more general setting as follows.

Definition 3.1. Let $\mathcal{A} = \{a_{n,i} > 0 : 1 \le i \le k_n, n \ge 1\}$ be a triangular array of positive numbers and, for $n \ge 1$, $d_n^{\mathcal{A}}$ be a function defined by

$$d_n^{\mathcal{A}}(t) = \sum_{i=1}^{k_n} e^{-2ta_{n,i}} \quad \forall t \in \mathbb{R}.$$

Then $\mathcal{A}$ is said to present:

(1) A cut-off with critical time $t_n > 0$ if

$$\lim_{n\to\infty} d_n^{\mathcal{A}}((1+\epsilon)t_n) = \begin{cases} 0 & \text{if } \epsilon \in (0,1) \\ \infty & \text{if } \epsilon \in (-1,0) \end{cases}.$$

(2) A (t_n, b_n) cut-off if there exist positive numbers t_n, b_n such that $b_n = o(t_n)$ and the following functions

$$\overline{f}(c) = \limsup_{n\to\infty} d_n^{\mathcal{A}}(t_n + cb_n), \quad \underline{f}(c) \liminf_{n\to\infty} d_n^{\mathcal{A}}(t_n + cb_n)$$

satisfy

$$\lim_{c\to\infty} \overline{f}(c) = 0, \lim_{c\to-\infty} \underline{f}(c) = \infty.$$

Remark 3.2. Note that a necessary condition for a triangular array $\mathcal{A}$ to have a cutoff(in Definition 3.1) is $k_n \to \infty$.

Definition 3.2. Let $\mathcal{A}$ and $d_n^{\mathcal{A}}$ be the same as in Definition 3.1. For $\epsilon > 0$, the mixing time of $\mathcal{A}$ is defined to be a sequence of nonnegative numbers $\{t_n^{\mathcal{A}}(\epsilon)\}_{n=1}^{\infty}$, where

$$t_n^{\mathcal{A}}(\epsilon) = \inf\{s \ge 0 : d_n^{\mathcal{A}}(s) \le \epsilon\}.$$

One can easily derive the following relation between the cutoff and the mixing time. This is similar to Proposition 1.10 and Theorem 3.1.

Proposition 3.1. *Let $\mathcal{A} = \{a_{n,i} : 1 \leq i \leq k_n, n \geq 1\}$ be a triangular array of positive numbers. Assume that $k_n \to \infty$. Then:*

(1) $\mathcal{A}$ presents a cutoff with critical time t_n if and only if

$$t_n^{\mathcal{A}}(\epsilon) \sim t_n, \quad \forall \epsilon > 0.$$

(2) $\mathcal{A}$ has a (t_n, b_n) cutoff if and only if

$$|t_n^{\mathcal{A}}(\epsilon) - t_n| = O(b_n), \quad \forall \epsilon > 0.$$

In particular, if $a_{n,i} \leq a_{n,i+1}$ for all $1 \leq i < k_n$ and $n \geq 1$, then

$$\mathcal{A} \text{ presents a cutoff} \quad \Leftrightarrow \quad \text{for some } \epsilon > 0,\ a_{n,1}^{-1} = o\left(t_n^{\mathcal{A}}(\epsilon)\right).$$

Furthermore, if $\mathcal{A}$ presents a cutoff, then, for any $\epsilon > 0$ and fixed $j \geq 1$, the family $\mathcal{A}$ has a (t_n, b_n) cutoff, where

$$t_n = t_n^{\mathcal{A}}(\epsilon), \quad b_n = a_{n,j\wedge k_n}^{-1}.$$

Definition 3.3. Let $\mathcal{A}$, $\overline{f}$ and $\underline{f}$ be the same as in Definition 3.1. Then a (t_n, b_n) cutoff for $\mathcal{A}$ is called

(1) weakly optimal if, for any (t_n, c_n) cutoff, one has $b_n = O(c_n)$.

(2) optimal if, for any (s_n, c_n) cutoff, one has $b_n = o(c_n)$. In this case, b_n is called an optimal window size of the cutoff.

(3) strongly optimal, if

$$\overline{f}(c) < \infty,\ \forall c < 0, \quad \underline{f}(c) > 0,\ \forall c > 0.$$

By Proposition 3.1, one can derive the same result as in Corollary 1.6.

Corollary 3.1. *Let A be a triangular array of positive numbers. Assume that $\mathcal{A}$ presents a cutoff. Then the following are equivalent.*

(1) The cutoff for $\mathcal{A}$ has an optimal window size b_n.

(2) For some $\epsilon > 0$, the family $\mathcal{A}$ presents a weakly optimal $\left(t_n^{\mathcal{A}}(\epsilon), b_n\right)$ cutoff.

In particular, if $\mathcal{A}$ presents a weakly optimal $\left(t_n^{\mathcal{A}}(\epsilon), b_n\right)$ cutoff, then it is optimal.

The following is a useful fact to bound the optimal window size of a cutoff.

Corollary 3.2. *Let $\mathcal{A} = \{a_{n,i} : 1 \le i \le k_n, n \ge 1\}$ be a triangular array of positive numbers satisfying $a_{n,i} \le a_{n,i+1}$ for $1 \le i < k_n$ and $n \ge 1$. Assume that $\mathcal{A}$ presents a cutoff with an optimal window size b_n. Then there are constants $c_1 > 0, c_2 > 0$ such that*

$$c_1 a_{n,k_n}^{-1} \le b_n \le c_2 a_{n,1}^{-1} \quad \forall n \ge 1.$$

Proof. The second inequality is proved by the definition of an optimal window size(in Definition 3.3(2)) and Proposition 3.1. For the first one, by Corollary 3.1, $\mathcal{A}$ has an optimal $(t_n^{\mathcal{A}}(\epsilon), b_n)$ cutoff. Note that if $t_n^{\mathcal{A}}(\epsilon) > 0$, then

$$d_n^{\mathcal{A}}(t_n^{\mathcal{A}}(\epsilon) + s) \ge \epsilon e^{-2sa_{n,k_n}}, \quad \forall s \ge 0.$$

Replacing $s = cb_n$ with $c > 0$ implies

$$0 = \lim_{c\to\infty} \limsup_{n\to\infty} e^{-2ca_{n,k_n}b_n} = \lim_{c\to\infty} \exp\left\{-c \liminf_{n\to\infty} a_{n,k_n} b_n\right\},$$

or equivalently,

$$\liminf_{n\to\infty} a_{n,k_n} b_n > 0.$$

□

The following is a fact similar to Lemma 1.3 which compares the optimal window sizes when two triangular arrays present cutoffs with the same critical time.

Lemma 3.2. *Let $\mathcal{A}$ and $\mathcal{B}$ be two triangular arrays that present cutoffs with the same critical time. Assume that both arrays have optimal window sizes. Then the following are equivalent.*

(1) $\mathcal{A}$ and $\mathcal{B}$ have the same optimal window size(in the sense of order), if any.

(2) $\mathcal{A}$ has a (t_n, b_n) cutoff if and only if $\mathcal{B}$ has a (t_n, b_n) cutoff.

Proof. Immediate from Definition 3.3. □

3.3 Cutoff for triangular arrays

Let $\mathcal{A} = \{a_{n,i} : 1 \le i \le k_n, n \ge 1\}$ be a triangular array of positive numbers and

$$d_n^{\mathcal{A}}(t) = \sum_{i=1}^{k_n} e^{-2ta_{n,i}} \quad \forall t \in \mathbb{R}, n \ge 1. \tag{3.2}$$

For $n \ge 1$, let $\mathcal{P}_n = \{x_{n,0}, x_{n,1}, ..., x_{n,l_n}\}$ be a partition of $(0, \infty)$ satisfying

$$x_{n,0} \le \min_{1\le i\le k_n} a_{n,i} \le \max_{1\le i\le k_n} a_{n,i} < x_{n,l_n} \tag{3.3}$$

and, for $1 \le j < l_n$, define

$$N_j^{\mathcal{A}}(\mathcal{P}_n) = \{1 \le i \le k_n : x_{n,j} \le a_{n,i} < x_{n,j+1}\}$$

and

$$\tau_{\mathcal{A}}(\mathcal{P}_n) = \sup_{0\le j<l_n} \left\{ \frac{\log |N_j^{\mathcal{A}}(\mathcal{P}_n)|}{2x_{n,j}} \right\}, \quad \xi_{\mathcal{A}}(\mathcal{P}_n) = \sup_{0\le j<l_n} \left\{ \frac{\log |N_j^{\mathcal{A}}(\mathcal{P}_n)|}{2x_{n,j+1}} \right\}. \tag{3.4}$$

A relation between the above two quantities is

$$\frac{\tau_{\mathcal{A}}(\mathcal{P}_n)}{1 + \epsilon_{\mathcal{A}}(\mathcal{P}_n)} \le \xi_{\mathcal{A}}(\mathcal{P}_n) \le \frac{\tau_{\mathcal{A}}(\mathcal{P}_n)}{1 + \eta_{\mathcal{A}}(\mathcal{P}_n)} \le \tau_{\mathcal{A}}(\mathcal{P}_n), \tag{3.5}$$

where

$$\epsilon_{\mathcal{A}}(\mathcal{P}_n) = \sup_{0\le j<l_n} \frac{x_{n,j+1}}{x_{n,j}} - 1, \quad \eta_{\mathcal{A}}(\mathcal{P}_n) = \inf_{0\le j<l_n} \frac{x_{n,j+1}}{x_{n,j}} - 1. \tag{3.6}$$

By the above definitions, one has

$$d_n^{\mathcal{A}}(t) \le \sum_{j=0}^{l_n} |N_j^{\mathcal{A}}(\mathcal{P}_n)| e^{-2tx_{n,j}} \le \sum_{j=0}^{l_n} \exp\{2x_{n,j}(\tau_{\mathcal{A}}(\mathcal{P}_n) - t)\} \tag{3.7}$$

and

$$d_n^{\mathcal{A}}(t) \ge \sum_{j=0}^{l_n} |N_j^{\mathcal{A}}(\mathcal{P}_n)| e^{-2tx_{n,j+1}} \ge \exp\{2x_{n,m_n+1}(\xi_{\mathcal{A}}(\mathcal{P}_n) - t)\}, \tag{3.8}$$

where $0 \le m_n < l_n$ satisfies

$$\xi_{\mathcal{A}}(\mathcal{P}_n) = \frac{\log |N_{m_n}^{\mathcal{A}}(\mathcal{P}_n)|}{2x_{n,m_n+1}}. \tag{3.9}$$

Remark 3.3. Note that if $\mathcal{P}'_n = \mathcal{P}_n \cup \mathcal{P}$ is another partition where $\mathcal{P} \cap [x_{n,0}, x_{n,l_n}) = \phi$, then $\tau_{\mathcal{A}}(\mathcal{P}'_n) = \tau_{\mathcal{A}}(\mathcal{P}_n)$ and $\xi_{\mathcal{A}}(\mathcal{P}'_n) = \xi_{\mathcal{A}}(\mathcal{P}_n)$.

By this remark, it suffices to consider partitions $(\mathcal{P}_n)_1^\infty$ satisfying

$$x_{n,0} \le \min\{a_{n,i} : 1 \le i \le k_n\} \le x_{n,1} \quad \forall n \ge 1. \tag{3.10}$$

The following two lemmas derive respectively an upper and a lower bound on $d_n^{\mathcal{A}}$ by partitioning the entries of $\mathcal{A}$ in a proper way.

Lemma 3.3. *Let $\mathcal{A} = \{a_{n,i} : 1 \le i \le k_n, n \ge 1\}$ be a triangular array of positive numbers with $a_{n,1} = \min\{a_{n,i} : 1 \le i \le k_n\}$, and $d_n^{\mathcal{A}}, \tau_{\mathcal{A}}, \eta_{\mathcal{A}}$ be quantities defined in* (3.2), (3.4) *and* (3.6). *For $n \ge 1$, let $\mathcal{P}_n = \{x_{n,0}, x_{n,1}, ..., x_{n,l_n}\}$ be a partition of $(0, \infty)$ satisfying* (3.3) *and* (3.10). *Assume that $a_{n,1} = O(x_{n,0})$. Then for any function $f : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying*

$$\inf_{t>0} f(t) > 0, \quad \log \frac{1}{t} = O(f(t)) \quad \textit{as } t \downarrow 0,$$

one has

$$\lim_{c\to\infty} \limsup_{n\to\infty} d_n^{\mathcal{A}}(t_n + cb_n) = 0,$$

where

$$t_n = \tau_{\mathcal{A}}(\mathcal{P}_n), \quad b_n = \frac{f(\eta_{\mathcal{A}}(\mathcal{P}_n))}{a_{n,1}}.$$

Proof. For convenience, let $t_n = \tau_{\mathcal{A}}(\mathcal{P}_n)$, $\epsilon_n = \epsilon_{\mathcal{A}}(\mathcal{P}_n)$ and $\eta_n = \eta_{\mathcal{A}}(\mathcal{P}_n)$. Note that $x_{n,j} \geq x_{n,0}(1+\eta_n)^j$ for $j \geq 0$. By (3.7), if $t > t_n$, then

$$d_n^{\mathcal{A}}(t) \leq \sum_{j=0}^{\infty} e^{2x_{n,0}(1+\eta_n)^j(t_n-t)} \leq \frac{e^{-2x_{n,0}(t-t_n)}}{\min\{\eta_n x_{n,0}(t-t_n), 1/2\}}, \tag{3.11}$$

where the last inequality is based on the following facts.

$$e^{-2t} \leq \max\{1-t, 1/2\}, \quad (1+t)^j \geq 1+tj \quad \forall t \geq 0.$$

Let $c_1 > 0$ be a constant such that

$$x_{n,0} \geq c_1 x_{n,1} \quad \forall n \geq 1, \quad f(t) \geq c_1 \max\{1, \log t^{-1}\} \quad \forall t > 0.$$

Replacing $t = t_n + cb_n$ with $b_n = f(\eta_n)/a_{n,1}$ in the above inequality, we get

$$\begin{aligned} d_n^{\mathcal{A}}(t_n + cb_n) &\leq \max\left\{\frac{e^{-2cc_1 f(\eta_n)}}{cc_1^2\eta_n}, 2e^{-2cc_1^2}\right\} \\ &\leq (cc_1^2)^{-1}\min\{(\eta_n)^{2cc_1^2-1}, e^{-2cc_1^2}\eta_n^{-1}\} + 2e^{-2cc_1^2}. \end{aligned}$$

Note that for $c > (2c_1^2)^{-1}$, $\min\{(\eta_n)^{2cc_1^2-1}, e^{-2cc_1^2}\eta_n^{-1}\} \leq e^{-2cc_1^2+1}$, which implies

$$d_n^{\mathcal{A}}(t_n + cb_n) \leq (cc_1^2 e + 2)e^{-2cc_1^2} \quad \forall c > (2c_1^2)^{-1}.$$

This proves the desired property. □

Lemma 3.4. *Let $\mathcal{A} = \{a_{n,i} : 1 \leq i \leq k_n, n \geq 1\}$ be a triangular array of positive numbers with $a_{n,1} = \min\{a_{n,i} : 1 \leq i \leq k_n\}$, and $d_n^{\mathcal{A}}, \tau_{\mathcal{A}}, \epsilon_{\mathcal{A}}$ be quantities defined in* (3.2), (3.4) *and* (3.6)*. For $n \geq 1$, let $\mathcal{P}_n = \{x_{n,0}, x_{n,1}, ..., x_{n,l_n}\}$ be a partition of* $(0, \infty)$ *satisfying* (3.3) *and* (3.10)*, and denote m_n to be a number such that* (5.13) *holds. Then one has*

$$\lim_{c \to -\infty} \liminf_{n \to \infty} d_n^{\mathcal{A}}(t_n + cb_n) = \infty, \tag{3.12}$$

where

$$t_n = \tau_{\mathcal{A}}(\mathcal{P}_n),\ \ b_n = \max\{x_{n,m_n+1}^{-1}, \epsilon_{\mathcal{A}}(\mathcal{P}_n)\tau_{\mathcal{A}}(\mathcal{P}_n)\}.$$

Proof. For convenience, we denote $t_n = \tau_{\mathcal{A}}(\mathcal{P}_n)$, $\xi_n = \xi_{\mathcal{A}}(\mathcal{P}_n)$ and $\epsilon_n = \epsilon_{\mathcal{A}}(\mathcal{P}_n)$. By (3.8), one can easily check that

$$\lim_{c\to-\infty} \liminf_{n\to\infty} d_n^{\mathcal{A}}(\xi_n + c x_{n,m_n+1}^{-1}) = \infty,$$

and by (3.5), we have $\xi_n \geq (1+\epsilon_n)^{-1} t_n$. Letting $b_n = \max\{x_{n,m_n+1}^{-1}, \epsilon_n t_n\}$ implies

$$\xi_n + c x_{n,m_n+1}^{-1} \geq t_n + (c-1) b_n \quad \text{for } c < 0.$$

Hence, (3.12) is proved by the above discussion and the monotonicity of $d_n^{\mathcal{A}}(\cdot)$. $\square$

Remark 3.4. Note that in Lemma 3.4 the term $t_n + c b_n$ in (3.12) need not be positive.

The next theorem is a combination of Lemma 3.3 and Lemma 3.4.

Theorem 3.3. *Let $\mathcal{A} = \{a_{n,i} : 1 \leq i \leq k_n, n \geq 1\}$ be a triangular array of positive numbers satisfying*

$$a_{n,1} = \min\{a_{n,i} : 1 \leq i \leq k_n\}, \quad k_n \to \infty,$$

and $\tau_{\mathcal{A}}, \epsilon_{\mathcal{A}}, \eta_{\mathcal{A}}$ be quantities defined in (3.4) and (3.6). For $n \geq 1$, let $\mathcal{P}_n = \{x_{n,0}, x_{n,1}, ..., x_{n,l_n}\}$ be a partition satisfying

$$a_{n,1} < x_{n,1}, \quad a_{n,1} = O(x_{n,0}), \quad \max_{1\leq i\leq k_n} a_{n,i} < x_{n,l_n}. \tag{3.13}$$

Assume that

$$\lim_{n\to\infty} \epsilon_{\mathcal{A}}(\mathcal{P}_n) = 0, \quad \lim_{n\to\infty} \frac{-\log(\eta_{\mathcal{A}}(\mathcal{P}_n))}{a_{n,1}\tau_{\mathcal{A}}(\mathcal{P}_n)} = 0.$$

Then $\mathcal{A}$ presents a (t_n, b_n) cutoff, where

$$t_n = \tau_{\mathcal{A}}(\mathcal{P}_n), \quad b_n = \max\left\{\frac{-\log\left(\eta_{\mathcal{A}}(\mathcal{P}_n)\right)}{a_{n,1}}, \epsilon_{\mathcal{A}}(\mathcal{P}_n)\tau_{\mathcal{A}}(\mathcal{P}_n)\right\}.$$

Proof. By applying Lemma 3.3 with $f(t) = \max\{-\log t, 1\}$ for $t > 0$ and Lemma 3.4. □

Remark 3.5. Note that the first two of (3.13) are not too hard to be satisfied. For example, one may choose $x_{n,0} = a_{n,1}$.

The following theorem says that the assumption and the result in the above theorem are roughly equivalent.

Theorem 3.4. *Let $\mathcal{A}, \tau_{\mathcal{A}}$ be the same as in Theorem 3.3 and $a_{n,1} = \min\{a_{n,i} : 1 \le i \le k_n\}$. Assume that $\mathcal{A}$ presents a cutoff, then there exists a sequence $(\epsilon_n)_1^\infty$ of positive numbers converging to 0 such that*

$$\lim_{n\to\infty} \frac{-\log(\epsilon_n)}{a_{n,1}\tau_{\mathcal{A}}(\mathcal{P}_n)} = 0,$$

where $\mathcal{P}_n = \{a_{n,1}(1+\epsilon_n)^i : i \ge 0\}$.

In particular, if $\{a_{n,i} : 1 \le i \le k_n, n \ge 1\}$ are bounded, then the above statement also holds for partitions of the form $\mathcal{P}_n = \{a_{n,1}(1 + i\epsilon_n) : i \ge 0\}$.

Proof. For $n \ge 1$, let $\mathcal{P}_n(\epsilon) = \{a_{n,1}(1+\epsilon)^i : i \ge 0\}$ and define

$$\delta_n = \sup\left\{\epsilon \in (0,1) : 1 + a_{n,1}\tau_{\mathcal{A}}(\mathcal{P}_n(\epsilon)) \le (e\epsilon)^{-1}\log \epsilon^{-1}\right\}.$$

Note that δ_n is defined with $0 < \delta_n < e^{-1}$ and, by the left-continuity of $\tau_{\mathcal{A}}(\mathcal{P}_n(\cdot))$, one has

$$1 + a_{n,1}\tau_{\mathcal{A}}(\mathcal{P}_n(\delta_n)) \le (e\delta_n)^{-1}\log \delta_n^{-1}.$$

Set $\epsilon_n = 2\delta_n$, then $\epsilon_n \in (0,1)$ and

$$1 + a_{n,1}\tau_{\mathcal{A}}(\mathcal{P}_n(\epsilon_n)) > (e\epsilon_n)^{-1}\log \epsilon_n^{-1}.$$

It remains to show that $\epsilon_n \to 0$, or equivalently $\delta_n \to 0$. To prove this, consider partitions $\mathcal{P}_n(\delta_n)$ and the function $f(t) = \max\{\log t^{-1}, 1\}$. By Lemma 3.3, one may choose $C > 0$ such that

$$t_n^{\mathcal{A}}(1) \le \tau_{\mathcal{A}}(\mathcal{P}_n(\delta_n)) + Cb_n = \tau_{\mathcal{A}}(\mathcal{P}_n(\delta_n)) + C(\log \delta_n^{-1})a_{n,1}^{-1}.$$

This implies

$$a_{n,1}t_n^{\mathcal{A}}(1) \le (e^{-1}\delta_n^{-1} + C)\log \delta_n^{-1} - 1.$$

Note also that by Proposition 3.1, we have $a_{n,1}t_n^{\mathcal{A}}(1) \to \infty$, which derives $\delta_n = o(1)$.

The last part can be proved almost word for word. □

Note that Theorem 3.3 gives a sufficient condition(which is also necessary in some sense by Theorem 3.4) for the cutoff of $\mathcal{A}$. However, it is not easy to compute the quantity $\tau_{\mathcal{A}}(\mathcal{P}_n)$, let alone to select an optimal partition. The following theorem provides another method to determine the cutoff which also gives the same window size(up to a constant) as Theorem 3.3.

Theorem 3.5. *Let $\mathcal{A} = \{a_{n,i} : 1 \le i \le k_n, n \ge 1\}$ be a triangular array of positive numbers with $a_{n,i} \le a_{n,i+1}$ for $1 \le i < k_n$ and $n \ge 1$. Assume that $k_n \to \infty$. Then $\mathcal{A}$ has a cutoff if and only if*

$$a_{n,1}^{-1} = o(\tau_n),$$

where $\tau_n = \max\left\{\frac{\log i}{2a_{n,i}} : 1 \le i \le k_n\right\}$. If the above identity holds, then the ℓ^2-cutoff has a critical time τ_n.

Moreover, if $\mathcal{A}$ presents a cutoff and $\mathcal{P}_n$ is a partition satisfying (3.13) *such that*

$$\lim_{n\to\infty} \epsilon_{\mathcal{A}}(\mathcal{P}_n) = 0, \quad \lim_{n\to\infty} \frac{-\log(\eta_{\mathcal{A}}(\mathcal{P}_n))}{a_{n,1}\tau_{\mathcal{A}}(\mathcal{P}_n)} = 0.$$

Then $\mathcal{A}$ has a (t_n, b_n) cutoff, where

$$t_n = \tau_n, \quad b_n = \max\left\{\frac{-\log\left(\eta_{\mathcal{A}}(\mathcal{P}_n)\right)}{a_{n,1}}, \epsilon_{\mathcal{A}}(\mathcal{P}_n)\tau_n\right\}.$$

Proof. Let $\mathcal{P}_n$ be any partition satisfying (3.13). A simple computation shows

$$[1+\epsilon_{\mathcal{A}}(\mathcal{P}_n)]\tau_{\mathcal{A}}(\mathcal{P}_n) \le \xi_{\mathcal{A}}(\mathcal{P}_n) \le \tau_n \le t_n^{\mathcal{A}}(1).$$

Then the first part is proved by Proposition 3.1 and Theorem 3.4. Note that if $\mathcal{A}$ has a cutoff, then

$$\tau_{\mathcal{A}}(\mathcal{P}_n) \sim t_n^{\mathcal{A}}(1) \sim \tau_n$$

and $|t_n - \tau_{\mathcal{A}}(\mathcal{P}_n)| = O(b_n)$. This proves the second part. □

Remark 3.6. Note that one can easily check the existence of a cutoff for a triangular array $\mathcal{A}$ and obtain a critical time from Theorem 3.5, but it's still hard to find a proper partition for a window size. Theorem 3.5 gives a clue on choosing partitions, that is, a necessary and sufficient condition for such a partition $\mathcal{P}_n$ is

$$\lim_{n\to\infty} \epsilon_{\mathcal{A}}(\mathcal{P}_n) = 0, \quad \lim_{n\to\infty} \frac{-\log(\eta_{\mathcal{A}}(\mathcal{P}_n))}{a_{n,1}\tau_n} = 0, \quad \tau_{\mathcal{A}}(\mathcal{P}_n) \sim \tau_n.$$

It is clear that none of the above theorems yields a weakly optimal cutoff. To fill in this gap, one may use the following proposition.

Proposition 3.2. *Let $\mathcal{A}$ be a triangular of positive numbers satisfying $a_{n,i} \le a_{n,i+1}$ for $1 \le i < k_n, n \ge 1$, and $d_n^{\mathcal{A}}$ be the function defined in (3.2). Assume that $\mathcal{A}$ presents a cutoff and t_n is a sequence of positive numbers such that*

$$0 < \liminf_{n\to\infty} d_n^{\mathcal{A}}(t_n) \le \limsup_{n\to\infty} d_n^{\mathcal{A}}(t_n) < \infty.$$

Then, for any fixed $k > 0$, $\mathcal{A}$ has a $\left(t_n, a_{n,k\wedge k_n}^{-1}\right)$ cutoff.

Proof. Let

$$c_1 = \frac{1}{2}\liminf_{n\to\infty} d_n^{\mathcal{A}}(t_n), \quad c_2 = 2\limsup_{n\to\infty} d_n^{\mathcal{A}}(t_n).$$

Then one may choose $N \ge 1$ such that

$$t_n^{\mathcal{A}}(c_2) \le t_n \le t_n^{\mathcal{A}}(c_1) \quad \forall n \ge N.$$

By Proposition 3.1, the above proves the desired cutoff. □

Proposition 3.1 gives a necessary and sufficient condition for a triangular array $\mathcal{A}$ to have a cutoff of (that is, $a_{n,1}^{-1} = o(t_n^{\mathcal{A}}(\epsilon))$). Note that it allows us to get rid of k, any fixed number, entries in $\mathcal{A}$ and not to change the result itself. The following is a fact based on this idea.

Proposition 3.3. *Let $\mathcal{A} = \{a_{n,i} : 1 \le i \le k_n, n \ge 1\}$ be a triangular array of positive numbers and $d_n^{\mathcal{A}}$ be the function defined in* (3.2). *Assume that $\mathcal{A}$ has a cutoff with critical time s_n and*

$$\lim_{n\to\infty} \left[d_n^{\mathcal{A}}(\delta s_n) - d_n^{\mathcal{A}'}(\delta s_n) \right] = 0, \quad \textit{for some } \delta \in (0,1],$$

where $\mathcal{A}'$ is a sub-array of $\mathcal{A}$ and $d_n^{\mathcal{A}'}$ is the function associated to $\mathcal{A}'$ and defined in (3.2). *Then:*

(1) If $\delta \in (0,1)$, then $\mathcal{A}'$ has a (t_n, b_n) cutoff if and only if $\mathcal{A}$ has a (t_n, b_n) cutoff. In this case, both $\mathcal{A}$ and $\mathcal{A}'$ have the same(up to a constant) optimal window size.

(2) If $\delta = 1$ and $\mathcal{A}'$ has a (t_n, b_n) cutoff with $|t_n - s_n| = O(b_n)$, then $\mathcal{A}$ has a (t_n, b_n) cutoff.

Proof. For two triangular arrays $\mathcal{A}_1 = \{a_{n,i}^{(1)} : 1 \le i \le k_n^{(1)}, n \ge 1\}$ and $\mathcal{A}_2 = \{a_{n,i}^{(2)} : 1 \le i \le k_n^{(2)}, n \ge 1\}$, we define $\mathcal{A}_1 \sqcup \mathcal{A}_2 = \{b_{n,i} : 1 \le i \le k_n^{(1)} + k_n^{(2)}, n \ge 1\}$ by letting

$$b_{n,i} = a_{n,i}^{(1)},\ \forall 1 \le i \le k_n^{(1)}, \quad b_{n,k_n^{(1)}+i} = a_{n,i}^{(2)},\ \forall 1 \le i \le k_n^{(2)}.$$

Then the corresponding distance function satisfies

$$d_n^{\mathcal{A}_1 \sqcup \mathcal{A}_2}(t) = d_n^{\mathcal{A}_1}(t) + d_n^{\mathcal{A}_2}(t).$$

This implies that, for $\epsilon > 0$ and $\delta > 0$,

$$\max\left\{t_n^{\mathcal{A}_1}(\epsilon+\delta), t_n^{\mathcal{A}_2}(\epsilon+\delta)\right\} \le t_n^{\mathcal{A}_1 \sqcup \mathcal{A}_2}(\epsilon+\delta) \le \max\left\{t_n^{\mathcal{A}_1}(\epsilon), t_n^{\mathcal{A}_2}(\delta)\right\}.$$

Letting $\mathcal{A}_1 = \mathcal{A}'$ and $\mathcal{A}_2$ be a triangular array such that $\mathcal{A} = \mathcal{A}_1 \sqcup \mathcal{A}_2$(regardless of the arrangement of elements in each row). For (1), one may choose $N \ge 1$ such that for $\epsilon > 0$ and $n \ge N$,

$$t_n^{\mathcal{A}'}(2\epsilon) \le t_n^{\mathcal{A}}(2\epsilon) \le t_n^{\mathcal{A}'}(\epsilon).$$

This proves the first result.

For (2), since $\mathcal{A}'$ presents (t_n, b_n) cutoff and $|s_n - t_n| = O(b_n)$, we may choose $C > 0$ and $N \ge 1$ such that, for $\epsilon > 0$ and $n \ge N$,

$$t_n - Cb_n \le t_n^{\mathcal{A}_1}(2\epsilon) \le t_n^{\mathcal{A}_1}(\epsilon) \le t_n + Cb_n,$$

and

$$t_n^{\mathcal{A}_2}(\epsilon) \le s_n \le t_n + Cb_n.$$

This implies $|t_n^{\mathcal{A}}(2\epsilon) - t_n| = O(b_n)$ for all $\epsilon > 0$ and hence proves the (t_n, b_n) cutoff for $\mathcal{A}$. □

Remark 3.7. Let $\mathcal{A}$ and $\mathcal{A}'$ be arrays in Proposition 3.3 and k_n, k_n' be numbers of entries in the respective nth rows. Assume $\mathcal{A}$ presents a cutoff. Then the identity in Proposition 3.3 always holds if $|k_n - k_n'| = O(1)$.

Because this section is not our goal, we only use the following simple examples to illustrate how these theorems go.

Example 3.1. Consider a triangular array $\mathcal{A}$ of positive numbers $a_{n,i}$, where

$$a_{n,1} = \frac{1}{\log\log n}, \quad a_{n,i} = \frac{1 + 1/\sqrt{\log n}}{\log\log n} \quad \text{for } 2 \le i \le n+1$$

and $d_n^{\mathcal{A}}(t) = \sum_{i=1}^{n+1} e^{-2ta_{n,i}}$. We deal with this problem in the following three ways.

(1) The first way is to bound $d_n(t)$ as follows.

$$(n+1)e^{-2ta_{n,2}} \leq d_n(t) \leq (n+1)e^{-2ta_{n,1}},$$

which implies

$$\frac{\log(n+1) - \log c}{2a_{n,2}} \leq t_n^{\mathcal{A}}(c) \leq \frac{\log(n+1) - \log c}{2a_{n,1}}.$$

Since $\left| \frac{\log(n+1)}{2a_{n,1}} - \frac{\log(n+1)}{2a_{n,2}} \right| \sim \frac{1}{2}\sqrt{\log n} \log\log n$, the array $\mathcal{A}$ has a (t_n, b_n) cut-off, where

$$t_n = \frac{1}{2}(\log n) \log\log n, \quad b_n = \sqrt{\log n} \log\log n.$$

Note that the above inequality is not enough to obtain a lower bound on the optimal window size.

(2) The second way is done by applying Theorem 3.5. It can be easily computed that

$$\tau_n = \frac{\log(n+1) \log\log n}{2(1 + 1/\sqrt{\log n})} = \frac{\log n}{2a_{n,2}} + o(1).$$

Since $a_{n,1}^{-1} = o(\tau_n)$, $\mathcal{A}$ has a cutoff. To get a window size, choosing $\epsilon_n = 1/\sqrt{\log n}$ and $\mathcal{P}_n = \{a_{n,1}(1+\epsilon_n)^i : i \geq 0\}$ derives

$$\tau_{\mathcal{A}}(\mathcal{P}_n) = \tau_n, \quad \epsilon_{\mathcal{A}}(\mathcal{P}_n) = \eta_{\mathcal{A}}(\mathcal{P}_n) = 1/\sqrt{\log n}.$$

This implies that $\mathcal{A}$ has a (t'_n, b'_n) cutoff, where

$$t'_n = \tau_n, \quad b'_n = \sqrt{\log n} \log\log n.$$

This gives the same result as in (1).

(3) The third way is by applying Proposition 3.2. It is reasonable to investigate whether t_n(in (1)) or t'_n(in (2)) satisfies the assumption. A simple computation then shows

$$d_n^{\mathcal{A}}(t_n) = e^{-\log n} + e^{-\sqrt{\log n}} = o(1), \quad d_n^{\mathcal{A}}(t'_n) \sim 1.$$

By Proposition 3.2 and Corollary 3.2, $\mathcal{A}$ presents an optimal (t''_n, b''_n) cutoff, where

$$t''_n = \tau_n, \quad b''_n = \log\log n.$$

Note that $|t_n - t''_n| \sim \frac{1}{2}b_n$, which is the reason why the window size in (1) can't be any smaller. In fact, $\mathcal{A}$ presents a weakly (t_n, b_n) cutoff, and neither optimality possesses by the (t'_n, b'_n) cutoff since t'_n is a correct choice of the critical time.

Example 3.2. Let $\mathcal{A} = \{a_{n,i} : 1 \le i \le n+1, n \ge 1\}$, where

$$a_{n,1} = \frac{1}{n\log\log n}, \quad a_{n,i} = \frac{1}{n} \quad \text{for } 2 \le i \le n+1.$$

Note that

$$\tau_n = \frac{n\log(n+1)}{2} = \frac{n\log n}{2} + O(1).$$

This implies $a_{n,1}^{-1} = o(\tau_n)$ and, by Theorem 3.5, $\mathcal{A}$ presents a cutoff. For a window size, consider a partition $\mathcal{P}_n = \{a_{n,1}(1+\epsilon_n)^i : i \ge 0\}$, where $\epsilon_n = 1/\log n$. A simple calculation shows that

$$\tau_{\mathcal{A}}(\mathcal{P}_n) \sim \tau_n, \quad \frac{-\log \epsilon_n}{a_{n,1}\tau_n} = o(1).$$

By Theorem 3.5, $\mathcal{A}$ has a (t_n, b_n) cutoff, where

$$t_n = \frac{n\log n}{2}, \quad b_n = n(\log\log n)^2.$$

Let $\mathcal{A}' = \{a'_{n,i} : 1 \le i \le n, n \ge 1\}$, where $a'_{n.i} = a_{n,2}$ for all $1 \le i \le n$. Then $\mathcal{A} = \{a_{n,1} : n \ge 1\} \cup \mathcal{A}'$. As mentioned in Remark 3.7, it can be easily proved that the requirement in Proposition 3.3 is satisfied with $\delta < 1$, and hence $\mathcal{A}$ and $\mathcal{A}'$ share the same optimal cutoff, which is the (t_n, n) cutoff. It is worth noting that $a_{n,1}^{-1}$ is not the optimal window size for the cutoff.

When using Theorem 3.5 to examine the cutoff for a triangular array, one needs to determine the value τ_n or at least provide a lower bound which is not too small.

In the computation of τ_n, one can find that the smallest term is always ignored. This is reasonable because τ_n is supposed to be a critical time and, if there is a cutoff, it makes the first term(in fact, any term or any k terms for fixed k) in the summation of $d_n^{\mathcal{A}}(c\tau_n)$ tend to 0 for all $c > 0$. The following provides some other choices on the selection of critical time and is useful in bounding the mixing time if a triangular array does not present a cutoff.

Proposition 3.4. *Let $\mathcal{A} = \{a_{n,i} : 1 \le i \le k_n, n \ge 1\}$ be a triangular array of positive numbers satisfying $a_{n,i} \le a_{n,i+1}$ for $1 \le i < k_n$ and $n \ge 1$. Let $d_n^{\mathcal{A}}(\cdot)$ be the function defined in (3.2) and $t_n^{\mathcal{A}}(\cdot)$ be the quantity defined in Definition 3.2. For $n \ge 1$ and $\alpha \ge 0$, set*

$$\tau_n(\alpha) = \max_{1 \le i \le k_n} \left\{ \frac{\log(i+\alpha)}{2a_{n,i}} \right\}.$$

Then $\tau_n(\alpha) \le \tau_n(\beta)$ for $0 \le \alpha < \beta$, and

$$c^{-1} t_n^{\mathcal{A}}\left((c-1)^{-1}\alpha^{1-c}\right) \le \tau_n(\alpha) \le t_n^{\mathcal{A}}\left((1+\alpha)^{-1}\right) \quad \forall c > 1, \alpha > 0,$$

where the second inequality also holds for $\alpha = 0$.

Proof. Let $\alpha \ge 0$ and, for $n \ge 1$, $1 \le i_n \le k_n$ be such that $\tau_n(\alpha) = \frac{\log(i_n+\alpha)}{2a_{n,i_n}}$. Then a calculation shows

$$\begin{aligned} d_n^{\mathcal{A}}(\tau_n(\alpha)) &= \sum_{i=1}^{k_n} e^{-2\tau_n(\alpha)a_{n,i}} \ge \sum_{i=1}^{i_n} \exp\left\{ -\frac{a_{n,i}\log(i_n+\alpha)}{a_{n,i_n}} \right\} \\ &\ge \frac{i_n}{i_n+\alpha} \ge \frac{1}{1+\alpha} \end{aligned}$$

This proves the second inequality.

For the first inequality, by definition, $\tau_n(\alpha) \ge \frac{\log(i+\alpha)}{2a_{n,i}}$ for all $1 \le i \le k_n$. This implies that for $c > 1$ and $\alpha > 0$,

$$d_n^{\mathcal{A}}(c\tau_n(\alpha)) \le \sum_{i=1}^{k_n} e^{-c\log(i+\alpha)} \le \int_\alpha^\infty t^{-c}dt = \frac{1}{(c-1)\alpha^{c-1}}.$$

□

Corollary 3.3. *For $\alpha \geq 0$, let $\tau_n(\alpha)$ be the quantity defined in Proposition 3.4. Then Theorem 3.5 remains true as τ_n is replaced by $\tau_n(\alpha)$.*

3.4 The ℓ^2-cutoff for normal Markov chains

In this section, we translate all results in the previous sections into the case of Markov chains through the following lemma.

Lemma 3.5. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of normal and irreducible Markov chains and set*

$$\Lambda = \{\lambda_{n,i} : 1 \leq i \leq |\mathcal{X}_n| - 1, n \geq 1\}, \quad \mathcal{B} = \{b_{n,i} : 1 \leq i \leq |\mathcal{X}_n| - 1, n \geq 1\}$$

be triangular arrays, where $b_{n,i} = -\log \beta_{n,i}$, $\lambda_{n,0} = 0 < \lambda_{n,1} \leq \cdots \leq \lambda_{n,|\mathcal{X}_n|-1}$ are the real parts of eigenvalues of $I - K_n$ and $\beta_{n,0} = 1 \geq \beta_{n,1} \geq \cdots \geq \beta_{n,|\mathcal{X}_n|-1}$ are the absolute values of eigenvalues of K_n. Assume that $|\mathcal{X}_n| \to \infty$ and, for $n \geq 1$, there exists a finite group G_n acting transitively on $\mathcal{X}_n$ such that

$$K_n(gx, gy) = K_n(x, y), \quad \forall x, y \in \mathcal{X}_n, g \in G_n. \tag{3.14}$$

Then:

1. *(1) The family $\mathcal{F}_c$ presents a ℓ^2-cutoff if and only if the array Λ presents a cutoff.*

2. *(2) Assume further that K_n is aperiodic and $t_n^{\mathcal{B}}(\epsilon) \to \infty$ for some $\epsilon > 0$. Then the family $\mathcal{F}_d$ presents a ℓ^2-cutoff if and only if the array $\mathcal{B}$ presents a cutoff.*

   In particular,

3. *(3) $\mathcal{F}_c$ presents a (t_n, b_n) ℓ^2-cutoff if and only if $\mathcal{B}$ has a (t_n, b_n) cutoff.*

4. *(4) Assume that $\inf_{n\geq 1} b_n > 0$. Then $\mathcal{F}_d$ presents a (t_n, b_n) ℓ^2-cutoff if and only if $\mathcal{B}$ has a (t_n, b_n) cutoff.*

Proof. (1) and (3) are immediate from the first result of Lemma 3.1. For (2), note that the ℓ^2-distance in discrete-time cases is given by

$$\|k_{n,x}^m - 1\|_2^2 = \sum_{i=1}^{|\mathcal{X}_n|-1} e^{-2mb_{n,i}} = d_n^{\mathcal{B}}(m).$$

One direction of the second part is then an immediate result of the above identity(that is, $\mathcal{B}$ presents a cutoff $\Rightarrow \mathcal{F}_d$ has a ℓ^2-cutoff). For the other direction, assume that $\mathcal{F}$ presents a ℓ^2-cutoff with critical time t_n. By assumption, one has $t_n \to \infty$ and then, for any $\delta \in (0,1)$, we may choose $N = N(\delta) > 0$ such that

$$\lceil t_n(1+\delta/2)\rceil \leq t_n(1+\delta),\ \lfloor t_n(1-\delta/2)\rfloor \geq t_n(1-\delta), \quad \forall n \geq N.$$

This implies that for $\delta \in (0,1)$,

$$\lim_{n\to\infty} d_n^{\mathcal{B}}(t_n(1+\delta)) \leq \lim_{n\to\infty} \|k_{n,x}^{\lceil t_n(1+\delta/2)\rceil} - 1\|_2^2 = 0$$

and

$$\lim_{n\to\infty} d_n^{\mathcal{B}}(t_n(1-\delta)) \leq \lim_{n\to\infty} \|k_{n,x}^{\lfloor t_n(1-\delta/2)\rfloor} - 1\|_2^2 = \infty.$$

(4) can be proved by a similar argument as above. □

For a detailed discussion of the cutoff for an array, please refer to section 3.3. The following theorem is implied by Lemma 3.5 and Theorem 3.5.

Theorem 3.6. *Let $\mathcal{F}$ and Λ be as in Lemma 3.5 and set*

$$t_n = \max\left\{\frac{\log i}{2\lambda_{n,i}} : 1 \leq i \leq |\mathcal{X}_n| - 1\right\}.$$

Assume that $|\mathcal{X}_n| \to \infty$ and, for $n \geq 1$, there exists a finite group acting transitively on G_n such that (3.14) holds. Then the family $\mathcal{F}_c$ presents a ℓ^2-cutoff if and only if

$$\lim_{n\to\infty} t_n\lambda_{n,1} = \infty.$$

If $\mathcal{F}_c$ presents a ℓ^2-cutoff, then the critical time is t_n.

Moreover, assume further that $(\epsilon_n)_1^\infty$ is a sequence converging to 0 such that, for some sequence $(j_n)_1^\infty$,

$$\lim_{n\to\infty} \frac{\lambda_{n,j_n}\log(\epsilon_n^{-1})}{\lambda_{n,1}\log N_{j_n}(n,\epsilon_n)} = 0,$$

where

$$N_j(n,\epsilon) = \{j \le i \le |\mathcal{X}_n| - 1 : \lambda_{n,i} < (1+\epsilon)\lambda_{n,j}\}.$$

Then $\mathcal{F}_c$ presents a (t_n, b_n) ℓ^2-cutoff with

$$b_n = \max\left\{\epsilon_n t_n, \frac{-\log \epsilon_n}{\lambda_{n,1}}\right\}.$$

The next theorem is a version of Theorem 3.6 for discrete-time cases.

Theorem 3.7. *Let $\mathcal{F}$ and $\mathcal{B}$ be as in Lemma 3.5 and set*

$$t_n = \max\left\{\frac{\log i}{2b_{n,i}} : 1 \le i \le |\mathcal{X}_n| - 1\right\}.$$

Assume that $|\mathcal{X}_n| \to \infty$, $T_2^d(K_n,\epsilon) \to \infty$ for some $\epsilon > 0$ and, for $n \ge 1$, there exists a finite group G_n acting transitively on $\mathcal{X}_n$ such that (3.14) *holds. Then Theorem 3.6 remains true for $\mathcal{F}_d$ if one replaces $\lambda_{n,i}$ with $b_{n,i}$ for $1 \le i \le |\mathcal{X}_n| - 1$ and $n \ge 1$.*

The following corollary treats the case where the Markov kernel possesses a large multiplicity of the spectral gap.

Corollary 3.4. *Let $\mathcal{F}$ and Λ be as in Lemma 3.5 and m_n be the multiplicity of the spectral gap of K_n. Assume that m_n tends to infinity. Then $\mathcal{F}_c$ presents a ℓ^2-cutoff with critical time $t_n = \max\left\{\frac{\log i}{2\lambda_{n,i}} : 1 \le i \le |\mathcal{X}_n| - 1\right\}$.*

In particular, for any sequence $\epsilon_n > 0$ satisfying

$$\epsilon_n \to 0, \quad \log(\epsilon_n^{-1}) = o(\log m_n),$$

the family $\mathcal{F}_c$ has a (t_n, b_n) ℓ^2-cutoff, where

$$b_n = \max\left\{\epsilon_n t_n, \frac{-\log \epsilon_n}{\lambda_{n,1}}\right\}.$$

The above conclusion also applies for $\mathcal{F}_d$ if one assumes $T_2^d(K_n, \epsilon) \to \infty$ for some $\epsilon > 0$, redefines m_n as the multiplicity of $b_{n,1}$ and replaces $\lambda_{n,i}$ with $b_{n,i}$, where $\mathcal{B} = \{b_{n,i} : 1 \le i \le |\mathcal{X}_n| - 1, n \ge 1\}$ is the array in Lemma 3.5.

In the following, we use the random walk on a hypercube as an illustration of the above results.

Example 3.3. For $n \ge 1$, let $\mathcal{X}_n = (\mathbb{Z}_2)^n$, $\pi_n \equiv 2^{-n}$ and K_n is a Markov kernel on $\mathcal{X}_n$ given by

$$K_n(x, y) = \begin{cases} \frac{1}{n+1} & \text{if } x - y = e_{n,i}, 0 \le i \le n \\ 0 & \text{otherwise} \end{cases}$$

where $e_{n,i}$ is an element in $\mathcal{X}_n$ whose entries are all zero except the ith coordinate and $e_{n,0} = 0$. By the theory of group representations, the functions $\{\phi_x\}_{x \in \mathcal{X}_n}$, which are defined by

$$\phi_x(y) = (-1)^{x \cdot y} \quad \forall y \in \mathcal{X}_n,$$

where $x \cdot y = \sum_{i=1}^n x_i y_i$ for $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$, are eigenfunctions of K_n with corresponding eigenvalues $\{\beta_x\}_{x \in \mathcal{X}_n}$ given by

$$1 - \beta_x = \frac{2(x_1 + \cdots + x_n)}{n+1}.$$

This implies that $\{\lambda_{n,i} : 1 \le i \le 2^n - 1\}$ contains $\frac{2j}{n+1}$ with multiplicity $\binom{n}{j}$ for $1 \le j \le n$. Since the spectral gap of K_n has multiplicity n, by Corollary 3.4, the family $\mathcal{F}_c$ presents a ℓ^2-cutoff. For a critical time and a window size, a simple computation shows $\binom{n}{1} + \cdots + \binom{n}{j} \le n^j$ and

$$t_n = \max_{1 \le i \le 2^n - 1}\left\{\frac{\log i}{2\lambda_{n,i}}\right\} = \frac{(n+1)\log n}{4}.$$

This implies

$$\|h^x_{n,t_n} - 1\|_2^2 = \sum_{i=1}^{2^n-1} e^{-2\lambda_{n,i}t_n} = \left(1+\frac{1}{n}\right)^n - 1 \sim e-1.$$

By Proposition 3.2, $\mathcal{F}_c$ has a (t_n, n) ℓ^2-cutoff.

For $\mathcal{F}_d$, note that, by the above computation, the absolute values of nonzero eigenvalues of K_n are contained in the set $\{1-\frac{2j}{n+1} : 0 \le j < \frac{n+1}{2}\}$, where $1-\frac{2j}{n+1}$ has multiplicity $\binom{n+1}{j}$ for $1 \le j < \frac{n+1}{2}$. One can easily check that $T_2^d(K_n, 1) \to \infty$. Then, by Corollary 3.4, $\mathcal{F}_d$ presents a ℓ^2-cutoff. By the concavity of the logarithmic function, one has

$$\log\left(1-\frac{2j}{n+1}\right) \le -\frac{2j}{n+1}.$$

Furthermore, it has been shown in the previous paragraph that $\binom{n+1}{1} + \binom{n+1}{2} + \cdots + \binom{n+1}{j} \le (n+1)^j$. Then, by Theorem 3.7, the ℓ^2-critical time for $\mathcal{F}_d$ is given by

$$\max_{1\le i\le 2^n-1}\left\{\frac{\log i}{2b_{n,i}}\right\} = \frac{(n+1)\log n}{4} + O(\log n).$$

For a window of the ℓ^2-cutoff for $\mathcal{F}_d$, let t_n be the quantity defined above. A simple computation shows that

$$\max_{x\in\mathcal{X}_n} \|k^{\lceil t_n\rceil}_{n,x} - 1\|_2^2 = \sum_{j=1}^{\lfloor\frac{n+1}{2}\rfloor} \binom{n+1}{j}\left(1-\frac{2j}{n+1}\right)^{t_n} \le \left(1+\frac{1}{n}\right)^n - 1 \sim e-1.$$

Since $\mathcal{F}_d$ is proved in [14] to present a strongly optimal (t_n, n) ℓ^1-cutoff. This implies that

$$\liminf_{n\to\infty} \max_{x\in\mathcal{X}_n} \|k^{\lfloor t_n\rfloor}_{n,x} - 1\|_2 \ge \liminf_{n\to\infty} \max_{x\in\mathcal{X}_n} \|K^{\lfloor t_n\rfloor}_{n,x} - \pi_n\|_{\mathrm{TV}} > 0.$$

Combining all above, one has $|T_2^d(K_n, e-1) - t_n| = O(1)$. By Theorem 2.2, the family $\mathcal{F}_d$ presents a (t_n, n) ℓ^2-cutoff.

From the above discussion, $\mathcal{F}_d$ and $\mathcal{F}_c$ both present a (t_n, n) ℓ^p-cutoff for $1 < p \le 2$. By the strong optimality of the ℓ^1-cutoff and Proposition 2.4, these cutoffs are optimal.

Example 3.4. For $n \ge 1$, let $(\mathcal{X}_n, K_n, \pi_n)$ be a finite Markov chain, where $\mathcal{X}_n = \mathbb{Z}_{2^n}$ and $\pi_n \equiv 2^{-n}$. Let κ_n be a probability measure on $\mathcal{X}_n$ defined by

$$\kappa_n(\pm 2^i) = \frac{1}{2n} \quad \forall 1 \le i \le n,$$

and K_n be a Markov kernel given by $K_n(x,y) = \kappa_n(x^{-1}y)$ for all $x, y \in \mathcal{X}_n$. By group representation theory, these functions $(\rho_x)_{x\in\mathcal{X}_n}$, where $\rho_x(y) = e^{2\pi ixy/2^n}$ for $y \in \mathcal{X}_n$ and $i = \sqrt{-1}$, are eigenvectors of K_n. For $x \in \mathcal{X}_n$, we denote β_x as the eigenvalue corresponding to the eigenvector ρ_x. Then β_x can be determined by

$$\beta_x = \sum_{y\in\mathcal{X}_n} \kappa_n(y)\rho_x(y) = \frac{1}{n}\sum_{j=1}^{n} \cos\left(2\pi x 2^{j-n}\right).$$

An observation from the above formula is that if $x, y \in \mathcal{X}_n$ and $x + y = 2^n$, then $\beta_x = \beta_y$. This implies that except β_0 and $\beta_{2^{n-1}}$, every eigenvalue has multiplicity at least 2.

For convenience, we identify $x \in \mathcal{X}_n$ with $(x_n, ..., x_1)$ if $x = \sum_{i=1}^n x_i 2^{i-1}$. Then, for $x = (x_n, ..., x_1) \in \mathcal{X}_n$,

$$\begin{aligned} 1 - \beta_x &= \frac{1}{n}\sum_{j=1}^{n}\left[1 - \cos\left(\pi\sum_{i=1}^{n} x_i 2^{i+j-n}\right)\right] \\ &= \frac{1}{n}\sum_{j=1}^{n-1}\left[1 - \cos\left(\pi\sum_{i=1}^{j} x_i 2^{i-j}\right)\right] \\ &= \frac{2}{n}\sum_{j=1}^{n-1}\sin^2\left(\pi\sum_{i=1}^{j} x_i 2^{i-j-1}\right) \end{aligned} \tag{3.15}$$

Note that for $x, y \in \mathcal{X}_n \setminus \{0, 2^{n-1}\}$, if $x + y = 2^n$, then either x or y(but not both) has the most right two 1s are contingent. In addition to the formula, one can see that for $x \in \mathcal{X}_n \setminus \{0, 2^{n-1}\}$, $1-\beta_x \ge \frac{3}{n}$. Since $1-\beta_0 = 0$ and $\beta_{2^{n-1}} = \frac{2}{n}$, the spectral gap of K_n is equal to $\frac{2}{n}$.

To determine the ℓ^2-cutoff, we need to study the distribution of eigenvalues. Note that if $x = (x_n, ..., x_1) \in \mathcal{X}_n$ satisfies $x_l = 1$ and $x_{l+1} = x_{l+2} = \cdots = x_{k-1} = x_k = 0$, then

$$2^{l-j-1} \le \sum_{i=1}^{j} x_i 2^{i-j-1} \le 2^{l-j} \quad \forall l+2 \le j \le k.$$

This implies, for $k \ge l+2$,

$$\frac{1}{8} \le \sum_{j=l+2}^{k} \sin^2\left(\pi \sum_{i=1}^{j} x_i 2^{i-j-1}\right) \le 1,$$

where the first inequality uses the concavity of the sine function on $(0, \pi/2)$ and the second inequality uses the convexity of $\sin^2 t$ on the region $(0, \pi/4)$. Hence we have, for $l \ge 1$,

$$\sum_{j=l+1}^{k} \sin^2\left(\pi \sum_{i=1}^{j} x_i 2^{i-j-1}\right) \in \begin{cases} (5/8, 2) & \text{if } k \ge l+2 \\ (1/2, 1) & \text{if } k = l+1 \end{cases}.$$

A similar proof also applies for the case $x_l = 0$ and $x_{l+1} = x_{l+2} = \cdots = x_{k-1} = x_k = 1$.

For $x \in \mathcal{X}_n$, let $N(x)$ to be the nonnegative number $\sum_{i=1}^{n-1} |x_{i+1} - x_i|$. Then the above calculations show that, for $x \ne 0$,

$$\frac{N(x) + 2x_1}{n} \le 1 - \beta_x \le \frac{4N(x) + 2x_1}{n}.$$

Let $\lambda_{n,0} = 0 < \lambda_{n,1} \le \cdots$ be an arrangement of $\{1 - \beta_x : x \in \mathcal{X}_n\}$. Since $\{x \in \mathcal{X}_n : N(x) = i\} = 2\binom{n-1}{i}$ for $0 \le i \le n-1$, one has $\lambda_{n,n} \le 4/n$. Combining all above, we get

$$\lambda_{n,1} \max_{i \ge 1} \left\{ \frac{\log i}{\lambda_{n,i}} \right\} \ge \frac{\log n}{2} \to \infty, \quad \text{as } n \to \infty.$$

By Theorem 3.6, the family $\mathcal{F}_c$, where $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$, presents a ℓ^2-cutoff.

For the discrete-time case, applying the fact in (2.20) implies $|\beta_x| \leq 1 - \frac{2}{n}$ for $x \neq 0$. Let $\mathcal{B} = \{b_{n,i} : 1 \leq i \leq |\mathcal{X}_n| - 1, n \geq 1\}$ be the triangular array defined in Lemma 3.5. Then, by the above fact and the computation in the previous paragraph, one has $b_{n,1} = -\log(1 - \frac{2}{n})$ and $b_{n,n} \geq -\log(1 - \frac{4}{n})$. Combining both facts, we get, by Lemma 3.1, $T_2^d(K_n, 1) \geq (\frac{n}{4} - 1) \log n$ for all $n \geq 4$ and

$$b_{n,1} \max_{i \geq 1} \left\{ \frac{\log i}{b_{n,i}} \right\} \geq \frac{\log(1 - 2/n) \log n}{\log(1 - 4/n)} \to \infty, \quad \text{as } n \to \infty.$$

By Theorem 3.7, $\mathcal{F}_d$ presents a ℓ^2-cutoff, and by Theorem 2.11, $\mathcal{F}_d$ and $\mathcal{F}_c$ have the same ℓ^2-critical time.

3.5 The continuous-time Random walk on a product space

In this section, the underlying Markov chains are of the following form. For $n \geq 1$, let k_n be a positive integer and, for $1 \leq i \leq k_n$, let $\mathcal{X}_{n,i}$ be a finite set and $K_{n,i}$ be an irreducible Markov kernel on $\mathcal{X}_{n,i}$ with stationary distribution $\pi_{n,i}$. Let $\mathcal{Y}_n = \prod_{i=1}^{k_n} \mathcal{X}_{n,i}$ and P_n be a Markov kernel on $\mathcal{Y}_n$ defined by

$$P_n(x, y) = p_{n,0}\delta(x, y) + \sum_{i=1}^{k_n} p_{n,i}\delta_i(x, y)K_{n,i}(x_i, y_i), \tag{3.16}$$

where

$$\sum_{i=0}^{k_n} p_{n,i} = 1, \quad \delta_i(x, y) = \prod_{\substack{j=1 \\ j \neq i}}^{k_n} \delta(x_j, y_j)$$

for $x = (x_1, ..., x_{k_n}), y = (y_1, ..., y_{k_n}) \in \mathcal{Y}_n, 1 \leq i \leq k_n$, and $\delta(u, v)$ equals to 1 if $u = v$ and 0 otherwise. In this case, the probability measure $\mu_n = \bigotimes_{i=1}^{k_n} \pi_{n,i}$ on $\mathcal{Y}_n$ is a stationary distribution of P_n. Note that P_n is irreducible if and only if $p_{n,i} > 0$ for all $1 \leq i \leq k_n$. The following proposition, which can be found in [30], is a useful fact in dealing with the ℓ^2-norm and the ℓ^2-distance for a Markov kernel satisfying (3.16).

Proposition 3.5. *Let* $\{(\mathcal{X}_i, K_i, \pi_i)\}_1^n$ *be irreducible Markov chains and set* $\mathcal{X} = \prod_1^n \mathcal{X}_i$, $\pi = \bigotimes_1^n \pi_i$ *and*

$$K(x,y) = p_0\delta(x,y) + \sum_{i=1}^n p_i\delta_i(x,y)K_i(x_i,y_i),$$

where $p_0 + \cdots + p_n = 1$, $\delta_i(x,y) = \prod_{j\neq i}\delta(x_j,y_j)$, *and* $\delta(u,v) = 1$ *if* $u = v$ *and* $\delta(u,v) = 0$ *otherwise. Let* $H_{i,t}$ *and* H_t *be the associated continuous-time semigroups of* K_i *and* K*. Then for* $t \geq 0$ *and* $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathcal{X}$,

$$H_t(x,y) = \prod_{i=1}^n H_{n,p_it}(x_i,y_i)$$

In particular, the ℓ^2*-norm of* h_t^x *satisfies* $\|h_t^x\|_2^2 = \prod_{i=1}^n \|h_{n,p_it}^{x_i}\|_2^2$, *or equivalently,*

$$\|h_t^x - 1\|_2^2 = \prod_{i=1}^n \left(1 + \|h_{n,p_it}^{x_i} - 1\|_2^2\right) - 1,$$

and the ℓ^2*-mixing time is bounded by*

$$T_2^c\left(K, \sqrt{(1+\epsilon)^n - 1}\right) \leq \max_{1\leq i\leq n}\left\{\frac{T_2^c(K_i,\epsilon)}{p_i}\right\} \leq T_2^c(K,\epsilon),$$

if $p_i > 0$ *for all* $1 \leq i \leq n$.

Proof. By a direct computation on $H_t(x,y)$. □

Remark 3.8. Note that the lower bound of the mixing time given in Proposition 3.5 can be much smaller than $T_2^c(K,\epsilon)$ if n is large.

3.5.1 The ℓ^2-cutoff for product chains

We use the following setting for the remaining of this chapter. For any finite sequences $A = \{a_1, .., a_n\}$ and $B = \{b_1, ..., b_m\}$, we define $A \sqcup B$ as a sequence

$\{c_1, .., c_{n+m}\}$, where $c_i = a_i$ for $1 \le i \le n$ and $c_{n+j} = b_j$ for $1 \le j \le m$. For finite sequences $(A_i)_1^n$, we define $A_1 \sqcup \cdots \sqcup A_n$ by iterating the following identity

$$A_1 \sqcup \cdots \sqcup A_n = (A_1 \sqcup \cdots \sqcup A_{n-1}) \sqcup A_n,$$

and for a short hand, we set

$$\bigsqcup_{i=1}^{n} A_i = A_1 \sqcup \cdots \sqcup A_n.$$

Let $c \in \mathbb{R}$ and $A = \{a_1, ..., a_n\}$. We define $cA = \{ca_1, ..., ca_n\}$. If $\mathcal{A}$ is a triangular array, we denote, for $n \ge 1$, $\mathcal{A}_n$ as the nth row of $\mathcal{A}$.

Note that, by the discussion before Lemma 3.1, the normality of a Markov kernel is sufficient for us to express the ℓ^2-distance as a function of its eigenvalues and eigenvectors. For $n \ge 1$ and $1 \le i \le k_n$, suppose that $K_{n,i}$ satisfies the assumption of Lemma 3.1 and let, for $0 \le j \le |\mathcal{X}_{n,i}| - 1$, $\lambda_{i,j}^n$ be the real parts of eigenvalues of $I - K_{n,i}$. Let P_n be the Markov kernel given in (3.16) and $\widetilde{H}_{n,t}$ be the continuous-time semigroup associated to P_n. Then, by Proposition 3.5, one has

$$\max_{y \in \mathcal{Y}_n} \|\widetilde{h}_{n,t}^y - 1\|_2^2 = \sum_{(j_1, \cdots, j_{k_n}) \in \prod_{i=1}^{k_n} \mathbb{Z}_{|\mathcal{X}_{n,i}|}} \exp\left\{ -2t \sum_{i=0}^{k_n} p_{n,i} \lambda_{i,j_i}^n \right\}.$$

Determining the ℓ^2-cutoff for the family $\{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ by using Theorem 3.5 with the above identity can be very complicated since there are $|\mathcal{X}_{n,1}| \times \cdots \times |\mathcal{X}_{n,k_n}| - 1$ terms needed to be considered for the chain P_n. The following theorem gives a reduction on the above summation by ignoring a bunch of eigenvalues of P_n, where in the end, there are only $|\mathcal{X}_{n,1}| + \cdots + |\mathcal{X}_{n,k_n}| - k_n$ terms remained.

Theorem 3.8. *Let $\mathcal{F} = \{(\mathcal{X}_{n,i}, K_{n,i}, \pi_{n,i}) : 1 \le i \le k_n, n \ge 1\}_1^\infty$ be a family of normal and irreducible Markov chains, and $\mathcal{G} = \{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ be a family induced from $\mathcal{F}$ by setting $\mathcal{Y}_n = \prod_{i=1}^n \mathcal{X}_{n,i}$, $\mu_n = \bigotimes_{i=1}^\infty \pi_{n,i}$ and defining P_n by*

(3.16) *with* $p_{n,i} > 0$ *for* $1 \le i \le k_n, n \ge 1$ *and* $\sum_{i=0}^{k_n} p_{n,i} = 1$. *For* $n \ge 1$ *and* $1 \le i \le k_n$, *let* $\Lambda_i^n = \{\lambda_{i,j}^n : 1 \le j \le |\mathcal{X}_{n,i}| - 1\}$ *be the set consisting of the real parts of nonzero eigenvalues of* $I - K_{n,i}$, *and* Γ *be a triangular array defined by*

$$\Gamma_n = \bigsqcup_{i=1}^{k_n} (p_{n,i}\Lambda_i^n) \quad \forall n \ge 1.$$

Assume that

$$\lim_{n\to\infty} \prod_{i=1}^{k_n} |\mathcal{X}_{n,i}| = \infty$$

and, for $n \ge 1$ *and* $1 \le i \le k_n$, *there is a finite group* $G_{n,i}$ *acting transitively on* $\mathcal{X}_{n,i}$ *such that*

$$K_{n,i}(gx, gy) = K_{n,i}(x, y), \quad \forall x, y \in \mathcal{X}_{n,i}, g \in G_{n,i}. \tag{3.17}$$

Then:

(1) $\mathcal{G}_c$ *presents a* ℓ^2*-cutoff if and only if* Γ *presents a cutoff. In particular, if both* $\mathcal{G}_c$ *and* Γ *presents a cutoff, then their critical times are the same.*

(2) $\mathcal{G}_c$ *has a* (t_n, b_n) ℓ^2*-cutoff if and only if* Γ *has a* (t_n, b_n) *cutoff.*

(3) $\mathcal{G}_c$ *has a strongly optimal* (t_n, b_n) ℓ^2*-cutoff if and only if* Γ *has a strongly optimal* (t_n, b_n) ℓ^2*-cutoff.*

Remark 3.9. Let $\mathcal{G}$ and Γ be as in Theorem 3.8. By Lemma 3.2, if one of $\mathcal{G}_c$ and Γ has an optimal window size, then the other has, too, and their optimal window sizes are of the same order.

Proof of Theorem 3.8. Let $H_{n,t}$ be the continuous-time semigroup associated to P_n and set $d_n(t) = \max_{y\in\mathcal{Y}_n} \|h_{n,t}^y - 1\|_2^2$. By Lemma 2.2 and Proposition 3.5, one has

$$d_n(t) = \prod_{i=1}^{k_n} \left(1 + \sum_{j=1}^{|\mathcal{X}_{n,i}|-1} e^{-2p_{n,i}\lambda_{i,j}^n t}\right) - 1.$$

Then a simple computation with the fact, $1+x \le e^x$ for $x \in \mathbb{R}$, implies that

$$\forall t \ge 0, \quad d_n^\Gamma(t) \le d_n(t) \le e^{d_n^\Gamma(t)} - 1,$$

where the function d_n^Γ is defined in (3.2). This is sufficient to (2) and (3). For (1), let $t_n^\Gamma(\cdot)$ be the mixing time defined in Definition 3.2. Then, for $\epsilon > 0$, one may choose $N(\epsilon) > 0$ such that

$$t_n^\Gamma(\epsilon^2) \le T_2^c(P_n, \epsilon) \le t_n^\Gamma(\log(1+\epsilon^2)) \quad \forall n \ge N(\epsilon).$$

Hence, by Theorem 2.4 and Proposition 3.1, the above inequality proves (1). □

Remark 3.10. Note that, in Theorem 3.8, the ℓ^2-distance of the chain $(\mathcal{Y}_n, P_n, \mu_n)$ and the function d_n^Γ defined in (3.2) are related as follows.

$$d_n^\Gamma(t) \le \max_{y \in \mathcal{Y}_n} d_{\mu_n,2}(H_{n,t}^y, \mu_n) \le e^{d_n^\Gamma(t)} - 1, \quad \forall t \ge 0,$$

where $H_{n,t} = e^{-t(I-P_n)}$.

Example 3.5 *(Continuation of Example 3.3).* Note that those Markov chains in Example 3.3 are of the form $(\mathcal{Y}_n, P_n, \mu_n)$ in Theorem 3.8, with $k_n = n$ and, for $1 \le i \le n$, $\mathcal{X}_{n,i} = \mathbb{Z}_2$, $p_{n,i} = \frac{1}{n+1}$ and

$$K_{n,i} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

In this case, $\Gamma_n = \{\gamma_{n,1}, ..., \gamma_{n,n}\}$, where $\gamma_{n,i} = \frac{2}{n+1}$ for all $1 \le i \le n$. Clearly, one has $d_n^\Gamma(t) = ne^{-4t/(n+1)}$. This implies that Γ has a strongly optimal $(\frac{n \log n}{4}, n)$ ℓ^2-cutoff and, by Theorem 3.8, so does $\mathcal{G}_c$.

In the following, we consider a generalization of the model in Example 3.3. Let $\mathcal{X}_{n,i} \equiv \mathbb{Z}_2$ and $\mathcal{Y}_n = \prod_{i=1}^n \mathcal{X}_{n,i} = (\mathbb{Z}_2)^n$. Note that, for any 2×2 stochastic matrix

K, the identity in (3.17) is satisfied if and only if K is symmetric. For $n \geq 1$ and $1 \leq i \leq n$, let $a_{n,i} \in (0,1]$, $b_{n,i} = 1 - a_{n,i}$ and

$$K_{n,i} = \begin{pmatrix} b_{n,i} & a_{n,i} \\ a_{n,i} & b_{n,i} \end{pmatrix}. \tag{3.18}$$

Let $\Lambda_i^n = \{\lambda_i^n : 1 \leq i \leq n,\ n \geq 1\}$ and $\Gamma = \{\gamma_{n,i} : 1 \leq i \leq n,\ n \geq 1\}$ be as in Theorem 3.8. Then a simple computation shows $\lambda_i^n = 2a_{n,i}$ and $\gamma_{n,i} = 2p_{n,i}a_{n,i}$ for $1 \leq i \leq n$.

Theorem 3.9. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : 1 \leq i \leq n,\ n \geq 1\}$ and $\mathcal{G} = \{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ be families of Markov chains, where*

$$\mathcal{X}_{n,i} \equiv \mathbb{Z}_2, \quad \mathcal{Y}_n = (\mathbb{Z}_2)^n, \quad \pi_n \equiv 2^{-1}, \quad \mu_n = 2^{-n}.$$

Let $K_{n,i}$ be a matrix of the form in (3.18) *with $a_{n,i} > 0$ and P_n be given by* (3.16) *with $\inf_i p_{n,i} > 0$ for $n \geq 1$. Assume that $b_{n,1} \leq \cdots \leq b_{n,n}$ is a rearrangement of $p_{n,1}a_{n,1}, ..., p_{n,n}a_{n,n}$. Then the family $\mathcal{G}_c$ presents a ℓ^2-cutoff if and only if*

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \left\{ \frac{b_{n,1}\log i}{b_{n,i}} \right\} = \infty.$$

In particular, if the above limit holds, then $\mathcal{G}_c$ presents a ℓ^2-cutoff with critical time

$$\max_{1\leq i\leq n} \left\{ \frac{\log i}{4b_{n,i}} \right\}.$$

Proof. By Theorem 3.5 and Theorem 3.8. □

Most of the time, the underlying Markov kernels are restricted to some specific cases. The following two corollaries concern two of them.

Corollary 3.5. *Let $\mathcal{G}$ be as in Theorem 3.9. Assume that $a_{n,i} \equiv a > 0$ and $p_{n,i} \leq p_{n,i+1}$ for $1 \leq i < n$ and $n \geq 1$. Then $\mathcal{G}_c$ presents a ℓ^2-cutoff if and only if*

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \left\{ \frac{p_{n,1}\log i}{p_{n,i}} \right\} = \infty.$$

Moreover, if the above limit holds, then the family $\mathcal{G}_c$ presents a ℓ^2-cutoff with critical time

$$\max\left\{\frac{\log i}{4ap_{n,i}} : 1 \le i \le n\right\}.$$

Proof. Immediate from Theorem 3.8. □

Corollary 3.6. *Let $\mathcal{G}$ be the same as in Theorem 3.9. Assume that*

$$\max_{1\le i\le n}\{p_{n,i}a_{n,i}\} = O\left(\min_{1\le i\le n}\{p_{n,i}a_{n,i}\}\right).$$

Then $\mathcal{G}_c$ has a ℓ^2-cutoff whose critical time is of the same order as $(\log n)/(p_{n,n}a_{n,n})$.

Proof. By Theorem 3.8. □

Remark 3.11. Note that the above theorem and corollaries do not provide any window for the cutoff. To obtain one, we may apply Theorem 3.5 or Proposition 3.2 on the triangular array Γ. To refine the size, one may use Proposition 3.3. For the optimality of a window size, Corollary 3.1, Corollary 3.2 and Proposition 2.4 provide some criterions to examine the obtained window size.

The following are two simple examples for an illustration of the previous results.

Example 3.6. Let $\mathcal{G}$ be the family in Theorem 3.9 and

$$p_{n,i} = \frac{1}{4n}\left(1+\frac{1}{n}\right)^{i-1} \quad \forall 1 \le i \le n,$$

where $p_{n,0} = 1-(p_{n,1}+\cdots+p_{n,n})$. Assume that $a_{n,i} \equiv a$ for some $a \in (0,1]$.

By Corollary 3.6, $\mathcal{G}_c$ presents a ℓ^2-cutoff. To find a critical time, let s_n be the solution of the following equation.

$$(t+1)^n = t^{n+1} \quad \forall t > 0.$$

Then $\lceil s_n \rceil = \frac{c_n n}{\log n}$, where c_n is a sequence bounded from above and below by positive numbers. A simple computation shows that the critical time is given by

$$\max_{1 \le i \le n} \left\{ \frac{\log i}{4ap_{n,i}} \right\} = \frac{\log(\lceil s_n \rceil)}{4ap_{n,s_n}} \sim \frac{n(\log n - \log\log n)}{a}.$$

Letting $d_n^\Gamma(t)$ be the function defined in (3.2) and $t_n = a^{-1}n(\log n - \log\log n)$ implies that, for $1 \le i \le n$,

$$e^{-4ap_{n,i}t_n} = \left(\frac{\log n}{n}\right)^{(1+n^{-1})^{i-1}} \begin{cases} \le \left(\frac{\log n}{n}\right)^{1+\frac{i-1}{n}} \\ \ge \left(\frac{\log n}{n}\right)^{1+\frac{2(i-1)}{n}} \end{cases}$$

where the inequalities use the facts $\log(1+t) \le t$ for $t > 0$ and $e^t \le 2t$ for $t \in (0,1)$. Note also that for $c > 0$,

$$\begin{aligned} \sum_{i=1}^n \left(\frac{\log n}{n}\right)^{1+\frac{c(i-1)}{n}} &= (\log n) \times \left\{ \frac{1}{n} \sum_{i=1}^n \left(\frac{\log n}{n}\right)^{c\frac{i-1}{n}} \right\} \\ &= (\log n) \left\{ \int_0^1 \left(\frac{\log n}{n}\right)^{ct} dt + O(n^{-1}) \right\} \\ &\sim c^{-1} \quad \text{as } n \to \infty. \end{aligned}$$

This implies

$$\frac{1}{2} \le \liminf_{n\to\infty} d_n^\Gamma(t_n) \le \limsup_{n\to\infty} d_n^\Gamma(t_n) \le 1.$$

By Proposition 3.2 and Corollary 3.2, $\mathcal{G}_c$ has an optimal (t_n, n) ℓ^2-cutoff, where

$$t_n = \frac{n(\log n - \log\log n)}{a}.$$

Example 3.7. Let $\mathcal{G}$ be the family in Theorem 3.9 with $p_{n,i} = n^{-1}$ for $2 \le i \le n$ and

$$n \le p_{n,1}^{-1} = o(n \log n), \quad p_{n,0} = 1 - (p_{n,1} + \cdots + p_{n,n}).$$

Assume that $a_{n,i}$(the quantity defined in (3.18)) and $(\log i)/a_{n,i}$ are both increasing in i for $1 \le i \le n$, and $\inf_n a_{n,1} > 0$. With these assumptions, we have

$$\max_{1 \le i \le n} \left\{ \frac{a_{n,1}p_{n,1}\log i}{a_{n,i}p_{n,i}} \right\} = \frac{a_{n,1}p_{n,1}n\log n}{a_{n,n}} \to \infty \quad \text{as } n \to \infty.$$

By Theorem 3.9, the family $\mathcal{G}_c$ presents a ℓ^2-cutoff with critical time

$$t_n = \frac{n \log n}{4a_{n,n}}.$$

To find a window for the ℓ^2-cutoff, we treat the simplest case. Assume that $a_{n,i} \equiv a$ for some $a \in (0,1]$. Let Γ be the same as in Theorem 3.9 and $d_n^\Gamma(\cdot)$ be the function defined in (3.2). A simple computation implies $d_n^\Gamma(t_n) \sim 1$. By Theorem 3.8, Proposition 3.2 and Corollary 3.2, the family $\mathcal{G}_c$ has an optimal (t_n, n) ℓ^2-cutoff.

Note that the spectral gap of P_n is $2ap_{n,1}$. If $p_{n,1} = (n \log\log n)^{-1}$, then the family $\mathcal{G}_c$ has an optimal window size for the ℓ^2-cutoff, which is of smaller order than the reciprocal of the spectral gap.

3.5.2 The ℓ^2 cutoff for some specific product chains

In the previous section, almost without any rigid assumption, Theorem 3.8 translates the ℓ^2-cutoff for a family of product chains(defined in (3.16)) to the cutoff of a triangular array Γ. The usefulness of Γ comes from that fact that it has fewer entries and is simpler than the triangular array containing the eigenvalues of $(P_n)_1^\infty$. That indeed saves us a lot of time on the determination of cutoffs, but not all families have the luck as in Example 3.5, where only the spectral gaps are involved. In this section, we will put some further assumption on the chains, which is not too difficult to examine, so that the cutoff can be determined by using only the spectral gaps.

Lemma 3.6. *For $n \geq 1$ and $1 \leq i \leq k_n$, let $A_i^n = \{a_{i,j}^n : 1 \leq j \leq l_{n,i}\}$ be a finite sequence of positive numbers, where k_n and $l_{n,i}$ are positive integers, and $\mathcal{B}$ be a* *triangular array defined by*

$$\mathcal{B}_n = \bigsqcup_{i=1}^{k_n} (p_{n,i} A_i^n), \quad \forall n \geq 1.$$

where $p_{n,1}, ..., p_{n,k_n}$ *are positive numbers. Assume that there exists a constant* $C \geq 1$ *such that, for* $n \geq 1$ *and* $1 \leq i \leq k_n$*, the following inequality*

$$f_{n,i} e^{-2a_{i,1}^n t} \leq \sum_{j=1}^{l_{n,i}} e^{-2a_{i,j}^n t} \leq C f_{n,i} e^{-2a_{i,1}^n t}, \tag{3.19}$$

holds for $t \geq (C + \log f_{n,i})(2a_{i,1}^n)^{-1}$*. Let* $\mathcal{C}$ *be a triangular array whose* n*th row has entries*

$$c_{n,i} = p_{n,j} a_{j,1}^n, \quad \text{for } \sum_{l=1}^{j-1} \lceil f_{n,l} \rceil + 1 \leq i \leq \sum_{l=1}^{j} \lceil f_{n,l} \rceil, \ 1 \leq j \leq k_n.$$

Assume that $\lceil f_{n,1} \rceil + \cdots + \lceil f_{n,k_n} \rceil \to \infty$ *as* $n \to \infty$*. Then:*

(1) $\mathcal{B}$ *presents a cutoff if and only if* $\mathcal{C}$ *presents a cutoff. In particular, if both* $\mathcal{B}$ *and* $\mathcal{C}$ *have a cutoff, then their critical time is the same.*

(2) If $\mathcal{C}$ *presents a* (t_n, b_n) *cutoff, then* $\mathcal{B}$ *has a* (t_n, b_n) *cutoff.*

Moreover, if, for some $\delta > 2$*,* (3.19) *holds for* $t \geq (C + \log f_{n,i})(\delta a_{i,1}^n)^{-1}$*, then:*

(3) $\mathcal{B}$ *presents a* (t_n, b_n) *cutoff if and only if* $\mathcal{C}$ *has a* (t_n, b_n) *cutoff.*

(4) $\mathcal{B}$ *has a strongly optimal* (t_n, b_n) *cutoff if and only if* $\mathcal{C}$ *has a strongly optimal* (t_n, b_n) *cutoff.*

Remark 3.12. (1) Let $m_{n,i}$ be the multiplicity of $a_{i,1}^n$ in A_i^n. Then the assumption in (3.19) implies that, for $1 \leq i \leq k_n$ and $n \geq 1$,

$$a_{i,1}^n = \min_j a_{i,j}^n, \quad f_{n,i} \leq m_{n,i} \leq C f_{n,i}.$$

(2) Note that the cutoff for a triangular array $\mathcal{A}$ is equivalent to the cutoff for $c\mathcal{A}$, where c is a positive number(This property can be seen from the equivalence of the ℓ^1-cutoff and the total variation cutoff). By this observation and the assumption $\inf\{f_{n,i} : 1 \le i \le k_n,\ n \ge 1\} > 0$, one may always restrict $f_{n,i}$ to positive integers.

(3) In Lemma 3.6, if the set $\{f_{n,i} : 1 \le i \le k_n,\ n \ge 1\}$ is bounded, then one may always choose $f_{n,i} \equiv 1$. In this case, the requirement (3.19) for $t \ge (C - \log f_{n,i})(\delta a^n_{i,1})^{-1}$ with $\delta > 2$ is equivalent of that with $\delta = 2$.

(4) For an example on the setting in Lemma 3.6, we consider the triangular Γ in Theorem 3.8. Recall that: For $n \ge 1$ and $1 \le i \le k_n$, the set of non-zero eigenvalues of $I - K_{n,i}$ is denoted by $\Lambda^n_i = \{\lambda^n_{i,j} : 1 \le j \le |\mathcal{X}_{n,i}| - 1\}$, where $\lambda^n_{i,1}$ is the spectral gap of $K_{n,i}$. Then the nth row of Γ is defined by

$$\Gamma_n = \{p_{n,i}\lambda^n_{i,j} : 1 \le j \le |\mathcal{X}_{n,i}| - 1,\ 1 \le i \le k_n\},$$

where $p_{n,i}$ is positive. If one replaces $\mathcal{B}$ with Γ in Lemma 3.6, then the nth row of the triangular array $\mathcal{C}$ is equal to

$$\{p_{n,i}\lambda^n_{i,1} : 1 \le i \le k_n\}.$$

This means that only the spectral gap of $K_{n,i}$ is considered.

Proof of Lemma 3.6. By the above remark, one has $a^n_{i,1} = \min_j a^n_{i,j}$ and then, for $n \ge 1$, the smallest entries of $\mathcal{B}_n$ and $\mathcal{C}_n$ are the same. Without loss of generality, we assume that $f_{n,i}$ is a positive integer for $1 \le i \le k_n$ and $n \ge 1$. Note that

$$\forall t \ge 0, \quad d^{\mathcal{C}}_n(t) = \sum_{i=1}^{k_n} f_{n,i} e^{-2p_{n,i}a^n_{i,1}t} \le \sum_{i=1}^{k_n}\sum_{j=1}^{l_{n,i}} e^{-2p_{n,i}a^n_{i,j}t} = d^{\mathcal{B}}_n(t). \tag{3.20}$$

To get an upper bound on $d^{\mathcal{B}}_n(t)$, we fix $\epsilon \in (0, e^{-C})$ and let $t_n = t^{\mathcal{B}}_n(\epsilon) > 0$ be the mixing time of $\mathcal{B}$ defined in Definition 3.2. By (3.20), one has $d^{\mathcal{C}}_n(t_n) \le \epsilon < e^{-C}$, which implies that $p_{n,i}t_n \geq (C+\log f(n,i))(2a^n_{i,1})^{-1}$ for all $1 \leq i \leq k_n$. Hence, we obtain from (3.19) that for $\epsilon \in (0, e^{-C})$,

$$\forall t \geq t^{\mathcal{B}}_n(\epsilon), \quad d^{\mathcal{B}}_n(t) \leq C d^{\mathcal{C}}_n(t). \tag{3.21}$$

By (3.20) and (3.21), one has

$$\forall \epsilon > 0, \quad t^{\mathcal{C}}_n(\epsilon) \leq t^{\mathcal{B}}_n(\epsilon) \quad \text{and} \quad \forall \epsilon \in (0, e^{-C}), \quad t^{\mathcal{B}}_n(\epsilon) \leq t^{\mathcal{C}}_n(\epsilon/C). \tag{3.22}$$

In addition to the fact $b_{n,1} = c_{n,1}$, (1) is proved by Proposition 3.1. Note that (2) can be easily obtained from (3.20) and (3.21).

For (3), assume that $\mathcal{B}$ has a (t_n, b_n) cutoff. By (3.20), one has

$$\lim_{c\to\infty} \limsup_{n\to\infty} d^{\mathcal{C}}_n(t_n + cb_n) = 0.$$

For the lower bound, since $\mathcal{B}$ has a (t_n, b_n) cutoff, we may choose $c_1 > 0$ and $N > 0$ such that

$$d^{\mathcal{B}}_n(t_n + cb_n) \leq e^{-C}, \quad \forall c > c_1,\ n \geq N.$$

This implies that for all $1 \leq i \leq k_n$,

$$p_{n,i}(t_n + cb_n) \geq (C + \log f_{n,i})(2a^n_{i,1})^{-1}.$$

In this case, since $b_n = o(t_n)$, we may choose a $N_1 \geq N$, such that for $c \geq c_1$ and $n \geq N_1$,

$$p_{n,i}(t_n - cb_n) \geq (1+o(1))(C+\log f_{n,i})(2a^n_{i,1})^{-1} \geq (C+\log f_{n,i})(\delta a^n_{i,1})^{-1},$$

for all $1 \leq i \leq k_n$. Hence,

$$d^{\mathcal{B}}_n(t_n - cb_n) \leq C d^{\mathcal{C}}_n(t_n - cb_n), \quad \forall n \geq N_1,\ c > c_1, \tag{3.23}$$

which implies

$$\lim_{c\to\infty} \liminf_{n\to\infty} d^{\mathcal{C}}_n(t_n - cb_n) \geq C^{-1} \lim_{c\to\infty} \liminf_{n\to\infty} d^{\mathcal{B}}_n(t_n - cb_n) = \infty.$$

For (4), note that, by Proposition 3.1, the inequalities, (3.20) and (3.21), are sufficient to prove the strongly optimal cutoff for $\mathcal{C}$ from that for $\mathcal{B}$. The inverse direction is obtained by applying (3.20) and (3.23). □

With the above lemma, we can improve the result of Theorem 3.8 as follows.

Theorem 3.10. *Let $\mathcal{G}$ and $\{\Lambda_i^n : 1 \le i \le k_n,\ n \ge 1\}$ be as in Theorem 3.8. Assume that* (3.17) *holds and there exist a positive number $C > 0$ and, for $1 \le i \le k_n$ and $n \ge 1$, such that*

$$f_{n,i}e^{-2\lambda_{i,1}^n t} \le \sum_{j=1}^{|\mathcal{X}_{n,i}|-1} e^{-2\lambda_{i,j}^n t} \le Cf_{n,i}e^{-2\lambda_{i,1}^n t}, \tag{3.24}$$

for $t \ge (C + \log f_{n,i})(2\lambda_{i,1}^n)^{-1}$. Let Σ be a triangular array whose nth row consists of entries

$$\sigma_{n,i} = p_{n,j}\lambda_{j,1}^n, \quad \textit{for } \sum_{l=1}^{j-1}\lceil f_{n,l}\rceil + 1 \le i \le \sum_{l=1}^{j}\lceil f_{n,l}\rceil,\ 1 \le j \le k_n.$$

Assume that $\lceil f_{n,1}\rceil + \cdots + \lceil f_{n,k_n}\rceil \to \infty$ as $n \to \infty$. Then:

(1) $\mathcal{G}_c$ presents a ℓ^2-cutoff if and only if Σ presents a cutoff. Furthermore, if any of $\mathcal{G}_c$ and Σ presents a cutoff, then their critical time is the same.

(2) If Σ has a (t_n, b_n) cutoff, then $\mathcal{G}_c$ has a (t_n, b_n) ℓ^2-cutoff.

If there exists $\delta > 2$ such that (3.24) *holds for $t \ge (C + \log f_{n,i})(\delta\lambda_{i,1}^n)^{-1}$ and for all $1 \le i \le k_n$, $n \ge 1$, then:*

(3) $\mathcal{G}_c$ has a (t_n, b_n) ℓ^2-cutoff if and only if Σ has a (t_n, b_n) cutoff.

(4) $\mathcal{G}_c$ has a strongly optimal (t_n, b_n) ℓ^2-cutoff if and only if Σ has a string optimal (t_n, b_n) cutoff.

Proof. By replacing $\mathcal{B}$ and $\mathcal{C}$ in Lemma 3.6 with Γ and Σ and applying Theorem 3.8. □

Remark 3.13. Note that, by (3.22) in the proof of Lemma 3.6, one has

$$t_n^{\Sigma}(\epsilon) \le t_n^{\Gamma}(\epsilon) \le t_n^{\Sigma}(\epsilon/C), \quad \forall \epsilon \in (0, e^{-C}).$$

By the above remark and Remark 3.10, we may bound the ℓ^2-mixing time of the product chain P_n from above and below by using the mixing of the triangular array Σ. The following states this fact and gives a correct order of the mixing time, which is useful when the family of product chains does not present a ℓ^2-cutoff.

Proposition 3.6. *Let $\mathcal{G} = \{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ and Σ be as in Theorem 3.10 satisfying (3.24) and C be the constant given there. Then, for $0 < \epsilon < \sqrt{e^{e^{-C}} - 1}$,*

$$t_n^{\Sigma}(\epsilon) \le T_2^c(P_n, \epsilon) \le t_n^{\Sigma}\left(\frac{1}{C}\log(1+\epsilon^2)\right).$$

In particular, for $0 < \epsilon < \sqrt{e^{e^{-C}} - 1}$, one may choose $c_2(\epsilon) > c_1(\epsilon) > 0$ such that

$$c_1(\epsilon) t_n \le T_2^c(P_n, \epsilon) \le c_2(\epsilon) t_n,$$

where

$$t_n = \max\left\{\frac{\log(i+1)}{\sigma_{n,i}} : 1 \le i \le k_n\right\}$$

and $\sigma_{n,1} \le \sigma_{n,2} \le \cdots \le \sigma_{n,k_n}$ is a rearrangement of the nth row of Σ.

Proof. Immediate from Remark 3.10, Remark 3.13 and Proposition 3.4. □

3.5.3 The ℓ^1-cutoff for products of random walks on finite abelian groups

From Theorem 3.10, one can see that, for specific Markov kernels, the criterion for testing the ℓ^2-cutoff depends only on the spectral gaps. In particular, if the coefficient $f_{n,i}$ in (3.24) is the multiplicity of an eigenvalue whose real part is $\lambda_{i,1}^n$, then one should be able to bound the total variation distance from below by using the spectrum. In this subsection, we will concentrate on products of chains on finite abelian groups. The following is an important fact for the lower bound of total variation.

Lemma 3.7. *Let $\mathcal{X}$ be a set and μ, ν be two probability measures on $\mathcal{X}$. Assume that ψ is a complex-valued function contained in $\ell^1(\mu) \cap \ell^1(\nu)$ with $E_\mu(\psi) \neq 0$ and $E_\nu(\psi) = 0$, where $E_\mu(f)$ is the expectation of f with respect to μ. Then*

$$\|\mu - \nu\|_{\text{TV}} \geq 1 - \frac{4(\text{Var}_\mu(\psi) + \text{Var}_\nu(\psi))}{|E_\mu(\psi)|^2} \tag{3.25}$$

Proof. Denote $s = |E_\mu(\psi)|/2$ and $A = \{x \in \mathcal{X} : |\psi(x)| \geq s\}$. Then, for $x \in A^c$, $|\psi(x) - E_\mu(\psi)| \geq s$. This implies

$$\mu(A^c) = E_\mu \mathbf{1}_{A^c} \leq E_\mu \left(\mathbf{1}_{A^c} \frac{|\psi(x) - E_\mu(\psi)|^2}{s^2} \right) \leq \frac{\text{Var}_\mu(\psi)}{s^2}.$$

By Chebyshev's inequality, we have

$$\|\mu - \nu\|_{\text{TV}} \geq \mu(A) - \nu(A) \geq 1 - \frac{\text{Var}_\mu(\psi) + \text{Var}_\nu(\psi)}{s^2}.$$

□

By the above lemma, we may bound the total variation distance from below by using the spectrum of a Markov kernel.

Proposition 3.7. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain, where $\mathcal{X}$ is a finite group, p is a probability measure on $\mathcal{X}$ and $K(x, y) = p(x^{-1}y)$ for all $x, y \in \mathcal{X}$. Assume that K is normal and $\beta_1, ..., \beta_n$ are eigenvalues of K whose corresponding orthonormal eigenvectors $\phi_1, ..., \phi_n$ satisfy*

$$|\phi_i| \equiv 1, \quad \phi_i(id) = 1, \quad \forall 1 \leq i \leq n,$$

and

$$\sum_{i\neq j} H_t(id, e^{-t(2-\overline{\beta_i}-\beta_j)}\phi_i\overline{\phi_j}) \le \left(\sum_{i=1}^n e^{-2t(1-\mathrm{Re}\beta_i)}\right)^2, \tag{3.26}$$

where id denotes the identity of G*. Then, for* $x \in \mathcal{X}$*,*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \ge 1 - \frac{8}{\sum_{i=1}^n e^{-2t(1-\mathrm{Re}\beta_i)}}.$$

Proof. Let $\psi = e^{-t(1-\overline{\beta_1})}\phi_1 + \cdots + e^{-t(1-\overline{\beta_n})}\phi_n$. By the assumption $|\phi_i| \equiv 1$ and $\phi_i(id) = 1$, we have

$$H_t(id, |\psi|^2) = \sum_{i=1}^n e^{-2t(1-\mathrm{Re}\beta_i)} + \sum_{i\neq j} H_t(id, e^{-t(2-\overline{\beta_i}-\beta_j)}\phi_i\overline{\phi_j}),$$

and

$$|H_t(id,\psi)|^2 = \left(\sum_{i=1}^n e^{-2t(1-\mathrm{Re}\beta_i)}\right)^2, \quad \mathrm{Var}_\pi(\psi) = \sum_{i=1}^n e^{-2t(1-\mathrm{Re}\beta_i)}.$$

Then the desired identity is proved by (3.26) and Lemma 3.7. □

The following lemma is an interesting observation of the random walk on a finite abelian group.

Lemma 3.8. *Let* $\mathcal{X}$ *be a finite group and* K *be a Markov kernel on* $\mathcal{X}$ *given by* $K(x,y) = p(x^{-1}y)$ *for* $x, y \in \mathcal{X}$*, where* p *is a probability measure on* $\mathcal{X}$*. Assume that* $\mathcal{X}$ *is abelian. Then* K *is normal.*

Note that if $\mathcal{X}$ is a finite abelian group, then it is isomorphic to a direct sum of cycles. In this case, the group representation theory implies that $\mathcal{X}$ is self-dual, that is, there exists a group isomorphism from $\mathcal{X}$ to its characters(or irreducible representations). For instance, if $\mathcal{X}$ is isomorphic to $\prod_1^n \mathbb{Z}_{k_i}$, then, for $x = (x_1, ..., x_n) \in \mathcal{X}$, the character ρ_x associated to x is given by

$$\rho_x(y) = \prod_{j=1}^n e^{2\pi i x_j y_j / k_j}, \quad \forall y = (y_1, ..., y_n) \in \mathcal{X}.$$

It is worth noting that if a transition matrix K on $\mathcal{X}$ satisfies

$$K(x,y) = p(x^{-1}y), \quad \forall x, y \in \mathcal{X},$$

where p is a probability measure on $\mathcal{X}$, then the eigenvalues of K are given by

$$\beta_x = \widehat{p}(\rho_x) = \sum_{y \in \mathcal{X}} p(y)\rho_x(y), \quad \forall x \in \mathcal{X}. \tag{3.27}$$

Remark 3.14. Let $\{\phi_i = \phi_{x_i} : \forall 1 \le i \le n\}$ be eigenvectors of K whose corresponding eigenvalues are $\beta_{x_1}, ..., \beta_{x_n}$ defined in (3.27). Then a sufficient condition for (3.26) is

$$1 - \mathrm{Re}\beta_{x_i x_j^{-1}} \ge (1 - \mathrm{Re}\beta_{x_i}) + (1 - \mathrm{Re}\beta_{x_j}) \quad \forall i \ne j.$$

By Proposition 3.7 and Remark 3.14, we my bound the total variation distance for products of chains as follows.

Theorem 3.11. *Let $\{(\mathcal{X}_i, K_i, \pi_i)\}_1^n$ be irreducible Markov chains, where $\mathcal{X}_i$ is a finite group, κ_i is a probability measure on $\mathcal{X}_i$ and $K_i(x,y) = \kappa_i(x^{-1}y)$ for $x, y \in \mathcal{X}_i$. Let $(\mathcal{Y}, P, \mu)$ be a Markov chain, where $\mathcal{Y} = \prod_1^n \mathcal{X}_i$ and $\mu = \bigotimes_1^n \pi_i$. The transition matrix P is given by*

$$K(x,y) = p_0 \delta(x,y) + \sum_{i=1}^n p_i \delta_i(x,y) K_i(x_i, y_i),$$

where $p_0 + \cdots + p_n = 1$, $\delta_i(x,y) = \prod_{j \ne i} \delta(x_j, y_j)$, and $\delta(u,v) = 1$ if $u = v$ and $\delta(u,v) = 0$ otherwise. Assume that, for $1 \le i \le n$, $\mathcal{X}_i$ is abelian and β_i is an eigenvalue of K_i. Then the total variation distance for the continuous-time semigroup H_t associated to P satisfies

$$\forall y \in \mathcal{Y}, \quad \|H_t^y - \mu\|_{\mathrm{TV}} \ge 1 - \frac{8}{\sum_{i=1}^n e^{-2t(1 - \mathrm{Re}\beta_i)}}.$$

Proof. For $1 \leq i \leq n$, let ρ_i be a character such that $\beta_i = \widehat{\kappa}_i(\rho_i)$(defined in (3.27)) and set

$$u_i = \overbrace{\mathbf{1} \otimes \cdots \otimes \mathbf{1}}^{i-1} \otimes \rho_i \otimes \overbrace{\mathbf{1} \otimes \cdots \otimes \mathbf{1}}^{n-i}.$$

Then one may choose $y_1, ..., y_n \in \mathcal{Y}$ such that $\rho_{y_i} = u_i$. This implies that, for $i < j$,

$$\rho_{y_i y_j^{-1}} = \overbrace{\mathbf{1} \otimes \cdots \otimes \mathbf{1}}^{i-1} \otimes \rho_i \otimes \overbrace{\mathbf{1} \otimes \cdots \otimes \mathbf{1}}^{j-i-1} \otimes \overline{\rho_j} \otimes \overbrace{\mathbf{1} \otimes \cdots \otimes \mathbf{1}}^{n-j}.$$

Hence we have

$$1 - \beta_{y_i} = \widehat{P}(u_i) = p_{n,i} - p_{n,i}\beta_i$$

and

$$1 - \beta_{y_i y_j^{-1}} = \widehat{P}(\rho_{y_i y_j^{-1}}) = p_{n,i} + p_{n,j} - p_{n,i}\beta_i - p_{n,j}\overline{\beta_j}.$$

By the above computation, one has

$$1 - \mathrm{Re}\beta_{y_i y_j^{-1}} = 1 - \mathrm{Re}\beta_{y_i} + 1 - \mathrm{Re}\beta_{y_j}, \quad \forall i \neq j,$$

and, by Proposition 3.7, this proves the desired inequality. □

The following theorem says that, in some specific cases, the equivalence of cutoffs given in Theorem 3.10 can be applied to the ℓ^1-cutoff.

Theorem 3.12. *Let $\mathcal{G}$ and $\{\Lambda_i^n : 1 \leq i \leq k_n,\ n \geq 1\}$ be as in Theorem 3.8, where $\mathcal{X}_{n,i}$ is abelian, $\kappa_{n,i}$ is a probability measure on $\mathcal{X}_{n,i}$ and $K_{n,i}(x,y) = \kappa_{n,i}(x^{-1}y)$ for $x, y \in \mathcal{X}_{n,i}$. Assume that $k_n \to \infty$ and there exists a positive number $C \geq 1$ such that, for $1 \leq i \leq k_n$ and $n \geq 1$,*

$$e^{-2\lambda_{i,1}^n t} \leq \sum_{j=1}^{|\mathcal{X}_{n,i}|-1} e^{-2\lambda_{i,j}^n t} \leq Ce^{-2\lambda_{i,1}^n t}, \quad \forall t \geq C(\lambda_{i,1}^n)^{-1}. \tag{3.28}$$

Let Σ be a triangular array whose nth row consists of $p_{n,1}\lambda_{1,1}^n, ..., p_{n,k_n}\lambda_{k_n,1}^n$. Then the following are equivalent.

(1) $\mathcal{G}_c$ *presents a* ℓ^2*-cutoff.*

(2) $\mathcal{G}_c$ *presents a* ℓ^1*-cutoff.*

(3) Σ *presents a cutoff.*

In particular, if any of these conditions holds, then the ℓ^1 *and* ℓ^2 *critical time for* $\mathcal{G}_c$ *and the critical time for* Σ *are the same. Moreover, if* Σ *has a* (t_n, b_n) *cutoff, then* $\mathcal{G}_c$ *has a* (t_n, b_n) ℓ^1 *and* ℓ^2 *cutoff.*

Proof. Immediate from Theorem 3.10 and Theorem 3.11. □

The following is a fact based on Theorem 3.12 and the monotonicity of the ℓ^p-norm in p.

Corollary 3.7. *Let* Σ *and* $\mathcal{G}$ *be as in Theorem 3.12. Assume that* (3.28) *holds and* Σ *presents a* (t_n, b_n) *cutoff. Then, for* $1 \le p \le 2$, $\mathcal{G}_c$ *presents a* (t_n, c_n) ℓ^p*-cutoff, where*

$$c_n = \max\{b_n, \sigma_{n,1}^{-1}\}$$

and $\sigma_{n,1} \le \sigma_{n,2} \le \cdots \le \sigma_{n,k_n}$ *is a rearrangement of the* n*th row of* Σ *for* $n \ge 1$.

Proof. Immediate from Theorem 3.12 and Proposition 2.6. □

3.5.4 An application: A product of simple random walks on cycles

Here, we apply the results in the previous subsection to the special case, where the state spaces are products of cycles and transition matrices are products of simple random walks. First, consider the following setting. Let $\mathcal{F} = \{(\mathbb{Z}_{m_{n,i}}, K_{n,i}, \pi_{n,i}) :$

$1 \leq i \leq k_n,\ n \geq 1\}$ be a family of irreducible Markov chains, where the Markov kernel is given by

$$K_{n,i}(j, j+1) = a_{n,i}, \quad K_{n,i}(j, j-1) = 1 - a_{n,i}, \quad \forall 1 \leq j \leq m_{n,i}, \tag{3.29}$$

and $a_{n,i} \in (0, 1]$. Let $\mathcal{G} = \{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ be a family induced from $\mathcal{F}$, where $\mathcal{Y}_n = \prod_{i=1}^{k_n} \mathbb{Z}_{m_{n,i}}$, $\mu_n = \bigotimes_{i=1}^{k_n} \pi_{n,i}$ and P_n is defined in (3.16). The following is an observation on the Markov chain we consider.

Lemma 3.9. *For $a, b \in [0, 1]$ and $n \geq 2$, let $(\mathbb{Z}_n, K_{a,b}, \pi)$ be a Markov chain with $\pi \equiv 1/n$ and*

$$K_{a,b}(j, j+1) = a,\ K_{a,b}(j, j) = b,\ K_{a,b}(j, j-1) = 1 - a - b \quad \forall 1 \leq j \leq n.$$

Let $H_{a,b,t}$ be the continuous-time semigroup associated to K. Then, for fixed $b \in [0, 1]$,

(1) $\|H_{a,b,t}(x, \cdot)/\pi - 1\|_2$ is independent of a for $a \in [0, 1-b]$.

(2) $\|K^m_{a,b}(x, \cdot)/\pi - 1\|_2$ is decreasing for $a \in [0, (1-b)/2]$.

Proof. Note that all eigenvalues of $K_{a,b}$ are $ae^{2\pi ij/n} + (1-a-b)e^{-2\pi ij/n} + b$ for $1 \leq j \leq n$, where $i = \sqrt{-1}$. Since $K_{a,b}$ is normal, one can easily compute the ℓ^2-distance by using the spectrum of $K_{a,b}$ as follows.

$$\|H_{a,b,t}(x, \cdot)/\pi - 1\|_2^2 = \sum_{j=1}^{n-1} e^{-2t(1-b)(1-\cos(2\pi j/n))} \tag{3.30}$$

and

$$\begin{aligned}\|K^m_{a,b}(x, \cdot)/\pi - 1\|_2^2 &= \sum_{j=1}^{n-1} \left|ae^{2\pi ij/n} + (1-a-b)e^{-2\pi ij/n} + b\right|^{2m} \\ &= \sum_{j=1}^{n-1} \Big(1 + 2b(b-1)[1 - \cos(2\pi j/n)] \\ &\qquad + 2a[a - (1-b)][1 - \cos(4\pi j/n)]\Big)^m\end{aligned}$$

For fixed b, it is obvious that $\|H_{a,b,t}(x,\cdot)/\pi - 1\|_2$ is independent of a. For the second part, note that

$$1 + 2b(b-1)\left[1 - \cos(2\pi j/n)\right] + 2a[a - (1-b)]\left[1 - \cos(4\pi j/n)\right] \geq 0$$

and $a[a-(1-b)]$ is strictly decreasing on $[0,(1-b)/2]$. Hence $\|K^m_{a,b}(x,\cdot)/\pi - 1\|_2$ is decreasing on $[0,(1-b)/2]$. □

Remark 3.15. Note that one may generalize the kernel in (3.29) by adding up a weight on the identity like the kernel in Lemma 3.9. However, this does not change the ℓ^2-distance that much since the difference between them is the re-scaling of time by multiply a constant factor. This factor can be seen from (3.30).

The following corollary generalizes part of the result in Lemma 3.9 to a product of chains.

Corollary 3.8. *Let* $\mathcal{F} = \{(\mathbb{Z}_{m_{n,i}}, K_{n,i}, \pi_{n,i}) : 1 \leq i \leq k_n,\ n \geq 1\}$ *and* $\mathcal{G} = \{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ *be families defined by* (3.16) *and* (3.29)*. For* $n \geq 1$*, let* $\widetilde{H}_{n,t}$ *be the continuous-time semigroup associated to* P_n*. Then the* ℓ^2*-distance* $\|\widetilde{H}_{n,t}(x,\cdot)/\mu_n - 1\|_2$ *is independent of the set* $\{a_{n,i} : 1 \leq i \leq k_n,\ n \geq 1\}$*.*

Proof. Immediate from Lemma 3.9 and Proposition 3.5. □

By Corollary 3.8, there is no loss of generality in assuming that $a_{n,i} \equiv 1/2$ in Corollary 3.8. The following theorem is our main result for the application of Theorem 3.12.

Theorem 3.13. *Let* $\mathcal{F} = \{(\mathbb{Z}_{m_{n,i}}, K_{n,i}, \pi_{n,i}) : 1 \leq i \leq k_n,\ n \geq 1\}$ *and* $\mathcal{G} = \{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ *be families defined above whose kernels are defined respectively by* (3.29) *and* (3.16)*, where* $a_{n,i} \equiv 1/2$ *and* $m_{n,i} \geq 2$*. For* $n \geq 1$ *and* $1 \leq i \leq k_n$*, let* λ_i^n *be the spectral gap of* $K_{n,i}$ *and set* Σ *to be a triangular array whose* n*th row* Σ_n

consists of elements $p_{n,1}\lambda_1^n, ..., p_{n,k_n}\lambda_{k_n}^n$. *Assume that* $k_n \to \infty$. *Then the following are equivalent.*

(1) $\mathcal{G}_c$ *presents a* ℓ^2*-cutoff.*

(2) $\mathcal{G}_c$ *presents a* ℓ^1*-cutoff.*

(3) Σ *presents a cutoff.*

(4) $\sigma_{n,1}^{-1} = o(\tau_n)$, *where, for* $n \geq 1$, $\sigma_{n,1} \leq \sigma_{n,2} \leq \cdots \leq \sigma_{n,k_n}$ *is a rearrangement of* $p_{n,1}\lambda_1^n, ..., p_{n,k_n}\lambda_{k_n}^n$ *and*

$$\tau_n = \max\left\{\frac{\log i}{2\sigma_{n,i}} : 1 \leq i \leq k_n\right\}.$$

If any of these conditions holds, then, for $1 \leq p \leq 2$, *the family* $\mathcal{G}_c$ *presents a* ℓ^p*-cutoff with critical time* τ_n.

Moreover, one has

$$\Sigma \text{ has a } (t_n, b_n) \text{ cutoff} \Leftrightarrow \mathcal{G}_c \text{ has a } (t_n, b_n)\ \ell^2\text{-cutoff} \Rightarrow \mathcal{G}_c \text{ has a } (t_n, b_n)\ \ell^1\text{-cutoff}.$$

and

$$\Sigma \text{ has a strongly optimal } (t_n, b_n) \text{ cutoff} \Leftrightarrow \mathcal{G}_c \text{ has a strongly optimal } (t_n, b_n)\ \ell^2\text{-cutoff}.$$

In particular, if Σ *presents a* (t_n, b_n) *cutoff, then, for* $1 < p < 2$, *the family* $\mathcal{G}_c$ *has a* (t_n, c_n) ℓ^p*-cutoff, where* $c_n = \max\{b_n, \lambda_n^{-1}\}$ *and* $\lambda_n = \min\{p_{n,i}\lambda_{i,1}^n : 1 \leq i \leq k_n\}$ *is the spectral gap of* P_n.

Theorem 3.13 is proved by Theorem 3.5, Theorem 3.12 and the following lemma.

Lemma 3.10. *For* $n \geq 2$, *let* $\lambda_n = 1 - \cos(2\pi/n)$ *and*

$$d_n(t) = \sum_{i=1}^{n-1} e^{-2t(1-\cos(2\pi i/n))}.$$

Then one has, for $t \geq 0$,

$$d_2(t) = e^{-4t},\ d_3(t) = 2e^{-3t},\ d_4(t) = 2e^{-2t} + e^{-4t},$$

and for odd $n \geq 5$,

$$2\left(1 + f_1(n, 12\pi^2 t)\right) e^{-2t\lambda_n} \leq d_n(t) \leq 2\left(1 + f_1(n, \pi^2 t/2)\right) e^{-2t\lambda_n},$$

and for even $n \geq 6$,

$$\begin{aligned} 2\left(1 + f_1(n, 12\pi^2 t)\right)e^{-2t\lambda_n} + f_2(n,t) &\leq d_n(t) \\ &\leq 2\left(1 + f_1(n, \pi^2 t/2)\right) e^{-2t\lambda_n} + f_2(n,t), \end{aligned}$$

where $c_n = \frac{1}{n}\left(\lceil n/2 \rceil - 1\right)$ *and*

$$f_1(n,t) = n \int_{1/n}^{c_n} e^{-tx^2} dx, \quad f_2(n,t) = e^{-4t}.$$

In particular, there exist positive numbers c_1, c_2 *such that*

$$2(1 + c_1 g(n,t))e^{-2t\lambda_n} \leq d_n(t) \leq 2(1 + c_2 g(n,t))e^{-2t\lambda_n}, \tag{3.31}$$

for all $t > 0$ *and* $n \geq 5$, *where*

$$g(n,t) = \begin{cases} n & \text{if } t \in (0,1) \\ nt^{-1/2} & \text{if } t \in (1, n^2) \\ n^2 t^{-1} e^{-t/n^2} & \text{if } t > n^2 \end{cases} \cdot$$

Proof. Note that

$$\begin{aligned} d_n(t) - f_2(n,t) &= 2 \sum_{i=1}^{\lceil n/2 \rceil - 1} e^{-2t(1-\cos(2\pi i/n))} \\ &= 2e^{-2t\lambda_n} \left(1 + \sum_{i=2}^{\lceil n/2 \rceil - 1} e^{-2t(\cos(2\pi/n) - \cos(2\pi i/n))}\right) \end{aligned}$$

To bound the difference among values taking by a cosine function, we use the following fact. For $0 \le s \le \pi/2$ and $0 \le t \le \pi$, if $s < t$, then

$$\frac{1}{8}(t^2 - s^2) \le \cos s - \cos t \le \frac{1}{2}(t^2 - s^2). \tag{3.32}$$

This implies

$$\frac{\pi^2}{4}\left(\frac{i}{n}\right)^2 \le \cos(2\pi/n) - \cos(2\pi i/n) \le 6\pi^2\left(\frac{i-1}{n}\right)^2,$$

where both inequalities also use the fact $i^2/2 \le i^2 - 1 \le 3(i-1)^2$ for $i \ge 2$. Since e^{-x} is a decreasing function, the above computation derives

$$\sum_{i=2}^{\lceil n/2 \rceil - 1} e^{-2t(\cos(2\pi/n) - \cos(2\pi i/n))} \begin{cases} \le f_1(n, \pi^2 t/2) \\ \ge f_1(n, 12\pi^2 t) \end{cases}$$

For the last part, assume that $n \ge 5$. By changing the variable in the integral, one has

$$f_1(n,t) = \frac{n}{\sqrt{t}} \int_{\sqrt{t}/n}^{\sqrt{t}c_n} e^{-y^2} dy \begin{cases} \le \int_{\sqrt{t}/n}^{\sqrt{t}/2} e^{-x^2} dx \\ \ge \int_{\sqrt{t}/n}^{\sqrt{t}(1/2-1/n)} e^{-x^2} dx \end{cases}.$$

This implies that, for $n \ge 5$,

$$f_1(n,t) \sim \frac{n^2}{2t} e^{-t/n^2}, \quad \text{as } \frac{\sqrt{t}}{n} \to \infty,$$

and

$$\frac{n}{10e} \le f_1(n,t) \le \frac{n}{2}, \ \forall t \le 1, \quad n\int_1^{3/2} e^{-x} dx \le f_1(n,t) \le n, \ \forall 1 \le t \le n^2.$$

Combining the above computations, we may choose two positive numbers c_1, c_2 such that

$$c_1 g(n,t) \le f_1(n,t) \le c_2 g(n,t), \quad \forall n \ge 5, \ t > 0,$$

where $g(n,t)$ be the function defined in Lemma 3.10. Then, the desired inequality in (3.31) is proved by the following fact.

$$e^{-4t}e^{2t\lambda_n} \le e^{-2t} \le \begin{cases} 1 & \text{if } t \in (0, n^2) \\ e^{-t/n^2}e^{-\sqrt{t}/n} & \text{if } t > n^2 \end{cases}.$$

□

Proof of Theorem 3.13. Let c_2 be the quantity given by Lemma 3.10. A simple computation show that (3.28) is satisfied with $C = \max\{20, 2(1+c_2)\}$. By Theorem 3.12, one has (1)⇔(2)⇔(3), and by Theorem 3.5, we get (3)⇔(4). The last part is proved by Theorem 3.10 and Theorem 3.12. □

We consider a specific example for an illustration of Theorem 3.12. Let $\mathcal{F} = \{(\mathbb{Z}_{n+1}, K_n, \pi_n)\}_1^\infty$ be a family of finite Markov chains, where

$$K_n(j, j+1) = K_n(j, j-1) = \frac{1}{2}, \quad \forall 1 \le j \le n+1,\ n \ge 1. \tag{3.33}$$

Let $\mathcal{G} = \{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ be another family of Markov chains given by

$$\mathcal{Y}_n = \prod_{i=1}^n \mathbb{Z}_{i+1}, \quad \mu_n = \bigotimes_{i=1}^n \pi_i, \tag{3.34}$$

and for $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathcal{Y}_n$,

$$P_n(x, y) = p_{n,0}\delta(x, y) + \sum_{i=1}^n p_{n,i}\delta_i(x, y)K_i(x_i, y_i), \tag{3.35}$$

where $p_{n,0} + \cdots + p_{n,n} = 1$, $\delta_i(x, y) = \prod_{j \ne i} \delta(x_j, y_j)$, and $\delta(u, v) = 1$ if $u = v$ and $\delta(u, v) = 0$ otherwise.

Remark 3.16. It is well-known that the above family $\mathcal{F}$ does not have a ℓ^1 or ℓ^2 cutoff. In this dissertation, the introduced techniques are sufficient to prove that there is no ℓ^p-cutoff for $\mathcal{F}_c$ with $1 < p \le \infty$. In details, by Lemma 3.10, one may choose a constant $C > 0$ such that

$$T_2^c(K_n, 1) < C\lambda_n^{-1}, \quad \forall n \ge 1,$$

where λ_n is the spectral gap of K_n. By Theorem 2.4, this implies that $\mathcal{F}_c$ does not present a ℓ^2-cutoff. Then, by Corollary 2.6, $\mathcal{F}_c$ does not present any ℓ^p-cutoff for $1 < p \le \infty$.

An interpretation of the above Markov kernel P_n is that, for $n \ge 1$, one coordinate of $\mathcal{Y}_n$ is chosen according to a probability measure $(p_{n,i})_0^n$. Then changing the state by adding up or subtracting 1 from that coordinate with probability 1/2. In the following, we consider three examples whose probability measures $(p_{n,i})_0^n$ are in order uniform, decreasing geometrically in i and decreasing arithmetically in i. From the results, one can see how the mixing time evolves as $(p_{n,i})_0^n$ changes.

Example 3.8 *($\{p_{n,i}\}_{i=1}^n$ is a uniform probability measure).* In this example, the nth Markov chain of the family $\mathcal{G}$ defined by (3.33), (3.34) and (3.35) transits its current state by first randomly choosing a digit and then randomly adding up or subtracting 1 in that digit. In order to examine the cutoff for $\mathcal{G}_c$ by using Theorem 3.13, one needs to check the following quantity.

$$\max_{0 \le i \le n-1} \left\{ \frac{[1-\cos(2\pi/(n+1))]\log(i+1)}{1-\cos(2\pi/(n-i+1))} \right\}$$

By (3.32), letting $i = \lceil n/2 \rceil$ in the above implies that for $n \ge 4$,

$$\frac{1-\cos(2\pi/(n+1))}{1-\cos(2\pi/(n-i+1))} \ge \frac{1-\cos(2\pi/(n+1))}{1-\cos(4\pi/(n+1))} \ge \frac{1}{16}.$$

Then the above quantity tends to infinity as $n \to \infty$ and, by Theorem 3.13, $\mathcal{G}_c$ presents a ℓ^p-cutoff with critical time

$$t_n = \frac{n}{2} \max_{0 \le i \le n-1} \left\{ \frac{\log(i+1)}{1-\cos(2\pi/(n-i+1))} \right\},$$

for $1 \leq p \leq 2$.

To determine an asymptotic value of t_n, note that

$$t_n \leq s_n = \frac{n \log n}{2[1 - \cos(2\pi/(n+1))]}.$$

For the lower bound, let $j \in (0, 1)$ and replace i with $\lceil jn \rceil$. This implies that for n large,

$$t_n \geq s_n(j) = \frac{n \log(nj + 1)}{2[1 - \cos(2\pi/(n - nj))]}$$

Since $1 - \cos\theta \sim \frac{\theta^2}{2}$ as $\theta \to 0$, one can easily conclude that for $j \in (0, 1)$,

$$\liminf_{n\to\infty} \frac{t_n}{s_n} \geq \liminf_{n\to\infty} \frac{s_n(j)}{s_n} = (1 - j)^2.$$

Letting $j \to 0$ implies $t_n \sim s_n$.

To select a window size, let $\mathcal{P}_n = \{n^{-1}\lambda_n(1 + \epsilon_n)^i : i = 0, 1, ...\}$ be a partition of $(0, \infty)$ with

$$\epsilon_n = \frac{\cos(2\pi/(n+1)) - \cos(2\pi/[n - n(\log n)^{-1}])}{1 - \cos(2\pi/(n+1))}.$$

By Taylor's expansion of the cosine function at 0, one can compute that

$$\begin{aligned} \epsilon_n &\sim \frac{n^{-2} - [n - n(\log n)^{-1}]^{-2} + O(n^{-4})}{n^{-2}} \\ &= \frac{1}{\log n - 1} + O(n^{-2}) \sim \frac{1}{\log n} = o(1) \end{aligned}$$

Let $\Sigma_n = \{n^{-1}\lambda_1, ..., n^{-1}\lambda_n\}$ and $\tau_\Sigma(\mathcal{P}_n)$ be the quantity defined in (3.4). Then we have

$$\tau_\Sigma(\mathcal{P}_n) \geq s_n - \frac{n \log\log n}{2[1 - \cos(2\pi/(n+1))]},$$

which implies

$$\limsup_{n\to\infty} \frac{-\log \epsilon_n}{n^{-1}\lambda_n \tau_\Sigma(\mathcal{P}_n)} \leq \limsup_{n\to\infty} \frac{\log\log n}{\log n} = 0.$$

By Theorem 3.3 and Theorem 3.5, Σ presents a $(\tau_\Sigma(\mathcal{P}_n), b_n)$ and (t_n, b_n) cutoff, where

$$b_n = \frac{n \log\log n}{1-\cos(2\pi/n)}.$$

Since $t_n \le s_n \le \tau_\Sigma(\mathcal{P}_n) + b_n$, Σ also has a (s_n, b_n) cutoff. Finally, a standard computation shows that

$$\frac{1}{1-\cos(2\pi/(n+1))} = \frac{1}{2\pi^2(n+1)^{-2} + O(n^{-4})} = \frac{(n+1)^2}{2\pi^2} + O(1).$$

Because $n\log n \le n^2 \log n = o(b_n)$, Σ presents a (t_n', b_n') cutoff, where

$$t_n' = \frac{n^3 \log n}{4\pi^2}, \quad b_n' = n^3 \log\log n.$$

Since the spectral gap of P_n is equal to $\frac{1}{n}[1-\cos(2\pi/(n+1))] \sim \frac{4\pi^2}{n^3}$, by Theorem 3.13, the family $\mathcal{G}_c$ has a (t_n', b_n') ℓ^p cutoff for $1 \le p \le 2$. Note that the window (b_n') given above is not optimal, since, by Theorem 2.4, one can has a cutoff whose window has size n^3.

Example 3.9 *($\{p_{n,i}\}_{i=1}^n$ is a decreasing arithmetic sequence).* Let $\mathcal{F}$ and $\mathcal{G}$ be as in Example 3.8, with $p_{n,i} = a_n(n+1-i)$ for $1 \le i \le n$, where $a_n = (1+2+\cdots+n)^{-1} = \frac{2}{n(n+1)}$. In order to apply Theorem 3.13, we need to examine the following quantity.

$$s_n = \max_{0\le i\le n-1} \left\{ \frac{[1-\cos(2\pi/(n+1))]\log(i+1)}{(i+1)[1-\cos(2\pi/(n+1-i))]} \right\}$$

Note that for $t \ge 0$, $\log(1+t) \le t$. This implies that for $i \ge 3$,

$$\frac{\log(i+1)}{\log i} = 1 + \frac{\log(1+1/i)}{\log i} \le 1 + i^{-1},$$

that is, $\frac{\log(i+1)}{i+1} \le \frac{\log i}{i}$ for $i \ge 3$. Since $2^{1/2} \le 3^{1/3}$, we may choose $N > 0$ such that

$$s_n = \frac{[1-\cos(2\pi/(n+1))]\log 3}{3[1-\cos(2\pi/(n-1))]} \quad \forall n \ge N.$$

Hence $s_n \sim (\log 3)/3$. By Theorem 3.13, $\mathcal{G}_c$ does not present a ℓ^2-cutoff. By Proposition 3.6, we may choose $c > 0$ and, for $\epsilon \in (0, c)$, $c_2(\epsilon) > c_1(\epsilon) > 0$ such that

$$c_1(\epsilon)t_n \le T_2^c(P_n, \epsilon) \le c_2(\epsilon)t_n,$$

where

$$t_n = \frac{n(n+1)}{2} \max_{0\le i\le n-1} \left\{ \frac{\log(i+2)}{(i+1)[1-\cos(2\pi/(n+1-i))]} \right\}$$

A similar analysis as before, we have $\frac{\log(i+2)}{i+1} \le \frac{\log(i+1)}{i}$ for all $i \ge 1$. This implies

$$t_n = \frac{(\log 2)n(n+1)}{2[1-\cos(2\pi/(n+1))]} \sim \frac{\log 2}{4\pi^2} n^4.$$

Example 3.10 *($\{p_{n,i}\}_{i=1}^n$ is a decreasing geometric sequence).* Let $\mathcal{F}, \mathcal{G}$ be the same as in Example 3.8 except that $p_{n,i}$ is replaced by

$$p_{n,i} = a_n r^{i-1} \quad \forall 1 \le i \le n,$$

where $r \in (0, 1)$ is a fixed constant and $a_n = (1+r^1+\cdots+r^{n-1})^{-1} = (1-r)/(1-r^n)$. To determine the ℓ^2-cutoff for $\mathcal{G}_c$ via Theorem 3.13, we need to see whether the following limit is true.

$$\lim_{n\to\infty} \max_{1\le i\le n} \left\{ \frac{r^i[1-\cos(2\pi/(n+1))]\log(i+1)}{1-\cos(2\pi/(n-i+1))} \right\} = \infty.$$

Note that $1 - \cos(2\pi/(n+1)) \le 1 - \cos(2\pi/(n-i+1))$ for $0 \le i \le n-1$. Since the function $\log t + t \log r$ is concave and has its maximum at $t = (\log r^{-1})^{-1}$, we have

$$r^i \log(i+1) \le ir^i \le (e\log(1/r))^{-1} < \infty. \tag{3.36}$$

This implies that $\mathcal{G}_c$ has no ℓ^2-cutoff.

For a bound on the ℓ^2-mixing time, by Proposition 3.6, one may choose $c > 0$ and, for $\epsilon \in (0, c)$, $c_2(\epsilon) > c_1(\epsilon) > 0$ such that

$$c_1(\epsilon)t_n \le T_2^c(P_n, \epsilon) \le c_2(\epsilon)t_n, \quad \forall \epsilon \in (0, c),$$

where

$$t_n = \left(\frac{1 - r^n}{(1-r)r^{n-1}} \right) \max_{0 \le i \le n-1} \left\{ \frac{r^i \log(i+2)}{[1 - \cos(2\pi/(n-i+1))]} \right\}.$$

Since, for fixed $r \in (0,1)$, the map $t \mapsto tr^t$ is increasing in $(0, (\log r^{-1})^{-1})$ and decreasing in $((\log r^{-1})^{-1}, \infty)$, one may choose $C > 1$ such that

$$\frac{e^{n \log(1/r)}}{Cn^2} \le t_n \le \frac{Ce^{n \log(1/r)}}{n^2}, \quad \forall n \ge 1.$$

3.5.5 The cutoff for the product of random walks on abelian groups with a bounded number of generators

In this section, we consider a specific class of Markov kernels on finite abelian groups. Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain, where $\mathcal{X}$ is a finite abelian group, p is a probability measure on $\mathcal{X}$ with support E, and K is given by $K(x,y) = p(x^{-1}y)$ for $x, y \in \mathcal{X}$. By Lemma 3.8, the Markov kernel K is normal. Let λ be the spectral gap of K. Assume that E is a symmetric set, in the sense that $x^{-1} \in E$ if $x \in E$, which contains the identity and generates $\mathcal{X}$, and p is supported on E satisfying

$$p(x) = p(x^{-1}), \quad \forall x \in E.$$

In this case, K is reversible and there exist positive constants c_1, c_2, c_3, c_4 depending only on the cardinality of E and the minimum probability of p on its support such that

$$\|h_t^e - 1\|_2 \le c_1 e^{-s/(c_2\gamma^2)}, \quad \forall t \ge c_2\gamma^2 + s,$$

and

$$\frac{c_3}{\gamma^2} \le \lambda \le \frac{c_4}{\gamma^2},$$

where γ is the diameter of $\mathcal{X}$ with respect to E, that is, the smallest integer n such that $E^n = \mathcal{X}$. The above facts are obtained from the discussions of the moderate growth, the doubling property and the Poincaré inequality in [17, 20] and the local Poincaré inequality and the Nash inequality in [20]. For details, please confer Lemma 5.1, Theorem 5.2 and the proof of Theorem 3.1 in [17] and look at Lemma 5.4 and Theorem 5.3 in [20].

Note that, by the operator theory, one has

$$\|h^e_{t_1+t_2} - 1\|_2 = \|H_{t_1+t_2} - \pi\|_{2\to\infty} \begin{cases} \leq \|H_{t_1} - \pi\|_{2\to\infty} e^{-t_2\lambda} \\ \geq \|H_{t_1+t_2} - \pi\|_{2\to 2} = e^{-(t_1+t_2)\lambda} \end{cases}.$$

Letting $t \geq c\lambda^{-1}$, $t_1 = c_2\gamma^2$ and $t_2 = t - t_1$ with $c > c_2c_4$ implies that

$$\|h^e_t - 1\|_2 \leq c_1 e^{-(t-t_1)\lambda} \leq c_1 e^{t_1\lambda} e^{-t\lambda} \leq (c_1 e^{c_2c_4}) e^{-t\lambda}.$$

Hence, the identity in (3.28) is satisfied with $C = \max\{c_2c_4, c_1^2 e^{2c_2c_4}\}$. The following theorem is a consequence obtained from the above facts.

Theorem 3.14. *Let $\mathcal{F} = \{(\mathcal{X}_{n,i}, K_{n,i}, \pi_{n,i}) : 1 \leq i \leq k_n,\ n \geq 1\}$ be a family of irreducible Markov chains, where $\mathcal{X}_{n,i}$ is a finite abelian group, $\kappa_{n,i}$ is a probability measure on $\mathcal{X}_{n,i}$ which is supported on a symmetric set $E_{n,i}$ containing the identity and*

$$K_{n,i}(x, y) = \kappa_{n,i}(x^{-1}y), \quad \forall x, y \in \mathcal{X}_{n,i}.$$

Let $\mathcal{G} = \{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ be a family induced from $\mathcal{F}$, where $\mathcal{Y}_n = \prod_{i=1}^{k_n} \mathcal{X}_{n,i}$ and P_n is the Markov kernel defined in (3.16)*. Set Σ to be a triangular array whose nth row contains the following elements*

$$p_{n,1}\lambda_1^n, ..., p_{n,k_n}\lambda_{k_n}^n,$$

where λ_i^n is the spectral gap of $K_{n,i}$ for $1 \leq i \leq k_n$ and $n \geq 1$. Assume that $k_n \to \infty$ and, for $1 \leq i \leq k_n$ and $n \geq 1$, the probability measure $\kappa_{n,i}$ satisfies

$$\kappa_{n,i}(x) = \kappa_{n,i}(x^{-1}) \geq \epsilon,\ \forall x \in E_{n,i},$$

where ϵ is a positive constant. Then the following are equivalent.

(1) $\mathcal{G}_c$ presents a ℓ^2-cutoff.

(2) $\mathcal{G}_c$ presents a ℓ^1-cutoff.

(3) Σ presents a cutoff.

In particular, if any of these conditions holds, then the ℓ^1 and ℓ^2 critical time for $\mathcal{G}_c$ and the critical time for Σ are the same. Moreover, if Σ has a (t_n, b_n) cutoff, then $\mathcal{G}_c$ has a (t_n, b_n) ℓ^1 and ℓ^2 cutoff.

For an application, we consider the following example.

Example 3.11. Let $\mathcal{F} = \{(\mathcal{X}_{n,i}, K_{n,i}, \pi_{n,i}) : 1 \leq i \leq k_n,\ n \geq 1\}$ be a family of finite Markov chains, where $\mathcal{X}_{n,i} = \mathbb{Z}_i$ and $\pi_{n,i} \equiv 1/i$. For $n \geq 1$ and $1 \leq i \leq k_n$, let $\kappa_{n,i}$ be a probability measure on $\mathbb{Z}_i$ and $K_{n,i}(x, y) = \kappa_{n,i}(x^{-1}y)$ for $x, y \in \mathbb{Z}_i$. Assume that L is a positive integer and

$$\kappa_{n,i}(a_{i,\ell}) = \kappa_{n,i}(-a_{i,\ell}) > 0, \quad \forall 0 \leq \ell \leq L,$$

where $a_{i,\ell} = \lfloor i^{\ell/L} \rfloor$. In this setting, it can be easily checked that the order of the diameter of $\mathbb{Z}_i$ with respect to the support of $\kappa_{n,i}$, that is $\{\pm a_{i,j} : 0 \leq j \leq L\}$, is the same as $i^{1/L}$ as $i \to \infty$. Assume that

$$\inf\{\kappa_{n,i}(a_{i,\ell}) : 0 \leq \ell \leq L,\ 1 \leq i \leq k_n,\ n \geq 1\} > 0.$$

Then, by the discussion in front of Theorem 3.14, there exist constants $c_2 > c_1 > 0$ such that the spectral gap λ_i^n of $K_{n,i}$ satisfies

$$c_1 i^{-2/L} \leq \lambda_i^n \leq c_2 i^{-2/L}, \quad \forall 1 \leq i \leq k_n,\ n \geq 1.$$

Let $\mathcal{G} = \{(\mathcal{Y}_n, P_n, \mu_n)\}_1^\infty$ be a family induced from $\mathcal{F}$, where $\mathcal{Y}_n = \prod_{i=1}^{k_n} \mathcal{X}_{n,i}$ and P_n is the Markov kernel given by (3.16). Then, by Theorem 3.14, the family $\mathcal{G}_c$ presents a ℓ^2-cutoff if and only if it presents a ℓ^1-cutoff. Moreover, if any of the ℓ^1 and ℓ^2 cutoffs exists, then the ℓ^1-critical time and the ℓ^2-critical time are the same.(In fact, the ℓ^p-critical time is the same for $1 \le p \le 2$) For the special case $p_{n,i} = 1/k_n$ for $1 \le i \le k_n$, let $a_{n,1} \le a_{n,2} \le \cdots \le a_{n,k_n-1}$ be a rearrangement of $\lambda_2^n, ..., \lambda_{k_n}^n$ and

$$t_n = \max_{1 \le i \le k_n - 1} \left\{ \frac{k_n \log i}{2a_{n,i}} \right\}$$

Assume that $k_n > 1$ for all n and $k_n \to \infty$. It can be proved that there exist $c_2 > c_2 > 0$ such that

$$c_1 k_n^{-2/L} \le a_{n,1} \le a_{n,k_n/2} \le c_2 k_n^{-2/L}, \quad \forall n \ge 1,$$

which implies

$$\frac{k_n^{1+2/L} \log k_n}{4c_2} \le t_n \le \frac{k_n^{1+2/L} \log k_n}{2c_1}, \quad \forall n \ge 1.$$

For example, if $k_n = n$, then the family $\mathcal{G}_c$ presents ℓ^2 and ℓ^1 cutoffs with the same critical time whose correct order is $n^{1+2/L} \log n$.

Chapter 4

The total variation cutoff

In this chapter, we will compare the total variation mixing time and the total variation cutoff between discrete-time and continuous-time cases. In Definition 1.4 and 1.5, a family $\mathcal{F}$ is said to present a total variation cutoff(in any sense) if $U = 1$ and

$$\rho_n(A) = \max_{x \in \mathcal{X}_n} \|A(x, \cdot) - \pi_n\|_{\text{TV}}.$$

Note that a total variation cutoff is equivalent to a ℓ^1-cutoff, which is defined by letting $U = 2$ and

$$\rho_n(A) = \max_{x \in \mathcal{X}_n} \|A(x, \cdot)/\pi_n - 1\|_1.$$

In the above setting, the mixing time defined in Definition 1.3 is called respectively a total variation mixing time and a ℓ^1-mixing time. We denote them as $T_{\text{TV}}(K_n, \epsilon)$ and $T_1(K_n, \epsilon)$. Note also that the ℓ^1-distance and the total variation distance are related by

$$\|\mu/\pi - 1\|_1 = 2\|\mu - \pi\|_{\text{TV}},$$

for any probability measure μ. This implies that $T_{\text{TV}}(K, \epsilon) = T_1(K, 2\epsilon)$ for all $\epsilon > 0$. According to this fact and Proposition 1.10 and 1.11, we may identify a total variation cutoff and a ℓ^1-cutoff without any change on the critical time and its window size.

In section 4.1, we compare the total variation distance between the discrete-time and continuous-time Markov chains. In section 4.2, we introduce Peres' conjecture and construct a counterexample for that conjecture by following Aldous' idea.

4.1 The total variation cutoff for finite Markov chains

The following Lemma gives a simple relation of the total variation distances between discrete-time and continuous-time cases.

Lemma 4.1. *Let $(\mathcal{X}, K, \pi)$ be an irreducible finite Markov chain and $H_t = e^{-t(I-K)}$ be the associated continuous-time semigroup with respect to K. For $m, t \geq 0$, set $a(m,t) = e^{-t}\sum_{j=0}^{m} \frac{t^j}{j!}$. Then, for $t, m \geq 0$, the maximum total variation distance $\max_{x\in\mathcal{X}} \|H_t^x - \pi_n\|_{\text{TV}}$ is bounded from above by*

$$\max_{x\in\mathcal{X}} \|K_x^m - \pi\|_{\text{TV}} + a(m,t),$$

and bounded from below by

$$e^{-t}\sum_{j=0}^{m} \frac{t^j}{j!}\left(K^j(x,A) - \pi(A)\right) - [1-a(m,t)] \min\left\{\pi(A), \max_{x\in\mathcal{X}_n} \|K_x^m - \pi\|_{\text{TV}}\right\},$$

where $A \subset \mathcal{X}$.

Proof. Note that for $t \geq 0$, one has

$$H_t(x,y) - \pi(y) = e^{-t}\sum_{j=0}^{\infty} \frac{t^j}{j!}\left(K^j(x,y) - \pi(y)\right), \quad \forall x, y \in \mathcal{X}.$$

By this identity, the upper bound of $\max_{x\in\mathcal{X}} \|H_t(x,\cdot) - \pi\|_{\text{TV}}$ is proved by Lemma 2.1 and the triangle inequality.

For the lower bound, let A be a subset of $\mathcal{X}$. Note that

$$\begin{aligned}\|H_t^x - \pi\|_{\text{TV}} \geq & H_t(x,A) - \pi(A) = e^{-t}\sum_{j=0}^{m} \frac{t^j}{j!}\left(K^j(x,A) - \pi(A)\right) \\ & + e^{-t}\sum_{j=m+1}^{\infty} \frac{t^j}{j!}\left(K^j(x,A) - \pi(A)\right).\end{aligned}$$

Then the desired inequality is proved by the fact

$$K^j(x,A) - \pi(A) \geq \max\{-\pi(A), -\|K_x^j - \pi\|_{\text{TV}}\}.$$

∎

By the above lemma, the total variation distances for a family of finite Markov chains can be related in the following asymptotic way.

Proposition 4.1. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of finite Markov chains and $H_{n,t} = e^{-t(I-K_n)}$ be the associated continuous-time semigroup w.r.t. K_n. Assume that $(t_n)_{n=1}^\infty$ and $(s_n)_1^\infty$ are sequences of positive numbers and t_n tends to infinity. Then*

$$\limsup_{n\to\infty}\left(\max_{x\in\mathcal{X}_n}\|H^x_{n,t_n+s_n}-\pi_n\|_{\mathrm{TV}}-\max_{x\in\mathcal{X}_n}\|K^{[t_n]}_{n,x}-\pi_n\|_{\mathrm{TV}}\right)\le\Phi(-L),$$

where $[\cdot]$ denotes either the floor $\lfloor\cdot\rfloor$ or the ceiling $\lceil\cdot\rceil$ and

$$L=\liminf_{n\to\infty}\frac{s_n}{\sqrt{t_n+s_n}},\qquad \Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}e^{-x^2/2}dx.$$

In particular,

(1) For $c\in(0,1)$, one has

$$\limsup_{n\to\infty}\max_{x\in\mathcal{X}_n}\|H^x_{n,t_n}-\pi_n\|_{\mathrm{TV}}\le\limsup_{n\to\infty}\max_{x\in\mathcal{X}_n}\|K^{\lceil ct_n\rceil}_{n,x}-\pi_n\|_{\mathrm{TV}},$$

and

$$\liminf_{n\to\infty}\max_{x\in\mathcal{X}_n}\|H^x_{n,t_n}-\pi_n\|_{\mathrm{TV}}\le\liminf_{n\to\infty}\max_{x\in\mathcal{X}_n}\|K^{\lfloor ct_n\rfloor}_{n,x}-\pi_n\|_{\mathrm{TV}}.$$

(2) Let $(b_n)_{n=1}^\infty$ be a sequence of positive numbers satisfying $b_n=o(t_n)$ and $\sqrt{t_n}=O(b_n)$. Then there exists a constant $C>0$ such that, for $c>0$,

$$\begin{aligned}&\limsup_{n\to\infty}\max_{x\in\mathcal{X}_n}\|H^x_{n,t_n+2cb_n}-\pi_n\|_{\mathrm{TV}}\\&\qquad\le\limsup_{n\to\infty}\max_{x\in\mathcal{X}_n}\|K^x_{n,\lceil t_n+cb_n\rceil}-\pi_n\|_{\mathrm{TV}}+\Phi(-cC),\end{aligned}$$

and for $c<0$

$$\begin{aligned}&\liminf_{n\to\infty}\max_{x\in\mathcal{X}_n}\|H^x_{n,t_n+cb_n}-\pi_n\|_{\mathrm{TV}}\\&\qquad\le\liminf_{n\to\infty}\max_{x\in\mathcal{X}_n}\|K^x_{n,\lfloor t_n+2cb_n\rfloor}-\pi_n\|_{\mathrm{TV}}+\Phi(-cC).\end{aligned}$$

Proof. The first inequality is immediately implied by Lemma 4.1 and Lemma 2.5 and the others are proved by applying the first inequality and the fact that for bounded sequences $(p_n)_1^\infty$ and $(q_n)_1^\infty$,

$$\sup_{n\geq 1}\{p_n+q_n\} \geq \sup_{n\geq 1} p_n + \inf_{n\geq 1} q_n.$$

□

The next corollary is a simple application of Proposition 4.1.

Corollary 4.1. *Let $\mathcal{F}=\{(\mathcal{X}_n,K_n,\pi_n)\}_1^\infty$ be a family of irreducible Markov chains. Assume that, $T^c_{\mathrm{TV}}(K_n,\epsilon)$ tends to infinity for some $\epsilon>0$. Then, for $\delta\in(0,1)$ and $\eta\in(0,\epsilon)$, there exists an integer $N=N(\delta,\eta)$ such that*

$$(1-\delta)T^c_{\mathrm{TV}}(K_n,\epsilon)\leq T^d_{\mathrm{TV}}(K_n,\eta), \quad \forall n\geq N.$$

In particular, if $\mathcal{F}_c$ and $\mathcal{F}_d$ present a total variation cutoff with respective critical time t_n and s_n, and $t_n\to\infty$, then

$$\liminf_{n\to\infty}\frac{s_n}{t_n}\geq 1.$$

Proof. The first part is an immediate result of the second inequality in Proposition 4.1(1). For the second part, note that $t_n\to\infty$ implies $s_n\to\infty$. By Proposition 1.10 and Proposition 1.11, one has

$$t_n\sim T^c_{\mathrm{TV}}(K_n,\epsilon), \quad s_n\sim T^d_{\mathrm{TV}}(K_n,\epsilon/2).$$

Then by the first part, we get

$$\limsup_{n\to\infty}\frac{t_n}{s_n}\leq\frac{1}{1-\delta}, \quad \forall\delta\in(0,1).$$

This proves the desired inequality. □

Most of the results obtained above deal with the upper bound for the total variation distance of continuous-time Markov chains to their stationarity. In the following, we give a lower bound on the total variation distance by assuming further a limiting property on the discrete-time Markov chains, which is natural for cutoffs on $\mathcal{F}_d$.

Proposition 4.2. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of irreducible Markov chains. For $m, n \geq 1$, let $x_{m,n} \in \mathcal{X}_n$, $A_{m,n} \subset \mathcal{X}_n$ and $t_n > 0$, $b_n > 0$ be positive numbers satisfying $t_n \to \infty$, $\sqrt{t_n} = O(b_n)$, and $b_n = o(t_n)$. Set*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \pi_n(A_{m,n}) = \epsilon, \tag{4.1}$$

$$\liminf_{m\to\infty} \liminf_{n\to\infty} \min\left\{K_n^j(x_{m,n}, A_{m,n}) : (1 - \tfrac{2}{m})t_n \leq j \leq (1 - \tfrac{1}{m})t_n\right\} = \epsilon_1, \tag{4.2}$$

and

$$\liminf_{m\to\infty} \liminf_{n\to\infty} \min\left\{K_n^j(x_{m,n}, A_{m,n}) : t_n - 2mb_n \leq j \leq t_n - mb_n\right\} = \epsilon_2. \tag{4.3}$$

Then, we have

$$\forall \delta \in (0,1), \quad \liminf_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H^x_{n,(1-\delta)t_n} - \pi_n\|_{\mathrm{TV}} \geq \epsilon_1 - \epsilon$$

and

$$\liminf_{c\to-\infty} \liminf_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H^x_{n,t_n+cb_n} - \pi_n\|_{\mathrm{TV}} \geq \epsilon_2 - \epsilon.$$

Proof. Note that, by Lemma 4.1, one has

$$\begin{aligned} \|H^x_{n,t} - \pi_n\|_{\mathrm{TV}} \geq (a(j_2,t) - a(j_1+1,t)) \min\left\{K_n^j(x,A) : j_1 \leq j \leq j_2\right\} \\ -\pi_n(A), \end{aligned} \tag{4.4}$$

where $a(m,t) = e^{-t}\sum_{i=0}^m \frac{t^i}{i!}$.

For the first inequality, since $t_n \to \infty$, by Lemma 2.5, one has, for $c_1 > 0, c_2 > 0$

$$\lim_{n\to\infty} a(c_1 t_n, c_2 t_n) = \begin{cases} 1 & \text{if } c_1 > c_2 \\ 0 & \text{if } c_1 < c_2 \end{cases}.$$

Replacing $t = (1 - \frac{3}{2m})t_n$, $j_1 = \lfloor (1-\frac{2}{m})t_n \rfloor$, $j_2 = \lfloor (1-\frac{1}{m})t_n \rfloor$, $x = x_{m,n}$ and $A = A_{m,n}$ in (4.4) implies that

$$\liminf_{m\to\infty} \liminf_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H^x_{n,(1-\frac{3}{2m})t_n} - \pi_n\|_{\text{TV}} \geq \epsilon_1 - \epsilon.$$

The desired limit is then proved by the monotonicity of $\max_x \|H^x_{n,t} - \pi_n\|_{\text{TV}}$ in t.

For the second inequality, note that, by Lemma 2.5, one may choose $C > 0$ such that

$$\liminf_{n\to\infty} \Big(a(t_n - mb_n, \lfloor t_n - \tfrac{3m}{2} b_n \rfloor) - a(t_n - 2mb_n, \lfloor t_n - \tfrac{3m}{2} b_n \rfloor) \Big) \geq \Phi(Cm) - \Phi(-Cm),$$

where $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$. Letting $m \to \infty$ then proves the second identity. □

Remark 4.1. By Proposition 1.10, it is equivalent in Proposition 4.2 if one replaces t_n with s_n, where $s_n \sim t_n$ in (4.2) and $|s_n - t_n| = O(b_n)$ in (4.3).

Corollary 4.2. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of ergodic Markov chains and, for $n, m \geq 1$, $x_{m,n} \in \mathcal{X}_n$ and $A_{m,n} \subset \mathcal{X}_n$. Assume that t_n is a sequence of positive numbers tending to infinity and $\epsilon = 0$ in (4.1).*

(1) If $\mathcal{F}_d$ presents a total variation cutoff with critical time t_n and $\epsilon_1 = 1$ in (4.2), then $\mathcal{F}_c$ presents a total variation cutoff with critical time t_n.

(2) If $\mathcal{F}_d$ presents a (t_n, b_n) total variation cutoff and $\epsilon_2 = 1$ in (4.3), then $\mathcal{F}_c$ presents a (t_n, c_n) total variation cutoff, where $c_n = \max\{b_n, \sqrt{t_n}\}$.

Proof. Immediate from Proposition 4.1 and Proposition 4.2. □

Example 4.1. Recall Example 2.3: For $n \geq 1$, let $(\mathcal{X}_n, K_n, \pi_n)$ be an irreducible Markov chain, where $\mathcal{X}_n = \mathbb{Z}_{a_n}$ with $a_n > 1$, $\pi_n \equiv a_n^{-n}$ and the Markov kernel is given by

$$K_n(x,y) = \begin{cases} \frac{1}{a_n} & \text{if } y = s(x) + (0,...,0,i) \text{ for some } i \in \mathbb{Z}_{a_n} \\ 0 & \text{otherwise} \end{cases},$$

where $s(x) = (x_2, x_3, ..., x_n, x_1)$ for all $x = (x_1, ..., x_n) \in \mathcal{X}_n$. It is clear that $\mathcal{F}_d$ presents a $(n, 1)$ total variation cutoff.

To apply Corollary 4.2, we choose, for $m \geq 1, n \geq m$, $x_{m,n} = 0 \in \mathcal{X}_n$ and

$$A_{m,n} = \{(0,...,0,z_1,...,z_{n-m}) | z_i \in \mathbb{Z}_{a_n}, \forall 1 \leq i \leq n-m\}.$$

and set $b_n = \sqrt{n}$. In the above setting, we have

$$\lim_{n\to\infty} \pi_n(A_{m,n}) = a_n^{-m} \leq 2^{-m} \to 0,$$

as $m \to \infty$, and

$$\min\left\{K_n^j(0, A_{m,n}) : n - 2mb_n \leq j \leq n - mb_n\right\} = 1 \quad \forall n, m \geq 1.$$

Hence, by Corollary 4.2, the family $\mathcal{F}_c$ presents a $(n, \sqrt{n})$ total variation cutoff.

Concerning the optimality of the window size $\sqrt{n}$, we claim first that

$$\lim_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H_{n,n}^x - 1\|_{\text{TV}} = \frac{1}{2}.$$

By the first part of Lemma 4.1 and Lemma 2.5, one has

$$\limsup_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H_{n,n}^x - 1\|_{\text{TV}} \leq \frac{1}{2}.$$

For the lower bound, let $A_{m,n}$ be the set defined above. Then the second part of Lemma 4.1 implies

$$\max_{x\in\mathcal{X}_n} \|H^x_{n,t} - 1\|_{\mathrm{TV}} \geq a(n-m,t) - 2^{-m}. \tag{4.5}$$

Hence we have

$$\liminf_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H^x_{n,n} - 1\|_{\mathrm{TV}} \geq \frac{1}{2} - 2^{-m}, \quad \forall m \geq 1.$$

Letting $m \to \infty$ then derives the desired identity.

Now let c_n be a positive number satisfying $c_n \geq 1$ and $c_n = o(\sqrt{n})$. By (4.5), we get

$$\liminf_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H^x_{n,n+cc_n} - 1\|_{\mathrm{TV}} \geq \frac{1}{2} - 2^{-m}, \quad \forall m \geq 1, c > 0.$$

Letting $m \geq 2$ implies that $\mathcal{F}_c$ can't have a (n, c_n) total variation cutoff. By Corollary 1.6, the family $\mathcal{F}_c$ presents an optimal $(n, \sqrt{n})$ total variation cutoff.

4.2 Peres' conjecture and Aldous' counterexample

At an ARCC workshop held by AIM in Palo Alto, December 2004, Peres formulated a conjecture as follows. Consider a family of finite Markov chains $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ with $\ell^1(\pi_n)$-mixing time $T_1(K_n, \epsilon)$ and spectral gap λ_n of K_n. The conjecture is: the family $\mathcal{F}$ has a ℓ^1-cut-off if and only if

$$\lambda_n^{-1} = o(T_1(K_n, 1)). \tag{4.6}$$

As we mentioned in Chapter 2, if the distance is measured with the ℓ^p-norm for $1 < p < \infty$ and the transition matrix is normal, then this conjecture is true for continuous-time cases which is proved in Theorem 2.4. For $p = 1$, we know from Remark 2.7 and Lemma 2.4 that if $\mathcal{F}$ has a total variation cutoff, then (4.6) holds.

At the workshop mentioned above, Aldous presented a counterexample to Peres' conjecture. In the following, we build on Aldous' idea and describe a series of examples satisfying (4.6) but failing to present a ℓ^1-cut-off both in discrete-time and continuous-time cases. Note that there is no known counterexample for Peres' conjecture in the case of random walks on finite groups.

Consider a family $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ of finite Markov chains, where

$$\mathcal{X}_n = \{0, 1, x_1, ..., x_n, y_1, ..., y_{2n}, z_1, ..., z_{3n}\}, \tag{4.7}$$

and the Markov kernel is given by

$$\begin{aligned}
&K_n(z_i, z_{i+1}) = \frac{p_{n,1} + p_{n,2}}{2}, \quad K_n(z_{i+1}, z_i) = \frac{q_{n,1} + q_{n,2}}{2} \quad \forall 1 \le i < 3n, \\
&K_n(z_1, z_1) = K_n(0, z_{3n}) = \frac{q_{n,1} + q_{n,2}}{2}, \quad K_n(z_{3n}, 0) = \frac{p_{n,1} + p_{n,2}}{2} \\
&K_n(x_i, x_{i+1}) = p_{n,1}, \quad K_n(x_{i+1}, x_i) = q_{n,1} \quad \forall 1 \le i < n, \\
&K_n(0, x_1) = \frac{p_{n,1}}{2}, \quad K_n(x_1, 0) = q_{n,1}, \\
&K_n(x_n, 1) = p_{n,1}, \quad K_n(1, x_n) = \frac{q_{n,1}}{2}, \\
&K_n(y_i, y_{i+1}) = p_{n,2}, \quad K_n(y_{i+1}, y_i) = q_{n,2} \quad \forall 1 \le i < 2n, \\
&K_n(0, y_1) = \frac{p_{n,2}}{2}, \quad K_n(y_1, 0) = q_{n,2}, K_n(y_{2n}, 1) = p_{n,2}, \\
&K_n(1, y_{2n}) = \frac{q_{n,2}}{2}, \quad K_n(1, 1) = \frac{p_{n,1} + p_{n,2}}{2},
\end{aligned} \tag{4.8}$$

with $0 < p_{n,j} < 1$ and $p_{n,j} + q_{n,j} = 1$ for $j \in \{1, 2\}$ and $n > 0$. Clearly, K_n is ergodic. Assume that

$$\left(\frac{p_{n,1}}{q_{n,1}}\right)^{n+1} = \left(\frac{p_{n,2}}{q_{n,2}}\right)^{2n+1}, \quad \forall n \ge 1. \tag{4.9}$$

Then the stationary distribution π_n is given by

$$\begin{aligned}
\pi_n(z_i) &= Z^{-1}\left(\frac{p_{n,1}+p_{n,2}}{q_{n,1}+q_{n,2}}\right)^{i-1} \quad \forall 1 \le i \le 3n,\\
\pi_n(0) &= Z^{-1}\left(\frac{p_{n,1}+p_{n,2}}{q_{n,1}+q_{n,2}}\right)^{3n},\\
\pi_n(x_i) &= \frac{1}{2}\left(\frac{p_{n,1}}{q_{n,1}}\right)^{i}\pi_n(0), \quad \pi_n(y_j) = \frac{1}{2}\left(\frac{p_{n,2}}{q_{n,2}}\right)^{j}\pi_n(0) \quad \forall i,j \ge 1\\
\pi_n(1) &= \left(\frac{p_{n,1}}{q_{n,1}}\right)^{n+1}\pi_n(0) = \left(\frac{p_{n,2}}{q_{n,2}}\right)^{2n+1}\pi_n(0),
\end{aligned} \tag{4.10}$$

where Z is a normalizing constant for π_n. In the above setting, one can easily check that K_n is reversible and $p_{n,1} > p_{n,2}$ if $p_{n,1} > 1/2$ and $p_{n,1} < p_{n,2}$ if $p_{n,1} < 1/2$. Here, we restrict ourselves to the case $p_{n,1} > 1/2$ for all $n \ge 1$.

From Proposition 4.3 and Proposition 4.4(in the following), it is clear that $\mathcal{F}$ is the desired family for a counterexample of Peres' conjecture if one assumes (4.11).

Proposition 4.3. *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n)\}_1^\infty$ be a family of finite Markov chains satisfying* (4.7), (4.8) *and* (4.10). *Assume that*

$$\liminf_{n\to\infty} p_{n,2}^n > 0. \tag{4.11}$$

Then there is no total variation cutoff for $\mathcal{F}_c$ and $\mathcal{F}_d$.

The following remark says that the sequence of total variation mixing time for $\mathcal{F}$ is of order at least n.

Remark 4.2. Clearly, one has that for $0 \le j \le 4n$,

$$\max_{x\in\mathcal{X}_n} \|K_{n,x}^j - \pi_n\|_{\mathrm{TV}} \ge \pi_n(1) - K_n^j(z_1, 1) = \pi_n(1),$$

and for $t \ge 0$,

$$\max_{x\in\mathcal{X}_n} \|H_{n,t}^x - \pi_n\|_{\mathrm{TV}} \ge \pi_n(1) - H_{n,t}(z_1, 1) \ge \pi_n(1) - e^{-t}\sum_{j=4n+1}^{\infty}\frac{t^j}{j!}.$$

Under the assumption of (4.11), the above implies

$$\lim_{n\to\infty}\max_{x\in\mathcal{X}_n}\|K^{4n}_{n,x}-\pi_n\|_{\text{TV}}=1$$

and, by Lemma 2.5,

$$\forall c\in(0,4),\quad \lim_{n\to\infty}\max_{x\in\mathcal{X}_n}\|H^x_{n,cn}-\pi_n\|_{\text{TV}}=1.$$

Hence, we may choose, for $\epsilon\in(0,1)$, an integer $N(\epsilon)$ such that

$$\forall n\ge N(\epsilon),\quad T^d_{\text{TV}}(K_n,\epsilon)\ge n,\quad T^c_{\text{TV}}(K_n,\epsilon)\ge n.$$

Proposition 4.4. *Let $\mathcal{F}=\{(\mathcal{X}_n,K_n,\pi_n)\}_1^\infty$ be as in Proposition 4.3 and, for $n\ge 1$, λ_n and μ_n be the spectral gap and the second largest singular value of K_n. Assume that*

$$\inf_{n\ge 1}p_{n,2}\ge\frac{2}{3}. \tag{4.12}$$

Then λ_n and $1-\mu_n^2$ are bounded from below by a positive number.

Remark 4.3. It is worth noting that, by Lemma 2.4, the above two propositions are also sufficient for a counterexample of the following statement

$$b_n^{-1}=o\left(T^d_{\text{TV}}(K_n,\epsilon)\right)\ \Rightarrow\ \mathcal{F}_d \text{ presents a total variation cutoff,}$$

where $b_n=\min\{-\log\mu_n,1\}$.

Proof of Proposition 4.3. By considering the transition path

$$\underbrace{x_i,x_{i+1},...,x_n}_{n-i+1},\overbrace{1,1,...,1}^{3n+i+1},$$

one has

$$K_n^{4n+1}(x_i,1)\ge p_{n,1}^{n-i+1}\left(\frac{p_{n,1}+p_{n,2}}{2}\right)^{3n+i}\ge p_{n,2}^{4n+1}.$$

A similar argument by considering the following paths

$$\overbrace{0, x_1, ..., x_n}^{n+1}, \overbrace{1, ..., 1}^{3n+1}, \qquad \overbrace{1, ..., 1}^{4n+2}, \qquad \overbrace{z_k, .., z_{3n}}^{3n-k+1}, 0, \overbrace{x_1, ..., x_n}^{n}, \overbrace{1, ..., 1}^{k},$$

$$\overbrace{y_j, y_{j+1}, ..., y_{2n}}^{2n-j+1}, \overbrace{1, 1, ..., 1}^{2n+j+1}, \qquad \overbrace{z_1, ..., z_{3n}}^{3n}, 0, \overbrace{y_1, ..., y_{2n}}^{2n},$$

we have

$$\min_{x \in \mathcal{X}_n} \{K_n^{4n+1}(x, 1)\} \geq \frac{p_{n,2}^{4n+1}}{2}, \quad K_n^{5n}(z_1, 1) \leq 1 - \frac{p_{n,2}^{5n}}{2},$$

which implies

$$\max_{x \in \mathcal{X}_n} \|K_{n,x}^{4n+1} - \pi_n\|_{\text{TV}} \leq 1 - \min \left\{ \pi_n(1), \frac{p_{n,2}^{4n+1}}{2} \right\} \tag{4.13}$$

and

$$\max_{x \in \mathcal{X}_n} \|K_{n,x}^{5n} - \pi_n\|_{\text{TV}} \geq \pi_n(1) - 1 + \frac{p_{n,2}^{5n}}{2}. \tag{4.14}$$

Let $C = \liminf_{n\to\infty} p_{n,2}^n \in (0, 1]$. For discrete-time cases, it suffices to prove that

$$\limsup_{n\to\infty} \max_{x \in \mathcal{X}_n} \|K_{n,x}^{4n+1} - \pi_n\|_{\text{TV}} < 1, \quad \liminf_{n\to\infty} \max_{x \in \mathcal{X}_n} \|K_{n,x}^{5n} - \pi_n\|_{\text{TV}} > 0.$$

By (4.11), one has

$$\lim_{n\to\infty} p_{n,2} = 1, \quad \lim_{n\to\infty} \pi_n(1) = 1. \tag{4.15}$$

Then, in addition to the facts in (4.13) and (4.14), we have

$$\limsup_{n\to\infty} \max_{x \in \mathcal{X}_n} \|K_{n,x}^{4n+1} - \pi_n\|_{\text{TV}} \leq 1 - \frac{C^4}{2} < 1,$$

and

$$\liminf_{n\to\infty} \max_{x \in \mathcal{X}_n} \|K_{n,x}^{5n} - \pi_n\|_{TV} \geq C^5 > 0.$$

This proves that $\mathcal{F}_d$ does not present a total variation cutoff.

For continuous-time cases, by applying Proposition 4.1(1) with $t_n = 10(4n+1)/9$ and $c = 9/10$, one has

$$\limsup_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H^x_{n,9n/2} - \pi_n\|_{\mathrm{TV}} < 1.$$

To prove that $\mathcal{F}_c$ does not present a total variation cutoff, it remains to show that

$$\liminf_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H^x_{n,cn} - \pi_n\|_{\mathrm{TV}} > 0,$$

for some $c > 9/2$. By considering the following paths

$$\overbrace{z_1,...,z_{j+1}}^{j+1} \quad \text{for } 0 \le j \le 3n-1, \quad \overbrace{z_1,...,z_{3n},0}^{3n+1} \quad \text{for } j = 3n,$$
$$\overbrace{z_1,...,z_{3n}}^{3n},0,\overbrace{y_1,...,y_{j-3n}}^{j-3n} \quad \text{for } 3n+1 \le j \le 5n,$$

we get

$$K^j_n(z_1, 1) \le 1 - \frac{C^5}{2} \quad \forall 0 \le j \le 5n.$$

Then, applying Proposition 4.2 with $x_{n,m} \equiv z_1$, $A_{n,m} = \mathcal{X}_n \setminus \{1\}$, $t_n = 5n$ and $\delta = 1/20$ implies

$$\liminf_{n\to\infty} \max_{x\in\mathcal{X}_n} \|H^x_{n,19n/4} - \pi_n\|_{\mathrm{TV}} \ge \frac{C^5}{2} > 0.$$

□

Proof of Proposition 4.4. Here we use Cheeger's inequality to prove this proposition. To state that inequality, we need the following setting. Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain. For any set $A \subset \mathcal{X}$, we define

$$\partial A = \{(x, y) \in \mathcal{X} \times \mathcal{X} | x \in A, y \in A^c \text{ or } y \in A, x \in A^c\},$$

and

$$Q(\partial A) = \frac{1}{2} \sum_{x\in A, y\in A^c} [\pi(x)K(x, y) + \pi(y)K(y, x)] = \sum_{x\in A, y\in A^c} \pi(x)K(x, y).$$

An isoperimetric constant, or the conductance, of the chain $(\mathcal{X}, K, \pi)$ is given by

$$\mathcal{I} = \mathcal{I}(K, \pi) = \min_{\substack{A \subset \mathcal{X}: \\ \pi(A) \leq 1/2}} \left\{ \frac{Q(\partial A)}{\pi(A)} \right\}.$$

Cheeger's inequality. Let $(\mathcal{X}, K, \pi)$ be a finite Markov chain and λ and $\mathcal{I}$ be the spectral gap and the isoperimetric constant of K. Then, one has

$$\frac{\mathcal{I}^2}{8} \leq \lambda \leq \mathcal{I}.$$

Here we will use the first inequality in the above to give a lower bound on λ_n and $1 - \mu_n^2$. First, assume that $p_{n,1} > 1/2$ and set $r_n = p_{n,2}/q_{n,2}$. Then $r_n > 1$ and, by (4.9),

$$r_n = \min \left\{ \frac{p_{n,2}}{q_{n,2}}, \frac{p_{n,1}}{q_{n,1}}, \frac{p_{n,1} + p_{n,2}}{q_{n,1} + q_{n,2}} \right\}.$$

By the formula of the stationary distribution π_n in (4.10), a bunch of computations implies that for $1 \leq j \leq 3n$,

$$\sum_{i=1}^{j} \pi_n(z_i) \leq \frac{r_n}{r_n - 1} \pi_n(z_j), \quad \sum_{i=1}^{3n} \pi_n(z_i) + \pi_n(0) \leq \frac{r_n}{r_n - 1} \pi_n(0),$$

and for $1 \leq j \leq n, 1 \leq k \leq 2n$,

$$\begin{aligned} &\sum_{i=1}^{3n} \pi_n(z_i) + \pi_n(0) + \sum_{i=1}^{j} \pi_n(x_i) + \sum_{i=1}^{k} \pi_n(y_i) \\ \leq &\frac{\pi_n(x_j)}{2} \left(\frac{r_n}{r_n - 1} \left(\frac{q_{n,1}}{p_{n,1}} \right)^j + \sum_{i=0}^{j-1} \left(\frac{q_{n,1}}{p_{n,1}} \right)^i \right) \\ &\qquad + \frac{\pi_n(y_k)}{2} \left(\frac{r_n}{r_n - 1} \left(\frac{q_{n,2}}{p_{n,2}} \right)^k + \sum_{i=0}^{k-1} \left(\frac{q_{n,2}}{p_{n,1}} \right)^i \right) \\ \leq &\frac{r_n}{r_n - 1} \max\{\pi_n(x_j), \pi_n(y_k)\}. \end{aligned}$$

Now assume further that $r_n \geq 2$ or equivalently $p_{n,2} \geq 2/3$, which implies $\pi_n(1) > 1/2$. Let $A \subset \mathcal{X}_n$ with $\pi_n(A) \leq 1/2$ and $x \in A$ be such that $\pi_n(x) =$

$\max\{\pi_n(z)|z \in A\}$. Since $1 \notin A$, there is always a vertex y connecting to x such that $\pi_n(y) > \pi_n(x)$. By this observation and the above computation, one has

$$Q(\partial A) \geq \pi_n(x)K_n(x,y) \geq \frac{2}{3}\pi_n(x) \geq \frac{1}{3}\pi_n(A),$$

and then, $\mathcal{I}_n = \mathcal{I}(K_n, \pi_n) \geq 1/3$. Hence, by Cheeger's inequality, we get

$$\lambda_n \geq \frac{\mathcal{I}_n^2}{8} \geq \frac{1}{72}.$$

For the lower bound of $1 - \mu_n^2$, note that this quantity is the spectral gap of $KK^* = K^2$. Let $A \subset \mathcal{X}_n$ be such that $\pi_n(A) \leq 1/2$ and x, as before, the element in A maximizing $\pi_n(z)$ for $z \in A$. Note that one can always choose a one-step neighbor $y \neq A$ of x under the transition kernel K_n^2 such that $\pi_n(y) > \pi_n(x)$. By this observation, a similar computation as before derives $Q(\partial A) \geq \frac{2}{9}\pi_n(A)$. Then the isoperimetric constant satisfies $\mathcal{I}(K_n^2, \pi_n) \geq 2/9$ and

$$1 - \mu_n^2 \geq \frac{\mathcal{I}_n^2}{8} \geq \frac{1}{182}.$$

$\square$

Chapter 5

Randomized riffle shuffle

In this chapter we consider some generalizations of the standard riffle shuffle of Gilbert, Shannon and Reeds (GSR-shuffle for short). The GSR-shuffle models the way typical card players shuffle cards. First, the deck is cut into two packs according to an $(n, \frac{1}{2})$-binomial random variable where n is the number of cards in the deck. Next, cards are dropped one by one from one or the other pack with probability proportional to the relative sizes of the packs. Hence, if the left pack contains a cards and the right pack b cards, the next card drops from the left pack with probability $a/(a+b)$.

The history of this model is described in [9, Chap. 4D] where the reader will also find other equivalent definitions and a discussion of how the model relates to real life card shuffling. The survey [12] gives pointers to the many developments that arose from the study of the GSR model.

Early results concerning the mixing time (i.e., how many shuffles are needed to mix up the deck) are described in [1, 2, 9]. In particular, using ideas of Reeds, Aldous proved in [1] that, asymptotically as the number n of cards tends to infinity, it takes $\frac{3}{2}\log_2 n$ shuffles to mix up the deck if convergence is measured in total variation (we use $\log_a$ to denote base a logarithms and log for natural, i.e., base e, logarithms).

In [6], Bayer and Diaconis obtained an exact useful formula for the probability distribution describing the state of the deck after k GSR-shuffles. Namely, suppose that cards are numbered 1 through n and that we start with the deck in order. Let σ denote a given arrangement of the cards and let $Q_n^k(\sigma)$ be the probability that the deck is in state σ after k GSR-shuffles. Then

$$Q_n^k(\sigma) = 2^{-kn}\binom{n+2^k-r}{n} \tag{5.1}$$

where r is the number of rising sequences in σ. Given an arrangement of the deck, a rising sequence is a maximal subset of cards consisting of successive face values displayed in order. For instance, the arrangement $3, 1, 4, 5, 7, 2, 8, 9, 6$ has rising sequences $(1, 2), (3, 4, 5, 6), (7, 8, 9)$. See [2, 6] for details. Using this formula, Bayer and Diaconis gave a very sharp version of the fact that the total variation mixing time is $\frac{3}{2}\log_2 n$ for the GSR-shuffle.

Theorem 5.1 (Bayer and Diaconis [6]). *Fix $c \in (-\infty, +\infty)$. For a deck of n cards, the total variation distance between the uniform distribution and the distribution of a deck after $k = \frac{3}{2}\log_2 n + c$ GSR-shuffles is*

$$\frac{1}{\sqrt{2\pi}}\int_{-2^{-c}/4\sqrt{3}}^{2^{-c}/4\sqrt{3}} e^{-t^2/2}dt + O_c(n^{-1/4}).$$

This result illustrates beautifully the so-called cutoff phenomenon discussed in [1, 2, 3, 9, 11, 29, 33]. Namely, there is a sharp transition in convergence to stationarity. Indeed, the integral above becomes small very fast as c tends to $+\infty$ and gets close to 1 even faster as c tends to $-\infty$.

The aim of this chapter is to illustrate further the notion of cutoff using some generalizations of the GSR-shuffle. Along this way we will observe several phenomena that have not been, to the best of our knowledge, noticed before. For a deck of n cards and a given integer m, a m-riffle shuffle is defined as follows. Cut the deck into m packs whose sizes $(a_1, \ldots, a_m)$ form a multinomial random vector. In other words, the probability of having packs of sizes $a_1, \ldots, a_m$ is $m^{-n}\frac{n!}{a_1! \ldots a_m!}$. Then form a new deck by dropping cards one by one from these packs with probability proportional to the relative sizes of the packs. Thus, if the packs have sizes $(b_1, \ldots, b_m)$

then the next card will drop from pack i with probability $b_i/(b_1 + \cdots + b_m)$. We will refer to an m-riffle shuffle simply as an m-shuffle in what follows. Obviously the GSR-shuffle is the same as a 2-shuffle.

These shuffles were considered in [6] where the following two lemmas are proved.

Lemma 5.1. *In distribution, an a-shuffle followed by an independent b-shuffle equals an ab-shuffle.*

Lemma 5.2. *For a deck of n cards in order, the probability that after an m-shuffle the deck is in state σ depends only of the number $r = r(\sigma)$ of rising sequences of σ and equals $Q_{n,m}(r)$ where*

$$Q_{n,m}(r) = m^{-n}\binom{n+m-r}{n}.$$

For instance, formula (5.1) for the distribution of the deck after k GSR-shuffles follows from a direct application of these two lemmas since k consecutive independent 2-shuffles equal a 2^k-shuffle in distribution. These lemmas will play a crucial role in this paper as well.

The model we consider is as follows. Let $p = (p(1), p(2), \dots)$ be the probability distribution of an integer valued random variable X, i.e.,

$$P(X = k) = p(k), \quad k = 1, 2, \dots.$$

A p-shuffle proceeds by picking an integer m according to p and performing an m-shuffle. In other words, the distribution of a p-shuffle is the p-mixture of the m-shuffle distributions. Note that Casinos use multiple decks for some games and that these are shuffled in various ways (including by shuffling machines). The model above (for some appropriate p) is not entirely unrealistic in this context.

Because of Lemma 5.2, the probability that starting from a deck in order we obtain a deck in state σ depends only on the number of rising sequences in σ and is given by

$$Q_{n,p}(r) = \sum_{1}^{\infty} p(m) Q_{n,m}(r) = E(Q_{n,X}(r)). \tag{5.2}$$

Abusing notation, if σ denotes a deck arrangement of n cards with r rising sequences, we write

$$Q_{n,p}(\sigma) = Q_{n,p}(r).$$

Very generally, if Q is a probability measure on deck arrangements (hence describes a shuffling method), we denote by Q^k the distribution of the deck after k such shuffles, starting from a deck in order. For instance, Lemma 5.1 yields

$$Q_{n,m}^k = Q_{n,m^k}.$$

Let U_n be the uniform distribution on the set of deck arrangements of n cards. Although this will not really play a role in this work, recall that deck arrangements can be viewed as elements of the symmetric group S_n in such a way that Q^k, the distribution after k successive Q-shuffles, is the k-fold convolution of Q by itself. See, e.g., [1, 6, 9, 30]. Each of the measure $Q_{n,p}$ generates a Markov chain on deck arrangements (i.e., on the symmetric group S_n) whose stationary distribution is U_n. These chains are ergodic if p is not concentrated at 1. They are not reversible. Note that [13] studies a similar but different model based on top m to random shuffles. See [13, Section 2].

The goal of this paper is to study the convergence of $Q_{n,p}^k$ to the uniform distribution in total variation as k tends to infinity and, more precisely, the occurrence of a total variation cutoff for families of shuffles $\{(S_n, Q_{n,p_n}, U_n)\}_1^\infty$ as the number n of cards grows to infinity and p_n is a fixed sequence of probability measures on the integers. To illustrate this, we state the simplest of our results.

Theorem 5.2. *Let p be a probability on the integers such that*

$$\mu = \sum_{1}^{\infty} p(k) \log k < \infty. \tag{5.3}$$

Fix $\epsilon \in (0,1)$. Then, for any $k_n > (1+\epsilon)\frac{3}{2\mu}\log n$, we have

$$\lim_{n\to\infty} \|Q_{n,p}^{k_n} - U_n\|_{\mathrm{TV}} = 0$$

whereas, for $k_n < (1-\epsilon)\frac{3}{2\mu}\log n$,

$$\lim_{n\to\infty} \|Q_{n,p}^{k_n} - U_n\|_{\mathrm{TV}} = 1.$$

In words, this theorem establishes a total variation cutoff at time $\frac{3}{2\mu}\log n$ (see the definition of cutoff in Section 5.1 below). If p is concentrated at 2, i.e., $Q_{n,p}$ represents a GSR-shuffle, then $\mu = \log 2$ and $\frac{3}{2\mu}\log n = \frac{3}{2}\log_2 n$ in accordance with the results of Aldous [1] and Bayer-Diaconis [6] (e.g., Theorem 5.1).

The results we obtain are more general and more precise than Theorem 5.2 in several directions. First, we will consider the case where the probability distribution $p = p_n$ depends on the size n of the deck. This is significant because we will not impose that the sequence p_n converges as n tends to infinity. Second, and this may be a little surprising at first, (5.3) is not necessary for the existence of a cutoff and we will give sufficient conditions that are weaker than (5.3). Third, under stronger moment assumptions, we will describe the optimal window size of the cutoff. For instance, Theorem 1 says that, for the GSR-shuffle, the window size is of order 1 with a normal shape. This result generalizes easily to any m-shuffle where m is a fixed integer greater or equal to 2. See Remark 5.1 and Theorem 5.10 below. Suppose now that instead of the GSR-shuffle we consider the p-shuffle with $p(2) = p(3) = 1/2$. In this case, $\mu = \log\sqrt{6}$. Theorem 5.2 gives a total variation cutoff at time $\frac{3}{2}\log_{\sqrt{6}} n$. We will show that this cutoff has optimal window size of order $\sqrt{\log n}$. Thus picking at random between 2 and 3 shuffles changes the window size significantly when compared to either pure 2-shuffles or pure 3-shuffles.

We close this introduction with a remark concerning the spectrum of these generalized riffle shuffles and how it relates to the window of the cutoff. As Lemma 5.1 makes clear, all riffle shuffles commute. Although riffle shuffles are not reversible, they are all diagonalizable with real positive eigenvalues and their spectra can be computed explicitly (this is another algebraic "miracle" attached to these shuffles!). See [6, 7, 8]. In particular, the second largest eigenvalue of an m-shuffle is $1/m$. Thus, the second largest eigenvalue of a p-shuffle is $\beta = \sum k^{-1}p(k)$. By definition, the relaxation time of a finite Markov chain is the inverse of the spectral gap $(1-\beta)^{-1}$ and one might expect that, quite generally, for families of Markov chains presenting a cutoff, this quantity would give a good control of the window of the cutoff. The generalized riffle shuffles studied here provided interesting (albeit non-reversible) counterexamples: Take, for instance, the case discussed earlier where $p(2) = p(3) = 1/2$. Then $\beta = \frac{5}{12}$ and $(1-\beta)^{-1} = \frac{12}{7}$, independently of the number n of cards. However, as mentioned above, the optimal window size of the cutoff for this family is $\sqrt{\log n}$. For generalized riffle shuffles, the window size of the cutoff and the relaxation time appear to be disconnected.

5.1 The cutoff phenomenon

Given two probability distributions μ, ν on a set S, the total variation distance between μ and ν is defined by

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A\subset S}\{\mu(A) - \nu(A)\}.$$

The next definition introduces the notion of cutoff for a family of ergodic Markov chains.

Definition 5.1. Let $\{(S_n, K_n, \pi_n)\}_1^\infty$ be a family of ergodic Markov chains where S_n denotes the state space, K_n the Markov kernel, and π_n the stationary distribution. This family satisfies a total variation cutoff with critical time $t_n > 0$ if, for any fixed $\epsilon \in (0, 1)$,

$$\lim_{n\to\infty} \sup_{x\in S_n} \|K_n^{k_n}(x, \cdot) - \pi_n\|_{\mathrm{TV}} = \begin{cases} 0 & \text{if } k_n > (1+\epsilon)t_n \\ 1 & \text{if } k_n < (1-\epsilon)t_n. \end{cases}$$

This definition was introduced in [2]. A more thorough discussion is in [11] where many examples are described. Note that this definition does not require that the critical time t_n tends to infinity (in [11], the corresponding definition requires that t_n tends to infinity). The positive times t_n can be arbitrary and thus can have several limit points in $[0, \infty]$. Examples of families having a cutoff with a bounded critical time sequence will be given below. Theorem 5.2 above states that, under assumption (5.3), a p-shuffle has a total variation cutoff with critical time $t_n = \frac{3}{2\mu}\log n$.

Informally, a family has a cutoff if convergence to stationarity occurs in a time interval of size $o(t_n)$ around the critical time t_n. The size of this time interval can be thought of as the "window" of the cutoff. The next definition carefully defines the notion of the window size of a cutoff.

Definition 5.2. Let $\{(S_n, K_n, \pi_n)\}_1^\infty$ be a family of ergodic Markov chains as in Definition 5.1. We say that this family presents a (t_n, b_n) total variation cutoff if the following conditions are satisfied:

1. For all $n = 1, 2, \ldots,$ we have $t_n > 0$ and $\lim_{n\to\infty} b_n/t_n = 0$.

2. For $c \in \mathbb{R} - \{0\}$ and $n \geq 1$, set

$$k = k(n,c) = \begin{cases} \lceil t_n + cb_n \rceil & \text{if } c > 0 \\ \lfloor t_n + cb_n \rfloor & \text{if } c < 0 \end{cases}.$$

The functions $\overline{f}, \underline{f}$ defined by

$$\overline{f}(c) = \limsup_{n\to\infty} \sup_{x\in S_n} \|K_n^k(x,\cdot) - \pi_n\|_{\mathrm{TV}} \text{ for } c \neq 0$$

and

$$\underline{f}(c) = \liminf_{n\to\infty} \sup_{x\in S_n} \|K_n^k(x,\cdot) - \pi_n\|_{\mathrm{TV}} \text{ for } c \neq 0$$

satisfy

$$\lim_{c\to\infty} \overline{f}(c) = 0, \quad \lim_{c\to-\infty} \underline{f}(c) = 1.$$

Definition 5.3. Referring to Definition 5.2, a (t_n, b_n) total variation cutoff is said to be optimal if the functions $\overline{f}, \underline{f}$ satisfy $\underline{f}(c) > 0$ and $\overline{f}(-c) < 1$ for all $c > 0$.

Note that this definition is the same as the strong optimality given in Definition 1.6. As mentioned in Remark 1.8, any family having a (t_n, b_n) cutoff (Definition 5.2) has a cutoff with critical time t_n (Definition 5.1). The sequence $(b_n)_1^\infty$ in Definition 5.2 describes an upper bound on the optimal window size of the cutoff. For instance the main result of Bayer and Diaconis [6], i.e., Theorem 5.1 above, shows that the GSR-shuffle family presents a (t_n, b_n) total variation cutoff with $t_n = \frac{3}{2}\log_2 n$ and $b_n = 1$. Theorem 5.1 actually determines exactly "the shape" of the cutoff, that is, the two functions $\overline{f}, \underline{f}$ of Definition 5.2. Namely, for the GSR-shuffle family and $t_n = \frac{3}{2}\log_2 n$, $b_n = 1$, we have

$$\overline{f}(c) = \underline{f}(c) = \frac{1}{\sqrt{2\pi}} \int_{-2^{-c}/4\sqrt{3}}^{2^{-c}/4\sqrt{3}} e^{-t^2/2} dt.$$

This shows that this cut-off is optimal (Definition 5.3).

The optimality introduced in Definition 5.3 is very strong. If a family presents an optimal (t_n, b_n) total variation cut-off and also a (s_n, c_n) total variation cut-off, then $t_n \sim s_n$ and $b_n = O(c_n)$. In words, if (t_n, b_n) is an optimal cut-off then there are no cut-offs with a window significantly smaller than b_n. For a more detailed discussion of the cutoff phenomena and their optimality, see Chapter 1 and Chapter 2.

5.2 Cutoffs for generalized riffle shuffles

In this section we state our main results and illustrate them with simple examples. They describe total variation cutoffs for generalized riffle shuffles, that is, for the p-shuffles defined in the introduction. More precisely, for each n (n is the number of cards), fix a probability distribution $p_n = (p_n(1), p_n(2), \dots)$ on the integers and consider the family of Markov chains (i.e., shuffles)

$$\{(S_n, Q_{n,p_n}, U_n)\}_1^\infty.$$

Here S_n is the set of all deck arrangements (i.e., the symmetric group) and U_n is the uniform measure on S_n, i.e., $U_n(A) = \frac{|A|}{n!}$ where $|A|$ is the number of elements in the set $A \subset S_n$. For any $x \in [0, \infty]$, set

$$\Psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-x/4\sqrt{3}}^{x/4\sqrt{3}} e^{-t^2/2} dt.$$

We start with the simple case where the probability distributions p_n is concentrated on exactly one integer m_n and use the notation Q_{n,m_n} for an m_n-shuffle.

Theorem 5.3. *Let $(m_n)_1^\infty$ be any sequence of integers all greater than 1 and set*

$$\mu_n = \log m_n, \quad t_n = \frac{3 \log n}{2\mu_n}.$$

Then the family $\{(S_n, Q_{n,m_n}, U_n)\}_1^\infty$ presents a (t_n, μ_n^{-1}) total variation cutoff.

Remark 5.1. When $m_n = m$ is constant Theorem 5.3 gives a $(\frac{3}{2}\log_m n, 1)$ total variation cutoff. In this case, for $k = \frac{3}{2}\log_m n + c$, one has the more precise result that $\|Q^k_{n,m} - U_n\|_{\text{TV}} = \Psi(m^{-c}) + O_c(n^{-1/4})$. In particular, for $m = 2$, this is the Theorem of Bayer and Diaconis stated as Theorem 1 in the introduction.

Next we give a more explicit version of Theorem 5.3 which requires some additional notation. For any real $t > 0$, set

$$\{t\} = \begin{cases} 1/2 & \text{if } 0 < t < 1/2 \\ k & \text{if } k - 1/2 \le t < k + 1/2 \text{ for some } k = 1, 2, \dots, \end{cases}$$

(this is a sort of "integer part" of t) and

$$d(t) = \begin{cases} 1/2 & \text{if } 0 < t < 1/2 \\ t - \{t\} & \text{if } 1/2 \le t < \infty. \end{cases}$$

Theorem 5.4. *Let $(m_n)_1^\infty$ be any sequence of integers all greater than 1. Consider the family of shuffles $\{(S_n, Q_{n,m_n}, U_n)\}_1^\infty$ and let μ_n, t_n be as in Theorem 5.3.*

(A) Assume that $\lim_{n\to\infty} m_n = \infty$, that is, $\lim_{n\to\infty} \mu_n = \infty$. Then, we have:

(1) The family $\{(S_n, Q_{n,m_n}, U_n)\}_1^\infty$ always has a $(\{t_n\}, b_n)$ cutoff for any positive $b_n = o(1)$, that is,

$$\lim_{n\to\infty} \inf_{k<\{t_n\}} \|Q^k_{n,m_n} - U_n\|_{\text{TV}} = 1, \quad \lim_{n\to\infty} \sup_{k>\{t_n\}} \|Q^k_{n,m_n} - U_n\|_{\text{TV}} = 0.$$

(2) If $\lim_{n\to\infty} |d(t_n)|\mu_n = \infty$ then there is a $(t_n, 0)$ cutoff, that is,

$$\lim_{n\to\infty} \inf_{k\le t_n} \|Q^k_{n,m_n} - U_n\|_{\text{TV}} = 1, \quad \lim_{n\to\infty} \sup_{k\ge t_n} \|Q^k_{n,m_n} - U_n\|_{\text{TV}} = 0.$$

(3) If $\liminf_{n\to\infty} |d(t_n)|\mu_n < \infty$ then there exists a sequence $(n_i)_1^\infty$ tending to infinity such that

$$0 < \liminf_{i\to\infty} \|Q^{\{t_{n_i}\}}_{n_i,m_{n_i}} - U_{n_i}\|_{\text{TV}} \le \limsup_{i\to\infty} \|Q^{\{t_{n_i}\}}_{n_i,m_{n_i}} - U_{n_i}\|_{\text{TV}} < 1.$$

In particular, there is no $(t_n, 0)$ total variation cutoff.

(4) If $\lim_{n\to\infty} d(t_n)\mu_n = L \in [-\infty, \infty]$ *exists then*

$$\lim_{n\to\infty} \|Q_{n,m_n}^{\lfloor\{t_n\}\rfloor} - U_n\|_{\mathrm{TV}} = \Psi(e^L). \tag{5.4}$$

(B) Assume that $(m_n)_1^\infty$ *is bounded. Then* t_n *tends to infinity, there is a* $(t_n, 1)$ *total variation cutoff and, for any fixed* $k \in \mathbb{Z}$*, we have*

$$0 < \liminf_{n\to\infty} \|Q_{n,m_n}^{\{t_n\}+k} - U_n\|_{\mathrm{TV}} \le \limsup_{n\to\infty} \|Q_{n,m_n}^{\{t_n\}+k} - U_n\|_{\mathrm{TV}} < 1.$$

In particular, the $(t_n, 1)$ *cutoff is optimal.*

Example 5.1. To illustrate this result, consider the case where $m_n = \lfloor n^\alpha \rfloor$ for some fixed $\alpha > 0$. In this case, we have

$$\mu_n \sim \alpha \log n, \;\; t_n = \frac{3\log n}{2\mu_n} \sim \frac{3}{2\alpha} \text{ as } n \text{ tends to infinity.}$$

(a) Assume that $\frac{3}{2\alpha} \in (k, k+1)$ for some $k = 0, 1, 2, \ldots$. Then $|d(t_n)|\mu_n \to \infty$ and

$$\lim_{n\to\infty} \|Q_{n,m_n}^{k} - U_n\|_{\mathrm{TV}} = 1, \;\; \lim_{n\to\infty} \|Q_{n,m_n}^{k+1} - U_n\|_{\mathrm{TV}} = 0.$$

(b) Assume that $\frac{3}{2\alpha} = k$ for some integer $k = 1, 2, \ldots$. Then $|d(t_n)| = O(n^{-\alpha})$. Hence $|d(t_n)|\mu_n \to 0$ as n tends to infinity. Theorem 5.4(1) shows that we have a (k, b_n) cutoff where b_n is an arbitrary sequence of positive numbers tending to 0. That means that

$$\lim_{n\to\infty} \|Q_{n,m_n}^{k-1} - U_n\|_{\mathrm{TV}} = 1, \;\; \lim_{n\to\infty} \|Q_{n,m_n}^{k+1} - U_n\|_{\mathrm{TV}} = 0.$$

Moreover Theorem 5.4(4) gives $\lim_{n\to\infty} \|Q_{n,m_n}^{k} - U_n\|_{\mathrm{TV}} = \Psi(1)$.

Example 5.2. Consider the case where $m_n = \lfloor (\log n)^\alpha \rfloor$, $\alpha > 0$. Then

$$\mu_n \sim \alpha \log\log n, \;\; t_n \sim \frac{3\log n}{2\alpha \log\log n} \text{ as } n \text{ tends to infinity.}$$

Note that t_n tends to infinity and the window size μ_n^{-1} goes to zero.

We now state results concerning general p-shuffles. We will need the following notation. For each n, let p_n be a probability distribution on the integers. Let X_n be a random variable with distribution p_n. Assume that p_n is not supported on a single integer and set

$$\mu_n = E(\log X_n), \quad \sigma_n^2 = \mathrm{Var}(\log X_n), \quad \xi_n = \frac{\log X_n - \mu_n}{\sigma_n}.$$

Consider the following conditions which may or may not be satisfied by p_n:

$$\lim_{n\to\infty} \frac{\log n}{\mu_n} = \infty. \tag{5.5}$$

$$\forall\, \epsilon > 0, \quad \lim_{n\to\infty} E\left(\xi_n^2 \mathbf{1}_{\{\xi_n^2 > \epsilon \mu_n^{-1} \log n\}}\right) = 0. \tag{5.6}$$

Condition (5.6) should be understood as a Lindeberg type condition. We will prove in Lemma 5.9 that (5.6) implies (5.5).

Theorem 5.5. *Referring to the notation introduced above, assume that*

$$0 < \mu_n, \sigma_n < \infty$$

and set

$$t_n = \frac{3\log n}{2\mu_n}, \quad b_n = \frac{1}{\mu_n} \max\left\{1, \sqrt{\frac{\sigma_n^2 \log n}{\mu_n}}\right\}.$$

Assume that the sequence (p_n) *satisfies* (5.6)*. Then the family* $\{(S_n, Q_{n,p_n}, U_n)\}_1^\infty$ *presents a* (t_n, b_n) *total variation cutoff. Moreover, if the window size* b_n *is bounded from below by a positive real number, then the* (t_n, b_n) *total variation cut-off is optimal.*

Example 5.3. Assume $p_n = p$ is independent of n and

$$\mu = \sum_1^\infty p(k) \log k < \infty, \quad \sigma^2 = \sum_1^\infty |\mu - \log k|^2 p(k) < \infty.$$

Then condition (5.6) holds and

$$t_n = \frac{3}{2\mu}\log n, \quad b_n \approx \sqrt{\log n}$$

where $b_n \approx \sqrt{\log n}$ means that the ratio $b_n/\log n$ is bounded above and below by positive constants. Thus Theorem 5.5 yields an optimal $(\frac{3}{2\mu}\log n, \sqrt{\log n})$ total variation cutoff.

Example 5.4. Assume that p_n is concentrated equally on two integers $m_n < m'_n$ and write $m'_n = m_n k_n^2$. Thus $p_n(m_n) = p_n(m_n k_n^2) = 1/2$ and

$$\mu_n = \log m_n k_n, \quad \sigma_n = \log k_n.$$

In this case, Condition (5.6) is equivalent to (5.5), that is

$$\mu_n = \log(m_n k_n) = o(\log n).$$

Assuming that (5.5) holds true, Theorem 5.5 yields a total variation cutoff at time

$$t_n = \frac{3\log n}{2\log m_n k_n}$$

with window size

$$b_n = \frac{1}{\log m_n k_n}\max\left\{1, \sqrt{\frac{(\log k_n)^2\log n}{\log m_n k_n}}\right\}.$$

For instance, assume that $m'_n = m_n + 1$ with m_n tending to infinity. Then (5.5) becomes $\log m_n = o(\log n)$ and we have

$$b_n = \frac{1}{\log m_n}\max\left\{1, \frac{(\log n)^{1/2}}{m_n(\log m_n)^{1/2}}\right\}.$$

For instance, take $m_n \approx (\log n)^\alpha$ with $\alpha \in (0,\infty)$. Then

$$t_n \sim \frac{3\log n}{2\alpha\log\log n}$$

and

$$b_n \approx \begin{cases} (\log\log n)^{-1} & \text{if } \alpha \in [1/2, \infty) \\ (\log n)^{1/2-\alpha}(\log\log n)^{-3/2} & \text{if } \alpha \in (0, 1/2). \end{cases}$$

In particular, $b_n = o(1)$ when $\alpha \geq 1/2$ but tends to infinity when $\alpha \in (0, 1/2)$. Compare with Example 5.2 above.

Regarding Theorem 5.5, one might want to remove the hypothesis of existence of a second moment concerning the random variables $\log X_n$. It turns out that it is indeed possible but at the price of losing control of the window of the cutoff. What may be more surprising is that one can also obtain results without assuming that the first moment μ_n is finite.

Theorem 5.6. *Referring to the notation introduced above, assume that* $\mu_n > 0$ *(including possibly* $\mu_n = \infty$*). Assume further that there exists a sequence* a_n *tending to infinity and satisfying*

$$a_n = O(\log n), \quad \lim_{n\to\infty} \frac{(\log n) E Z_n^2}{a_n^2 E Y_n} = 0, \quad \lim_{n\to\infty} \frac{\log n}{E Y_n} = \infty, \tag{5.7}$$

where $Y_n = Z_n = \log X_n$ *if* $\log X_n \leq a_n$*, and* $Y_n = 0$*,* $Z_n = a_n$ *if* $\log X_n > a_n$*. Then the family* $\{(S_n, Q_{n,p_n}, U_n)\}_1^\infty$ *presents a total variation cutoff with critical time*

$$t_n = \frac{3 \log n}{2 E Y_n}.$$

Remark 5.2. In Theorem 5.6, if (5.7) holds for some sequence (a_n) then it also holds for any sequence $(d a_n)$ with $d > 0$. Moreover, for all $d > 0$,

$$E\left((\log X_n) \mathbf{1}_{\{\log X_n \leq d a_n\}}\right) \sim E Y_n.$$

This is proved in Lemma 5.10 below.

Example 5.5. Assume $p_n(\lfloor e^i \rfloor) = c_n^{-1} i^{-2}$ for all $1 \leq i \leq \lfloor \log n \rfloor$, where $c_n = 1 + 2^{-2} + 3^{-2} + \cdots + (\lfloor \log n \rfloor)^{-2}$. Note that $c_n \to c = \pi^2/6$ as $n \to \infty$. In this case, $\mu_n \sim c^{-1} \log\log n$, $\sigma_n^2 \sim c^{-1}\log n$ and for $\epsilon > 0$

$$E\left[\xi_n^2 \mathbf{1}_{\{\xi_n^2 < \epsilon\mu_n^{-1}\log n\}}\right] \sim \sqrt{\frac{\epsilon}{\log\log n}}.$$

Hence the Lindeberg type condition (5.6) does not hold and Theorem 5.5 does not apply. However, if we consider $a_n = \log n$ and try to apply Theorem 5.6, we have $EY_n = \mu_n \sim c^{-1}\log\log n$ and $EZ_n^2 \sim c^{-1}\log n$. This implies that (5.7) holds and yields a total variation cutoff with critical time $\frac{\pi^2 \log n}{4\log\log n}$.

The untruncated version of this example is $p_n(\lfloor e^i \rfloor) = p(\lfloor e^i \rfloor) = c^{-1}i^{-2}$, $i = 1, 2, \ldots$ and $c = \pi^2/6$. In this case, $\mu_n = \mu = \infty$. Theorem 5.6 applies with $a_n = \log n$ and yields a total variation cutoff with critical time $\frac{\pi^2 \log n}{4\log\log n}$.

We end this section with a result which is a simple corollary of Theorem 5.6 and readily implies Theorem 5.2.

Theorem 5.7. *Let X_n, p_n, μ_n be as above. Assume that*

$$\mu_n = E(\log X_n) = o(\log n) \tag{5.8}$$

and that, for any fixed $\eta > 0$,

$$E[(\log X_n)\mathbf{1}_{\{\log X_n > \eta \log n\}}] = o_\eta(\mu_n). \tag{5.9}$$

Then the family $\{(S_n, Q_{n,p_n}, U_n)\}_0^\infty$ has a total variation cutoff at time $t_n = \frac{3\log n}{2\mu_n}$.

Example 5.6. Suppose $p_n = p$ and $0 < \mu_n = \mu < \infty$ as in Theorem 5.2. Then condition (5.8)-(5.9) are obviously satisfied. Thus Theorem 5.2 follows immediately from Theorem 5.7 as mentioned above.

Remark 5.3. Condition (5.9) holds true if X_n satisfies the (logarithmic) moment condition that there exists $\epsilon > 0$ such that

$$\frac{E([\log X_n]^{1+\epsilon})}{(\log n)^\epsilon} = o(\mu_n).$$

5.3 An application: Continuous-time card shuffling

In this section, we consider the continuous-time version of the previous card shuffling models where the waiting times between two successive shuffles are independent exponential(1) random variables. Thus, the distribution of card arrangements at time t starting from the deck in order is given by the probability measure $H_{n,t} = e^{-t(I-Q_{n,p_n})}$ defined by

$$H_{n,t}(\sigma) = H_{n,t}(r) = e^{-t}\sum_{k=0}^{\infty}\frac{t^k}{k!}Q^k_{n,p_n}(r) \quad \text{for } \sigma \in S_n, \tag{5.10}$$

where r is the number of rising sequences of σ.

The definition of total variation cutoff and its optimality for continuous time families is the same as in Definitions 5.1, 5.2 and 5.3 except that all times are now taken to be non-negative reals. To state our results concerning the family $\{(S_n, H_{n,t}, U_n)\}_1^\infty$ of continuous time Markov chains associated with p_n-shuffles, $n = 1, 2, \ldots$, we keep the notation introduced in Section 5.2. In particular, we set

$$\mu_n = E(\log X_n), \;\; \sigma_n^2 = \mathrm{Var}(\log X_n), \;\; t_n = \frac{3\log n}{2\mu_n},$$

where X_n denotes a random variable with distribution p_n, and, if $\mu_n, \sigma_n \in (0,\infty)$,

$$\xi_n = \frac{\log X_n - \mu_n}{\sigma_n}.$$

We will obtain the following theorems as corollaries of the discrete time results of Section 5.2. Our first result concerns the case where each p_n is concentrated on one integer as in Theorem 5.3.

Theorem 5.8. *Assume that for each n there is an integer m_n such that $p(m_n) = 1$ Then $\mu_n = \log m_n$, $t_n = \frac{3\log n}{2\log m_n}$ and the family $\mathcal{F} = \{(S_n, H_{n,t}, U_n)\}_1^\infty$ presents a total variation cutoff if and only if*

$$\lim_{n\to\infty}\frac{\log n}{\log m_n} = \infty.$$

Moreover, if this condition is satisfied then $\mathcal{F}$ has an optimal $\left(t_n, \sqrt{t_n}\right)$ total variation cutoff.

Compare with the discrete time result stated in Theorem 5.3 and with Example 5.1 which we now revisit.

Example 5.7. Assume that $P(X_n = \lfloor n^\alpha \rfloor) = 1$ for a fixed $\alpha > 0$ as in Example 5.1. According to Theorem 5.8, the continuous time family $\mathcal{F}$ does not present a total variation cutoff in this case since $\lim_{n\to\infty} \frac{\log n}{\mu_n} = \alpha < \infty$. Recall from Example 5.1 that the corresponding discrete time family has a cutoff.

Assume that $P(X_n = \lfloor (\log n)^\alpha \rfloor) = 1$ for some fixed $\alpha > 0$ as in Example 5.3. In this case, the family $\mathcal{F}$ presents a $(t_n, \sqrt{t_n})$ total variation cutoff with $t_n = \frac{3\log n}{2\alpha \log\log n}$. Note that the window of the continuous time cutoff differs greatly from the window of the discrete time cutoff in this case.

Next we consider the general case under various hypotheses paralleling Theorems 5.5 and 5.6.

Theorem 5.9. *Consider the continuous time family $\mathcal{F} = \{(S_n, H_{n,t}, U_n)\}_1^\infty$ associated to a sequence $(X_n)_1^\infty$ of integer valued random variables with probability distributions $(p_n)_1^\infty$.*

(1) Assume that $\mu_n, \sigma_n \in (0,\infty)$ for all $n \geq 1$ and that (5.6) holds. Then the family $\mathcal{F}$ presents an optimal (t_n, b_n) total variation cutoff, where

$$t_n = \frac{3\log n}{2\mu_n},\ b_n = \frac{1}{\mu_n}\max\left\{(\mu_n+\sigma_n)\sqrt{\frac{\log n}{\mu_n}}, 1\right\}.$$

(2) Assume that $\mu_n > 0$ (including possibly $\mu_n = \infty$) and there exists a sequence $(a_n)_1^\infty$ tending to infinity such that (5.7) holds. Then $\mathcal{F}$ presents a total

variation cutoff with critical time

$$t_n = \frac{3 \log n}{2EY_n}$$

where $Y_n = (\log X_n)\mathbf{1}_{\{\log X_n \le a_n\}}$.

Remark 5.4. Theorem 5.9(2) applies when $p_n = p$ is independent of n and $\mu = \sum_1^\infty p(k) \log k < \infty$. In this case, the family $\mathcal{F} = \{(S_n, H_{n,t}, U_n)\}_1^\infty$ presents a total variation cutoff with critical time $t_n = \frac{3 \log n}{2\mu}$ as in Theorem 5.2. If in addition we assume that $\sigma^2 = \sum_1^\infty |\mu - \log k|^2 p(k) < \infty$ then Theorem 5.9(1) applies and shows that $\mathcal{F}$ has a $(t_n, \sqrt{\log n})$ total variation cutoff. Compare with Example 5.3.

We now describe how Theorem 5.9 applies to Examples 5.4-5.5 of Section 5.2.

Example 5.8. Assume, as in Example 5.4, that $p_n(m_n) = p_n(m_n k_n^2) = 1/2$. Assume further that $\mu_n = \log(m_n k_n) = o(\log n)$. Then, by Theorem 5.9(2), $\mathcal{F}$ presents a $(t_n, \sqrt{t_n})$ total variation cutoff, where

$$t_n = \frac{3 \log n}{2 \log m_n k_n}.$$

Finally, for Example 5.5, both in truncated and untruncated cases, Theorem 5.9(2) implies that the family presents a total variation cutoff with critical time $\frac{\pi^2 \log n}{4 \log \log n}$.

5.4 Technical tools

Two of the main technical tools we will use have already been stated as Lemma 5.1 and 5.2 in the introduction. In particular, Lemma 5.2 gives the probability distribution describing a deck of n cards after an m-shuffle, namely,

$$Q_{n,m}(r) = m^{-n} \binom{n + m - r}{n}$$

where r is the number of rising sequences in the arrangement of the deck. The next three known lemmas give further useful information concerning this distribution.

Lemma 5.3 (Tanny, [35]). *Let $R_{n,h}$ be the number of deck arrangements of n cards having $r = n/2 + h$ rising sequences, $1 \le r \le n$. Then, uniformly in h,*

$$\frac{R_{n,h}}{n!} = \frac{e^{-6h^2/n}}{\sqrt{\pi n/6}}\left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right)$$

Lemma 5.4 (Bayer and Diaconis, [6, Proposition 1]). *Fix $a \in (0, \infty)$. For any integers n, m such that $c = c(n,m) = mn^{-3/2} > a$ and any $r = \frac{n}{2} + h \in \{1, 2, \ldots, n\}$, we have*

$$Q_{n,m}\left(\frac{n}{2} + h\right) = \frac{1}{n!}\exp\left\{\frac{1}{c\sqrt{n}}\left(-h + \frac{1}{2} + O_a\left(\frac{h}{n}\right)\right)\right.$$
$$\left. - \frac{1}{24c^2} - \frac{1}{2}\left(\frac{h}{cn}\right)^2 + O_a\left(\frac{1}{cn}\right)\right\}$$

as n goes to infinity.

Lemma 5.5 (Bayer and Diaconis, [6, Proposition 2]). *Let h^* be the unique integer such that $Q_{n,m}\left(\frac{n}{2} + h\right) \ge \frac{1}{n!}$ if and only if $h \le h^*$. Fix $a \in (0, \infty)$. For any integers n, m such that $c = c(n,m) = mn^{-3/2} > a$, we have*

$$h^* = \frac{-\sqrt{n}}{24c} + O_a(1)$$

as n tends to ∞.

The statements of Lemmas 5.4 and 5.5 are somewhat different from the statement in Propositions 1 and 2 in [6] but the same proofs apply. The following theorem generalizes [6, Theorem 4], that is, Theorem 5.1 of the introduction. The proof, based on the three lemmas above, is the same as in [6]. It is omitted.

Theorem 5.10. *Fix $a \in (0,\infty)$. For any integers n, m such that $c = c(n,m) = mn^{-3/2} > a$ we have*

$$\|Q_{n,m} - U_n\|_{\mathrm{TV}} = \frac{1}{\sqrt{2\pi}} \int_{-1/(4\sqrt{3}\,c)}^{1/(4\sqrt{3}\,c)} e^{-t^2/2} dt + O_a\left(n^{-1/4}\right).$$

Theorem 5.10 provides sufficient information to obtain good upper bounds on the cutoff times of generalized riffle shuffles. It is however not sufficient to obtain matching lower bounds and study the cutoff phenomenon. The reminder of this section is devoted to results that will play a crucial role in obtaining sharp lower bounds on cutoff times for generalized riffle shuffles. It is reasonable to guess that shuffling cards with an $(m+1)$-shuffle is more efficient than shuffling cards with an m-shuffle . The following Proposition which is crucial for our purpose says that this intuition is correct when convergence to stationarity is measured in total variation.

Proposition 5.1. *For any integers n, m, we have*

$$\|Q_{n,m+1} - U_n\|_{\mathrm{TV}} \leq \|Q_{n,m} - U_n\|_{\mathrm{TV}}.$$

Proof. Let $A_m = \{\sigma \in S_n | Q_{n,m}(\sigma) < \frac{1}{n!}\}$ for $m \geq 1$. By Lemma 5.6 below, we have $A_{m+1} \subset A_m$ and $Q_{n,m}(\sigma) \leq Q_{n,m+1}(\sigma)$ for $\sigma \in A_{m+1}$. This implies

$$\begin{aligned}\|Q_{n,m} - U_n\|_{\mathrm{TV}} &= U_n(A_m) - Q_{n,m}(A_m) \geq U_n(A_{m+1}) - Q_{n,m}(A_{m+1}) \\ &\geq U_n(A_{m+1}) - Q_{n,m+1}(A_{m+1}) = \|Q_{n,m+1} - U_n\|_{\mathrm{TV}}.\end{aligned}$$

□

Lemma 5.6. *For any integers, n, m and $r \in \{1, \ldots, n\}$, we have:*

(1) $Q_{n,m}(r) \leq Q_{n,m+1}(r)$, if $Q_{n,m}(r) \leq \frac{1}{n!}$.

(2) $Q_{n,k}(r) > \frac{1}{n!}$ *for all* $k \geq m$*, if* $Q_{n,m}(r) > \frac{1}{n!}$.

In particular, if n, m, r *are such that* $Q_{n,m}(r) \leq \frac{1}{n!}$*, then*

$$k \mapsto Q_{n,k}(r)$$

is non-decreasing on $\{1, \ldots, m\}$.

Proof. We prove this lemma by fixing n and $1 \leq r \leq n$, and considering all possible cases of m. For $1 \leq m < r$, the first claim holds immediately from Lemma 5.2 since $Q_{n,m}(r) = 0$, and no $Q_{n,m}(r)$ satisfies the assumption of the second claim.

For $m \geq r$, consider the following map

$$x \xmapsto{f} n \log\left(\frac{x+1}{x}\right) + \log\left(\frac{x-r+1}{x-r+1+n}\right) \ \forall x \in [r, \infty).$$

The formula of the distribution of deck arrangements in Lemma 5.2 implies

$$f(m) = \log\left(\frac{Q_{n,m}(r)}{Q_{n,m+1}(r)}\right).$$

A direct computation on the derivative of f shows that

$$f'(x) = \frac{n[(2r-n-1)x - (r-1)(r-1-n)]}{x(x+1)(x-r+1)(x-r+1+n)}.$$

Here we consider all possible relation between r and n. If r, n satisfy $\frac{n+1}{2} \leq r \leq n$, then the derivative f' is positive on $[r, \infty)$. This implies that $f(x)$ is strictly increasing for $x \geq r$. As

$$\lim_{x \to \infty} f(x) = 0, \tag{5.11}$$

it follows that the function f is negative for $x \geq r$ and hence $Q_{n,m}(r) \leq Q_{n,m+1}(r)$ for $m \geq r$. This proves the first claim. Moreover, as

$$\lim_{m \to \infty} Q_{n,m}(r) = \frac{1}{n!}, \tag{5.12}$$

we have $Q_{n,m}(r) \le \frac{1}{n!}$ for all $m \ge r$ and $\frac{n+1}{2} \le r \le n$.

If r, n satisfy $1 \le r < \frac{n+1}{2}$, let $x_0 = \frac{(r-1)(r-1-n)}{2r-n-1}$. In this case, the derivative f' satisfies

$$f'(x) \begin{cases} \ge 0 & \text{if } r \le x \le x_0 \\ < 0 & \text{if } x > x_0 \end{cases}.$$

This implies that f is either decreasing on $[r, \infty)$ or increasing on $[r, x_0]$ and decreasing on (x_0, ∞) according to whether $x_0 < r$ or $x_0 \ge r$.

On one hand, if $x_0 < r$, that is, f is decreasing on $[r, \infty)$, then (5.11) implies that f is positive on $[r, \infty)$, which means, in particular, that $Q_{n,m}(r) \ge Q_{n,m+1}(r)$ for $m \ge r$. In this case, (5.12) implies that $Q_{n,m}(r) \ge \frac{1}{n!}$ for $m \ge r$.

On the other hand, if $x_0 \ge r$, that is, f increases on $[r, x_0)$ and decreases on $[x_0, \infty)$, then (5.11) implies that f has at most one zero in $[r, \infty)$. If f has no zero, then f is positive on $[r, \infty)$ and thus (by (5.12))

$$Q_{n,m}(r) \ge Q_{n,m+1}(r) \ge \frac{1}{n!} \quad \forall m \ge r,$$

This proves claim (2) (claim (1) is empty in this case).

If f has a zero, say z, then (5.11) implies that $f < 0$ on $[r, z)$ and $f > 0$ on (z, ∞). Assume that $Q_{n,m}(r)$ attains its maximum at $m_0 = \lfloor z \rfloor + 1$. Then $Q_{n,m}(r)$ is increasing for $m \in [r, m_0]$ and decreasing for $m \in [m_0, \infty)$. In the region $[m_0, \infty)$, (5.12) implies as before that $Q_{n,m}(r) > \frac{1}{n!}$ for $m \ge m_0$. In the region $[r, m_0]$, let $m_1 \ge r$ be the largest integer m such that $Q_{n,m}(r) \le \frac{1}{n!}$. Then $m_1 < m_0$ and thus $Q_{n,m}(r)$ is increasing on $[r, m_1 + 1]$, which proves the first claim. Moreover, $Q_{n,m}(r) > \frac{1}{n!}$ on $[m_1 + 1, \infty)$, which proves the second claim. □

Lemma 5.7. *Consider all deck arrangements of a deck of n cards.*

(1) For $1 \le r \le n$, let A_r be the set of deck arrangements with number of rising

sequences in $\{r, \ldots, n\}$. *Then for all integers* n, m *and* $r \in \{1, \ldots, n\}$, *we have*

$$U_n(A_r) - Q_{n,m}(A_r) \geq 0.$$

(2) *Fix* $a > 0$. *For any integer* n, m, *let* $c = c(n,m) = mn^{-3/2} > a$. *Let* B_c *be the set of deck arrangements with number of rising sequences in* $[\frac{n}{2} - \frac{\sqrt{n}}{24c} + n^{\frac{1}{4}}, n]$. *Then*

$$\inf_{k \leq m} \Big(U_n(B_c) - Q_{n,k}(B_c) \Big) = \frac{1}{\sqrt{2\pi}} \int_{-1/(4\sqrt{3}\,c)}^{1/(4\sqrt{3}\,c)} e^{-t^2/2} dt + O_a\left(n^{-\frac{1}{4}}\right).$$

Proof. As $Q_{n,m}(r)$ is non-increasing in r, we have either $Q_{n,m}(\sigma) \leq \frac{1}{n!}$ for all $\sigma \in A_r$ or $Q_{n,m}(\sigma) \geq \frac{1}{n!}$ for all $\sigma \in S_n - A_r$. The inequality stated in (1) thus follows from the obvious identity

$$U_n(A_r) - Q_{n,m}(A_r) = Q_{n,m}(S_n - A_r) - U_n(S_n - A_r).$$

To prove (2), let $h_0 = -\frac{\sqrt{n}}{24c} + n^{\frac{1}{4}}$. By Lemma 5.5, since $h_0 \geq h^*$ for large n, we have $Q_{n,m}(\sigma) \leq \frac{1}{n!}$ for $\sigma \in B_c$. Lemma 5.6 then implies

$$\inf_{k \leq m} \Big(U_n(B_c) - Q_{n,k}(B_c) \Big) = U_n(B_c) - Q_{n,m}(B_c) \quad \text{for } n \text{ large.}$$

By Lemmas 5.3, 5.5, we have

$$\left| \Big(U_n(B_c) - Q_{n,m}(B_c) \Big) - \|Q_{n,m} - U_n\|_{\mathrm{TV}} \right| \leq \sum_{h=h^*}^{h_0} \frac{R_{n,h}}{n!}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{h^*\sqrt{12/n}}^{h_0\sqrt{12/n}} e^{-t^2/2} dt + O\left(n^{-\frac{1}{2}}\right) = O_a\left(n^{-\frac{1}{4}}\right).$$

The equality in (2) then follows from Theorem 5.10. □

5.5 Proof of Theorem 5.3, 5.4

The following lemma is a corollary of Theorem 5.10. It is the main tool used to prove Theorems 5.3 and 5.4.

Lemma 5.8. *For $n \in \mathbb{N}$, let $m_n \in \mathbb{N}$ and $c_n = m_n n^{-3/2}$. Set*

$$\liminf_{n\to\infty} c_n = L, \ \limsup_{n\to\infty} c_n = U.$$

(1) If $L > 0$(including possibly the infinity), then

$$\limsup_{n\to\infty} \|Q_{n,m_n} - U_n\|_{\mathrm{TV}} \le \Psi(L^{-1}).$$

(2) If $U < \infty$(including possibly 0), then

$$\liminf_{n\to\infty} \|Q_{n,m_n} - U_n\|_{\mathrm{TV}} \ge \Psi(U^{-1}).$$

(3) If $U = L \in [0,\infty]$, then

$$\lim_{n\to\infty} \|Q_{n,m_n} - U_n\|_{\mathrm{TV}} = \Psi(U^{-1}).$$

Proof. Note that (3) follows immediately from (1) and (2). As the proofs of (1) and (2) are similar, we only prove (1). Assume first that $0 < L < \infty$. Let $\epsilon \in (0, L)$ and choose $N = N(\epsilon)$ such that $c_n \ge L - \epsilon$ for $n \ge N$. This implies that for $n \ge N$,

$$\begin{aligned} \|Q_{n,m_n} - U_n\|_{\mathrm{TV}} &\le \sup_{k \ge (L-\epsilon)n^{3/2}} \|Q_{n,k} - U_n\|_{\mathrm{TV}} \\ &= \Psi((L-\epsilon)^{-1}) + O_L\left(n^{-1/4}\right), \end{aligned}$$

where the last equality follows from Theorem 5.10. Letting n tend to infinity first and then ϵ to 0 gives (1).

If $L = \infty$, let $C \in (0,\infty)$ and choose $N = N(C)$ so large that $c_n \ge C$ if $n \ge N$. As in the previous case, for $n \ge N$,

$$\|Q_{n.m_n} - U_n\|_{\mathrm{TV}} \le \Psi(C^{-1}) + O_C\left(n^{-1/4}\right).$$

Now letting n, C tend to infinity yields (1) again.

□

Proof of Theorem 5.3. For $n \geq 1$ and $c \in \mathbb{R}$, let $t_n = \frac{3\log n}{2\mu_n}$ and

$$m_n(c) = \begin{cases} \lceil t_n + c\mu_n^{-1} \rceil & \text{if } c > 0 \\ \lfloor t_n + c\mu_n^{-1} \rfloor & \text{if } c < 0 \end{cases}.$$

This implies

$$\mu_n^{m_n(c)} n^{-3/2} \begin{cases} \geq e^c & \text{if } c > 0 \\ \leq e^c & \text{if } c < 0 \end{cases}.$$

Let $\overline{f}, \underline{f}$ be the functions introduced in Definition 5.2. By Lemmas 5.1 and 5.8, we have

$$\overline{f}(c) \leq \Psi(e^{-c}) \quad \text{if } c > 0,$$

and

$$\underline{f}(c) \geq \Psi(e^{-c}) \quad \text{if } c < 0.$$

Letting c tend respectively to ∞ and $-\infty$ proves Theorem 5.3. $\square$

Proof of Theorem 5.4. In this proof, k always denotes a non-negative integer. We first assume that m_n tends to infinity. Note that

$$k \begin{cases} \geq t + 1/2 & \text{if } k > \{t\} \\ \leq t - 1/2 & \text{if } k < \{t\} \text{ and } t \in [1/2, \infty) \\ = 0 & \text{if } k < \{t\} \text{ and } t \in (0, 1/2) \end{cases}.$$

This implies

$$m_n^k n^{-3/2} \begin{cases} \geq m_n^{1/2} & \text{if } k > \{t_n\} \\ \leq m_n^{-1/2} & \text{if } k < \{t_n\} \text{ and } t_n \geq 1/2 \\ = n^{-3/2} & \text{if } k < \{t_n\} \text{ and } t_n \in (0, 1/2) \end{cases}.$$

Theorem 5.4(1) thus follows from Lemmas 5.1 and 5.8.

The proof of Theorem 5.4(2) is similar to the proof of (1) but depends on the observation that

$$k \begin{cases} \geq t + |d(t)| & \text{if } k > t \\ \leq t - |d(t)| & \text{if } k < t \end{cases} \quad \text{for } k \in \mathbb{N},$$

which implies

$$m_n^k n^{-3/2} \begin{cases} \geq \exp\{|d(t_n)|\mu_n\} & \text{if } k > t_n \\ \leq \exp\{-|d(t_n)|\mu_n\} & \text{if } k < t_n. \end{cases}$$

For Theorem 5.4(3), by assumptions

$$\liminf_{n\to\infty} |d(t_n)|\mu_n < \infty, \ \lim_{n\to\infty} \mu_n = \infty.$$

Thus we can choose $M > 0$ and a sequence $(n_i)_1^\infty$ tending to infinity such that $|d(t_{n_i})|\mu_{n_i} \leq M$ and $t_{n_i} \geq 1/2$ for all $i \geq 1$. Since $\{t\} = t - d(t)$ for $t \geq 1/2$, we have that for all $i \geq 1$,

$$e^{-M} \leq m_{n_i}^{\{t_{n_i}\}} n_i^{-3/2} \leq e^M.$$

By Lemmas 5.1 and 5.8, this implies that

$$\limsup_{i\to\infty} \|Q_{n_i,m_{n_i}}^{\{t_{n_i}\}} - U_{n_i}\|_{\text{TV}} \leq \Psi(e^M) < 1,$$

and

$$\liminf_{i\to\infty} \|Q_{n_i,m_{n_i}}^{\{t_{n_i}\}} - U_{n_i}\|_{\text{TV}} \geq \Psi(e^{-M}) > 0.$$

For Theorem 5.4(4), if $L < \infty$, then the fact, $\lim_{n\to\infty} \mu_n = \infty$, implies that $t_n \geq 1/2$ for large n. In this case, $\{t_n\} = t_n - d(t_n) \in \mathbb{Z}$ and

$$m_n^{\{t_n\}} = n^{3/2} e^{-d(t_n)\mu_n}. \tag{5.13}$$

Then the desired inequality (5.4) follows from Lemmas 5.1 and 5.8.

If $L = \infty$, let $(n_i)_1^\infty$ be a sequence such that $t_n \geq 1/2$ if and only if $n = n_i$ for some i. Observe that if $n \notin \{n_i | i = 1, 2, ...\}$, then $\lfloor \{t_n\} \rfloor = 0$, and hence (5.4) follows immediately. For the sequence $(t_{n_i})_1^\infty$, since (5.13) holds in this case, the discussion for $L < \infty$ is applicable for t_{n_i} and hence (5.4) holds. This finishes the proof of (4).

We now assume that $(m_n)_1^\infty$ is bounded and let N be an upper bound of m_n. The proof in this case is similar to the proof of (3) after observing that

$$t_n + k - 1 < \{t_n\} + k < t_n + k + 1,$$

and

$$\min\{2^{k-1}, N^{k-1}\} \leq m_n^{\{t_n\}+k} n^{-3/2} \leq \max\{2^{k+1}, N^{k+1}\}.$$

□

5.6 Proof of Theorem 5.5

We start with the following elementary but crucial lemma.

Lemma 5.9. *Let $\{Y_n\}_{n=1}^\infty$ be a sequence of nonnegative random variables. Set*

$$\mu_n = E[Y_n], \quad \sigma_n^2 = \mathrm{Var}(Y_n), \quad \xi_n = \frac{Y_n - \mu_n}{\sigma_n}.$$

Suppose that $(a_n)_{n=1}^\infty$ is a sequence of positive numbers such that the Lindeberg type condition

$$\forall \epsilon > 0, \lim_{n\to\infty} E\left[\xi_n^2 1_{\{\xi_n^2 > \epsilon a_n\}}\right] = 0, \tag{5.14}$$

holds. Then

$$\lim_{n\to\infty} a_n = \infty \quad \textit{and} \quad \lim_{n\to\infty} \frac{\sigma_n^2}{\mu_n^2 a_n} = 0.$$

Proof. Note that $E\left[\xi_n^2 1_{\{\xi_n^2 \leq \epsilon a_n\}}\right] \leq \epsilon a_n$ for all $\epsilon > 0$. By (5.14), this implies

$$\liminf_{n\to\infty} a_n \geq \epsilon^{-1} E[\xi_n^2] = \epsilon^{-1}.$$

Hence $\lim_{n\to\infty} a_n = \infty$. Next, fix $\epsilon > 0$. As Y_n is nonnegative, we have

$$E\left[\xi_n^2 1_{\{\xi_n<0\}}\right] \leq \frac{\mu_n^2}{\sigma_n^2} \leq \frac{\sqrt{\epsilon}\mu_n^2 a_n}{\sigma_n^2},$$

for all n large enough, and

$$E\left[\xi_n^2 1_{\{0<\xi_n\leq\sqrt{\epsilon a_n}\}}\right] \leq \sigma_n^{-1}\sqrt{\epsilon a_n}E\left[(Y_n-\mu_n)1_{\{0<\xi_n\leq\sqrt{\epsilon a_n}\}}\right] \leq \sqrt{\frac{\epsilon\mu_n^2 a_n}{\sigma_n^2}}.$$

Let $L = \liminf_{n\to\infty} \mu_n^2 a_n/\sigma_n^2 \in [0,\infty]$. Combining both inequalities and letting $n \to \infty$ imply

$$1 \leq \sqrt{\epsilon}(L + \sqrt{L}).$$

Letting $\epsilon \to 0$ shows that $L = \infty$, that is, $\sigma_n^2/(\mu_n^2 a_n) \to 0$. □

Besides Lemma 5.9, the central limit theorem for sums of independent but not necessarily equidistributed random variables also plays an important role in the proof of Theorem 5.5. The following version is taken from [32].

Theorem 5.11. *(Central Limit Theorem) For $n > 0$, let $\xi_{n,1},\ldots,\xi_{n,n}$ be a sequence of independent random variables with mean $m_{n,k} = E[\xi_{n,k}]$ and variance $\sigma_{n,k}^2 = Var(\xi_{n,k})$. Let*

$$\zeta_{n,k} = \frac{\xi_{n,k} - m_{n,k}}{\sqrt{\sigma_{n,1}^2 + \cdots + \sigma_{n,n}^2}}, \quad \forall 1 \leq k \leq n.$$

If the Lindeberg condition holds, that is,

$$\lim_{n\to\infty} \sum_{k=1}^{n} E\left[\zeta_{n,k}^2 1_{\{\zeta_{n,k}^2\geq\epsilon\}}\right] = 0, \quad \forall \epsilon > 0, \tag{L}$$

then

$$\sum_{k=1}^{n} \zeta_{n,k} \xrightarrow{D} \mathcal{N}(0,1).$$

Recall the generalized model of riffle shuffle defined in (5.2). For $n \geq 1$, let p_n be the distribution of an integer-valued random variable X_n and consider the family $\{(S_n, Q_{n,p_n}, U_n)\}_1^\infty$ where

$$Q_{n,p_n}(\cdot) = E(Q_{n,X_n}(\cdot)) = \sum_{m=1}^{\infty} p_n(m) Q_{n,m}(\cdot).$$

Let $X_{n,1}, X_{n,2}, ...$ be a sequence of i.i.d. random variables sharing the same distribution as X_n. Then, for $a, k > 0$,

$$\begin{aligned} \|Q^k_{n,p_n} - U_n\|_{\mathrm{TV}} \leq & \sum_{m=1}^{\infty} P\left(\prod_{i=1}^{k} X_{n,i} = m\right) \|Q_{n,m} - U_n\|_{\mathrm{TV}} \\ \leq & P\left(\prod_{i=1}^{k} X_{n,i} \leq n^{3/2} a\right) + P\left(\prod_{i=1}^{k} X_{n,i} \geq n^{3/2} a\right)\left(\Psi(a^{-1}) + O_a\left(n^{-1/4}\right)\right) \quad (5.15) \\ = & (\Psi(a^{-1}) - 1) P\left\{\prod_{i=1}^{k} X_{n,i} \geq n^{3/2} a\right\} + 1 + O_a\left(n^{-1/4}\right), \end{aligned}$$

where the first inequality comes from the triangle inequality and the second inequality follows from Theorem 5.10.

Consider the set B_a defined in Lemma 5.7, that is, the subset of S_n containing permutations with numbers of rising sequences in $[\frac{n}{2} - \frac{\sqrt{n}}{24a} + n^{1/4}, n]$. Lemma 5.7 then implies that

$$\begin{aligned} \|Q^k_{n,p_n} - U_n\|_{\mathrm{TV}} \geq & \sum_{m \leq n^{3/2} a} P\left\{\prod_{i=1}^{k} X_{n,i} = m\right\} (U_n(B_a) - Q_{n,m}(B_a)) \\ \geq & \Psi(a^{-1}) P\left\{\prod_{i=1}^{k} X_{n,i} \leq n^{3/2} a\right\} + O_a\left(n^{-1/4}\right). \end{aligned} \quad (5.16)$$

Proof of Theorem 5.5. For $c \in \mathbb{R} - \{0\}$, let

$$k = k(n,c) = \begin{cases} \lceil t_n + c b_n \rceil & \text{if } c > 0 \\ \lfloor t_n + c b_n \rfloor & \text{if } c < 0 \end{cases},$$

where $t_n = \frac{3\log n}{2\mu_n}$ and $b_n = \frac{1}{\mu_n}\max\left\{1, \sqrt{\frac{\sigma_n^2 \log n}{\mu_n}}\right\}$. By hypothesis, (5.6) holds. Thus Lemma 5.9 implies

$$\lim_{n\to\infty} t_n = \infty,\ \ b_n = o(t_n). \tag{5.17}$$

By Definition 5.2, to prove a (t_n, b_n) total variation cut-off, we have to show that

$$\lim_{c\to\infty} \overline{f}(c) = 0 \ \ \lim_{c\to-\infty} \underline{f}(c) = 1,$$

where

$$\overline{f}(c) = \limsup_{n\to\infty} \|Q^k_{n,p_n} - U_n\|_{\mathrm{TV}},\ \ \underline{f}(c) = \liminf_{n\to\infty} \|Q^k_{n,p_n} - U_n\|_{\mathrm{TV}}.$$

Note that $b_n \geq \frac{1}{2\mu_n}\left(1 + \sqrt{\frac{\sigma_n^2 \log n}{\mu_n}}\right)$. This implies

$$\log(n^{3/2}e^{c/2}) - k\mu_n + \frac{c}{2}\sqrt{\frac{\sigma_n^2 \log n}{\mu_n}} \begin{cases} \leq 0 & \text{if } c > 0 \\ \geq 0 & \text{if } c < 0 \end{cases}.$$

Hence, we have

$$P\left\{\prod_{i=1}^{k} X_{n,i} \geq n^{\frac{3}{2}}e^{c/2}\right\} \geq P\left\{\frac{\sum_{i=1}^{k}\log X_{n,i} - k\mu_n}{\sigma_n\sqrt{k}} \geq -\frac{c}{2}\sqrt{\frac{\log n}{k\mu_n}}\right\} \ \text{ for } c > 0$$

and

$$P\left\{\prod_{i=1}^{k} X_{n,i} \leq n^{\frac{3}{2}}e^{c/2}\right\} \geq P\left\{\frac{\sum_{i=1}^{k}\log X_{n,i} - k\mu_n}{\sigma_n\sqrt{k}} \leq -\frac{c}{2}\sqrt{\frac{\log n}{k\mu_n}}\right\} \ \text{ for } c < 0.$$

For fixed $c \in \mathbb{R} - \{0\}$, consider the sequence

$$\log X_{n,1}, \log X_{n,2}, ..., \log X_{n,k},$$

as the k-th row of a triangular array of random variables. As $k \sim t_n$ and (5.6) holds, the Lindeberg condition (L) is satisfied. Hence Theorem 5.11 yields

$$\liminf_{n\to\infty} P\left\{\prod_{i=1}^{k} X_{n,i} \geq n^{\frac{3}{2}}e^{c/2}\right\} \geq \frac{1}{2}\left(1 + \Psi(2\sqrt{2}c)\right) \quad \text{if } c > 0,$$

and

$$\liminf_{n\to\infty} P\left\{\prod_{i=1}^{k} X_{n,i} \leq n^{\frac{3}{2}} e^{c/2}\right\} \geq \frac{1}{2}\left(1+\Psi(-2\sqrt{2}c)\right) \quad \text{if } c<0.$$

Then, by (5.15) and (5.16), we have

$$\overline{f}(c) \leq 1-\frac{1}{2}\left(1-\Psi(e^{-c/2})\right)\left(1+\Psi(2\sqrt{2}c)\right) \quad \text{for } c>0,$$

and

$$\underline{f}(c) \geq \frac{1}{2}\Psi(e^{-c/2})\left(1+\Psi(-2\sqrt{2}c)\right) \quad \text{for } c<0.$$

Hence the (t_n, b_n)-cutoff is proved by letting c tend to ∞ and $-\infty$ respectively.

For the optimality of such total variation cutoff, we need to estimate $\overline{f}(c)$ for $c<0$ and $\underline{f}(c)$ for $c>0$. Assume that $b_n \geq b > 0$ for all $n \geq 1$. Then we have

$$k = \begin{cases} \lfloor t_n + cb_n \rfloor > t_n + cb_n - 1 \geq t_n + (c-b^{-1})b_n & \text{if } c<0 \\ \lceil t_n + cb_n \rceil < t_n + cb_n + 1 \leq t_n + (c+b^{-1})b_n & \text{if } c>0 \end{cases}.$$

Arguing as in the proof of cutoff above, we obtain

$$\liminf_{n\to\infty} P\left\{\prod_{i=1}^{k} X_{n,i} \geq n^{\frac{3}{2}} e^{(c-b^{-1})}\right\} \geq \frac{1}{2}\left(1-\Psi(4\sqrt{2}(b^{-1}-c))\right) \text{ for } c<0,$$

and

$$\liminf_{n\to\infty} P\left\{\prod_{i=1}^{k} X_{n,i} \leq n^{\frac{3}{2}} e^{(c+b^{-1})}\right\} \geq \frac{1}{2}\left(1-\Psi(4\sqrt{2}(b^{-1}+c))\right) \text{ for } c>0.$$

Hence, the functions $\overline{f}, \underline{f}$ are bounded by

$$\forall\, c<0,\ \ \overline{f}(c) \leq 1-\frac{1}{2}\left(1-\Psi(e^{(b^{-1}-c)})\right)\left(1-\Psi(4\sqrt{2}(b^{-1}-c))\right) < 1,$$

and

$$\forall\, c>0,\ \ \underline{f}(c) \geq \frac{1}{2}\Psi(e^{-(b^{-1}+c)})\left(1-\Psi(4\sqrt{2}(b^{-1}+c))\right) > 0.$$

By Definition 5.3, the family $\{(S_n, Q_{n,p_n}, U_n)\}_1^\infty$ has an optimal (t_n, b_n) total variation cutoff.

□

5.7 Proof of Theorems 5.6, 5.7

To work without assuming the existence of μ_n, we need the following weak law of large numbers for triangular arrays. See, e.g., [22].

Theorem 5.12. *(Weak law of large numbers) For each n, let $W_{n,k}$, $1 \le k \le n$, be independent. Let $b_n > 0$ with $b_n \to \infty$, and $\bar{W}_{n,k} = W_{n,k} 1_{\{|W_{n,k}| \le b_n\}}$. Suppose that*

(1) $\sum_{k=1}^{n} P\{|W_{n,k}| > b_n\} \to 0$, and

(2) $b_n^{-2} \sum_{k=1}^{n} E\bar{W}_{n,k}^2 \to 0$ as $n \to \infty$.

If we set $S_n = W_{n,1} + ... + W_{n,n}$ and put $s_n = \sum_{k=1}^{n} E\bar{W}_{n,k}$, then

$$\frac{S_n - s_n}{b_n} \to 0 \quad \text{in probability.}$$

Proof of Theorem 5.6. For $0 < |\epsilon| < 1$, let

$$k = k(n,\epsilon) = \begin{cases} \lceil (1+\epsilon)t_n \rceil & \text{if } \epsilon > 0 \\ \lfloor (1+\epsilon)t_n \rfloor & \text{if } \epsilon < 0 \end{cases}.$$

By (5.15) and (5.16), to prove a total variation cutoff with critical time t_n, it suffices to prove that for all $a > 0$

$$\lim_{n\to\infty} P\left\{\prod_{i=1}^{k} X_{n,i} \ge n^{\frac{3}{2}} a\right\} = 1, \text{ if } \epsilon > 0, \tag{5.18}$$

and

$$\lim_{n\to\infty} P\left\{\prod_{i=1}^{k} X_{n,i} \le n^{\frac{3}{2}} a\right\} = 1, \text{ if } \epsilon < 0. \tag{5.19}$$

Indeed, if these limits holds true then (5.15) and (5.16) give

$$\limsup_{n\to\infty} \|Q_{n,p_n}^k - U_n\|_{\text{TV}} \le \Psi(a^{-1}) \quad \text{for } \epsilon > 0$$

and

$$\liminf_{n\to\infty} \|Q_{n,p_n}^k - U_n\|_{\text{TV}} \ge \Psi(a^{-1}) \quad \text{for } \epsilon < 0.$$

The total variation cutoff is then proved by letting a tend to infinity and 0 respectively.

To prove (5.18)-(5.19), note that $EZ_n^2 = EY_n^2 + a_n^2 P\{\log X_n > a_n\}$. By the second part of assumption (5.7), we have

$$(1+\epsilon)t_n P\{\log X_n > a_n\} \to 0 \quad \text{and} \quad (1+\epsilon)t_n a_n^{-2} EY_n^2 \to 0, \quad \text{as } n \to \infty. \tag{5.20}$$

In order to apply Theorem 5.12, for fixed $\epsilon \in (-1, 1)$, consider

$$W_{k,1} = \log X_{n,1}, ..., W_{k,k} = \log X_{n,k}$$

as the k-th row of a triangular array of random variables. Then (5.20) shows that the hypotheses (1) and (2) in Theorem 5.12 hold. Hence

$$a_n^{-1}\left(\sum_{i=1}^{k} \log X_{n,i} - (1+\epsilon)t_n EY_n\right) \to 0 \text{ in probability.} \tag{5.21}$$

Note also that for $a > 0$, $a_n^{-1}\left(\log(n^{3/2}a) - (1+\epsilon)t_n EY_n\right) \sim \frac{-3\epsilon \log n}{2a_n}$. Hence the first part of assumption (5.7) implies that

$$\begin{aligned} &\limsup_{n\to\infty} a_n^{-1}\left(\log(n^{3/2}a) - (1+\epsilon)t_n EY_n\right) < 0 \quad \text{if } \epsilon > 0, \\ &\liminf_{n\to\infty} a_n^{-1}\left(\log(n^{3/2}a) - (1+\epsilon)t_n EY_n\right) > 0 \quad \text{if } \epsilon < 0. \end{aligned} \tag{5.22}$$

Combining both (5.21) and (5.22) proves (5.18) and (5.19). □

Proof of Theorem 5.7. Let X_n be integer valued random variables such that

$$P\{X_n = k\} = p_n(k) \quad \text{for } k = 1, 2, ...$$

and satisfying (5.8), (5.9). Let $a_n = \log n$ in Theorem 5.6 so that

$$Y_n = (\log X)\mathbf{1}_{\{\log X \le \log n\}}, \; Z_n = Y_n + (\log n)\mathbf{1}_{\{\log X > \log n\}}.$$

Set $L_n = \log X_n$. By (5.9), we have $E(L_n \mathbf{1}_{\{L_n > \log n\}}) = o(\mu_n)$. Hence $E(Y_n) \sim \mu_n$ and the third condition of (5.7) follows from (5.8). To apply Theorem 5.6, it remains to show

$$\lim_{n\to\infty} \frac{EY_n^2}{EY_n \log n} = 0, \quad \lim_{n\to\infty} \frac{P\{L_n > \log n\} \log n}{EY_n} = 0,$$

or equivalently,

$$\lim_{n\to\infty} \frac{EY_n^2}{\mu_n \log n} = 0, \quad \lim_{n\to\infty} \frac{P\{L_n > \log n\} \log n}{\mu_n} = 0.$$

The hypothesis (5.9) gives

$$\frac{P\{L_n > \log n\} \log n}{\mu_n} \leq \frac{E(L_n \mathbf{1}_{\{L_n > \log n\}})}{\mu_n} = o(1)$$

which proves the second desired limit. For the first limit, for any $\eta \in (0, 1)$, write

$$\begin{aligned} EY_n^2 &= E[L_n^2 \mathbf{1}_{\{L_n \leq \eta \log n\}}] + E(L_n^2 \mathbf{1}_{\{\eta \log n < L_n \leq \log n\}}) \\ &\leq \eta \mu_n \log n + E(L_n \mathbf{1}_{\{L_n > \eta \log n\}}) \log n \\ &\leq (\eta + o_\eta(1)) \mu_n \log n \end{aligned}$$

where we have used (5.9) again to obtain the last inequality. Thus

$$\frac{EY_n^2}{\mu_n \log n} \leq \eta + o_\eta(1).$$

Letting n tend to infinity and then η tend to 0 shows that the left-hand side tends to 0 as desired. □

The next lemma deals with condition (5.7) appearing in Theorem 5.6 and plays a role in the proof of Theorem 5.9(2).

Lemma 5.10. *For $n \geq 1$, let $a_n, b_n > 0$ and X_n be a non-negative random variable. According to the sequence $(a_n)_1^\infty$ and $c > 0$, set $Y_n = X_n \mathbf{1}_{\{X_n \leq c a_n\}}$ and $Z_n = Y_n + c a_n \mathbf{1}_{\{X_n > c a_n\}}$. Consider the following conditions.*

$$a_n = O(b_n), \quad \lim_{n\to\infty} \frac{b_n E Z_n^2}{a_n^2 E Y_n} = 0, \quad \lim_{n\to\infty} \frac{b_n}{EY_n} = \infty. \tag{5.23}$$

Then (5.23) holds for some $c > 0$ if and only if it holds for any $c > 0$.

Proof. On direction is obvious. For the other direction, we assume that (5.23) holds for some $c > 0$. The second condition in (5.23) implies

$$P\{X_n > ca_n\} = o\left(\frac{EY_n}{b_n}\right), \quad \frac{EY_n^2}{a_n^2} = o\left(\frac{EY_n}{b_n}\right). \tag{5.24}$$

Let $d > 0$ and $Y_n' = X_n \mathbf{1}_{\{X_n \le da_n\}}$ and $Z_n' = Y_n' + da_n \mathbf{1}_{\{X_n > da_n\}}$. Then (5.24) and Chebyshev inequality imply

$$\begin{aligned} |EY_n' - EY_n| &\le \begin{cases} ca_n P\{Y_n > da_n\} & \text{if } d < c \\ \\ da_n P\{X_n > ca_n\} & \text{if } d > c \end{cases} \\ &= o\left(\frac{a_n EY_n}{b_n}\right) = o(EY_n), \end{aligned}$$

and

$$\begin{aligned} |EZ_n'^2 - EZ_n^2| &\le |d^2 - c^2| a_n^2 P\{X_n > (d \wedge c) a_n\} \\ =& |d^2 - c^2| a_n^2 \left(P\{Y_n > (d \wedge c) a_n\} + P\{X_n > ca_n\}\right) \\ \le& |d^2 - c^2| a_n^2 \left(\frac{EY_n^2}{(d \wedge c)^2 a_n^2} + P\{X_n > ca_n\}\right) = o\left(\frac{a_n^2 EY_n}{b_n}\right). \end{aligned}$$

Hence we have $EY_n' \sim EY_n$ and $\frac{b_n E(Z_n')^2}{d^2 a_n^2 EY_n'} \to 0$. □

5.8 Proofs of Theorems 5.8 and 5.9

In this section we are concerned with the continuous time process whose distribution at time t, $H_{n,t}$, is given by (5.10), that is

$$H_{n,t} = e^{-t} \sum_{k=0}^{\infty} \frac{t^k}{k!} Q_{n,p_n}^k.$$

Let $X_{n,1}, X_{n,2}, \ldots$ be a sequence of independent random variables with probability distribution p_n. Let $\tilde{X}_n$ be an integer valued random variable whose probability distribution $\tilde{p}_n$ is given by

$$\tilde{p}_n(l) = P\{\tilde{X}_n = l\} = \begin{cases} e^{-P\{X_{n,1}\neq 1\}} & \text{if } l = 1 \\ e^{-1}\sum_1^\infty \frac{1}{j!}P\left\{\prod_1^j X_{n,i} = l\right\} & \text{if } l > 1 \end{cases}. \tag{5.25}$$

With this notation , we have

$$H_{n,1} = E(Q_{n,\tilde{X}_n}) = Q_{n,\tilde{p}_n}$$

and

$$H_{n,k} = E(Q^k_{n,\tilde{X}_n}) = Q^k_{n,\tilde{p}_n}, \quad k = 1, 2, \ldots.$$

Let h be any nonnegative function defined on $[0,\infty)$ satisfying $h(0) = 0$. Fubini's Theorem yields

$$E(h(\log \tilde{X}_n)) = e^{-1}\sum_{j=1}^{\infty}\frac{1}{j!}E(h(\bar{X}_{n,j})), \tag{5.26}$$

where $\bar{X}_{n,j} = \log X_{n,1} + \cdots + \log X_{n,j}$. Thus, if we assume that $\mu_n, \sigma_n < \infty$ and let $h(t) = t$ (resp. $h(t) = t^2$), we obtain

$$E(\log \tilde{X}_n) = \mu_n \quad \text{and} \quad \mathrm{Var}(\log \tilde{X}_n) = \sigma_n^2 + \mu_n^2.$$

Proof of Theorem 5.8. Here, we deal with the case where, for each n, $p_n(m_n) = 1$ for some integer m_n. Observe that for any integers n, M and time $t > 0$,

$$\|H_{n,t} - U_n\|_{\mathrm{TV}} \geq H_{n,t}(id) - \frac{1}{n!} \geq e^{-t} - \frac{1}{n!}$$
$$\|H_{n,t} - U_n\|_{\mathrm{TV}} \leq e^{-t}\sum_{i=0}^{M}\frac{t^i}{i!} + \|Q^M_{n,p_n} - U_n\|_{\mathrm{TV}},$$

where id is the identity of S_n, that is, represents the deck in order.

Assume that

$$\liminf_{n\to\infty}\frac{\log n}{\mu_n} < \infty.$$

Let M be an integer and $(n_k)_1^\infty$ be an increasing sequence such that $\sup_{k\geq 1} \frac{2\log n_k}{\mu_{n_k}} < M$. Let $(t_k)_1^\infty$ be an arbitrary sequence of positive numbers. Then, by Theorem 5.3 and the observation above, we have

$$\lim_{k\to\infty} \|H_{n_k,t_k} - U_{n_k}\|_{\mathrm{TV}} = 0 \Longleftrightarrow \lim_{k\to\infty} t_k = \infty.$$

This means that the subfamily $\{(S_{n_k}, H_{n_k,t}, U_{n_k})\}_1^\infty$, and thus $\mathcal{F}$ itself, does not present a total variation cutoff.

Assume now that

$$\lim_{n\to\infty} \frac{\log n}{\mu_n} = \infty.$$

Then $t_n = \frac{3\log n}{2\mu_n}$ tends to infinity and thus $t_n \sim \lfloor t_n \rfloor$. Clearly, a $(t_n, \sqrt{t_n})$ cutoff for $H_{n,t}$ is equivalent to a $(t_n, \sqrt{t_n})$ cutoff for $Q^k_{n,\tilde{p}_n}$. We now prove the desired cutoff by applying Theorem 5.5 to $Q_{n,\tilde{p}_n}$. To this end, we need to show that (5.6) holds for $\tilde{X}_n$. Set $\tilde{\xi}_n = \frac{\log \tilde{X}_n - \mu_n}{\sqrt{\sigma_n^2+\mu_n^2}}$. Then (5.26) implies

$$E\left(\tilde{\xi}_n^2 \mathbf{1}_{\{\tilde{\xi}_n^2 > \epsilon \frac{\log n}{\mu_n}\}}\right) = \sum_{j > \sqrt{\frac{\epsilon \log n}{\mu_n}}}^{\infty} \frac{e^{-1} j^2}{(j+1)!} \to 0 \quad \text{as } n \to \infty,$$

for any $\epsilon > 0$ and $n \geq m_n^{1/\epsilon}$. Hence (5.6) holds for $\tilde{X}_n$ and, by Theorem 5.5, the family $\{(S_n, Q_{n,\tilde{p}_n}, U_n)\}_1^\infty$ presents, as desired, an optimal (t_n, b_n) total variation cutoff with $b_n = \sqrt{\log n/\mu_n}$. □

Proof of Theorem 5.9(1). As in the proof of Theorem 5.8, the desired cutoff for the family $\{(S_n, H_{n,t}, U_n)\}_1^\infty$ is equivalent to the same cutoff for $\{(S_n, Q_{n,\tilde{p}_n}, U_n)\}_1^\infty$ because cutoff time and window size tend to infinity. Hence, the desired conclusion will follow from Theorem 5.5 if we can show that $\tilde{X}_n$ at (5.25) satisfies (5.6). Set $\tilde{\xi}_n = \frac{\log \tilde{X}_n - \mu_n}{\sqrt{\sigma_n^2+\mu_n^2}}$. Then (5.26) implies

$$E\left(\tilde{\xi}_n^2 \mathbf{1}_{\{\tilde{\xi}_n^2 > \epsilon \frac{\log n}{\mu_n}\}}\right) = e^{-1} \sum_{j=1}^{\infty} \frac{1}{j!} E\left(\frac{\left(\bar{X}_{n,j} - \mu_n\right)^2}{\sigma_n^2 + \mu_n^2} \mathbf{1}_{\left\{\frac{\left(\bar{X}_{n,j}-\mu_n\right)^2}{\sigma_n^2+\mu_n^2} > \epsilon \frac{\log n}{\mu_n}\right\}}\right), \tag{5.27}$$

if $\epsilon\mu_n^{-1}\log n > 1$. Fix $\epsilon, \delta > 0$ and let $M = M(\delta) \in \mathbb{N}$, $N = N(\epsilon, M) \in \mathbb{N}$ such that $2\sum_{M+1}^{\infty}\frac{j^2}{j!} < \delta$ and $\sqrt{\frac{\epsilon\log n}{\mu_n}} \geq 2M$ if $n \geq N$. In this case, (5.27) implies that

$$E\left(\tilde{\xi}_n^2\mathbf{1}_{\{\tilde{\xi}_n^2>\epsilon\frac{\log n}{\mu_n}\}}\right) \leq \delta + e^{-1}\sum_{j=1}^{M}\frac{1}{j!}E\left(\frac{\left(\bar{X}_{n,j}-\mu_n\right)^2}{\sigma_n^2+\mu_n^2}\mathbf{1}_{\left\{\frac{\bar{X}_{n,j}-\mu_n}{\sqrt{\sigma_n^2+\mu_n^2}}>\sqrt{\frac{\epsilon\log n}{\mu_n}}\right\}}\right). \quad (5.28)$$

To bound the expectation in the right hand side, we consider the following sets. For $1 \leq i \leq j \leq M$, let

$$A_{n,i,j} = \left\{\log X_{n,i} > \frac{1}{j}\left(\mu_n + \sqrt{\frac{\epsilon(\sigma_n^2+\mu_n^2)\log n}{\mu_n}}\right)\right\}$$
$$B_{n,i} = \left\{\frac{(\log X_{n,i}-\mu_n)^2}{\sigma_n^2} > \frac{\epsilon\log n}{4M^2\mu_n}\right\}.$$

Then

$$\left\{\frac{\bar{X}_{n,j}-\mu_n}{\sqrt{\sigma_n^2+\mu_n^2}} > \sqrt{\frac{\epsilon\log n}{\mu_n}}\right\} \subset \bigcup_{i=1}^{j}A_{n,i,j} \quad (5.29)$$

and

$$A_{n,i,j} \subset B_{n,i} \quad \text{if } \sqrt{\frac{\epsilon\log n}{\mu_n}} \geq 2M.$$

This implies that for $n \geq N, 1 \leq i \leq j \leq M$,

$$\begin{aligned}
&E\left(\frac{\left(\bar{X}_{n,j}-\mu_n\right)^2}{\sigma_n^2+\mu_n^2}\mathbf{1}_{A_{n,i,j}}\right)\\
\leq& 2E\left(\frac{(\bar{X}_{n,j}-\log X_{n,i})^2}{\sigma_n^2+\mu_n^2}\mathbf{1}_{B_{n,i}}\right) + 2E\left(\frac{(\log X_{n,i}-\mu_n)^2}{\sigma_n^2}\mathbf{1}_{B_{n,i}}\right)\\
=&\frac{2\left((j-1)\sigma_n^2+(j-1)^2\mu_n^2\right)}{\sigma_n^2+\mu_n^2}P\{B_{n,i}\} + 2E\left(\xi_n^2\mathbf{1}_{\{\xi_n^2>\frac{\epsilon\log n}{4M^2\mu_n}\}}\right)\\
\leq& 3E\left(\xi_n^2\mathbf{1}_{\{\xi_n^2>\frac{\epsilon\log n}{4M^2\mu_n}\}}\right) \quad \text{if } n \text{ is large.}
\end{aligned}$$

Now, using (5.29) and these estimates in (5.28), and applying the hypothesis that X_n satisfies (5.6), we obtain

$$\limsup_{n\to\infty} E\left(\tilde{\xi}_n^2\mathbf{1}_{\{\tilde{\xi}_n^2>\epsilon\frac{\log n}{\mu_n}\}}\right) \leq \delta \quad \forall\delta, \epsilon > 0.$$

Hence (5.6) holds for $\tilde{X}_n$. By Theorem 5.5, the family $\{(S_n, H_{n,t}, U_n)\}_1^\infty$ presents an optimal $\left(\frac{3\log n}{2\mu_n}, b_n\right)$ total variation cutoff, where

$$b_n = \frac{1}{\mu_n}\max\left\{\sqrt{\frac{(\sigma_n^2+\mu_n^2)\log n}{\mu_n}}, 1\right\}$$

(note that b_n always tends to infinity). □

Proof of Theorem 5.9(2). The proof is similar to that of part (1) except that we will use Theorem 5.6 instead of Theorem 5.5. Let

$$\tilde{Y}_n = (\log \tilde{X}_n)\mathbf{1}_{\{\log \tilde{X}_n \le a_n\}}, \quad \tilde{Z}_n = \tilde{Y}_n + a_n\mathbf{1}_{\{\log \tilde{X}_n > a_n\}}.$$

By (5.26), we have

$$E\tilde{Y}_n = e^{-1}\sum_{j=1}^{\infty}\frac{1}{j!}E\left[\left(\sum_{i=1}^{j}\log X_{n,i}\right)\mathbf{1}_{\{\sum_1^j \log X_{n,i}\le a_n\}}\right].$$

It is apparent that $E\tilde{Y}_n \le EY_n$. For $j > 0$, we have

$$E\left[\left(\sum_{i=1}^{j}\log X_{n,i}\right)\mathbf{1}_{\{\sum_1^j \log X_{n,i}\le a_n\}}\right] \ge \sum_{i=1}^{j}\left\{E\left(\log X_{n,i}\mathbf{1}_{\left\{\log X_{n,i}\le \frac{a_n}{j}\right\}}\right) \times \prod_{\substack{k=1\\k\ne i}}^{j} P\left(\log X_{n,k} \le \frac{a_n}{j}\right)\right\}.$$

By Lemma 5.10 (or Remark 5.2) and (5.24), we have

$$\liminf_{n\to\infty} E\left[\left(\sum_{i=1}^{j}\log X_{n,i}\right)1_{\{\sum_1^j \log X_{n,i}\le a_n\}}\right]\Big/ EY_n \ge j.$$

Hence, for $k > 0$

$$\liminf_{n\to\infty}\frac{E\tilde{Y}_n}{EY_n} \ge e^{-1}\sum_{j=0}^{k}\frac{1}{j!}.$$

Letting $k\to\infty$ implies $E\tilde{Y}_n \sim EY_n$.

To apply Theorem 5.6, it remains to prove that the second part of (5.7) holds for $\tilde{Y}_n$ and $\tilde{Z}_n$, that is,

$$E(\tilde{Y}_n^2) = o\left(\frac{a_n^2 EY_n}{\log n}\right), \quad P\left\{\log \tilde{X}_n > a_n\right\} = o\left(\frac{EY_n}{\log n}\right).$$

Note that, by the hypothesis that X_n satisfies (5.7), we have

$$E(Y_n^2) = o\left(\frac{a_n^2 EY_n}{\log n}\right), \quad P\left\{\log X_n > a_n\right\} = o\left(\frac{EY_n}{\log n}\right).$$

Then (5.26), Lemma 5.10 and the above observation imply

$$\begin{aligned} E(\tilde{Y}_n^2) &= e^{-1}\sum_{j=1}^{\infty}\frac{1}{j!}E\left[\left(\sum_{i=1}^{j}\log X_{n,i}\right)^2 \mathbf{1}_{\{\sum_1^j \log X_{n,i}\le a_n\}}\right] \\ &\le e^{-1}\sum_{j=1}^{\infty}\frac{1}{j!}E\left(\sum_{i=1}^{j}(\log X_{n,i})\mathbf{1}_{\{\log X_{n,i}\le a_n\}}\right)^2 \\ &= EY_n^2 + (EY_n)^2 \le 2EY_n^2 = o\left(\frac{a_n^2 EY_n}{\log n}\right), \end{aligned}$$

and

$$\begin{aligned} P\{\log \tilde{X}_n > a_n\} &= e^{-1}\sum_{j=1}^{\infty}\frac{1}{j!}P\left\{\sum_{i=1}^{j}\log X_{n,i} > a_n\right\} \\ &\le e^{-1}\sum_{j=1}^{\infty}\frac{1}{(j-1)!}P\left\{\log X_n > \frac{a_n}{j}\right\}. \end{aligned}$$

Since, for $j \ge 1$,

$$\begin{aligned} P\left\{\log X_n > \frac{a_n}{j}\right\} &= P\{\log X_n > a_n\} + P\left\{Y_n > \frac{a_n}{j}\right\} \\ &= P\{\log X_n > a_n\} + \frac{j^2 EY_n^2}{a_n^2} = j^2 \times o\left(\frac{EY_n}{\log n}\right), \end{aligned}$$

we have

$$P\{\log \tilde{X}_n > a_n\} = o\left(\frac{EY_n}{\log n}\right).$$

By Theorem 5.6, the family $\{(S_n, Q_{n,\tilde{p}_n}, U_n)\}_1^\infty$ presents a total variation cutoff with critical time $\frac{3\log n}{2EY_n}$. Hence the same holds for $\{(S_n, H_{n,t}, U_n)\}_1^\infty$. □

Appendix A

Techniques and proofs

A.1 Fundamental results of analysis

Lemma A.1. *Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be measure spaces and $T : L^p(\mu) \to L^r(\nu)$ be a bounded linear operator with $1 \le p, r \le \infty$. Let $T^* : (L^r(\nu))^* \to (L^p(\mu))^*$ be the adjoint operator of T. Then the operator norms of T and T^*, denoted by $\|T\|_{p\to r}$ and $\|T^*\|_{s\to q}$ with $p^{-1} + q^{-1} = 1$ and $r^{-1} + s^{-1} = 1$, satisfy*

$$\|T^*\|_{s\to q} = \|T\|_{p\to r}.$$

Proof. Note that for $f \in (L^r(\nu))^*$ and $u \in L^p(\mu)$,

$$|(T^* f)(u)| = |f(Tu)| \le \|T\|_{p\to r}\|f\|_{(L^r(\nu))^*}\|u\|_p,$$

which implies $\|T^*\|_{s\to q} \le \|T\|_{p\to r}$.

Conversely, for $v \in L^s(\nu)$, define $T_v(w) = \int_{\mathcal{Y}} v(y)w(y)d\nu(y)$ for all $w \in L^r(\nu)$. It is obvious that $T_v \in (L^r(\nu))^*$, $\|T_v\|_{(L^r(\nu))^*} = \|v\|_s$ and for $u \in L^p(\mu)$,

$$\int_{\mathcal{Y}} v(y)(Tu)(y)d\nu(y) = T_v(Tu) = (T^*T_v)(u) \le \|T^*\|_{s\to q}\|v\|_s\|u\|_p,$$

which implies $\|T\|_{p\to r} \le \|T^*\|_{s\to q}$. □

Lemma A.2. *Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be measure spaces and $T : L^p(\mu) \to L^r(\nu)$ be a bounded linear operator with $1 \le p, r \le \infty$. Assume that K is the kernel of T, that is, K is a $\mathcal{Y} \times \mathcal{X}$-measurable function such that for $f \in L^p(\mu)$, $Tf(\cdot) = \int_{\mathcal{Y}} K(\cdot, x)f(x)d\mu(x)$ almost everywhere in ν. Set $h(y) = \|K(y,\cdot)\|_{L^q(\mu)}$ for $y \in \mathcal{Y}$. Then*

$$\|T\|_{p\to r} \le \|h\|_{L^r(\nu)}.$$

In particular, if $\mathcal{X}$ is a countable set, $\mu > 0$ and $r = \infty$, then

$$\|T\|_{p\to\infty} = \|h\|_{\infty}.$$

Proof. Note that for $f \in L^p(\mu)$ and $g \in L^s(\nu)$,

$$\begin{aligned}\int_{\mathcal{Y}} (Tf)(y)g(y)d\nu(y) &= \iint_{\mathcal{Y}\times\mathcal{X}} K(y,x)f(x)g(y)d\nu(y)d\mu(x) \\ &\le \|f\|_{L^p(\mu)}\|g\|_{L^s(\nu)}\|h\|_{L^r(\nu)}\end{aligned}$$

The inequality is then proved by taking supremum over the set $\{f \in L^p(\mu), g \in L^s(\nu) : \|f\|_p \le 1, \|g\|_s \le 1\}$.

For the second identity, by the definition of the operator norm, one has

$$\begin{aligned}\|T\|_{p\to\infty} &= \sup_{\|f\|_{L^p(\mu)}\le 1} \sup_{y\in\mathcal{Y}} |Tf(y)| \\ &= \sup_{y\in\mathcal{Y}} \sup_{\|f\|_{L^p(\mu)}\le 1} \int_{\mathcal{X}} K(y,x)f(x)d\mu(x) = \|h\|_{\infty}.\end{aligned}$$

□

Theorem A.1. *(Riesz-Thorin interpolation theorem) Let T be a bounder linear operator from $L^{p_1}(\mu)$ to $L^{q_1}(\nu)$ and from $L^{p_2}(\mu)$ to $L^{q_2}(\nu)$, where $1 \le p_1, p_2, q_1, q_2 \le \infty$. For $\theta \in (0,1)$, set $p^{-1} = \theta p_1^{-1} + (1-\theta)p_2^{-1}$ and $q^{-1} = \theta q_1^{-1} + (1-\theta)q_2^{-1}$. Then the operator $T : L^p(\mu) \to L^q(\nu)$ is bounded and its norm satisfies*

$$\|T\|_{p\to q} \le \|T\|_{p_1\to q_1}^{\theta}\|T\|_{p_2\to q_2}^{1-\theta}.$$

Proof. Refer to Theorem 1.3 on p.179 of [34]. □

Lemma A.3. *Let T be a bounded linear operator from $L^1(\mu)$ to $L^{s'}(\nu)$ and from $L^s(\mu)$ to $L^{\infty}(\nu)$, where $s^{-1} + s'^{-1} = 1$ and $1 \le s \le \infty$. Then, for any $1 \le r, q \le \infty$ satisfying $q^{-1} = r^{-1} + s^{-1}$, the norm $\|T\|_{q\to r}$ is bounded by*

$$\|T\|_{q\to r} \le \|T\|_{s\to\infty}^{s'/q'}\|T\|_{1\to s'}^{1-s'/q'},$$

where $q^{-1} + q'^{-1} = 1$.

Moreover, if S *is a bounded linear operator from* $L^r(\nu)$ *to* $L^\infty(\pi)$, *then* $\|ST\|_{q\to\infty}$ *is bounded and satisfies*

$$\|ST\|_{q\to\infty} \le \|T\|_{s\to\infty}^{s'/q'} \|T\|_{1\to s'}^{1-s'/q'} \|S\|_{r\to\infty}.$$

Proof. Note that for $\theta = s'/q'$,

$$\frac{\theta}{s} + \frac{1-\theta}{1} = \frac{1}{q}, \quad \frac{\theta}{\infty} + \frac{1-\theta}{s'} = \frac{1}{r}.$$

The first part is then proved by Theorem A.1. The second inequality is obtained by the follow fact.

$$\|ST\|_{q\to\infty} \le \|T\|_{q\to r} \|S\|_{r\to\infty}.$$

□

BIBLIOGRAPHY

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

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

[3] Aldous, D. and Diaconis, P. *Strong uniform times and finite random walks.* Adv. in Appl. Math. 8, no. 1, 69–97, 1987.

[4] Aldous, D. and Fill, J.A. *Preliminary version of a book on finite Markov chians.* http://www.stat.berkeley.edu/users/aldous

[5] Bacher, R. *Valeur Propre minimale du laplacien de Coxeter pour le gropue symétrique.* J. Algebra **167**, 460–472, 1994.

[6] Bayer, D. and Diaconis, P. *Trailing the Dovetail Shuffle to its Lair.* Ann. Appl. Probab. **2**, 294–313, 1992.

[7] Bidigare, P., Hanlon, P. and Rockmore, D. *A combinatorial description of the spectrum for the Tsetlin library and its generalization to hyperplane arrangements.* Duke Math. J. 99, 135–174, 1999.

[8] Brown, K. and Diaconis, P. *Random walks and hyperplane arrangements.* Ann. Probab. 26, 1813–1854, 1998.

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

[10] Diaconis, P. *Finite Fourier methods: Access to tools.*Proc. Symp. Appl. Math. **44**, 171–194, 1991.

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

[12] Diaconis, P. *Mathematical developments from the analysis of riffle shuffling.* In “Groups, combinatorics & geometry” (Durham, 2001), 73–97, World Sci. Publishing, 2003.

[13] Diaconis, P., Fill, J. A. and Pitman, J. *Analysis of top to random shuffles.* Combin. Probab. Comput. **1**, no. 2, 135–155, 1992.

[14] Diaconis, P., Graham, R. L. and Morrison, J. A. *Asymptotic analysis of a random walk on a hypercube with many dimensions.* Random Structures Algorithms **1**, no. 1, 51–72, 1990.

[15] Diaconis, P. and Saloff-Coste, L. *Comparison theorem for reversible Markov chains* Ann. Appl. Probab. **3** 696–730, 1993.

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

[17] Diaconis, P. and Saloff-Coste, L. *Moderate growth and random walk on finite groups.* Geom. Funct. Anal. **4**, no. 1, 1–36, 1994.

[18] Diaconis, P. and Saloff-Coste, L. *Random walks on finite groups: a survey of analytic techniques.* Probability measures on groups and related structures, XI (Oberwolfach, 1994), 44–75, World Sci. Publishing, River Edge, NJ, 1995.

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

[20] Diaconis, P. and Saloff-Coste, L. *Nash inequalities for finite Markov chains.* J. Theoret. Probab. **9**, no. 2, 459–510, 1996.

[21] Diaconis, P. and Shahshahani, M. *Generating a random permutation with random transpositions.* Z. Wahrsch. Verw. Gebiete **57**, no. 2, 159–179, 1981.

[22] Durrett, R. *Probability:Theorey and Examples*, second edition, Duxbury Press, 1996.

[23] Durrett, R. *Shuffling chromosomes.* J. Theoret. Probab. **16**, no. 3, 725–750, 2003.

[24] Feller, W. *An introduction to probability theory and its applications*, New York : Wiley, 3rd ed. 1968.

[25] Horn, P. and Johnson, C. *Matrix Analysis.* Cambridge Univ. Press, 1985.

[26] Horn, P. and Johnson, C. *Topics in Matrix Analysis.* Cambridge Univ. Press, 1990.

[27] Jerrum, M. and Sinclair, A. *Aproximating the perminant* SIAM Jorurnal on computing **18**, 1149–1178, 1989.

[28] Norris, J. R. *Markov chains.* New York : Cambridge University Press (1997).

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

[30] Saloff-Coste, L. *Random walks on finite groups.* in Probability on discrete structures, 263–346, Encyclopaedia Math. Sci., 110, Springer, Berlin, 2004. (H. Kesten, ed.)

[31] Saloff-Coste, L. *Total variation lower bounds for finite Markov chains: Wilson's lemma.* Random walks and geometry, 515–532, Walter de Gruyter GmbH & Co. KG, Berlin, 2004.

[32] Shiryaev, A.N. *Probability.* Springer Verlag, 2nd ed, 1996.

[33] Stark, D.; Ganesh, A. and O'Connell, N. *Information loss in riffle shuffling.* Combin. Probab. Comput. 11, 79–95, 2002.

[34] Stein E. and Weiss G. (1971) *Introduction to Fourier analysis in Euclidean spaces.*Princeton Univ. Press, Princeton.

[35] Tanny, S. *A probabilistic interpretation of the Eulerian numbers.* Duke Math. J. **40**, 717–722, 1973. [Correction **41** 689, 1974.]

[36] Wilson, D. *Mixing times of lozenge tiling and card shuffling Markov chains.* Ann. Appl. Probab. **14**, 2004, 274–325.