Electronic Journal of Probability

Vol. 13 (2008), Paper no. 3, pages 26–78.

Journal URL
http://www.math.washington.edu/~ejpecp/

The cutoff phenomenon for ergodic Markov processes

Guan-Yu Chen*
Department of Mathematics
National Chiao Tung University
Hsinchu 300, Taiwan
Email:gychen@math.nctu.edu.tw

Laurent Saloff-Coste†
Malott Hall, Department of Mathematics
Cornell University
Ithaca, NY 14853-4201
Email:lsc@math.cornell.edu

Abstract

We consider the cutoff phenomenon in the context of families of ergodic Markov transition functions. This includes classical examples such as families of ergodic finite Markov chains and Brownian motion on families of compact Riemannian manifolds. We give criteria for the existence of a cutoff when convergence is measured in L^p-norm, $1 < p < \infty$. This allows us to prove the existence of a cutoff in cases where the cutoff time is not explicitly known. In the reversible case, for $1 < p \leq \infty$, we show that a necessary and sufficient condition for the existence of a max-L^p cutoff is that the product of the spectral gap by the max-L^p mixing time tends to infinity. This type of condition was suggested by Yuval Peres. Illustrative examples are discussed.

Key words: Cutoff phenomenon, ergodic Markov semigroups.

AMS 2000 Subject Classification: Primary 60J05,60J25.

Submitted to EJP on May 29, 2007, final version accepted January 8, 2008.

*First author partially supported by NSF grant DMS 0306194 and NCTS, Taiwan
†Second author partially supported by NSF grants DMS 0603886

1 Introduction

Let K be an irreducible aperiodic Markov kernel with invariant probability π on a finite state space Ω. Let $K_x^l = K^l(x,\cdot)$ denote the iterated kernel. Then

$$\lim_{l\to\infty} K_x^l = \pi$$

and this convergence can be studied in various ways. Set $k_x^l = K_x^l/\pi$ (this is the density of the probability measure K_x^l w.r.t. π). Set $D_\infty(x,l) = \max_y\{|k_x^l(y)-1|\}$ and, for $1 \le p < \infty$,

$$D_p(x,l) = \left(\sum_y |k_x^l(y)-1|^p \pi(y)\right)^{1/p}.$$

For $p=1$, this is (twice) the total variation distance between K_x^l and π. For $p=2$, this is the so-called chi-square distance.

In this context, the idea of cutoff phenomenon was introduced by D. Aldous and P. Diaconis in [1; 2; 3] to capture the fact that some ergodic Markov chains converge abruptly to their invariant distributions. In these seminal works convergence was usually measured in total variation. See also [9; 21; 29] where many examples are described. The first example where a cutoff in total variation was proved (although the term cutoff was actually introduced in later works) is the random transposition Markov chain on the symmetric group studied by Diaconis and Shahshahani in [13]. One of the most precise and interesting cutoff result was proved by D. Aldous [1] and improved by D. Bayer and P. Diaconis [4]. It concerns repeated riffle shuffles. We quote here Bayer and Diaconis version for illustration purpose.

Theorem 1.1 ([4])**.** *Let P_n^l denote the distribution of a deck of n cards after l riffle shuffles (starting from the deck in order). Let u_n be the uniform distribution and set $l = (3/2)\log_2 n + c$. Then, for large n,*

$$\|P_n^l - u_n\|_{\mathrm{TV}} = 1 - 2\Phi\left(-\frac{2^{-c}}{4\sqrt{3}}\right) + O\left(\frac{1}{n^{1/4}}\right)$$

where

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

As this example illustrates, the notion of cutoff is really meaningful only when applied to a family of Markov chains (here, the family is indexed by the number of cards). Precise definitions are given below in a more general context.

Proving results in the spirit of the theorem above turns out to be quite difficult because it often requires a very detailed analysis of the underlying family of Markov chains. At the same time, it is believed that the cutoff phenomenon is widely spread and rather typical among fast mixing Markov chains. Hence a basic natural question is whether or not it is possible to prove that a family of ergodic Markov chains has a cutoff without studying the problem in excruciating detail and, in particular, without having to determine the cutoff time. In this spirit, Yuval Peres proposed a simple criterion involving only the notion of spectral gap and mixing time.

The aim of this paper is to show that, somewhat surprisingly, the answer to this question is yes and Peres criterion works as long as convergence is measured using the chi-square distance D_2

(and, to some extent, D_p with $1 < p \leq \infty$). Unfortunately, for the very natural total variation distance (i.e., the case $p = 1$), the question is more subtle and no good general answer is known (other measuring tools such as relative entropy and separation will also be briefly mentioned later).

Although the cutoff phenomenon has mostly been discussed in the literature for finite Markov chains, it makes perfect sense in the much more general context of ergodic Markov semigroups. For instance, on a compact Riemannian manifold, let μ_t be the distribution of Brownian motion at time t, started at some given fixed point x at time 0. In this case, the density of μ_t with respect to the underlying normalized Riemannian measure is the heat kernel $h(t, x, \cdot)$. Now, given a sequence of Riemannian Manifolds M_n, we can ask whether or not the convergence of Brownian motion to its equilibrium presents a cutoff. Anticipating the definitions given later in this paper, we state the following series of results which illustrates well the spirit of this work.

Theorem 1.2. *Referring to the convergence of Brownian motion to its stationary measure on a family (M_n) of compact Riemannian manifolds, we have:*

1. *If all manifolds in the family have the same dimension and non-negative Ricci curvature, then there is no max-L^p cutoff, for any p, $1 \leq p \leq \infty$.*
2. *If for each n, $M_n = \mathbb{S}_n$ is the (unit) sphere in $\mathbb{R}^{n+1}$ then, for each $p \in [1, \infty)$ (resp. $p = \infty$), there is a max-L^p cutoff at time $t_n = \frac{\log n}{2n}$ (resp. $t_n = \frac{\log n}{n}$) with strongly optimal window $1/n$.*
3. *If for each n, $M_n = \mathrm{SO}(n)$ (the special orthogonal group of n by n matrices, equipped with its canonical Killing metric) then, for each $p \in (1, \infty]$, there is a max-L^p cutoff at time t_n with window 1. The exact time t_n is not known but, for any $\eta \in (0, 1)$, t_n is asymptotically between $(1-\eta)\log n$ and $2(1+\eta)\log n$ if $p \in (1, \infty)$ and between $2(1-\eta)\log n$ and $4(1+\eta)\log n$ if $p = \infty$.*

The last case in this theorem is the most interesting to us here as it illustrates the main result of this paper which provides a way of asserting that a cutoff exists even though one is not able to determine the cutoff time. Note that $p = 1$ (i.e., total variation) is excluded in (3) above. It is believed that there is a cutoff in total variation in (3) but no proof is known at this writing.

Returning to the setting of finite Markov chains, it was mentioned above that the random transposition random walk on the symmetric group S_n was the first example for which a cutoff was proved. The original result of [13] shows (essentially) that random transposition has a cutoff both in total variation and in L^2 (i.e., chi-square distance) at time $\frac{1}{2}n \log n$. It may be surprising then that for great many other random walks on S_n (e.g., adjacent transpositions, random insertions, random reversals, ...), cutoffs and cutoff times are still largely a mystery. The main result of this paper sheds some light on these problems by showing that, even if one is not able to determine the cutoff times, all these examples present L^2-cutoffs.

Theorem 1.3. *For each n, consider an irreducible aperiodic random walk on the symmetric group S_n driven by a symmetric probability measure v_n. Let β_n be the second largest eigenvalue, in absolute value. Assume that $\inf_n \beta_n > 0$ and set*

$$\sigma_n = \sum_{x \in S_n} \mathrm{sgn}(x) v_n(x).$$

If $|\sigma_n| < \beta_n$ or $|\sigma_n| = \beta_n$ but $x \mapsto \operatorname{sgn}(x)$ is not the only eigenfunction associated to $\pm\beta_n$, then, for any fixed $p \in (1, \infty]$, this family of random walks on S_n has a max-L^p cutoff

To see how this applies to the adjacent transposition random walk, i.e., v_n is uniform on $\{\mathrm{Id}, (i, i+1), 1 \leq i \leq n-1\}$, recall that $1 - \beta_n$ is of order $1/n^3$ in this case whereas one easily computes that $\sigma_n = -1 + 2/n$. See [29] for details and further references to the literature.

The crucial property of the symmetric group S_n used in the theorem above is that most irreducible representations have high multiplicity. This is true for many family of groups (most simple groups, either in the finite group sense or in the Lie group sense). The following theorem is stated for special orthogonal groups but it holds also for any of the classical families of compact Lie groups. Recall that a probability measure v on a compact group G is symmetric if $v(A) = v(A^{-1})$. Convolution by a symmetric measure is a self-adjoint operator on $L^2(G)$. If there exists l such that the l-th convolution power $v^{(l)}$ is absolutely continuous with a continuous density then this operator is compact.

Theorem 1.4. *For each n, consider a symmetric probability measure v_n on the special orthogonal group $\mathrm{SO}(n)$ and assume that there exists l_n such that $v_n^{(l_n)}$ is absolutely continuous with respect to Haar measure and admits a continuous density. Let β_n be the second largest eigenvalue, in absolute value, of the operator of convolution by v_n. Assume that $\inf_n \beta_n > 0$. Then, for any fixed $p \in (1, \infty]$, this family of random walks on $\mathrm{SO}(n)$ has a max-L^p cutoff.*

Among the many examples to which this theorem applies, one can consider the family of random planar rotations studied in [24] which are modelled by first picking up a random plane and then making a θ rotation. In the case $\theta = \pi$, [24] proves a cutoff (both in total variation and in L^2) at time $(1/4)n \log n$. Other examples are in [22; 23].

Another simple but noteworthy application of our results concerns expander graphs. For simplicity, for any fixed k, say that a family (V_n, E_n) of finite non-oriented k-regular graphs is a family of expanders if (a) the cardinality $|V_n|$ of the vertex set V_n tends to infinity with n, (b) there exists $\epsilon > 0$ such that, for any n and any set $A \subset V_n$ of cardinality at most $|V_n|/2$, the number of edges between A and its complement is at least $\epsilon|A|$. Recall that the lazy simple random walk on a graph is the Markov chain which either stays put or jumps to a neighbor chosen uniformly at random, each with probability $1/2$.

Theorem 1.5. *Let (V_n, E_n) be a family of k-regular expander graphs. Then, for any $p \in (1, \infty]$, the associated family of lazy simple random walks presents a max-L^p cutoff as well as an L^p cutoff from any fixed sequence of starting points.*

We close this introduction with remarks regarding practical implementations of Monte Carlo Markov Chain techniques. In idealized MCMC practice, an ergodic Markov chain is run in order to sample from a probability distribution of interest. In such situation, one can often identify parameters that describe the complexity of the task (in card shuffling examples, the number of cards). For simplicity, denote by n the complexity parameter. Now, in order to obtain a "good" sample, one needs to determine a "sufficiently large" running time T_n to be used in the sampling algorithm. Cost constraints (of various sorts) imply that it is desirable to find a reasonably low sufficient running time. The relevance of the cutoff phenomenon in this context is that it implies that there indeed exists an asymptotically optimal sufficient running time. Namely, if the family of Markov chains underlying the sampling algorithm (indexed by the complexity parameter n) presents a cutoff at time t_n then the optimal sufficient running time T_n is asymptotically equivalent to t_n. If there is a cutoff at time t_n with window size b_n (by definition, this implies that b_n/t_n tends to 0), then one gets the more precise result that the optimal running time T_n should satisfy $|T_n - t_n| = O(b_n)$ as n tends to infinity. The crucial point here is that, if there is a cutoff, these relations hold for any desired fixed admissible error size whereas, if there is no cutoff, the optimal sufficient running time T_n depends greatly of the desired admissible error size.

Now, in the discussion above, it is understood that errors are measured in some fixed acceptable way. The chi-square distance at the center of the present work is a very strong measure of convergence, possibly stronger than desirable in many applications. Still, what we show is that in any reversible MCMC algorithm, assuming that errors are measured in chi-square distance, if any sufficient running time is much longer than the relaxation time (i.e., the inverse of the spectral gap) then there is a cutoff phenomenon. This means that for any such algorithm, there is an asymptotically well defined notion of optimal sufficient running time as discussed above (with window size equal to the relaxation time). This work says nothing however about how to find this optimal sufficient running time.

Description of the paper

In Section 2, we introduce the notion of cutoff and its variants in a very general context. We also discuss the issue of optimality of the window of a cutoff. This section contains technical results whose proofs are in the appendix.

In Section 3, we introduce the general setting of Markov transition functions and discuss L^p cutoffs from a fixed starting distribution.

Section 4 treats max-L^p cutoffs under the hypothesis that the underlying Markov operators are normal operators. In this case, a workable necessary and sufficient condition is obtained for the existence of a max-L^p cutoff when $1 < p < \infty$.

In Section 5, we consider the reversible case and the normal transitive case (existence of a transitive group action). In those cases, we show that the existence of a max-L^p cutoff is independent of $p \in (1, \infty)$. In the reversible case, $p = +\infty$ is included.

Finally, in Section 6, we briefly describe examples proposed by David Aldous and by Igor Pak that show that the criterion obtained in this paper in the case $p \in (1, \infty)$ does not work for $p = 1$ (i.e., in total variation).

2 Terminology

This section introduces some terminology concerning the notion of cutoff. We give the basic definitions and establish some relations between them in a context that emphasizes the fact that no underlying probability structure is needed.

2.1 Cutoffs

The idea of a cutoff applies to any family of non-increasing functions taking values in $[0, \infty]$.

Definition 2.1. For $n \geq 1$, let $D_n \subset [0,\infty)$ be an unbounded set containing 0. Let $f_n : D_n \to [0,\infty]$ be a non-increasing function vanishing at infinity. Assume that

$$M = \limsup_{n\to\infty} f_n(0) > 0. \tag{2.1}$$

Then the family $\mathcal{F} = \{f_n : n = 1, 2, ...\}$ is said to present

(c1) a precutoff if there exist a sequence of positive numbers t_n and $b > a > 0$ such that

$$\lim_{n\to\infty} \sup_{t>bt_n} f_n(t) = 0, \quad \liminf_{n\to\infty} \inf_{t<at_n} f_n(t) > 0.$$

(c2) a cutoff if there exists a sequence of positive numbers t_n such that

$$\lim_{n\to\infty} \sup_{t>(1+\epsilon)t_n} f_n(t) = 0, \quad \lim_{n\to\infty} \inf_{t<(1-\epsilon)t_n} f_n(t) = M,$$

for all $\epsilon \in (0,1)$.

(c3) a (t_n, b_n) cutoff if $t_n > 0$, $b_n \geq 0$, $b_n = o(t_n)$ and

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

where, for $c \in \mathbb{R}$,

$$\overline{F}(c) = \limsup_{n\to\infty} \sup_{t>t_n+cb_n} f_n(t), \quad \underline{F}(c) = \liminf_{n\to\infty} \inf_{t<t_n+cb_n} f_n(t). \tag{2.2}$$

Regarding (c2) and (c3), we sometimes refer informally to t_n as a cutoff sequence and b_n as a window sequence.

Remark 2.1. In (c3), since f_n might not be defined at $t_n + cb_n$, we have to take the supremum and the infimum in (2.2). However, if $D_n = [0,\infty)$ and $b_n > 0$, then a (t_n, b_n) cutoff is equivalent to ask $\lim_{c\to\infty} \overline{G}(c) = 0$ and $\lim_{c\to-\infty} \underline{G}(c) = M$, where for $c \in \mathbb{R}$,

$$\overline{G}(c) = \limsup_{n\to\infty} f_n(t_n + cb_n), \quad \underline{G}(c) = \liminf_{n\to\infty} f_n(t_n + cb_n).$$

Remark 2.2. To understand and picture what a cutoff entails, let $\mathcal{F}$ be a family as in Definition 2.1 with $D_n \equiv [0,\infty)$ and let $(t_n)_1^\infty$ be a sequence of positive numbers. Set $g_n(t) = f_n(t_n t)$ for $t > 0$ and $n \geq 1$. Then $\mathcal{F}$ has a precutoff if and only if there exist $b > a > 0$ and $t_n > 0$ such that

$$\lim_{n\to\infty} g_n(b) = 0, \quad \liminf_{n\to\infty} g_n(a) > 0.$$

Similarly, $\mathcal{F}$ has a cutoff with cutoff sequence t_n if and only if

$$\lim_{n\to\infty} g_n(t) = \begin{cases} 0 & \text{for } t > 1 \\ M & \text{for } 0 < t < 1 \end{cases}$$

Equivalently, the family $\{g_n : n = 1, 2, ...\}$ has a cutoff with cutoff sequence 1.

Remark 2.3. Obviously, (c3)⇒(c2)⇒(c1). In (c3), if, for $n \ge 1$, $D_n = [0, \infty)$ and f_n is continuous, then the existence of a (t_n, b_n) cutoff implies that $b_n > 0$ for n large enough.

Remark 2.4. It is worth noting that another version of cutoff, called a weak cutoff, is introduced by Saloff-Coste in [28]. By definition, a family $\mathcal{F}$ as above is said to present a weak cutoff if there exists a sequence of positive numbers t_n such that

$$\forall \epsilon > 0, \lim_{n\to\infty} \sup_{t>(1+\epsilon)t_n} f_n(t) = 0, \quad \liminf_{n\to\infty} \inf_{t<t_n} f_n(t) > 0.$$

It is easy to see that the weak cutoff is stronger than the precutoff but weaker than the cutoff. The weak cutoff requires a positive lower bound on the left limit of f_n at t_n whereas the cutoffs in (c1)-(c3) require no information on the values of f_n in a small neighborhood of t_n. This makes it harder to find a cutoff sequence for a weak cutoff and differentiates the weak cutoff from the notions considered above.

The following examples illustrate Definition 2.1. Observe that the functions f_n below are all sums of exponential functions. Such functions appear naturally when the chi-square distance is used in the context of ergodic Markov processes.

Example 2.1. Fix $\alpha > 0$. For $n \ge 1$, let f_n be an extended function on $[0, \infty)$ defined by $f_n(t) = \sum_{k\ge1} e^{-tk^\alpha/n}$. Note that $f_n(0) = \infty$ for $n \ge 1$. This implies $M = \limsup_{n\to\infty} f_n(0) = \infty$. We shall prove that $\mathcal{F}$ has no precutoff. To this end, observe that since $k^\alpha \ge 1 + \alpha \log k$, $k \ge 1$, we have

$$e^{-t/n} \le f_n(t) \le e^{-t/n} \sum_{k=1}^{\infty} k^{-\alpha t/n}, \quad \forall t \ge 0, n \ge 1. \tag{2.3}$$

Let t_n and b be positive numbers such that $f_n(bt_n) \to 0$. The first inequality above implies $t_n/n \to \infty$. The second inequality gives $f_n(at_n) = O(e^{-at_n/n})$ for all $a > 0$. Hence, we must have $f_n(at_n) \to 0$ for all $a > 0$. This rules out any precutoff.

Example 2.2. Let $\mathcal{F} = \{f_n : n \ge 1\}$, where $f_n(t) = \sum_{k\ge1} n^k e^{-tk/n}$ for $t \ge 0$ and $n \ge 1$. Then $f_n(t) = \infty$ for $t \in [0, n \log n]$ and $f_n(t) = ne^{-t/n}/(1 - ne^{-t/n})$ for $t > n \log n$. This implies that $M = \infty$. Setting $t_n = n \log n$ and $b_n = n$, the functions $\overline{F}$ and $\underline{F}$ defined in (2.2) are given by

$$\overline{F}(c) = \underline{F}(c) = \begin{cases} \frac{e^{-c}}{1-e^{-c}} & \text{if } c > 0 \\ \infty & \text{if } c \le 0 \end{cases} \tag{2.4}$$

Hence, $\mathcal{F}$ has a $(n \log n, n)$ cutoff.

Example 2.3. Let $\mathcal{F} = \{f_n : n \ge 1\}$, where $f_n(t) = (1 + e^{-t/n})^n - 1$ for $t \ge 0$, $n \ge 1$. Obviously, $M = \infty$. In this case, setting $t_n = n \log n$ and $b_n = n$ yields

$$\overline{F}(c) = \underline{F}(c) = e^{e^{-c}} - 1, \quad \forall c \in \mathbb{R}. \tag{2.5}$$

This proves that $\mathcal{F}$ has the $(n \log n, n)$ cutoff.

2.2 Window optimality

It is clear that the quantity b_n in (c3) reflects the sharpness of a cutoff and may depend on the choice of t_n. We now introduce different notions of optimality for the window sequence in (c3).

Definition 2.2. Let $\mathcal{F}$ and M be as in Definition 2.1. Assume that $\mathcal{F}$ presents a (t_n, b_n) cutoff. Then, the cutoff is

(w1) weakly optimal if, for any (t_n, d_n) cutoff for $\mathcal{F}$, one has $b_n = O(d_n)$.

(w2) optimal if, for any (s_n, d_n) cutoff for $\mathcal{F}$, we have $b_n = O(d_n)$. In this case, b_n is called an optimal window for the cutoff.

(w3) strongly optimal if, for all $c > 0$,

$$0 < \liminf_{n\to\infty} \sup_{t > t_n + cb_n} f_n(t) \le \limsup_{n\to\infty} \inf_{t < t_n - cb_n} f_n(t) < M.$$

Remark 2.5. Obviously, (w3)⇒(w2)⇒(w1). If $\mathcal{F}$ has a strongly optimal (t_n, b_n) cutoff, then $b_n > 0$ for n large enough. If D_n is equal to $\mathbb{N}$ for all $n \ge 1$, then a strongly optimal (t_n, b_n) cutoff for $\mathcal{F}$ implies $\liminf\limits_{n\to\infty} b_n > 0$.

Remark 2.6. Let $\mathcal{F}$ be a family of extended functions defined on $\mathbb{N}$. If $\mathcal{F}$ has a (t_n, b_n) cutoff with $b_n \to 0$, it makes no sense to discuss the optimality of the cutoff and the window. Instead, it is worthwhile to determine the limsup and liminf of the sequences

$$f_n(\lfloor t_n \rfloor + k) \quad \text{for } k = -1, 0, 1.$$

See [7] for various examples.

Remark 2.7. Let $\mathcal{F}$ be a family of extended functions presenting a strongly optimal (t_n, b_n) cutoff. If $T = [0, \infty)$ then there exist $N > 0$ and $0 < c_1 < c_2 < M$ such that $c_1 \le f_n(t_n) \le c_2$ for all $n > N$. In the discrete time case where $T = \mathbb{N}$, we have instead $c_1 \le f_n(\lceil t_n \rceil) \le f_n(\lfloor t_n \rfloor) \le c_2$ for all $n > N$.

The following lemma gives an equivalent definition for (w3) using the functions in (2.2).

Lemma 2.1. *Let $\mathcal{F}$ be a family as in Definition 2.1 with $D_n = \mathbb{N}$ for all $n \ge 1$ or $D_n = [0, \infty)$ for all $n \ge 1$. Assume (2.1) holds. Then a family presents a strongly optimal (t_n, b_n) cutoff if and only if the functions, $\overline{F}$ and $\underline{F}$, defined in (2.2) with respect to t_n, b_n satisfy $\overline{F}(-c) < M$ and $\underline{F}(c) > 0$ for all $c > 0$.*

Proof. See the appendix. □

Our next proposition gives conditions that are almost equivalent to the various optimality conditions introduced in Definition 2.2. These are useful in investigating the optimality of a window.

Proposition 2.2. *Let $\mathcal{F} = \{f_n, n = 1, 2, ...\}$ be a family of non-increasing functions $f_n : [0, \infty) \to [0, \infty]$ vanishing at infinity. Set $M = \limsup_n f_n(0)$. Assume that $M > 0$ and that $\mathcal{F}$ has a (t_n, b_n) cutoff with $b_n > 0$. For $c \in \mathbb{R}$, let*

$$\overline{G}(c) = \limsup_{n\to\infty} f_n(t_n + cb_n), \quad \underline{G}(c) = \liminf_{n\to\infty} f_n(t_n + cb_n). \tag{2.6}$$

(i) If there exists $c > 0$ such that either $\underline{G}(c) > 0$ or $\overline{G}(-c) < M$ holds, then the (t_n, b_n) cutoff is weakly optimal. Conversely, if the (t_n, b_n) cutoff is weakly optimal, then there is $c > 0$ such that either $\overline{G}(c) > 0$ or $\underline{G}(-c) < M$.

(ii) If there exist $c_2 > c_1$ such that $0 < \underline{G}(c_2) \le \overline{G}(c_1) < M$, then the (t_n, b_n) cutoff is optimal. Conversely, if the (t_n, b_n) cutoff is optimal, then there are $c_2 > c_1$ such that $\overline{G}(c_2) > 0$ and $\underline{G}(c_1) < M$.

(iii) The (t_n, b_n) cutoff is strongly optimal if and only if $0 < \underline{G}(c) \le \overline{G}(c) < M$ for all $c \in \mathbb{R}$. In particular, if (t_n, b_n) is an optimal cutoff and there exists $c \in \mathbb{R}$ such that $\overline{G}(c) = M$ or $\underline{G}(c) = 0$, then there is no strongly optimal cutoff for $\mathcal{F}$.

Remark 2.8. Proposition 2.2 also holds if one replaces $\overline{G}, \underline{G}$ with $\overline{F}, \underline{F}$ defined at (2.2).

Remark 2.9. Consider the case when f_n has domain $\mathbb{N}$. Assume that $\liminf_n b_n > 0$ and replace $\overline{G}, \underline{G}$ in Proposition 2.2 with $\overline{F}, \underline{F}$ defined at (2.2). Then (iii) remains true whereas the first parts of (i),(ii) still hold if, respectively,

$$\liminf_{n\to\infty} b_n > 2/c, \quad \liminf_{n\to\infty} b_n > 4/(c_2 - c_1).$$

The second parts of (i),(ii) hold if we assume $\limsup_n b_n = \infty$.

Example 2.4 (Continuation of Example 2.2). Let $\mathcal{F}$ be the family in Example 2.2. By (2.4), $\mathcal{F}$ has a (t_n, n) cutoff with $t_n = n \log n$. Suppose $\mathcal{F}$ has a (t_n, c_n) cutoff. By definition, since $f_n(t_n + n) = 1/(e-1)$, we may choose $C > 0, N > 0$ such that

$$f_n(t_n + Cc_n) < f_n(t_n + n), \quad \forall n \ge N.$$

This implies $n = O(c_n)$ and, hence, the $(n \log n, n)$ cutoff is weakly optimal. We will prove later in Example 2.6 that such a cutoff is optimal but that no strongly optimal cutoff exists.

Example 2.5 (Continuation of Example 2.3). For the family $\mathcal{F}$ in Example 2.3, (2.5) implies that the $(n \log n, n)$ cutoff is strongly optimal.

2.3 Mixing time

The cutoff phenomenon in Definition 2.1 is closely related to the way each function in $\mathcal{F}$ tends to 0. To make this precise, consider the following definition.

Definition 2.3. Let f be an extended real-valued non-negative function defined on $D \subset [0, \infty)$. For $\epsilon > 0$, set

$$T(f, \epsilon) = \inf\{t \in D : f(t) \le \epsilon\}$$

if the right hand side above is non-empty and let $T(f, \epsilon) = \infty$ otherwise.

In the context of ergodic Markov processes, $T(f_n, \epsilon)$ appears as the mixing time. This explains the title of this subsection.

Proposition 2.3. *Let $\mathcal{F} = \{f_n : [0, \infty) \to [0, \infty] | n = 1, 2, ...\}$ be a family of non-increasing functions vanishing at infinity. Assume that (2.1) holds. Then:*

(i) $\mathcal{F}$ has a precutoff if and only if there exist constants $C \ge 1$ and $\delta > 0$ such that, for all $0 < \eta < \delta$,

$$\limsup_{n\to\infty} \frac{T(f_n, \eta)}{T(f_n, \delta)} \le C. \tag{2.7}$$

(ii) $\mathcal{F}$ has a cutoff if and only if (2.7) *holds for all* $0 < \eta < \delta < M$ *with* $C = 1$.

(iii) For $n \geq 1$, *let* $t_n > 0$, $b_n \geq 0$ *be such that* $b_n = o(t_n)$. *Then* $\mathcal{F}$ *has a* (t_n, b_n) *cutoff if and only if, for all* $\delta \in (0, M)$,

$$|t_n - T(f_n, \delta)| = O_\delta(b_n). \tag{2.8}$$

Proof. The proof is similar to that of the next proposition. □

Remark 2.10. If (2.7) holds for $0 < \eta < \delta < M$ with $C = 1$, then $T(f_n, \eta) \sim T(f_n, \delta)$ for all $0 < \eta < \delta < M$, where, for two sequences of positive numbers (t_n) and (s_n), $t_n \sim s_n$ means that $t_n/s_n \to 1$ as $n \to \infty$.

Proposition 2.4. *Let* $\mathcal{F} = \{f_n : \mathbb{N} \to [0, \infty] | n = 1, 2, ...\}$ *be a family of non-increasing functions vanishing at infinity. Let* M *be the limit defined in* (2.1). *Assume that there exists* $\delta_0 > 0$ *such that*

$$\lim_{n\to\infty} T(f_n, \delta_0) = \infty. \tag{2.9}$$

Then (i) *and* (ii) *in Proposition 2.3 hold. Furthermore, if* b_n *satisfies*

$$\liminf_{n\to\infty} b_n > 0, \tag{2.10}$$

then (iii) *in Proposition 2.3 holds.*

Proof. See the appendix. □

Remark 2.11. A similar equivalent condition for a weak cutoff is established in [6]. In detail, referring to the setting of Proposition 2.3 and 2.4, a family $\mathcal{F} = \{f_n : n = 1, 2, ...\}$ has a weak cutoff if and only if there exists a positive constant $\delta > 0$ such that (2.7) holds for $0 < \eta < \delta$ with $C = 1$.

Remark 2.12. More generally, if $\mathcal{F}$ is the family introduced in Definition 2.1, then Proposition 2.3 holds when D_n is dense in $[0, \infty)$ for all $n \geq 1$. Proposition 2.4 holds when $[0, \infty) = \bigcup_{x \in D_n} (x - r, x + r)$ for all $n \geq 1$, where r is a fixed positive constant. This fact is also true for the equivalence of the weak cutoff in Remark 2.11.

A natural question concerning cutoff sequences arises. Suppose a family $\mathcal{F}$ has a cutoff with cutoff sequence $(s_n)_1^\infty$ and a cutoff with cutoff sequence $(t_n)_1^\infty$. What is the relation between s_n and t_n? The following corollary which follows immediately from Propositions 2.3 and 2.4 answers this question.

Corollary 2.5. *Let* $\mathcal{F}$ *be a family as in Proposition 2.3 satisfying* (2.1) *or as in Proposition 2.4 satisfying* (2.9).

(i) If $\mathcal{F}$ *has a cutoff, then the cutoff sequence can be taken to be* $(T(f_n, \delta))_1^\infty$ *for any* $0 < \delta < M$.

(ii) $\mathcal{F}$ *has a cutoff with cutoff sequence* $(t_n)_1^\infty$ *if and only if* $t_n \sim T(f_n, \delta)$ *for all* $0 < \delta < M$.

(iii) Assume that $\mathcal{F}$ *has a cutoff with cutoff sequence* $(t_n)_1^\infty$. *Then* $\mathcal{F}$ *has a cutoff with cutoff sequence* $(s_n)_1^\infty$ *if and only if* $t_n \sim s_n$.

In the following, if $\mathcal{F}$ is the family in Proposition 2.4, we assume further that the sequence $(b_n)_1^\infty$ satisfies (2.10)*.*

(iv) If $\mathcal{F}$ has a (t_n, b_n) cutoff, then $\mathcal{F}$ has a $(T(f_n, \delta), b_n)$ cutoff for any $0 < \delta < M$.

(v) Assume that $\mathcal{F}$ has a (t_n, b_n) cutoff. Let $s_n > 0$ and $d_n \geq 0$ be such that $d_n = o(s_n)$ and $b_n = O(d_n)$. Then $\mathcal{F}$ has a (s_n, d_n) cutoff if and only if $|t_n - s_n| = O(d_n)$.

The next corollaries also follow immediately from Propositions 2.3 and 2.4. They address the optimality of a window.

Corollary 2.6. *Let $\mathcal{F}$ be a family in Proposition 2.3 satisfying* (2.1)*. Assume that $\mathcal{F}$ has a cutoff. Then the following are equivalent.*

(i) b_n is an optimal window.

(ii) $\mathcal{F}$ has an optimal $(T(f_n, \delta), b_n)$ cutoff for some $0 < \delta < M$.

(iii) $\mathcal{F}$ has a weakly optimal $(T(f_n, \delta), b_n)$ cutoff for some $0 < \delta < M$.

Proof. (ii)$\Rightarrow$(iii) is obvious. For (i)$\Rightarrow$(ii), assume that $\mathcal{F}$ has a (t_n, b_n) cutoff. Then, by Proposition 2.3, $\mathcal{F}$ has a $(T(f_n, \delta), b_n)$ cutoff for all $\delta \in (0, M)$. The optimality is obvious from that of the (t_n, b_n) cutoff. For (iii)$\Rightarrow$(i), assume that $\mathcal{F}$ has a (s_n, c_n) cutoff. By Proposition 2.3, $\mathcal{F}$ has a $(T(f_n, \delta), c_n)$ cutoff. Consequently, the weak optimality implies that $b_n = O(c_n)$. ☐

Remark 2.13. In the case where $\mathcal{F}$ consists of functions defined on $[0, \infty)$, there is no difference between a weakly optimal cutoff and an optimal cutoff if the cutoff sequence is selected to be $(T(f_n, \delta))_1^\infty$ for some $0 < \delta < M$.

Corollary 2.7. *Let $\mathcal{F}$ be as in Proposition 2.4 satisfying* (2.9) *and $(b_n)_1^\infty$ be such that $\liminf_{n\to\infty} b_n > 0$. Assume that $\mathcal{F}$ has a cutoff. Then the following are equivalent.*

(i) b_n is an optimal window.

(ii) For some $\delta \in (0, M)$, the family $\mathcal{F}$ has both weakly optimal $(T(f_n, \delta), b_n)$ and $(T(f_n, \delta) - 1, b_n)$ cutoffs.

Proof. See the appendix. ☐

Example 2.6 (Continuation of Example 2.2). In Example 2.2, the family $\mathcal{F}$ has been proved to have a $(n \log n, n)$ cutoff and the functions $\underline{F}, \overline{F}$ are computed out in (2.4). We noticed in Example 2.4 that this is weakly optimal. By Lemma 2.8, we may conclude from (2.4) that n is an optimal window and also that no strongly optimal cutoff exists. Indeed, the forms of $\underline{F}, \overline{F}$ show that the optimal "right window" is of order n but the optimal "left window" is 0. Since our definition for an optimal cutoff is symmetric, the optimal window should be the larger one and no strongly optimal window can exist. The following lemma generalizes this observation.

Lemma 2.8. *Let $\mathcal{F}$ be a family as in Proposition 2.3 that satisfies* (2.1)*. Assume that $\mathcal{F}$ has a (t_n, b_n) cutoff and let $\overline{F}, \underline{F}$ be functions in* (2.2) *associated with t_n, b_n.*

(i) Assume that either $\underline{F} > 0$ or $\overline{F} < M$. Then the (t_n, b_n) cutoff is optimal.

(ii) Assume that either $\underline{F} > 0$ with $\overline{F}(c) = M$ for some $c \in \mathbb{R}$ or $\overline{F} < M$ with $\underline{F}(c) = 0$ for some $c \in \mathbb{R}$. Then there is no strongly optimal cutoff for $\mathcal{F}$.

The above is true for a family as in Proposition 2.4 if we assume further $\liminf_{n\to\infty} b_n > 0$.

Proof. See the appendix. ☐

The following proposition compares the window of a cutoff between two families. This is useful in comparing the sharpness of cutoffs when two families have the same cutoff sequence.

Proposition 2.9. *Let $\mathcal{F} = \{f_n : n \geq 1\}$ and $\mathcal{G} = \{g_n : n \geq 1\}$ be families both as in Proposition 2.3 or 2.4 and set*

$$\limsup_{n\to\infty} f_n(0) = M_1, \quad \limsup_{n\to\infty} g_n(0) = M_2.$$

Assume that $M_1 > 0$ and $M_2 > 0$. Assume further that $\mathcal{F}$ has a strongly optimal (t_n, b_n) cutoff and that $\mathcal{G}$ has a (s_n, c_n) cutoff with $|s_n - t_n| = O(b_n)$. Then:

(i) If $f_n \leq g_n$ for all $n \geq 1$, then $b_n = O(c_n)$.

(ii) If $M_1 = M_2$ and, for $n \geq 1$, either $f_n \geq g_n$ or $f_n \leq g_n$, then $b_n = O(c_n)$.

Proof. See the appendix. ☐

3 Ergodic Markov processes and semigroups

3.1 Transition functions, Markov processes

As explained in the introduction, the cutoff phenomenon was originally introduced in the context of finite Markov chains. However, it makes sense in the much larger context of ergodic Markov processes. In what follows, we let time be either continuous $t \in [0, \infty)$ or discrete $t \in \{0, 1, 2, \dots, \infty\} = \mathbb{N}$.

A Markov transition function on a space Ω equipped with a σ-algebra $\mathcal{B}$, is a family of probability measures $p(t, x, \cdot)$ indexed by $t \in T$ ($T = [0, \infty)$ or $\mathbb{N}$) and $x \in \Omega$ such that $p(0, x, \Omega \setminus \{x\}) = 0$ and, for each $t \in T$ and $A \in \mathcal{B}$, $p(t, x, A)$ is $\mathcal{B}$-measurable and satisfies

$$p(t + s, x, A) = \int_\Omega p(s, y, A) p(t, x, dy).$$

A Markov process $X = (X_t, t \in T)$ with filtration $\mathcal{F}_t = \sigma(X_s : s \leq t) \subset \mathcal{B}$ has $p(t, x, \cdot)$, $t \in T$, $x \in \Omega$, as transition function provided

$$E(f \circ X_s \,|\mathcal{F}_t) = \int_\Omega f(y) p(s - t, X_t, dy)$$

for all $0 < t < s < \infty$ and all bounded measurable f. The measure $\mu_0(A) = P(X_0 \in A)$ is called the initial distribution of the process X. All finite dimensional marginals of X can be expressed in terms of μ_0 and the transition function. In particular,

$$\mu_t(A) = P(X_t \in A) = \int p(t,x,A)\mu_0(dx).$$

Given a Markov transition function $p(t,x,\cdot)$, $t \in T, x \in \Omega$, for any bounded measurable function f, set

$$P_t f(x) = \int f(y)p(t,x,dy).$$

For any measure ν on $(\Omega, \mathcal{B})$ with finite total mass, set

$$\nu P_t(A) = \int p(t,x,A)\nu(dx).$$

We say that a probability measure π is invariant if $\pi P_t = \pi$ for all $t \in T$. In this general setting, invariant measures are not necessarily unique.

Example 3.1 (Finite Markov chains). A (time homogeneous) Markov chain on finite state space Ω is often described by its Markov kernel $K(x,y)$ which gives the probability of moving from x to y. The associated discrete time transition function $p^d(t,\cdot,\cdot)$ is defined inductively for $t \in \mathbb{N}$, $x,y \in \Omega$, by $p^d(0,x,y) = \delta_x(y)$ and

$$p^d(1,x,y) = K(x,y), \quad p^d(t,x,y) = \sum_{z \in \Omega} p^d(t-1,x,z)p^d(1,z,y). \tag{3.1}$$

The associated continuous time transition function $p^c(t,\cdot,\cdot)$ is defined for $t \geq 0$ and $x,y \in \Omega$ by

$$p^c(t,x,y) = e^{-t}\sum_{j=0}^{\infty} \frac{t^j}{j!} p^d(j,x,y). \tag{3.2}$$

One says that K is irreducible if, for any $x,y \in \Omega$, there exists $l \in \mathbb{N}$ such that $p^d(l,x,y) > 0$. For irreducible K, there exists a unique invariant probability π such that $\pi K = \pi$ and $p^c(t,x,\cdot)$ tends to π as t tends to infinity.

3.2 Measure of ergodicity

Our interest here is in the case where some sort of ergodicity holds in the sense that, for some initial measure μ_0, $\mu_0 P_t$ converges (in some sense) to a probability measure. By a simple argument, this limit must be an invariant probability measure.

In order to state our main results, we need the following definition.

Definition 3.1. Let $p(t,x,\cdot)$, $t \in T, x \in \Omega$, be a Markov transition function with invariant measure π. We call spectral gap (of this Markov transition function) and denote by λ the largest $c \geq 0$ such that, for all $t \in T$ and all $f \in L^2(\Omega,\pi)$,

$$\|(P_t - \pi)f\|_2 \leq e^{-tc}\|f\|_2 \tag{3.3}$$

Remark 3.1. If $T=[0,\infty)$ and P_tf tends to f in $L^2(\Omega,\pi)$ as t tends to 0 (i.e., P_t is a strongly continuous semigroup of contractions on $L^2(\Omega,\pi)$) then λ can be computed in term of the infinitesimal generator A of $P_t=e^{tA}$. Namely,

$$\lambda=\inf\left\{\langle -Af,f\rangle : f\in \mathrm{Dom}(A),\ \text{real valued},\ \pi(f)=0,\ \|f\|_2=1\right\}.$$

Note that A is not self-adjoint in general and thus λ is not always in the spectrum of A (it is in the spectrum of the self-adjoint operator $\frac{1}{2}(A+A^*)$). If A is self-adjoint then λ measures the gap between the smallest element of the spectrum of $-A$ (which is the eigenvalue 0 with associated eigenspace the space of constant functions) and the rest of the spectrum of $-A$ which lies on the positive real axis.

Remark 3.2. If $T=\mathbb{N}$ then λ is simply defined by

$$\lambda=-\log\left(\|P_1-\pi\|_{L^2(\Omega,\pi)\to L^2(\Omega,\pi)}\right).$$

In other words, $e^{-\lambda}$ is the second largest singular value of the operator P_1 on $L^2(\Omega,\pi)$.

Remark 3.3. If $\lambda>0$ then $\lim_{t\to\infty}p(t,x,A)=\pi(A)$ for π almost all x. Indeed (assuming for simplicity that $T=\mathbb{N}$), for any bounded measurable function f, we have

$$\pi(\sum_n|P_nf-\pi(f)|^2)=\sum_n\|(P_n-\pi)(f)\|^2_{L^2(\Omega,\pi)}\le\left(\sum_n e^{-2\lambda n}\right)\|f\|^2_\infty.$$

Hence $P_nf(x)$ converges to $\pi(f)$, π almost surely.

Remark 3.4. As $\|P_t-\pi\|_{L^1(\Omega,\pi)\to L^1(\Omega,\pi)}$ and $\|P_t-\pi\|_{L^\infty(\Omega,\pi)\to L^\infty(\Omega,\pi)}$ are bounded by 2, the Riesz-Thorin interpolation theorem yields

$$\|P_t-\pi\|_{L^p(\Omega,\pi)\to L^p(\Omega,\pi)}\le 2^{|1-2/p|}e^{-t\lambda(1-|1-2/p|)}. \tag{3.4}$$

We now introduce the distance functions that will be used throughout this work to measure convergence to stationarity. First, set

$$D_{\mathrm{TV}}(\mu_0,t)=\|\mu_0P_t-\pi\|_{\mathrm{TV}}=\sup_{A\in\mathcal{B}}\{|\mu_0P_t(A)-\pi(A)|\}.$$

This is the total variation distance between probability measures.

Next, fix $p\in[1,\infty]$. If t is such that the measure μ_0P_t is absolutely continuous w.r.t. π with density $h(t,\mu_0,y)$, set

$$D_p(\mu_0,t)=\left(\int_\Omega|h(t,\mu_0,y)-1|^p\pi(dy)\right)^{1/p}, \tag{3.5}$$

(understood as $D_\infty(\mu_0,t)=\|h(t,\mu_0,\cdot)-1\|_\infty$ when $p=\infty$). If μ_0P_t is not absolutely continuous with respect to π, set $D_1(\mu_0,t)=2$ and, for $p>1$, $D_p(\mu_0,t)=\infty$.

When $\mu_0=\delta_x$, we write

$$h(t,x,\cdot)\text{ for }h(t,\delta_x,\cdot)\text{ and }D_p(x,t)\text{ for }D_p(\delta_x,t).$$

Note that $t\mapsto D_p(x,t)$ is well defined for every starting point x.

The main results of this paper concern the functions D_p with $p\in(1,\infty]$. For completeness, we mention three other traditional ways of measuring convergence.

- Separation: use $\mathrm{sep}(\mu_0, t) = \sup_y\{1 - h(t, \mu_0, y)\}$ if the density exists, $\mathrm{sep}(\mu_0, t) = 1$, otherwise.
- Relative entropy: use $\mathrm{Ent}_\pi(\mu_0, t) = \int h(t, \mu_0, y) \log h(t, \mu_0, y)\pi(dy)$ if the density exists, $\mathrm{Ent}_\pi(\mu_0, t) = \infty$ otherwise.
- Hellinger: use $H_\pi(\mu_0, t) = 1 - \int \sqrt{h(t, \mu_0, y)}\pi(dy)$ if the density exists, $H_\pi(\mu_0, t) = 1$ otherwise.

Proposition 3.1. *Let $p(t, x, \cdot)$, $t \in T, x \in \Omega$, be a Markov transition function with invariant measure π. Then, for any $1 \le p \le \infty$, and any initial measure μ_0 on Ω, the function $t \mapsto D_p(\mu_0, t)$ from T to $[0, \infty]$ is non-increasing.*

Proof. Fix $1 \le p \le \infty$. Consider the operator P_t acting on bounded functions. Since π is invariant, Jensen inequality shows that P_t extends as a contraction on $L^p(\Omega, \pi)$. Given an initial measure μ_0, the measure $\mu_0 P_t$ is absolutely continuous w.r.t. π with a density in $L^p(\Omega, \pi)$ if and only if there exists a constant C such that

$$|\mu_0 P_t(f)| \le C\|f\|_q$$

for all $f \in L^q(\Omega, \pi)$ where $1/p + 1/q = 1$ (for $p \in (1, \infty]$, this amounts to the fact that L^p is the dual of L^q whereas, for $p = 1$, it follows from a slightly more subtle argument). Moreover, if this holds then the density $h(t, \mu_0, \cdot)$ has $L^p(\Omega, \pi)$-norm

$$\|h(t, \mu_0, \cdot)\|_p = \sup\{\mu_0 P_t(f) : f \in L^q(\Omega, \pi),\ \|f\|_q \le 1\}.$$

Now, observe that $\mu_{t+s} = \mu_t P_s$ with $\mu_t = \mu_0 P_t$. Also, by the invariance of π, $\mu_{t+s} - \pi = (\mu_t - \pi)P_s$. Finally, for any $f \in L^q(\Omega, \pi)$,

$$|[\mu_{t+s} - \pi](f)| = |[\mu_t - \pi]P_s(f)|.$$

Hence, if μ_t is absolutely continuous w.r.t. π with a density $h(t, \mu_0, \cdot)$ in $L^p(\Omega, \pi)$ then

$$|[\mu_{t+s} - \pi](f)| \le \|h(t, \mu_0, \cdot) - 1\|_p \|P_s f\|_q \le \|h(t, \mu_0, \cdot) - 1\|_p \|f\|_q.$$

It follows that μ_{t+s} is absolutely continuous with density $h(t + s, \mu_0, \cdot)$ in $L^p(\Omega, \pi)$ satisfying

$$D_p(\mu_0, t + s) = \|h(t + s, \mu_0, \cdot) - 1\|_p \le \|h(t, \mu_0, \cdot) - 1\|_p = D_p(\mu_0, t)$$

as desired. □

Remark 3.5. Somewhat different arguments show that total variation, separation, relative entropy and the Hellinger distance all lead to non-increasing functions of time.

Next, given a Markov transition function with invariant measure π, we introduce the maximal L^p distance over all starting points (equivalently, over all initial measures). Namely, for any fixed $p \in [1, \infty]$, set

$$\overline{D}_p(t) = \overline{D}_{\Omega,p}(t) = \sup_{x \in \Omega} D_p(x, t). \tag{3.6}$$

Obviously, the previous proposition shows that this is a non-increasing function of time. Let us insist on the fact that the supremum is taken over all starting points. In fact, let us introduce also

$$\widetilde{D}_{\pi,p}(t) = \pi\text{- }\operatorname*{ess\,sup}_{x\in\Omega} D_p(x,t). \tag{3.7}$$

Obviously $\widetilde{D}_{\pi,p}(t) \le \overline{D}_p(t)$. Note that, in general, $\widetilde{D}_{\pi,p}(t)$ cannot be used to control $D_p(\mu_0,t)$ unless μ_0 is absolutely continuous w.r.t. π. However, if Ω is a topological space and $x \mapsto D_p(x,t)$ is continuous then $\widetilde{D}_{\pi,p}(t) = \overline{D}_p(t)$.

Proposition 3.2. *Let $p(t,x,\cdot)$, $t \in T, x \in \Omega$, be a Markov transition function with invariant measure π. Then, for any $p \in [1,\infty]$, the functions $t \mapsto \overline{D}_p(t)$ and $t \mapsto \widetilde{D}_{\pi,p}(t)$ are non-increasing and sub-multiplicative.*

Proof. Assume that $t,s \in T$ are such that $h(s,x,\cdot)$ and $h(t,x,\cdot)$ exist and are in $L^p(\Omega,\pi)$, for a.e. x (otherwise there is nothing to prove). Fix such an x and observe that, for any $f \in L^q(\Omega,\pi)$ with $1/p+1/q=1$,

$$p(t+s,x,f) - \pi(f) = [p(s,x,\cdot)-\pi][P_t-\pi](f).$$

It follows that

$$\begin{aligned}
|p(t+s,x,f)-\pi(f)| &\le \|h(s,x,\cdot)-1\|_p \|[P_t-\pi](f)\|_q \\
&\le \|h(s,x,\cdot)-1\|_p \|(P_t-\pi)f\|_\infty \\
&\le \|h(s,x,\cdot)-1\|_p \operatorname*{ess\,sup}_{y\in\Omega} \|h(t,y,\cdot)-1\|_p \|f\|_q \\
&\le \|h(s,x,\cdot)-1\|_p \widetilde{D}_{\pi,p}(t) \|f\|_q.
\end{aligned}$$

Hence

$$D_p(x,t+s) \le D_p(x,s)\widetilde{D}_{\pi,p}(t).$$

This is a slightly more precise result than stated in the proposition. □

Remark 3.6. One of the reasons behind the sub-multiplicative property of $\overline{D}_p$ and $\widetilde{D}_{\pi,p}$ is that these quantities can be understood as operator norms. Namely,

$$\begin{aligned}
\overline{D}_p(t) &= \sup\left\{ \sup_\Omega\{|(P_t-\pi)f|\} :\ f \in L^q(\Omega,\pi),\ \|f\|_q = 1 \right\} \\
&= \|P_t - \pi\|_{L^q(\Omega,\pi)\to B(\Omega)}
\end{aligned} \tag{3.8}$$

where $B(\Omega)$ is the set of all bounded measurable functions on Ω equipped with the sup-norm, and

$$\begin{aligned}
\widetilde{D}_{\pi,p} &= \sup\left\{ \pi\text{- }\operatorname*{ess\,sup}_\Omega\{|(P_t-\pi)f|\} :\ f \in L^q(\Omega,\pi),\ \|f\|_q = 1 \right\} \\
&= \|P_t - \pi\|_{L^q(\Omega,\pi)\to L^\infty(\Omega,\pi)}.
\end{aligned} \tag{3.9}$$

See [14, Theorem 6, p.503].

3.3 L^p-cutoffs

Fix $p \in [1, \infty]$. Consider a family of spaces Ω_n indexed by $n = 1, 2, \dots$ For each n, let $p_n(t, x, \cdot)$, $t \in [0, \infty)$, $x \in \Omega_n$ be a transition function with invariant probability π_n. Fix a subset E_n of probability measures on Ω_n and consider the supremum of the corresponding L^p distance between $\mu P_{n,t}$ and π_n overall $\mu \in E_n$, that is,

$$f_n(t) = \sup_{\mu \in E_n} D_p(\mu, t)$$

where $D_p(\mu, t)$ is defined by (3.5). One says that the sequence (p_n, E_n) presents an L^p-cutoff when the family of functions $\mathcal{F} = \{f_n, n = 1, 2, \dots\}$ presents a cutoff in the sense of Definition 2.1. Similarly, one defines L^p precutoff and L^p (t_n, b_n)-cutoff for the sequence (p_n, E_n).

We can now state the first version of our main result.

Theorem 3.3 (L^p-cutoff, $1 < p < \infty$)**.** *Fix $p \in (1, \infty)$. Consider a family of spaces Ω_n indexed by $n = 1, 2, \dots$ For each n, let $p_n(t, \cdot, \cdot)$, $t \in T$, $T = [0, \infty)$ or $T = \mathbb{N}$, be a transition function on Ω_n with invariant probability π_n and spectral gap λ_n. For each n, let E_n be a set of probability measures on Ω_n and consider the supremum of the corresponding L^p distance to stationarity*

$$f_n(t) = \sup_{\mu \in E_n} D_p(\mu, t),$$

where $D_p(\mu, t)$ is defined at (3.5). Assume that each f_n tends to zero at infinity, fix $\epsilon > 0$ and consider the ϵ-L^p-mixing time

$$t_n = T_p(E_n, \epsilon) = T(f_n, \epsilon) = \inf\{t \in T : D_p(\mu, t) \le \epsilon, \forall \mu \in E_n\}.$$

1. *When $T = [0, \infty)$, assume that*
$$\lim_{n \to \infty} \lambda_n t_n = \infty. \tag{3.10}$$
Then the family of functions $\mathcal{F} = \{f_n, n = 1, \dots,\}$ presents a (t_n, λ_n^{-1}) cutoff.

2. *When $T = \mathbb{N}$, set $\gamma_n = \min\{1, \lambda_n\}$ and assume that*
$$\lim_{n \to \infty} \gamma_n t_n = \infty. \tag{3.11}$$
Then the family of functions $\mathcal{F} = \{f_n, n = 1, \dots,\}$ presents a (t_n, γ_n^{-1}) cutoff.

If $E_n = \{\mu\}$, we write $T_p(\mu, \epsilon)$ for $T_p(E_n, \epsilon)$. If $\mu = \delta_x$, we write $T_p(x, \epsilon)$ for $T_p(\delta_x, \epsilon)$. It is obvious that

$$T_p(E_n, \epsilon) = \sup_{\mu \in E_n} T_p(\mu, \epsilon).$$

In particular, if $\mathcal{M}_n$ is the set of probability measures on Ω_n, then

$$T_p(\mathcal{M}_n, \epsilon) = \sup_{x \in \Omega_n} T_p(x, \epsilon).$$

Proof of Theorem 3.3. Set $1/p + 1/q = 1$. Let $\mu_{n,t} = \mu_{n,0} P_{n,t}$. Fix $\epsilon > 0$ and set $t_n = T_p(E_n, \epsilon)$ and assume (as we may) that t_n is finite for n large enough. For $f \in L^q(\Omega_n, \pi_n)$ and $t = u + v$, $u, v > 0$, we have

$$[\mu_{n,t} - \pi_n](f) = [\mu_{n,u} - \pi_n][P_{n,v} - \pi_n](f).$$

Hence, using (3.4),

$$\begin{aligned} |[\mu_{n,t} - \pi_n](f)| &\leq D_p(\mu_{n,0}, u) \|[P_{n,v} - \pi_n](f)\|_q \\ &\leq D_p(\mu_{n,0}, u) 2^{|1-2/p|} e^{-v\lambda_n (1-|1-2/p|)} \|f\|_q. \end{aligned}$$

Taking the supremum over all f with $\|f\|_q = 1$ and over all $\mu_{n,0} \in E_n$ yields

$$f_n(u+v) \leq 2^{|1-2/p|} f_n(u) e^{-v\lambda_n(1-|1-2/p|)}.$$

Using this with either $u > t_n$, $v = \lambda_n^{-1} c$, $c > 0$, or $0 < u < t_n + \lambda_n^{-1} c$, $v = -\lambda_n^{-1} c$, $c < 0$, (the latter u can be taken positive for n large enough because, by hypothesis, $t_n \lambda_n$ tends to infinity), we obtain

$$\overline{F}(c) = \limsup_{n \to \infty} \sup_{t > t_n + c\lambda_n^{-1}} f_n(t) \leq \epsilon 2^{|1-2/p|} e^{-c(1-|1-2/p|)}, \quad c > 0,$$

and

$$\underline{F}(c) = \liminf_{n \to \infty} \inf_{t < t_n + c\lambda_n^{-1}} f_n(t) \geq \epsilon 2^{|1-2/p|} e^{-c(1-|1-2/p|)}, \quad c < 0.$$

This proves the desired cutoff. □

Remark 3.7. In Theorem 3.3 the stated sufficient condition, i.e., $\lambda_n t_n \to \infty$ (resp. $\gamma_n t_n \to \infty$) is also obviously necessary for a (t_n, λ_n^{-1}) cutoff (resp. a (t_n, γ_n^{-1}) cutoff). However, it is important to notice that these conditions are not necessary for the existence of a cutoff with cutoff time t_n and unspecified window. See Example 3.2 below.

Remark 3.8. The reason one needs to introduce $\gamma_n = \min\{1, \lambda_n\}$ in order to state the result in Theorem 3.3(2) is obvious. In discrete time, it makes little sense to talk about a window of width less than 1.

Remark 3.9. The conclusion of Theorem 3.3 (in both cases (1) and (2)) is false for $p = 1$, even under an additional self-adjoiness assumption. Whether or not it holds true for $p = \infty$ is an open question in general. It does hold true for $p = \infty$ when self-adjoiness is assumed.

The following result is an immediate corollary of Theorem 3.3. It indicates one of the most common ways Theorem 3.3 is applied to prove an L^2-cutoff.

Corollary 3.4 (L^2-cutoff)**.** *Consider a family of spaces Ω_n indexed by $n = 1, 2, \ldots$ For each n, let $p_n(t, \cdot, \cdot)$, $t \in T$, $T = [0, \infty)$ or $T = \mathbb{N}$, be a transition function on Ω_n with invariant probability π_n and spectral gap λ_n. Assume that there exists $c > 0$ such that, for each n, there exist $\phi_n \in L^2(\Omega_, \pi_n)$ and $x_n \in \Omega_n$ such that*

$$|(P_{n,t} - \pi_n)\phi_n(x_n)| \geq e^{-c\lambda_n t} |\phi_n(x_n)|.$$

1. *If $T = [0, \infty)$ and*

$$\lim_{n \to \infty} (|\phi_n(x_n)| / \|\phi_n\|_{L^2(\Omega_n, \pi_n)}) = \infty$$

then the family of functions $D_2(x_n, t)$ presents a cutoff.

2. If $T = \mathbb{N}$, $\sup_n \lambda_n < \infty$ and

$$\lim_{n\to\infty} (|\phi_n(x_n)|/\|\phi_n\|_{L^2(\Omega_n,\pi_n)}) = \infty$$

then the family of functions $D_2(x_n, t)$ presents a cutoff.

3.4 Some examples of cutoffs

This section illustrates Theorem 3.3 with several examples. We start with an example showing that the sufficient conditions of Theorem 3.3 are not necessary.

Example 3.2. For $n \geq 1$, let K_n be a Markov kernel on the finite set $\Omega_n = \{0,1\}^n$ defined by

$$K_n(x,y) = \begin{cases} 1/2 & \text{if } y_{i+1} = x_i \text{ for } 1 \leq i \leq n-1 \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

for all $x = x_n x_{n-1} \cdots x_1, y = y_n \cdots y_1 \in \Omega_n$. In other words, if we identify $(\mathbb{Z}_2)^n$ with $\mathbb{Z}_{2^n}$ by mapping $x = x_n \cdots x_1$ to $\sum_i x_i 2^{i-1}$, then $K_n(x,y) > 0$ if and only if $y = 2x$ or $y = 2x+1 (\text{mod } 2^n)$. For such a Markov kernel, let $p_n^d(t,\cdot,\cdot)$, $p_n^c(t,\cdot,\cdot)$, be, respectively, the discrete and continuous Markov transition functions defined at (3.1) and (3.2). Obviously, the unique invariant probability measure π_n for both Markov transition functions is uniform on Ω_n. It is worth noting that, for $n \geq 1$, $1 \leq p \leq \infty$ and $t \geq 0$, the L^p distance between $\delta_x P_{n,t}^c$(resp. $\delta_x P_{n,t}^d$) and π_n is independent of $x \in \Omega_n$. Hence we fix the starting point to be $\mathbf{0}$ (the string of n 0s) and set (with $* = d$ or c)

$$f_{n,p}^*(t) = \left(\sum_y |(p_n^*(t,\mathbf{0},y)/\pi_n(y)) - 1|^p \pi_n(y) \right)^{1/p}.$$

The following proposition shows that, for any $1 \leq p \leq \infty$, there are cutoffs in both discrete and continuous time with $t_n = T(f_n^*, 1/2)$ of order n and spectral gap bounded above by $1/n$. This provides examples with a cutoff even so $t_n\lambda_n$ (or $t_n\gamma_n$) stays bounded.

Proposition 3.5. *Referring to the example and notation introduced above, let λ_n^d and λ_n^c be respectively the spectral gaps of p_n^d and p_n^c. Then $\lambda_n^d = 0$ and $\lambda_n^c \leq 1/n$.*
Moreover, for any fixed p, $1 \leq p \leq \infty$, we have:

(i) The family $\{f_n^d\}$ has an optimal $(n, 1)$ cutoff. No strongly optimal cutoff exists.

(ii) The family $\{f_n^c\}$ has an $(t_n(p), b_n(p))$ cutoff, where

$$t_n(p) = \frac{(1-1/p)n\log 2}{1-2^{1/p-1}}, \quad b_n(p) = \log n, \quad \textit{for } 1 < p < \infty,$$

and

$$t_n(1) = n,\ b_n(1) = \sqrt{n},\ t_n(\infty) = (2\log 2)n,\ b_n(\infty) = 1.$$

For $p = 1, \infty$, these cutoffs are strongly optimal.

Proof. See the appendix. ∎

Example 3.3 (Riffle shuffle). The aim of this example is to point out that applying Theorem 3.3 requires some non-trivial information that is not always easy to obtain. For a precise definition of the riffle shuffle model, we refer the reader to [1; 8; 29]. Theorem 1.1 (due to Bayer and Diaconis) gives a sharp cutoff in total variation (i.e., L^1). The argument uses the explicit form of the distribution of the deck of cards after l riffle shuffles. It can be extended to prove an L^p cutoff for each $p \in [1, \infty]$ with cutoff time $(3/2)\log_2 n$ for each finite p and $2\log_2 n$ for $p = \infty$. See [6].

The question we want to address here is whether or not Theorem 3.3 easily yields these L^p cutoffs in the restricted range $p \in (1, \infty)$. The answer is no, at least, not easily. Indeed, to apply Theorem 3.3, we basically need two ingredients: (a) a lower bound on $T_{n,p}(\mu_{n,0}, \epsilon)$ for some fixed ϵ; (b) a lower bound on λ_n.

Here $\mu_{n,0}$ is the Dirac mass at the identity and we will omit all references to it. As $T_{n,p}(\epsilon) \geq T_{n,1}(\epsilon)$, a lower bound on $T_{n,p}(\epsilon)$ of order $\log n$ is easily obtained from elementary entropy consideration as in [1, (3.9)]. The difficulty is in obtaining a lower bound on λ_n. Note that $\lambda_n = -\log \beta_n$ where β_n is the second largest singular value of the riffle shuffle random walk. It is known that the riffle shuffle walk is diagonalizable with eigenvalue 2^{-i} but its singular values are not known. This problem amounts to study the walk corresponding to a riffle shuffle followed by the inverse of a riffle shuffle.

Example 3.4 (Top in at random). Recall that top in at random is the walk on the symmetric group corresponding to inserting the top card at a uniform random position. Simple elegant arguments (using either coupling or stationary time) can be used to prove a total variation (and a separation) cutoff at time $n \log n$. In particular, $T_{n,p}(\epsilon)$ is at least of order $n \log n$ for all $p \in [1, \infty]$. To prove a cutoff in L^p, it suffices to bound β_n, the second largest singular value of the walk, from above (note that this example is known to be diagonalizable with eigenvalues i/n but this is not what we need). Fortunately, β_n^2 is actually the second largest eigenvalue of the walk called random insertion which is bounded using comparison techniques in [11]. This shows that $\lambda_n = -\log \beta_n \geq c/n$. Hence top in at random presents a cutoff for all $p \in (1, \infty)$. The cutoff time is not known although one might be able to find it using the results in [10]. Note that Theorem 3.3 does not treat the case $p = \infty$.

Example 3.5 (Random transposition). In the celebrated random transposition walk, two positions i, j are chosen independently uniformly at random and the cards at these positions are switched (hence nothing changes with probability $1/n$). This example was first studied using representation theory in [13]. A simple argument (coupon collector problem) shows that $T_{n,1}(\epsilon)$ is at least of order $n \log n$ for $\epsilon > 0$ small enough. This example is reversible so that $\beta_n = e^{-\lambda_n}$ is the second largest eigenvalue. Representation theory easily yields all eigenvalues and $\beta_n = 1 - 2/n$ so that $\lambda_n \sim 2/n$. Hence, Theorem 3.3 yields a cutoff in L^p for $p \in (1, \infty)$. This is well known for $p \in [1, 2]$ with cutoff time $(1/2)n \log n$ (and also for $p = \infty$ with cutoff time $n \log n$) but the L^p cutoff for $p \in (2, \infty)$ is a new result. The cutoff time for $p \in (2, \infty)$ is not known!

Example 3.6 (Regular expander graphs). Expander graphs are graphs with very good "expansion properties". For simplicity, for any fixed k, say that a family (V_n, E_n) of finite non-oriented k-regular graphs is a family of expanders if (a) the cardinality $|V_n|$ of the vertex set V_n tends to infinity with n, (b) there exists $\epsilon > 0$ such that, for any n and any set $A \subset V_n$ of cardinality at most $|V_n|/2$, the number of edges between A and its complement is at least $\epsilon|A|$. Recall that the lazy simple random walk on a graph is the Markov chain which either stays put or jumps to a neighbor chosen uniformly at random, each with probability 1/2.

A simple entropy like argument shows that, for any family of k-regular graphs we have $T_{n,1}(\eta) \geq c_k \log V_n$ if $\eta > 0$ is small enough. The lazy walk on a regular graph is reversible with the uniform probability as reversible measure. Hence, $\beta_n = e^{-\lambda_n}$ is the second largest eigenvalue of the walk. By the celebrated Cheeger type inequality, the expansion property implies that $\beta_n \leq 1 - \epsilon^2/(64k^2)$. Hence λ_n is bounded below by a constant independent of n. This shows that Theorem 3.3 applies and gives a L^p cutoff for $p \in (1, \infty)$. The L^∞ cutoff follows by Theorem 5.4 below because these walks are reversible. This proves Theorem 1.5 of the introduction. Whether or not there is always a total variation (i.e., L^1) cutoff is an open problem.

Example 3.7 (Birth and death chains). In this example, for simplicity, we consider a single positive recurrent lazy birth and death chain on the non-negative integers $\mathbb{N}$ with invariant probability measure π. We will prove that, under minimal hypotheses, there is a cutoff for the family of functions $D_2(x,t)$ as x tends to infinity. Thus this result deals with a single Markov chain and consider what happens when the starting point is taken as the parameter. Our main hypothesis will be the existence of a non-trivial spectral gap (i.e., $\lambda > 0$), a hypothesis that can be cast as an explicit condition on the coefficients of the birth and death chain.

For each i, fix $p_i, q_i \in (0,1)$, $p_i + q_i = 1$ and consider the (lazy) birth and death chain with with transition probabilities $P(X_{l+1} = 0 | X_l = 0) = (1+q_0)/2$, $P(X_{l+1} = 1 | X_l = 0) = p_0/2$ and, for $i \geq 1$,

$$P(X_{l+1} = j | X_l = i) = \begin{cases} 1/2 & \text{if } j = i, \\ p_i/2 & \text{if } j = i+1, \\ q_i/2 & \text{if } j = i-1 \\ 0 & \text{otherwise.} \end{cases}$$

For $t \in \mathbb{N}$, $x, y \in \mathbb{N}$, let $p(t,x,y)$ be the probability of moving from x to y in t steps (this is our transition function). The associated process is an irreducible aperiodic reversible Markov chain with reversible measure (not necessary a probability measure)

$$\pi(0) = c, \quad \pi(i) = c \prod_0^i (p_i/q_{i+1}).$$

Let us assume that

$$\sum_i \prod_0^i (p_i/q_{i+1}) < \infty$$

and pick c such that π is a probability measure. Consider the self-adjoint Markov operator $P_1 : L^2(\mathbb{N}, \pi) \to L^2(\mathbb{N}, \pi)$. Because of the laziness that is built into the chain, the spectrum of P_1 is contained in $[0,1]$. Let us further make the hypothesis that

$$M = \sup_j \left\{ \left(\sum_{i<j} \frac{1}{\pi(i) p_i} \right) \left(\sum_{i \geq j} \pi(i) \right) \right\} < \infty. \tag{3.13}$$

This is our main hypothesis. It implies that there is a gap in the spectrum of P_1 between the eigenvalue 1 and the rest of the spectrum. In other words, the spectrum of $P_1 - \pi$ is contained in an interval of the form $[0, \mu]$ with

$$\|P_1 - \pi\|_{L^2(\mathbb{N},\pi) \to L^2(\mathbb{N},\pi)} = \mu < 1.$$

Obviously, this is equivalent to

$$\|P_t - \pi\|_{L^2(\mathbb{N},\pi)\to L^2(\mathbb{N},\pi)} = \mu^t = e^{-\lambda t} \text{ with } \lambda = -\log \mu > 0.$$

For an elegant proof of sharp spectral estimates of birth and death chains, see [20].

Because the underlying space is countable (i.e., for each x, the probability measure concentrated at x, δ_x, has density $\pi(x)^{-1}\mathbf{1}_{\{x\}}$ w.r.t. π), it follows that

$$\forall\, x \in \mathbb{N}, \;\; \forall\, t \in \mathbb{N}, \;\; D_2(x,t) \le \pi(x)^{-1/2} e^{-t\lambda}.$$

In particular, $D_2(x,t)$ tends to zero when t tends to infinity. Next, observe that

$$D_2(x,t) \ge c \text{ for } t < x$$

because, if $t < x$, then $p(t,x,0) = 0$ whereas $\pi(0) = c$. This implies $T_2(x, c/2) \ge x$. Applying Theorem 3.3(2), we find that the family of functions $\mathcal{F} = \{D_2(x,t) : x \in \mathbb{N}\}$ presents a cutoff (as x tends to infinity) with window 1 (the cutoff time is unknown and it would be extremely difficult to describe it in this generality).

This example can be generalized to allow the treatment of families of birth and death chains $p_n(t,x,y), \pi_n$ with starting points x_n that may vary or not. The simplest case occurs when exists $\epsilon > 0$ such that

$$\pi_n(\{0\}) > \epsilon \tag{3.14}$$

(i.e., the point 0 has minimal mass ϵ for all chains in the family). Assuming (3.14), there is a constant $C(\epsilon) \in (0,\infty)$ such that $1/(8M_n) \le 1 - \beta_n \le C(\epsilon)/M_n$ with M_n defined at (3.13) and we have the obvious mixing time lower bound $T_{n,2}(x_n, \epsilon/2) \ge x_n$. This implies a cutoff as long as x_n/M_n tends to infinity.

4 Normal ergodic Markov operators

Although it is customary in the subject to work with self-adjoint operators, there are many interesting examples that are normal but not self-adjoint. The simplest ones are non-symmetric random walks on abelian groups.

4.1 Max-L^p cutoffs

A bounded operator P on a Hilbert space H is normal if it commutes with its adjoint, i.e., $PP^* = P^*P$. The spectral theorem for normal operator implies that, to any continuous function f on the spectrum $\sigma(P)$ of P, one can associate in a natural way a normal operator $f(P)$ which satisfies

$$\|f(P)\|_{H\to H} = \max\{|f(s)| : s \in \sigma(P)\}.$$

An important special case is the case of self-adjoint operators where $P = P^*$.

The notion of normal operator is relevant to us here because, if $p(t,x,\cdot)$, $t \in T$, $x \in \Omega$, is a Markov transition function with invariant probability π such that, for each $t \in T \cap [0,1]$, $P_t : L^2(\Omega,\pi) \to L^2(\Omega,\pi)$ is normal, then the spectral gap defined at (3.3) satisfies

$$\|P_t - \pi\|_{L^2(\Omega,\pi)\to L^2(\Omega,\pi)} = e^{-t\lambda}. \tag{4.1}$$

First, observe that P_t preserves the space $L^2_0(\Omega,\pi)=\{f\in L^2(\Omega,\pi):\pi(f)=0\}$ and that

$$\|P_t-\pi\|_{L^2(\Omega,\pi)\to L^2(\Omega,\pi)}=\|P_t\|_{L^2_0(\Omega,\pi)\to L^2_0(\Omega,\pi)}.$$

Now, the case when $T=\mathbb{N}$ is clear since $P_t=(P_1)^t$ and P_1 is normal. When $T=[0,\infty)$, observe that, by the semigroup property, for any rational $a/b>0$, $P^b_{a/b}=P^a_1$. If we set

$$\|P_1\|_{L^2_0(\Omega,\pi)\to L^2_0(\Omega,\pi)}=e^{-\rho},$$

it follows that $\|P_{a/b}\|_{L^2_0(\Omega,\pi)\to L^2_0(\Omega,\pi)}=e^{-(a/b)\rho}$. As $t\mapsto\|P_t\|_{L^2_0(\Omega,\pi)\to L^2_0(\Omega,\pi)}$ is non-increasing, this implies $\|P_t\|_{L^2_0(\Omega,\pi)\to L^2_0(\Omega,\pi)}=e^{-t\rho}$ for all $t\ge 0$ and $\rho=\lambda$.
The following lemma is crucial for our purpose.

Lemma 4.1. *Consider a Markov transition function $p(t,x,\cdot)$, $t\in T$, $x\in\Omega$ with invariant probability measure π and spectral gap λ. Assume that for each $t\in T\cap[0,1]$, P_t is normal on $L^2(\Omega,\pi)$. Then, for any $r\in[1,\infty]$, there exists $\theta_r\in[1/2,1]$ such that*

$$\|P_t-\pi\|_{L^r(\Omega,\pi)\to L^r(\Omega,\pi)}\ge 2^{-1+\theta_r}e^{-\theta_r\lambda t}.$$

Proof. By the Riesz-Thorin interpolation theorem, we have

$$\|P_t-\pi\|^{\theta}_{L^r(\Omega,\pi)\to L^r(\Omega,\pi)}\|P_t-\pi\|^{1-\theta}_{L^\infty(\Omega,\pi)\to L^\infty(\Omega,\pi)}\ge\|P_t-\pi\|_{L^2(\Omega,\pi)\to L^2(\Omega,\pi)}$$

if $r\in[1,2]$ and $\theta=r/2$ and

$$\|P_t-\pi\|^{\theta}_{L^r(\Omega,\pi)\to L^r(\Omega,\pi)}\|P_t-\pi\|^{1-\theta}_{L^1(\Omega,\pi)\to L^1(\Omega,\pi)}\ge\|P_t-\pi\|_{L^2(\Omega,\pi)\to L^2(\Omega,\pi)}$$

if $r\in[2,\infty]$ and $\theta=r'/2$, $1/r+1/r'=1$. Since the operator norms on L^1 and L^∞ are bounded by 2 and since P_t, $t\in T\cap[0,1]$, is normal, this shows that for any $r\in(1,\infty)$, there exists $\theta_r\in[1/2,1]$ such that

$$\|P_t-\pi\|_{L^r(\Omega,\pi)\to L^r(\Omega,\pi)}\ge 2^{-1+\theta_r}e^{-\theta_r\lambda t}.$$

□

We can now state and prove our main theorems.

Theorem 4.2 (Max-L^p cutoff, continuous time normal case)**.** *Fix $p\in(1,\infty)$. Consider a family of spaces Ω_n indexed by $n=1,2,\dots$ For each n, let $p_n(t,\cdot,\cdot)$, $t\in[0,\infty)$, be a transition function on Ω_n with invariant probability π_n and with spectral gap λ_n. Assume that $P_{n,t}$ is normal on $L^2(\Omega_n,\pi_n)$, for each $t\in[0,1]$.*
Consider the Max-L^p distance to stationarity $f_n(t)=\overline{D}_{\Omega_n,p}(t)$ defined at (3.6) *and set*

$$\mathcal{F}=\{f_n:n=1,2,\dots\}.$$

Assume that each f_n tends to zero at infinity, fix $\epsilon>0$ and consider the ϵ-max-L^p-mixing time

$$t_n=T_{n,p}(\epsilon)=T(f_n,\epsilon)=\inf\{t>0:\overline{D}_{\Omega_n,p}(t)\le\epsilon\}.$$

The following properties are equivalent:

1. *$\lambda_n t_n$ tends to infinity;*

2. *The family $\mathcal{F}$ presents a precutoff;*

3. *The family $\mathcal{F}$ presents a cutoff;*

4. *The family $\mathcal{F}$ presents a (t_n, λ_n^{-1})-cutoff.*

Proof. It suffices to show that (2) implies (1). Fix $p \in (1, \infty)$, define q by $1 = 1/p + 1/q$ and observe that

$$\begin{aligned} \overline{D}_{\Omega_n,p}(t) &= \sup_{x \in \Omega_n} \{D_p(x,t)\} \geq \|P_{n,t} - \pi_n\|_{L^q(\Omega_n,\pi_n) \to L^\infty(\Omega_n,\pi_n)} \\ &\geq \|P_{n,t} - \pi_n\|_{L^q(\Omega_n,\pi_n) \to L^q(\Omega_n,\pi_n)} \geq 2^{-1+\theta_q} e^{-t\theta_q \lambda_n}, \end{aligned}$$

where we have used Lemma 4.1 to obtain the last inequality. Now, suppose that there is a precutoff at time s_n. Then there exists positive reals $a < b$ such that

$$\liminf_{n \to \infty} \overline{D}_{\Omega_n,p}(as_n) = 2\delta > 0.$$

and

$$0 = \limsup_{n \to \infty} \overline{D}_{\Omega_n,p}(bs_n) \geq 2^{-1+\theta_q} \limsup_{n \to \infty} e^{-bs_n \lambda_n \theta_q}$$

The first inequality implies $s_n = O(T_{n,p}(\delta))$ and the second one implies that $\lambda_n s_n$ tends to infinity. A fortiori, $\lambda_n T_{n,p}(\delta)$ tends to infinity. By Theorem 3.3, this proves the $(T_{n,p}(\delta), \lambda_n^{-1})$ cutoff and, by Corollary 2.5(ii), $\lambda_n t_n$ tends to infinity. □

Theorem 4.3 (Max-L^p cutoff, discrete time normal case)**.** *Fix $p \in (1, \infty)$. Consider a family of spaces Ω_n indexed by $n = 1, 2, \ldots$ For each n, let $p_n(t, \cdot, \cdot)$, $t \in \mathbb{N}$, be a transition function on Ω_n with invariant probability π_n and spectral gap λ_n. Assume that $P_{n,1}$ is normal on $L^2(\Omega_n, \pi_n)$. Consider the Max-L^p distance to stationarity $f_n(t) = \overline{D}_{\Omega_n,p}(t)$ defined at (3.6) and set*

$$\mathcal{F} = \{f_n : n = 1, 2, \ldots\}.$$

Assume that each f_n tends to zero at infinity, fix $\epsilon > 0$ and consider the ϵ-max-L^p-mixing time

$$t_n = T_{n,p}(\epsilon) = T(f_n, \epsilon) = \inf\{t > 0 : \overline{D}_{\Omega_n,p}(t) \leq \epsilon\}.$$

Assume further that $t_n \to \infty$. Setting $\gamma_n = \min\{1, \lambda_n\}$, the following properties are equivalent:

1. *$\gamma_n t_n$ tends to infinity;*

2. *The family $\mathcal{F}$ presents a precutoff;*

3. *The family $\mathcal{F}$ presents a cutoff;*

4. *The family $\mathcal{F}$ presents a (t_n, γ_n^{-1})-cutoff.*

Proof. The proof is similar to that of the continuous time case and is omitted. □

4.2 Examples of max-L^p cutoffs

This section describes a number of interesting situations where either Theorem 4.2 or Theorem 4.3 applies.

4.2.1 High multiplicity

Consider a family of spaces Ω_n indexed by $n = 1, 2, \dots$ For each n, let $p_n(t, \cdot, \cdot)$, $t \in T$, be a Markov transition function on Ω_n with invariant probability π_n and spectral gap λ_n. Assume that $P_{n,1}$ is normal on $L^2(\Omega_n, \pi_n)$ and that there is an eigenvalue ζ_n of modulus $|\zeta_n| = e^{-\lambda_n}$ with multiplicity at least m_n (i.e., the space of functions $\psi \in L^2(\Omega_n, \pi_n)$ such that $P_t\psi = \zeta_n^t \psi$, $t \in T$, is of dimension at least m_n). We claim that the following hold:

(1) If $T = (0, \infty)$ and m_n tends to infinity then there is a max-L^2 cutoff.

(2) If $T = \mathbb{N}$, $\sup_n \lambda_n < \infty$ and m_n tends to infinity then there is a max-L^2 cutoff.

We give the proof for the continuous time case (the discrete time case is similar). Let $\psi_{n,i}$, $i = 1, \dots, m_n$ be orthonormal eigenfunctions such that $P_{n,t}\psi_{n,i} = \zeta_n^t \psi_{n,i}$. For $x \in \Omega_n$, $\psi_n(x, y) = \sum_1^{m_n} \overline{\psi_{n,i}(x)} \psi_{n,i}(y)$. Observe that, for each $x \in \Omega_n$,

$$\|\psi_n(x, \cdot)\|^2_{L^2(\Omega_n,\pi_n)} = \sum_i |\psi_{n,i}(x)|^2 = \psi_n(x, x)$$

and $\max_x \psi_n(x, x) \geq \pi_n(\psi_n(x, x)) = m_n$. By hypothesis

$$\begin{aligned} D_{n,2}(x, t) &= \sup \left\{ |(P_{n,t} - \pi_n)f(x)| : f \in L^2(\Omega_n, \pi_n), \|f\|_{L^2(\Omega_n,\pi_n)} = 1 \right\} \\ &\geq |\zeta_n^t| \frac{|\psi_n(x, x)|}{\|\psi_n(x, \cdot)\|_{L^2(\Omega_n,\pi_n)}} = e^{-t\lambda_n} |\psi_n(x, x)|^{1/2}. \end{aligned}$$

It follows that

$$\overline{D}_{\Omega_n,2}(t) \geq e^{-t\lambda_n} m_n^{1/2}.$$

In particular, if $t_n = T_{n,2}(\epsilon) = \inf\{t > 0 : \overline{D}_{\Omega_n,2}(t) \leq \epsilon\}$, we get

$$e^{-2t_n\lambda_n} \leq \epsilon / m_n^{1/2}.$$

If m_n tends to infinity, this shows that $t_n\lambda_n$ tends to infinity and it follows from Theorem 4.2 that there is a max-L^2 cutoff.

This result can be extended using Corollary 3.4 as follows. If $x_n \in \Omega_n$ is such that $\psi_n(x_n, x_n)$ tends to infinity then there is an L^2 cutoff starting from x_n.

4.2.2 Brownian motion examples

Let (M_n, g_n) be a family of compact Riemannian manifolds (for simplicity, without boundary) where each manifold is equipped with its normalized Riemannian measure π_n. The heat semigroup $P_{n,t}$ on M_n is the Markov semigroup with infinitesimal generator the Laplace-Beltrami operator on (M_n, g_n). It corresponds to Brownian motion and has π_n as invariant measure. It is self-adjoint on $L^2(M_n, \pi_n)$ and ergodic. We denote by λ_n the spectral gap of (M_n, g_n) and set

$$T_{n,p} = T_{M_n,p}(\epsilon) = \inf\{t : \overline{D}_{M_n,p}(t) \leq \epsilon\}$$

for some fixed ϵ, e.g., $\epsilon = 1$. Here, if $h_n(t, x, y)$ denotes the heat kernel on (M_n, g_n) with respect to π_n, we have

$$\overline{D}_{M_n,p}(t) = \sup_{x \in M_n} \left(\int_{M_n} |h_n(t, x, y) - 1|^p \pi_n(dy) \right)^{1/p}.$$

Examples with fixed dimension

We first consider two different situations where all the manifolds have the same dimension d. For details and further references concerning background material, we refer to [25].

Example 4.1 (Non-negative Ricci curvature). Consider the case where all the manifold M_n have non-negative Ricci curvature. Let δ_n be the diameter of (M_n, g_n). In this case, well-known spectral estimates show that there are constants $c(d), C(d)$ such that

$$c(d)\delta_n^{-2} \leq \lambda_n \leq C(d)\delta_n^{-2}.$$

Moreover, [25] shows that there are constants $a(d), A(d)$ such that

$$a(d)\delta_n^{-2} \leq T_{n,p} \leq A(d)\delta_n^{-2}.$$

By Theorem 4.2, there is no max-L^p precutoff, $1 < p < \infty$. In fact, there is no max-L^1 precutoff either. See [25, Theorem 5]. This proves Theorem 1.2(1).

Example 4.2 (Compact coverings). Consider a fixed compact manifold (N, g) with non-compact universal cover $\widetilde{N}$ and fundamental group $\Gamma = \pi_1(N)$. Assume that Γ admits a countable family Γ_n of subgroups and, for each n, consider the manifold $M_n = \widetilde{N}/\Gamma_n$, equipped with the Riemannian structure g_n induced by g. Again, let δ_n be the diameter of (M_n, g_n). Now, Theorem 4.2 offers the following dichotomy: either (a) $\lambda_n T_{n,2}$ tends to infinity and there is a $(T_{n,2}, \lambda_n^{-1})$ max-L^2 cutoff, or (b) $\lambda_n T_{n,2}$ does not tend to infinity and there is no max-L^2 precutoff. The result of [25, Theorem 3] relates this to properties of Γ as follows. If Γ is a group of polynomial volume growth (for instance, Γ is nilpotent) then we must be in case (a). If Γ has Kazhdan's property (T) (for instance $\Gamma = \mathrm{SL}(2, \mathbb{R}^n)$, $n > 2$) then we must be in case (b).

Examples with varying dimension

The unit spheres $\mathbb{S}_n$ (in $\mathbb{R}^{n+1}$) provides one of the most obvious natural family of compact manifolds with increasing dimension. Theorem 1.2(2) describes the Brownian motion cutoff on spheres. Details are in [26]. The infinite families of classical simple compact Lie groups, $\mathrm{SO}(n)$, $\mathrm{SU}(n)$, $\mathrm{Sp}(n)$ yield natural examples of families of Riemannian manifolds with increasing dimensions (the dimension of each of these groups is of order n^2).

Example 4.3 (Classical simple compact Lie groups). On each simple compact Lie group G, there is, up to time change, a unique bi-invariant Brownian motion which corresponds to the canonical bi-invariant Riemannian metric obtained by considering on the Lie algebra $\mathfrak{g}$ of G

the bilinear form $B(X,Y) = -\mathrm{trace}(\mathrm{ad}X\mathrm{ad}Y)$ where $\mathrm{ad}X$ is the linear map defined on $\mathfrak{g}$ by $Z \mapsto \mathrm{ad}X(Z) = [X,Z]$ (on compact simple Lie groups, this bilinear form – equals to minus the Killing form — is positive definite). In what follows, we consider that each simple compact Lie group is equipped with this canonical Riemannian structure. We let $d(G)$ and $\lambda(G)$ be the dimension and spectral gap of G. See [26; 27] for further relevant details. It is well-known that there exist constants c_1, c_2 (one can take $c_1 = 1/4$, $c_2 = 1$) such that

$$c_1 \le \lambda(G) \le c_2.$$

Moreover, [27, Theorem 1.1] shows that, for any $\epsilon \in (0,2)$, there exist $c_3 = c_3(\epsilon) > 0$ such that, for any $1 \le p \le \infty$,

$$T_{G,p}(\epsilon) \ge c_3 \log d(G).$$

From this and Theorem 4.2 we deduce that, for any fixed $p \in (1,\infty)$ and for any sequence (G_n) of simple compact Lie groups, Brownian motion on G_n presents a max-L^p cutoff if and only if the dimension $d(G_n)$ of G_n tends to infinity. This proves Theorem 1.2(3) in the case $p \neq \infty$. The case $p = \infty$ follows from Theorem 5.3.

4.2.3 Random walks on the symmetric and alternating groups

For further background on random walks on finite groups, see [8; 29]. Let G be a finite group and let u be the uniform probability measure on G. The (left-invariant) random walk driven by a given probability measure v on G is the discrete time Markov process with transition function

$$p^d(l,x,A) = v^{(l)}(x^{-1}A)$$

where $v^{(l)}$ is the l-th convolution power of v by itself. A walk is irreducible if the support of v generates G. It is aperiodic if the support v is not contained in any coset of a proper normal subgroup of G. If the walk driven by v is irreducible and aperiodic then its limiting distribution as l tends to infinity is u. The adjoint walk is driven by the $\check{v}$ where $\check{v}(A) = v(A^{-1})$. Hence, the walk is normal if and only if $\check{v} * v = v * \check{v}$ where $*$ denotes convolution. To ease the comparison with the literature, when the walk is normal, let us denote by μ the second largest singular value of the operator of convolution by v on $L^2(G,u)$. If λ^d is the spectral gap as defined at (3.3) for the discrete time transition function $p^d(l,x,\cdot) = v^{(l)}(x^{-1}\cdot)$ then

$$\mu = e^{-\lambda^d}.$$

The following theorem shows that most families of normal walks on the symmetric group S_n or the alternating group A_n have a max-L^2 cutoff. For clarity, recall that, in the present context, the notion of max-L^2 cutoff is based on the chi-square distance function

$$\overline{D}_{G,2}(l) = \left(|G|^{-1} \sum_{y \in G} [|G| v^{(l)}(y) - 1]^2 \right)^{1/2}.$$

Theorem 4.4. *Let $G_n = S_n$ or A_n. For each n, let v_n be a probability measure on G_n such that the associated (left-invariant) random walk is irreducible, aperiodic and normal. Let $\beta_n = e^{-\lambda_n^d}$ be the corresponding second largest singular value and assume that $\inf_n \beta_n > 0$.*

- *If $G_n = A_n$, $n = 1, 2, \ldots$, then the family of random walks driven by these v_n presents a max-L^2 cutoff with window size $(1-\beta_n)^{-1}$.*
- *If $G = S_n$, set*

$$\sigma_n = \sum_{x \in S_n} \operatorname{sgn}(x) v_n(x).$$

If $|\sigma_n| < \beta_n$ then the family of random walks driven by these v_n presents a max-L^2 cutoff with window size $(1-\beta_n)^{-1}$.

Proof. Let $T_{n,2} = \inf\{l > 0 : f_n(l) \leq 1\}$ be the L^2-mixing time. Observe that $\inf_n \beta_n > 0$ implies $\min\{-\log \beta_n, 1\} \asymp (1-\beta_n) \asymp \lambda_n^d$ where $\asymp$ is used to indicate that the ratio of the two sides is uniformly bounded away form 0 and infinity. To prove that there is a $(T_{n,2}, (1-\beta_n)^{-1})$-max-$L^2$ cutoff, it suffices by Theorem 4.3 to show that $\lambda_n T_{n,2}$ tends to infinity. Let ρ_n be an irreducible representation of G_n at which the singular value β_n is attained. Let $d(\rho_n)$ be the dimension of the representation ρ_n. Then, since each irreducible representation appears with multiplicity equal to its dimension, the Plancherel formula yields

$$\overline{D}^2_{G_n,2}(l) \geq d(\rho_n)\beta_n^{2l}.$$

Hence we obtain

$$\lambda_n^d T_{n,2} = (-\log \beta_n) T_{n,2} \geq (1/2) \log d(\rho_n). \tag{4.2}$$

By Theorem 4.3, the family has a $(T_{n,2}, \lambda_n^{-1})$ max-L^2 cutoff if $d(\rho_n)$ tends to infinity. Now, for $n \geq 5$, the irreducible representations of S_n all have dimensions at least $n-1$ except the trivial and sign representations which both have dimension 1. The irreducible representations of A_n are obtained in a simple way by restriction of the irreducible representations of S_n and it follows that the non-trivial irreducible representations of A_n have dimension at least $(n-1)/2$. If $G = A_n$ this together with Theorem 4.3 and (4.2) proves that there is a max-L^2 cutoff as desired. If $G = S_n$, in order to obtain the announced max-L^2 cutoff, it suffices to rule out the possibility that $\beta_n = |\sigma_n|$. $\square$

Remark 4.1. The proof of Theorem 4.4 shows that, in the case of S_n, we can replace the hypothesis that $|\sigma_n| < \beta_n$ by weaker hypothesis, for instance, that the singular value β_n is attained at least by one representation different from the sign representation. The celebrated random transposition walk is such an example: it has $\beta_n = -1 + 2/n$ and $\beta_n = 1 - 2/n$ so that $|\sigma_n| = \beta_n$ but β_n is also attained at the natural dimension $n-1$ representation.

Remark 4.2. Instead of the discrete time random walk associated to v_n, consider the associated continuous time random walk with transition function

$$p_n^c(t, x, A) = e^{-t} \sum_0^\infty \frac{t^l}{l!} v_n^{(l)}(x^{-1}A).$$

Let λ_n^c be the spectral gap of this transition function. Dropping the irrelevant hypothesis $\inf \beta_n > 0$ in Theorem 4.4, there is always a max-L^2 cutoff in continuous time on A_n and there is one on S_n if $\beta_n < 1 - \lambda_n^c$.

Remark 4.3. There are many explicit random walks on the symmetric group for which the max-L^2 cutoff time is not known explicitly. One celebrated example is the adjacent transposition random walk where v is the uniform probability on the generating set $E = \{\mathrm{Id}, (i, i+1), i = 1, \ldots, n-1\}$. Many other examples, (e.g., random insertions, random reversals, ...) are described in [29].

Remark 4.4. The proof of theorem 1.4 is the same as that of the case $G = A_n$ in Theorem 4.4: It is well known that the irreducible representations of $\mathrm{SO}(n)$ have dimension at least n.

5 Comparison of max-L^p mixing times and cutoffs when p varies

The aim of this section is to show that, for some families of Markov transition functions, the existence of a max-L^p cutoff is a property that is independent of p for $1 < p \le \infty$. Two cases are of special interests, namely, the reversible case and the normal transitive case (all operators are normal and invariant under a transitive group action). For this purpose we present comparison results for L^p-mixing times of Markov transition functions.

5.1 Good adjoint

Let $p(t, x, \cdot)$, $t \in T, x \in \Omega$ be a Markov transition function with invariant measure π. The adjoint of P_t on $L^2(\Omega, \pi)$ is given formally by

$$p^*(t, x, A) = \frac{\int_A p(t, z, dx)\pi(dz)}{\pi(dx)},$$

i.e., the Radon-Nikodym derivative of the measure $B \mapsto \nu_A(B) = \int_A p(t, z, B)\pi(dz)$ with respect to π. Observe that $\nu_\Omega(B) = \int_\Omega p(t, z, B)\pi(dz) = \pi(B)$ so that ν_A is absolutely continuous w.r.t. π. However this defines $p^*(t, x, A)$ only for π almost all x. Thus we consider the following technical hypothesis (very often satisfied in practice). We say that a Markov transition function $p(t, x, \cdot)$, $t \in T$, $x \in \Omega$, with invariant measure π *admits a good adjoint with respect to* π if there is a Markov transition function $p^*(t, x, \cdot)$, $t \in T, x \in \Omega$, such that the adjoint P_t^* of P_t on $L^2(\Omega, \pi)$ is the Markov operator associated to the Markov transition function $p^*(t, x, \cdot)$. This allows us to act with P_t^* on measures by $\mu P_t^*(A) = \int_\Omega p^*(t, x, A)\mu(dx)$. In particular, πP_t^* is well defined and we have

$$\pi P_t^*(A) = \int \mathbf{1}_\Omega(x) P_t^* \mathbf{1}_A(x)\pi(dx) = \int P_t \mathbf{1}_\Omega(x)\mathbf{1}_A(x)\pi(dx) = \pi(A).$$

That is, π is an invariant measure for the transition function p^*.

There are at least three important instances when this property is automatically satisfied. These instances are described below.

The first case is when Ω is a topological space and the Markov transition function $p(t, x, \cdot)$ is of the form

$$p(t, x, dy) = h(t, x, y)\pi(dy), \;\; h(t, x, y) \in \mathcal{C}(\Omega \times \Omega, \pi \otimes \pi), \; 0 < t \in T.$$

In this case,

$$p^*(t, x, dy) = h(t, y, x)\pi(dy).$$

The second case is when P_t, $t \in T$, is self-adjoint on $L^2(\Omega, \pi)$, in which case $P_t^* = P_t$ and

$$p^*(t, x, dy) = p(t, x, dy).$$

The third case is when there is a compact group acting continuously transitively on Ω and preserving the transition function. This case will be discussed in detail below. The simplest instance is when $\Omega = G$ is a compact group and $p(t, x, A) = p(t, e, x^{-1}A)$, $t \in T$, $x \in G$. Then the (normalized) Haar measure is an invariant measure and the adjoint P_t^* of P_t has transition function

$$p^*(t, x, A) = p(t, e, A^{-1}x).$$

5.2 Mixing time comparisons

Let $p(t, x, \cdot)$ be a Markov transition function on Ω with invariant probability measure π. This section is devoted to comparisons between the various max-L^p mixing times, $1 \leq p \leq \infty$. Hence, we set

$$T_p(\epsilon) = \inf\{t \in T : \overline{D}_p(t) \leq \epsilon\}$$

with $\overline{D}_p$ defined at (3.6). We will also need to use

$$\widetilde{T}_p(\epsilon) = \inf\{t \in T : \widetilde{D}_{\pi,p}(t) \leq \epsilon\}$$

The advantage of $\widetilde{T}$ over T is that, since it requires only $P_t f(x)$ be defined for π almost all x, it is well-defined for the adjoint $p^*(t, x, A)$ without the technical hypothesis that a good adjoint exists. In particular, we set

$$\widetilde{T}_p^*(\epsilon) = \inf\{t \in T : \widetilde{D}_{\pi,p}^*(t) \leq \epsilon\}$$

where

$$\widetilde{D}_{\pi,p}^*(t) = \pi\text{-}\operatorname*{ess\,sup}_{x \in \Omega} \left(\int_\Omega |h^*(t, x, y) - 1|^p \pi(dy) \right)^{1/p}$$

if $p^*(t, x, \cdot) = h^*(t, x, y)\pi(dy)$, π-almost surely, and $\widetilde{D}_{\pi,p}^*(t) = \infty$ otherwise. In terms of operator norms, setting $1/p + 1/q = 1$, we have

$$\widetilde{D}_{\pi,p}^*(t) = \|P_t^*\|_{L^q(\Omega,\pi) \to L^\infty(\Omega,\pi)} = \|P_t\|_{L^1(\Omega,\pi) \to L^p(\Omega,\pi)}.$$

In cases when we assume that there is a good adjoint, we set

$$T_p^*(\epsilon) = \inf\{t \in T : \overline{D}_p^*(t) \leq \epsilon\}$$

with $\overline{D}_p^*$ is defined at (3.6) but with $p(t, x, \cdot)$ replaced by $p^*(t, x, \cdot)$.

Our main mixing time comparisons are stated in the following proposition.

Proposition 5.1. *Let $p(t, x, \cdot)$ be a Markov transition function with invariant measure π. Referring to the mixing times introduced above, the following inequalities hold:*

1. *For $1 \leq p \leq q \leq \infty$ and any fixed $\epsilon > 0$, $T_p(\epsilon) \leq T_q(\epsilon)$ and $\widetilde{T}_p(\epsilon) \leq \widetilde{T}_q(\epsilon)$.*

2. *For $1 \le q, r, s, \le \infty$ with $1 + 1/q = 1/r + 1/s$ and any $\epsilon, \eta, \delta > 0$,*

$$T_q(\epsilon^{s/q}\eta^{1-s/q}\delta) \le \max\{\mathbf{1}_{[1,\infty)}(q)\widetilde{T}_s(\epsilon), \mathbf{1}_{(1,\infty]}(q)\widetilde{T}^*_s(\eta)\} + T_r(\delta).$$

3. *For $1 \le p \le \infty$, $1/p + 1/p' = 1$ and $\epsilon > 0$,*

$$T_\infty(\epsilon^2) \le T_p(\epsilon) + \widetilde{T}^*_{p'}(\epsilon).$$

4. *For $1 < p < q \le \infty$ and $\epsilon > 0$,*

$$T_q(\epsilon^{m_{p,q}}) \le m_{p,q} \max\left\{T_p(\epsilon), \widetilde{T}^*_p(\epsilon)\right\}$$

where $m_{p,q} = \lceil p'/q' \rceil$, $1/p + 1/p' = 1$, $1/q + 1/q' = 1$.

Proof. The inequalities stated in (1) follow readily from Jensen's inequality. In order to prove (2), let q', r', s' the dual exponents of q, r, s (e.g., $1/q + 1/q' = 1$). Observe that

$$\overline{D}_q(u+v) \le \|P_u - \pi\|_{L^{q'}(\Omega,\pi)\to L^{r'}(\Omega,\pi)}\overline{D}_r(v).$$

Next, note that $1/q' = 1/r' + 1/s'$ and recall the interpolation inequality

$$\begin{aligned} \|P_u - \pi\|_{L^{q'}(\Omega,\pi)\to L^{r'}(\Omega,\pi)} &\le \|P_u - \pi\|^{s/q}_{L^{s'}(\Omega,\pi)\to L^\infty(\Omega,\pi)}\|P_u - \pi\|^{1-s/q}_{L^1(\Omega,\pi)\to L^s(\Omega,\pi)} \\ &\le (\widetilde{D}_{\pi,s}(u))^{s/q}(\widetilde{D}^*_{\pi,s}(u))^{1-s/q}. \end{aligned}$$

From this we deduce that

$$\overline{D}_q(u+v) \le (\widetilde{D}_{\pi,s}(u))^{s/q}(\widetilde{D}^*_{\pi,s}(u))^{1-s/q}\overline{D}_r(v).$$

The mixing time inequality stated in (2) follows. The inequality in (3) follows from (2) with $q = \infty$, $r = p$, $s = p'$ $\eta = \delta = \epsilon$.
To prove (4), set $1/p_j = (1 - 1/p)j + 1/q$, $j = 0, 1, 2 \dots$. As

$$1 + 1/p_i = 1/p_{i+1} + 1/p,$$

the result in (2) yields

$$T_{p_i}(\epsilon\eta) \le \max\{\widetilde{T}_p(\epsilon), \widetilde{T}^*_p(\epsilon)\} + T_{p_{i+1}}(\eta).$$

Hence, for each $i = 1, 2, \dots,$

$$T_q(\epsilon^{i+1}) \le i \max\{\widetilde{T}_p(\epsilon), \widetilde{T}^*_p(\epsilon)\} + T_{p_i}(\epsilon).$$

Now, $p_j \le p$ if and only if $j + 1 \ge p(q-1)/(q(p-1)) = p'/q'$. By (1), it follows that

$$\begin{aligned} T_q(\epsilon^{m_{p,q}}) &\le (m_{p,q} - 1)\max\{\widetilde{T}_p(\epsilon), \widetilde{T}^*_p(\epsilon)\} + T_p(\epsilon) \\ &\le m_{p,q}\max\{T_p(\epsilon), \widetilde{T}^*_p(\epsilon)\} \end{aligned}$$

□

Remark 5.1. Proposition 5.1(1) and Theorem 4.2 show that, for normal transition functions with $T = [0, \infty)$, a max-L^p cutoff always implies a max-L^q cutoff for $1 \le p \le q < \infty$. In the discrete time case $T = \mathbb{N}$, the same conclusion holds assuming $T_{n,p}(\epsilon)$ tends to infinity for ϵ small enough.

Remark 5.2. In the reversible case, Proposition 5.1 shows that the max-L^p mixing time controls the max-L^q mixing time if $1 < p \le q \le \infty$. See the next section. The interesting examples studied in [21] show that there can be no such control when $p = 1$.

5.3 Max-L^p cutoffs: the reversible case

Let $p(t,x,\cdot)$ be a Markov transition function with invariant distribution π. Assume that for each $t \in T$, P_t is self-adjoint on $L^2(\Omega,\pi)$ (hence, has a good adjoint!). Then, obviously, we have

$$D_p(x,t) = D_p^*(x,t), \;\; \overline{D}_p(t) = \overline{D}_p^*(t), \;\; T_p(\epsilon) = T_p^*(\epsilon).$$

Using the notation introduced in Proposition 5.1, it follows that for $1 < p \le q \le \infty$,

$$T_p(\epsilon) \le T_q(\epsilon) \le m_{p,q} T_p(\epsilon^{m_{p,q}}).$$

Furthermore, when $T = [0,\infty)$, we have

$$\overline{D}_\infty(t) = [\overline{D}_2(t/2)]^2.$$

When $T = \mathbb{N}$, if t is even,

$$\overline{D}_\infty(t) = [\overline{D}_2(t/2)]^2.$$

whereas, if t is odd,

$$\overline{D}_\infty(t) \le \overline{D}_2((t-1)/2)\overline{D}_2((t+1)/2) \le [\overline{D}_2((t+1)/2)]^2.$$

Together with Theorems 4.2-4.3, this yields the following statements.

Lemma 5.2. *Consider a family of spaces Ω_n indexed by $n = 1,2,\dots$ For each n, let $p_n(t,\cdot,\cdot)$, $t \in T$, be a transition function on Ω_n with invariant probability π_n. Assume that $P_{n,t}$ is reversible on $L^2(\Omega_n,\pi_n)$, for each $t \in T$. The following properties holds:*

1. *For each n, the function $t \mapsto \overline{D}_{\Omega_n,p}(t)$ tends to zero at infinity for some $p \in (1,\infty]$ if and only if it tends to zero for any $p \in (1,\infty]$.*
2. *Assume that for each n, the functions in* (1) *tend to zero. Then $T_{n,p}(\epsilon)$ tends to infinity with n for some $p \in (1,\infty]$ and $\epsilon > 0$ if and only if it does for all such p and ϵ.*

Theorem 5.3 (Max-L^p cutoff, continuous time reversible case)**.** *Consider a family of spaces Ω_n indexed by $n = 1,2,\dots$ For each n, let $p_n(t,\cdot,\cdot)$, $t \in [0,\infty)$, be a transition function on Ω_n with invariant probability π_n and with spectral gap λ_n. Assume that $P_{n,t}$ is reversible on $L^2(\Omega_n,\pi_n)$, for each $t \in (0,\infty)$.*
Assume that for each n and $p \in (1,\infty]$, the function $t \mapsto \overline{D}_{\Omega_n,p}(t)$ tends to zero at infinity. Then the following properties are equivalent:

1. *For some $p \in (1,\infty]$ and some $\epsilon > 0$, $\lambda_n T_{n,p}(\epsilon)$ tends to infinity;*
2. *For any $p \in (1,\infty]$ and any $\epsilon > 0$, $\lambda_n T_{n,p}(\epsilon)$ tends to infinity;*
3. *For some $p \in (1,\infty]$ there is a max-L^p precutoff.*
4. *For any $p \in (1,\infty]$ there is a max-L^p cutoff.*
5. *For any $p \in (1,\infty]$ and any $\epsilon > 0$, there is a $(T_{n,p}(\epsilon), \lambda_n^{-1})$ max-L^p cutoff.*

Finally, for any $\epsilon > 0$, $T_{n,\infty}(\epsilon^2) = 2T_{n,2}(\epsilon)$, and there is a t_n (resp. (t_n, b_n)) max-L^2 cutoff if and only if there is a $2t_n$ (resp. $(2t_n, b_n)$) max-L^∞ cutoff.

Theorem 5.4 (Max-L^p cutoff, discrete time reversible case)**.** *Consider a family of spaces Ω_n indexed by $n = 1, 2, \ldots$ For each n, let $p_n(t, \cdot, \cdot)$, $t \in \mathbb{N}$, be a transition function on Ω_n with invariant probability π_n and with spectral gap λ_n. Assume that $P_{n,1}$ is reversible on $L^2(\Omega_n, \pi_n)$. Assume that for each n and some $p \in (1, \infty]$, the function $t \mapsto \overline{D}_{\Omega_n,p}(t)$ tends to zero at infinity. Assume further that, for some $\epsilon > 0$ and $p \in (1, \infty]$, $T_{n,p}(\epsilon)$ tends to infinity with n. Then, setting $\gamma_n = \min\{1, \lambda_n\}$, the following properties are equivalent:*

1. *For some $p \in (1, \infty]$ and some $\epsilon > 0$, $\gamma_n T_{n,p}(\epsilon)$ tends to infinity;*
2. *For any $p \in (1, \infty]$ and any $\epsilon > 0$, $\gamma_n T_{n,p}(\epsilon)$ tends to infinity;*
3. *For some $p \in (1, \infty]$ there is max-L^p precutoff.*
4. *For any $p \in (1, \infty]$ there is max-L^p cutoff.*
5. *For any $p \in (1, \infty]$ and any $\epsilon > 0$, there is max-L^p $(T_{n,p}(\epsilon), \gamma_n^{-1})$ max-L^p cutoff.*

Finally, there is a t_n (resp. (t_n, b_n)) max-L^2 cutoff if and only if there is a $2t_n$ (resp. $(2t_n, b_n)$) max-L^∞ cutoff.

Remark 5.3. The precise relation between max-L^2 and max-L^∞ cutoffs stated in Theorems 5.3 and 5.4 is specific to maximum cutoffs and to the reversible case. We do not know how to treat max-L^∞ cutoffs in the normal non-reversible case. We now show that the precise relation between L^2 and L^∞ cutoff times does not hold in general for reversible chains in the case of a fixed starting distribution. The simplest example may be the lazy birth and death chain on $\{0, \ldots, n\}$ with $p + q = 1$, $p > 1/2$, and

$$P(X_{t+1} = y | X_t = x) = \begin{cases} p/2 & \text{if } y = x+1, x = 0, \ldots, n-1 \\ q/2 & \text{if } y = x+1, x = 1, \ldots, n \\ 1/2 & \text{if } y = x, x = 1, \ldots, n-1 \\ (1+q)/2 & \text{if } y = x = 0 \\ (1+p)/2 & \text{if } y = x = n \end{cases}$$

This chain is reversible with reversible measure $\pi(x) = c(p/q)^x$. It can be diagonalized explicitly (see [15, Pages 436-438]). In any case, it is easy to see that the spectral gap λ_n is bounded above and below by positive constants independent of n. It is also clear that, starting from 0, the L^r mixing time is of order n, for any $r \in [1, \infty]$ (this is because most of the mass of the stationary distribution is near n). Using Theorem 5.4 we deduce that there is a max-L^r cutoff, for any $r \in (1, \infty]$. Moreover the max-L^∞ cutoff time is twice the max-L^2 cutoff time. Now, consider the chain started at n. Then, for any $r \in [1, \infty)$ and $\epsilon \in (0, 1)$, one easily shows that the mixing time $T_r(n, \epsilon)$ is bounded above independently of n (use $D_r(n, t) \le e^{-c_r \lambda_n t} \pi_n(n)^{-1+1/r}$, $r \in (1, \infty)$, and, for $r = 1$, remember that $D_1(n, t) \le D_2(n, t)$). However, $D_\infty(n, n-1) \ge 1$ because, starting from n, we cannot reach 0 in less than n steps. Hence $T_\infty(n, 1/2) \ge n$. In particular, $T_\infty(n, 1/2)$ is much bigger than $2T_2(n, 1/2)$.

5.4 Max-L^p cutoffs for birth and death chains

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

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

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

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

with $c = \nu(0)$ a normalizing constant. Birth and death chains are in fact reversible (i.e., satisfy $\nu(x)K(x, y) = \nu(y)K(y, x)$), hence diagonalizable with real eigenvalues in $[-1, 1]$. Let α_i, $i = 0, \dots, m$, be the eigenvalues of $I - K$ in non-decreasing order (I denotes the identity operator). Thus $\alpha_0 = 0 < \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_m \leq 2$. The irreducibility of the chain is reflected in the fact that $\alpha_1 > 0$. It is also well known that $\alpha_m = 2$ if and only if the chain is periodic (of period 2) which happens if and only if $r_x = 0$ for all x. In fact, because we are dealing here with irreducible birth and death chains, it is known that the α_i's are all distinct (e.g., [5; 19]). Karlin and McGregor [18; 17] observed that the spectral analysis of any given birth and death chain can be treated as an orthogonal polynomial problem. This sometimes leads to the exact computation of the spectrum. See, e.g., [16; 18; 17; 31].

Given a finite birth and death chain as above, we consider the discrete and continuous time transition functions $p^*(t, x, y)$, $* = d, c$, defined at (3.1)-(3.2). Moreover, in what follows we consider families of birth and death chains indexed by n with $\Omega_n = \{0, \dots, m_n\}$. For any $1 \leq p \leq \infty$, we set

$$f^*_{n,p}(t) = D^*_{\Omega_n,p}(0, t) = \left(\sum_y |(p^*_n(t, 0, y)/\nu_n(y)) - 1|^p \nu_n(y) \right)^{1/p}$$

and

$$f^*_{n,\mathrm{sep}}(t) = \max_y \{1 - (p^*_n(t, 0, y)/\nu_n(y))\}.$$

The following theorem is proved in [12]. It gives a spectral necessary and sufficient condition for a family of birth and death chains (all started at 0) to have a cutoff in separation.

Theorem 5.5 ([12, Theorems 5.1–5.2])**.** *Referring to a sequence of birth and death chains as described above, set*

$$\chi_n = \alpha_{n,1}, \quad \theta_n = \sum_{i=1}^{m_n} \alpha_{n,i}^{-1}.$$

1. *The family $\{f^c_{n,\mathrm{sep}}\}$ has a cutoff if and only if $\chi_n \theta_n$ tends to infinity.*
2. *Assume that $p_{n,x} + q_{n,x+1} \leq 1$ for all n and all $x \in \{0, \dots, m_n - 1\}$. Then the family $\{f^d_{n,\mathrm{sep}}\}$ has a cutoff if and only if $\chi_n \theta_n$ tends to infinity.*

In both cases, whenever a cutoff occurs, it occurs at time θ_n.

Using this result and Theorems 5.3-5.4, we will prove the following statement.

Theorem 5.6. *Referring to a sequence of birth and death chains as described above, assume that $\chi_n\theta_n$ tends to infinity. Then, for any $1 < p \leq \infty$,*

1. *the family has a max-L^p cutoff in continuous time;*
2. *assuming that $p_{n,x} + q_{n,x+1} \leq 1$ for all n and all $x \in \{0, \ldots, m_n - 1\}$, the family has a max-$L^p$ cutoff in discrete time.*

Proof. Consider the continuous time case. Then, because of reversibility, we have $\lambda_n = \chi_n = \alpha_{n,1}$. By Theorem 5.3, for reversible chains the max-L^p cutoffs, $1 < p \leq \infty$, are all equivalent. Let $T_{n,\infty}(0, \epsilon)$ be the ϵ-L^∞-mixing time starting from 0, with ϵ fixed small enough, say $\epsilon = 1/4$. Assume that $\chi_n\theta_n$ tends to infinity. By Theorem 5.3, to prove the desired L^∞-cutoff it suffices to prove that $\chi_n T_{n,\infty}(1/4)$ tends to infinity. By Theorem 5.5 (θ_n) is a separation cutoff sequence. Thus, there are constants $0 < a < 1 < b < \infty$ and a sequence θ'_n such that

$$a\theta_n < \theta'_n < b\theta_n$$

and

$$\mathrm{sep}_n(\theta'_n) = \max_y \{1 - (p_n^c(\theta'_n, 0, y)/\nu_n(y))\} = 1/4.$$

Now, obviously,

$$\mathrm{sep}_n(\theta'_n) \leq \max_{x,y} \{|1 - (p_n^c(\theta'_n, x, y)/\nu_n(y))|\}$$

Hence, $T_{n,\infty}(1/4) \geq \theta'_n$. This shows that $\chi_n T_{n,\infty}(1/4)$ tends to infinity. Hence there is a max-L^∞ cutoff as desired. The proof of the discrete case is similar (note that the hypothesis $\chi_n\theta_n$ tends to infinity implies that m_n tends to infinity. For any birth and death chain family, this implies that $T_{n,\infty}(1/4)$ tends to infinity). □

5.5 Max-L^p cutoff: the transitive normal case

This section is devoted to the important case where a group acts continuously transitively on the underlying space and this action preserves the transition function.

5.5.1 Transitive group action

Assume Ω is a compact topological space on which a compact group G acts continuously transitively. In this case, $\Omega = G/G_o$ (as topological spaces) where G_o is the stabilizer of any fixed point $o \in \Omega$. In this case, $\Omega = G/G_o$ carries a unique G invariant measure π given by $\pi(A) = u(\phi^{-1}(A))$ where $\phi : G \to G/G_o$ denotes the canonical projection map.

We assume that this action preserves a given transition function $p(t, x, \cdot)$, $t \in T$, $x \in \Omega$, defined on the Borel σ-algebra $\mathcal{B}$ of Ω. By definition, this means that, for all $t \in T$, $x \in \Omega$, $A \in \mathcal{B}$ and $g \in G$,

$$p(t, x, A) = p(t, gx, gA) \tag{5.1}$$

This implies that π is an invariant measure for this transition function.

A crucial observation is that we can lift $p(t,x,\cdot)$ to a transition function on G (for further details, see, e.g., [30]). Namely, for any $x \in \Omega$, let g_x be an element in G such that $g_x o = x$. For any $g \in G$ and any Borel subset $A \subset G$

$$\tilde{p}(t,g,A) = \int_\Omega \int_{G_o} \mathbf{1}_A(g_y h) d_{G_o}h \, p(t,go,dy) = \int_\Omega |g_y^{-1}A \cap G_o|_{G_o} p(t,go,dy)$$

where $|g_y^{-1}A \cap G_o|_{G_o}$ is the measure of $g_y^{-1}A \cap G_o$ with respect to normalized Haar measure $d_{G_o}h$ on G_o. Note that $|g_y^{-1}A \cap G_o|_{G_o}$ is independent of the choice of g_y. Note also that if A has the property that $A = AG_o$ (i.e., $\phi^{-1}[\phi(A)] = A$), then

$$g_y^{-1}A \cap G_o \neq \emptyset \Leftrightarrow y \in \phi(A).$$

Hence, if $A = AG_o$ then

$$\tilde{p}(t,g,A) = p(t,go,\phi(A)).$$

This transition function on G obviously satisfies

$$\tilde{p}(t,hg,hA) = \tilde{p}(t,g,A) = \tilde{p}(t,e,g^{-1}A), \quad g,h \in G$$

This means that the Markov operators

$$\mu \mapsto \mu\widetilde{P}_t, \quad f \mapsto \widetilde{P}_t f$$

are in fact convolution operators. Namely, setting

$$q_t(A) = \tilde{p}(t,e,A),$$

we have

$$\mu\widetilde{P}_t(A) = \int_G q_t(g^{-1}A)\mu(dg), \quad \widetilde{P}_t f(g) = \int_G f(gh) q_t(dh).$$

These measures form a convolution semigroup and have the property that for any $g \in G_o$, $q_t(Ag) = q_t(gA) = q_t(A)$. Conversely, if we start with a convolution semigroup of probability measures q_t on G satisfying this last property (bi-invariance under G_o), we obtain a G invariant transition function on $\Omega = G/G_o$ by setting

$$p(t,x,A) = q_t(g_x^{-1}\phi^{-1}(A)).$$

To clarify the relations between operating on G and operating on Ω, it suffices to consider the operators $S : L^p(G,dg) \to L^p(\Omega,\pi)$ and $T : L^p(\Omega,\pi) \to L^p(G,dg)$ defined by

$$Sf(x) = \int_{G_o} f(g_x h) d_{G_o}h, \quad Tf(g) = f(go).$$

It is clear that this operators, originally defined on continuous functions, extends uniquely to operators with norm equal to 1 between the relevant L^p spaces. They are, formally, adjoint of each other. Moreover, we clearly have

$$\widetilde{P}_t = TP_tS \text{ and } P_t = S\widetilde{P}_tT.$$

This structure also allows us to see that the adjoint P_t^* of P_t is of the same form with transition function given by

$$p^*(t,x,A) = \check{q}_t(g_x^{-1}\phi^{-1}(A)).$$

Indeed,

$$P_t^* = S\widetilde{P}_t^*T$$

and P^* is convolution by $\check{q}_t$ where $\check{q}_t(A) = q_t(A^{-1})$.

For any fixed $t \in T$, the measures $p(t,x,\cdot), x \in \Omega$, all are absolutely continuous w.r.t π if and only if $p(t,o,\cdot)$ is, if and only if the measure q_t on G is absolutely continuous w.r.t. the Haar measure dg on G. In this case, if we set

$$p(t,x,dy) = h(t,x,y)\pi(dy), \quad \widetilde{p}(t,g,dg') = \widetilde{h}(t,g,g')dg', \quad q_t(dg) = \phi_t(g)dg$$

we have

$$h(t,x,y) = \widetilde{h}(t,g_x,g_y) = \phi_t(g_x^{-1}g_y)$$

and

$$p^*(t,x,dy) = h(t,y,x)\pi(dy).$$

With this preparation, we can state and prove the following result.

Proposition 5.7. *Let Ω be a compact space equipped with its Borel σ-algebra. Let $p(t,x,\cdot)$, $t \in T$, $x \in \Omega$, be a transition function . Assume that there exists a compact group G that acts continuously and transitively on Ω and such that $p(t,gx,gA) = p(t,x,A)$, for all $t \in T$, $x \in \Omega$, $g \in G$. Let π be the unique G-invariant probability measure on Ω as above. For any $1 \le p \le \infty$, we have:*

1. *For all $x,y \in \Omega$, $D_p(x,t) = D_p(y,t) = \overline{D}_p(t)$.*
2. *For all $x \in \Omega$, $D_p(x,t) = D_p^*(x,t)$ where D_p^* corresponds to the adjoint transition function $p^*(t,x,\cdot)$ on Ω.*

Proof. The first assertion is obvious. To prove the second, observe that

$$\begin{aligned} D_p^p(x,t) &= \int_\Omega |h(t,o,y)-1|^p\pi(dy) = \int_\Omega |\phi_t(g_y)-1|^p\pi(dy) \\ &= \int_\Omega\int_{G_o} |\phi_t(g_yh)-1|^p d_{G_o}h\, \pi(dy) = \int_G |\phi_t(g)-1|^p dg. \end{aligned}$$

Similarly,

$$D_p^*(x,t) = \int_G |\check{\phi}_t(g)-1|^p dg = \int_G |\phi_t(g^{-1})-1|^p dg.$$

The desired result follows since Haar measure is preserved by the transformation $g \mapsto g^{-1}$ $\square$

5.5.2 Mixing time and Max-L^p cutoffs in the transitive case

In the transitive case, Proposition 5.7 allows us to simplify the mixing time comparisons of Proposition 5.1 as follows.

Proposition 5.8. *Let Ω be a compact space equipped with its Borel σ-algebra. Let $p(t,x,\cdot)$, $t \in T$, $x \in \Omega$, be a transition function. Assume that there exists a compact group G that acts continuously and transitively on Ω and such that $p(t,gx,gA) = p(t,x,A)$, for all $t \in T$, $x \in \Omega$, $g \in G$. Let π be the unique G-invariant probability measure on Ω. Referring to the mixing times $T_p(\epsilon) = \inf\{t > 0 : \overline{D}_p(t) \le \epsilon\}$, the following inequalities hold:*

1. *For $1 \le p \le q \le \infty$ and any fixed $\epsilon > 0$, $T_p(\epsilon) \le T_q(\epsilon)$.*
2. *For $1 \le q, r, s, \le \infty$ with $1 + 1/q = 1/r + 1/s$ and any $\epsilon, \delta > 0$,*
$$T_q(\epsilon\delta) \le T_s(\epsilon) + T_r(\delta).$$
3. *For $1 < p < q \le \infty$ and $\epsilon > 0$,*
$$T_q(\epsilon^{m_{p,q}}) \le m_{p,q} T_p(\epsilon)$$
where $m_{p,q} = \lceil p'/q' \rceil$, $1/p + 1/p' = 1/q + 1/q' = 1$.

In the next three statements we consider a family of compact spaces Ω_n indexed by $n = 1, 2, \dots$. For each n, assume that there exists a compact group G_n that acts continuously and transitively on Ω_n and let π_n be the unique G_n-invariant measure on Ω_n. Let $p_n(t,\cdot,\cdot)$, $t \in T$, be a transition function on Ω_n Assume that $p_n(t,gx,gA) = p_n(t,x,A)$, for all $t \in T$, $x \in \Omega_n$, $g \in G_n$. Assume also that $P_{n,t}$ is normal on $L^2(\Omega_n, \pi_n)$, for each $t \in T \cap (0,1]$ and let λ_n be the spectral gap as defined at (3.3). We will refer to this as the *normal transitive* setup. This of course includes the case where $\Omega = G$ is a compact group and the transition function is given by convolution. The proof of the following statements easily follows from Theorems 4.2-4.3 and Proposition 5.8. Details are omitted.

Lemma 5.9. *Referring to the normal transitive setup introduced above, the following properties holds:*

1. *For each n, the function $t \mapsto \overline{D}_{\Omega_n,p}(t)$ tends to zero at infinity for some $p \in (1,\infty]$ if and only if it tends to zero for any $p \in (1,\infty]$.*
2. *Assume that for each n, the functions in* (1) *tend to zero. Then $T_{n,p}(\epsilon)$ tends to infinity with n for some $p \in (1,\infty]$ and $\epsilon > 0$ if and only if it does for all such p and ϵ.*

Theorem 5.10 (Max-L^p cutoff, normal transitive continuous case)**.** *Referring to the normal transitive setup introduced above, assume that $T = [0,\infty)$ and that, for each n and $p \in (1,\infty]$, the function $t \mapsto \overline{D}_{\Omega_n,p}(t)$ tends to zero at infinity. Then the following properties are equivalent:*

1. *For some $p \in (1,\infty]$ and some $\epsilon > 0$, $\lambda_n T_{n,p}(\epsilon)$ tends to infinity;*
2. *For any $p \in (1,\infty]$ and any $\epsilon > 0$, $\lambda_n T_{n,p}(\epsilon)$ tends to infinity;*

3. *For some $p \in (1, \infty]$ there is a max-L^p precutoff.*

4. *For any $p \in (1, \infty)$ there is a max-L^p cutoff.*

5. *For any $p \in (1, \infty)$ and any $\epsilon > 0$, there is a $(T_{n,p}(\epsilon), \lambda_n^{-1})$ max-L^p cutoff.*

Theorem 5.11 (Max-L^p cutoff, normal transitive discrete case)**.** *Referring to the normal transitive setup introduced above, assume that $T = \mathbb{N}$ and that, for each n, $p \in (1, \infty]$ and $\epsilon > 0$, the function $t \mapsto \overline{D}_{\Omega_n,p}(t)$ tends to zero as t tends to infinity. Assume further that $T_{n,p}(\epsilon)$ tends to infinity with n. Then, setting $\gamma_n = \min\{1, \lambda_n\}$, the following properties are equivalent:*

1. *For some $p \in (1, \infty]$ and some $\epsilon > 0$, $\gamma_n T_{n,p}(\epsilon)$ tends to infinity;*

2. *For any $p \in (1, \infty]$ and any $\epsilon > 0$, $\gamma_n T_{n,p}(\epsilon)$ tends to infinity;*

3. *For some $p \in (1, \infty]$ there is a max-L^p precutoff.*

4. *For any $p \in (1, \infty)$ there is a max-L^p cutoff.*

5. *For any $p \in (1, \infty)$ and any $\epsilon > 0$, there is a $(T_{n,p}(\epsilon), \gamma_n^{-1})$ max-L^p cutoff.*

Note that in statements (4)-(5) of these theorems, the case $p = \infty$ is excluded. This is a important difference between the above result and the similar statement in the reversible case. We do not know whether or not a max-L^2 cutoff implies a max-L^∞ cutoff in this setting (note that it does imply a max-L^∞ precutoff).

6 Total variation examples

This section discusses examples showing that the L^p results described in this paper for $1 < p < \infty$ do not hold true for $p = 1$.

6.1 Aldous' example

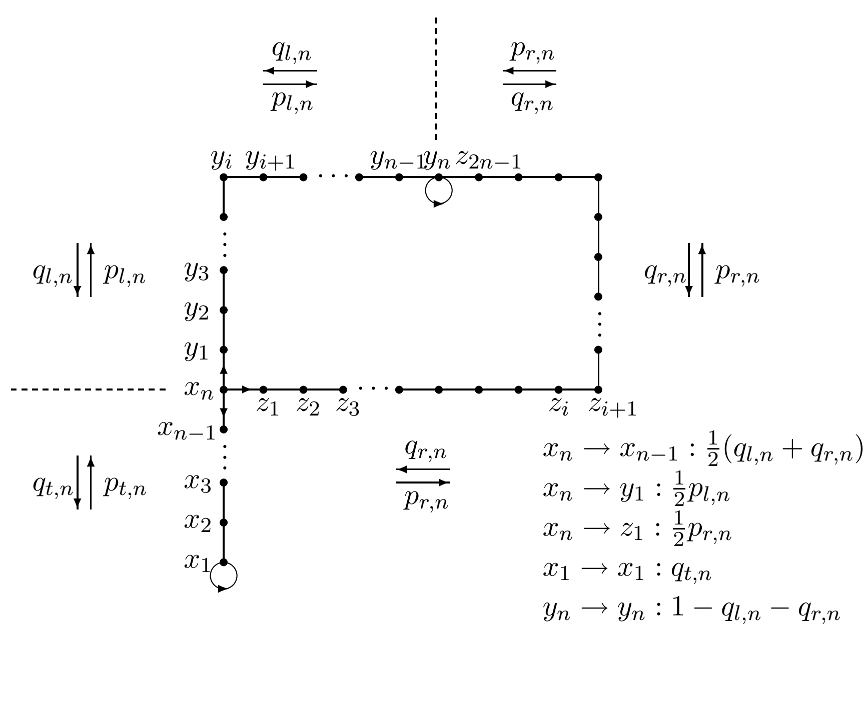
At the ARCC workshop "Sharp Thresholds for Mixing Times" organized at AIM, Palo Alto, in December 2004, David Aldous proposed the following example of a reversible Markov chain with the property that the product "spectral gap $\times$ maximum total variation mixing time" tends to infinity but which does not have a total variation cutoff. The proposed chain is made of three parts: a tail and two arms. The two arms are attached to the tail and are joined together at the other end. The tail is a finite segment of length n, say $\{x_1, \dots x_n\}$. The left arm also has length n, $\{y_1, \dots, y_n\}$. The right arm has length $2n$ $\{z_1, \dots z_{2n}\}$ with $z_{2n} = y_n$. Transitions are essentially like a birth and death chain with a constant upward drift. More precisely, pick $p_{i,n} > 1/2$, $q_{i,n} < 1/2$, $p_{i,n} + q_{i,n} = 1$, $i = t, l, r$ (t for tail, l, r for left and right). Along the tail, go up with probability $p_{t,n}$ and down with probability $q_{t,n}$. At the top of the tail, go down with probability $(q_{l,n} + q_{r,n})/2$, left with probability $p_{l,n}/2$, right with probability $p_{r,n}/2$. Along each arm go up or down with the corresponding probability $p_{i,n}, q_{i,n}$, $i = l$ or r. At the point $y_n = z_{2n}$, go to y_{n-1} with probability $q_{l,n}$, to z_{2n-1} with probability $q_{r,n}$, or stay put with probability $1 - q_{l,n} - q_{r,n}$. See Figure 1. To make this chain reversible, we must choose the left and right arm probabilities so that

$$p_{l,n}/q_{l,n} = (p_{r,n}/q_{r,n})^2.$$

The stationary measure is easy to compute and concentrates mostly around the point $y_n = z_{2n}$. The idea now is as follows. Assuming that $p_{i,n} > 2/3$, $i = t, l, r$, it is easy to show using an isoperimetric (i.e., conductance) type argument that the spectral gap of this chain is bounded away from 0. The claim is that there exists $1 < a < b < \infty$ and $\epsilon \in (0, 1)$ such that, for the chain started at x_1, the variation distance is less than $1 - \epsilon$ at time an but still more than ϵ at time bn, uniformly over all large n. In addition, the starting point with the slowest total variation mixing is the end of the tail x_1. This implies that there is no maximum total variation cutoff. Nevertheless the product "spectral gap × maximum total variation mixing time" is of order n and tends to infinity. The reason why total variation is less than $1-\epsilon$ at time an is that one has a good chance to reach y_n through the left arm at that time. The reason why total variation is greater than ϵ at time bn ($b > a$) is that it takes longer to reach $y_n = z_{2n}$ along the right arm.

The computations required to prove these claims are somewhat technical. They become quite simple if one assumes that $p_{i,n} \geq 1 - 1/n$, $i = t, l, r$. In this case the stationary measure concentrates strongly at $y_n = z_{2n}$. See [6].

Figure 1: Aldous' example. In this figure, each edge has two directions which denote the neighboring transitions and have weights specified by the side notations except those five marked directly on the graph with probability described in the right-bottom corner.



6.2 Pak's example

Igor Pak suggested the following type of examples. Let Ω_n be a sequence of finite sets, each equipped with a transition function $p_n(t,x,\cdot)$, $t \in \mathbb{N}$, $x \in \Omega_n$, with stationary distribution π_n. For each n, consider the new Markov chain with one step kernel

$$q_n(1,x,y) = (1-a_n)p_n(1,x,y) + a_n\pi_n(y).$$

A simple calculation (using the fact that π_n is stationary for p_n) shows that this has transition function

$$q_n(t,x,y) = (1-a_n)^t p_n(t,x,y) + (1-(1-a_n)^t)\pi_n(y).$$

Hence,

$$\|q_n(t,x,\cdot) - \pi_n\|_{\mathrm{TV}} = (1-a_n)^t\|p_n(t,x,\cdot) - \pi_n\|_{\mathrm{TV}}.$$

The same relation holds for other norms and other starting distributions. In particular, for $1 \le p \le \infty$,

$$\|(q_n(t,x,\cdot)/\pi_n) - 1\|_{L^p(\Omega_n,\pi_n)} = (1-a_n)^t\|(p_n(t,x,\cdot)/\pi_n) - 1\|_{L^p(\Omega_n,\pi_n)}.$$

Furthermore (using obvious notation)

$$\|Q_{n,t} - \pi_n\|_{L^2(\Omega_n,\pi_n)\to L^2(\Omega_n,\pi_n)} = (1-a_n)^t\|P_{n,t} - \pi_n\|_{L^2(\Omega_n,\pi_n)\to L^2(\Omega_n,\pi_n)}.$$

This implies that the spectral gap $\lambda(Q_n)$ (relative to q_n) is related to the spectral gap $\lambda(P_n)$ (relative to p_n) by

$$\lambda(Q_n) = \lambda(P_n) - \log(1-a_n).$$

Now, let us assume that the family $p_n(t,x_n,\cdot)$ admits a total variation cutoff with cutoff sequence t_n tending to infinity and spectral gap $\lambda(P_n)$ such that $\lambda(P_n)t_n$ tends to infinity and $\lambda(P_n) \le 1$. Pick a sequence a_n tending to 0 such that $\lambda(P_n)a_n^{-1}$ and t_na_n tends to infinity (that is a_n^{-1} tends to infinity faster than $\lambda(P_n)^{-1}$ but slower than t_n). Then, the family $q_n(t,x_n,\cdot)$ (indexed by n) has the following properties;

1. Its spectral gap $\lambda(Q_n)$ satisfies $\lambda(Q_n) \sim \lambda(P_n)$;

2. For any $\epsilon \in (0,1)$, its total variation mixing time $T_{\mathrm{TV}}(q_n,x_n,\epsilon)$ satisfies $T_{\mathrm{TV}}(q_n,x_n,\epsilon) \sim a_n^{-1}\log(1/\epsilon)$;

3. The product $\lambda_n(Q_n)T_{\mathrm{TV}}(q_n,x_n,\epsilon) \sim \lambda(P_n)a_n^{-1}\log(1/\epsilon)$ tends to infinity

4. There is no total variation cutoff and, in fact, no total variation precutoff.

5. For each $p \in (1,\infty)$, there is a L^p cutoff with window $\lambda(P_n)^{-1}$ and cutoff time $s_n(p)$ of order at least the order of a_n^{-1}.

The assertions (1)-(2)-(3) are clear. Assertion (4) follows from Proposition 2.3(i). For $p \in (1,\infty)$, $T_p(q_n,x_n,2\epsilon) \ge T_{\mathrm{TV}}(q_n,x_n,\epsilon)$. Hence (3) and Theorem 3.3 imply an L^p cutoff as stated in (5). If $s_n(p)$ is an L^p cutoff time then, for any $\eta \in (0,1)$, we must have $(1+\eta)s_n(p) \ge a_n^{-1}$ for n large enough.

The same type of computations holds for max total variation cutoff if the original chain has a max total variation cutoff. Similar computations work in continuous time. This general class of examples clearly shows that the conclusions of Theorems 3.3, 4.2, 4.3, 5.10 and 5.11 do not hold true in the case $p=1$. Namely, the condition $\lambda(Q_n)T_{\mathrm{TV}}(q_n,x_n,\epsilon)\to\infty$ is not sufficient for a total variation cutoff.

It is interesting to consider the case (excluded above) when $a_n^{-1}=t_n$. In this case, (1) still holds true. Assertion (2) can be replaced by $(a_n/2)\le T_{\mathrm{TV}}(q_n,x_n,\epsilon)\le a_n^{-1}\log(1/\epsilon)$, for n large enough. It follows that the product $\lambda_n(Q_n)T_{\mathrm{TV}}(q_n,x_n,\epsilon)$ still tends to infinity. Furthermore, for any $\eta\in(0,1)$, we have

$$\lim_{n\to\infty}\|q_n((1-\eta)t_n,x_n,\cdot)-\pi_n\|_{\mathrm{TV}}=e^{-1+\eta}$$

and

$$\lim_{n\to\infty}\|q_n((1+\eta)t_n,x_n,\cdot)-\pi_n\|_{\mathrm{TV}}=0$$

This shows that this family does not have a total variation cutoff but does have a total variation precutoff and even a total variation weak cutoff in the the sense of Remark 2.4. This yields examples with a total variation weak cutoff but no total variation cutoff. A similar analysis also applies to max total variation.

Finally, it is worth pointing out that this same class of examples shows that the condition $\lambda(Q_n)T_{\mathrm{sep}}(q_n,x_n,\epsilon)\to\infty$ is not sufficient for a separation cutoff.

A Techniques and proofs

Proof of Lemma 2.1. For the case $D_n=[0,\infty)$ for all $n\ge 1$, we may choose, by Remark 2.5, an integer N such that $b_n>0$ for $n\ge N$. This implies the following inequalities which are sufficient to show the desired equivalence.

$$\limsup_{n\to\infty}\inf_{t<t_n-cb_n}f_n(t)\le\overline{F}(-2c)\le\limsup_{n\to\infty}\inf_{t<t_n-2cb_n}f_n(t),\ \forall c>0,$$

and

$$\liminf_{n\to\infty}\sup_{t>t_n+2cb_n}f_n(t)\le\underline{F}(2c)\le\liminf_{n\to\infty}\sup_{t>t_n+cb_n}f_n(t),\ \forall c>0.$$

For the case $D_n=\mathbb{N}$ for $n\ge 1$, we let, by Remark 2.5, N,b be positive numbers such that $b_n\ge b$ for $n\ge N$. Then the above two inequalities become, for $c>0$,

$$\limsup_{n\to\infty}\inf_{t<t_n-cb_n}f_n(t)\le\overline{F}(-c-2/b)\le\limsup_{n\to\infty}\inf_{t<t_n-(c+2/b)b_n}f_n(t),$$

and

$$\liminf_{n\to\infty}\sup_{t>t_n+(c+2/b)b_n}f_n(t)\le\underline{F}(c+2/b)\le\liminf_{n\to\infty}\sup_{t>t_n+cb_n}f_n(t).$$

This proves the second case. □

Proof of Proposition 2.2. For part (i), consider the case where $\underline{G}(c) > 0$ (the other case is similar). Assume that $\mathcal{F}$ has a (t_n, d_n) cutoff. Then, by definition, we may choose positive integers C, N such that $f_n(t_n + Cd_n) < \underline{G}(c)$ for $n \geq N$. Observe that we may enlarge N so that $f_n(t_n + Cd_n) < f_n(t_n + cb_n)$ for $n \geq 1$. The weak optimality is then given by the monotonicity of f_n.

In the other direction, assume that $\overline{G}(c) = 0$ and $\underline{G}(c) = M$ for all $c > 0$. Set $n_1 = 1$ and, for $k \geq 1$, let n_{k+1} be an integer greater than n_k such that, for $1 \leq i \leq k$,

$$f_{n_k}(t_{n_k} + 2^{1-i} b_{n_k}) \leq 2^{i-k}, \ f_{n_k}(t_{n_k} - 2^{i-k} b_{n_k}) \geq \min\{(1 - 2^{-i})M, 2^i\}.$$

For $n \geq 1$, let $c_n = b_n$ if $n \notin \{n_k : k \geq 1\}$ and let $c_{n_k} = 2^k b_{n_k}$ for $k \geq 1$. It is easy to see that, for $j \geq 1$,

$$\limsup_{n\to\infty} f_n(t_n + 2^j c_n) \leq \limsup_{n\to\infty} f_{n_k}(t_{n_k} + 2^{j-k} b_{n_k}) \leq 2^{1-j}$$

and

$$\liminf_{n\to\infty} f_n(t_n - 2^j c_n) \geq \liminf_{n\to\infty} f_{n_k}(t_{n_k} - 2^{j-k} b_{n_k}) \geq \min\{(1 - 2^{-j})M, 2^j\}.$$

This implies there is a (t_n, c_n) cutoff (let j tend to infinity). Hence, the (t_n, b_n) cutoff is not weakly optimal.

For part (ii), assume that $0 < \underline{G}(c_2) \leq \overline{G}(c_1) < M$ for some $c_2 > c_1$ and that $\mathcal{F}$ has a (s_n, d_n) cutoff. As before, we may choose $C_2 > C_1$ and $N > 0$ such that, for $n \geq N$,

$$f_n(s_n + C_2 d_n) \leq f_n(t_n + c_2 b_n) \leq f_n(t_n + c_1 b_n) \leq f_n(s_n + C_1 d_n).$$

The monotonicity of f_n then implies $(C_2 - C_1)d_n > (c_2 - c_1)b_n$ for $n \geq N$. This proves the optimality of the (t_n, b_n) cutoff.

In the other direction, assume that the (t_n, b_n) cutoff is optimal. Let $A = \{c \in \mathbb{R} : \overline{G}(c) = 0\}$. If A is empty, there is nothing to prove. If A is nonempty, let $c_0 = \inf\{c \in A\}$. For $n \geq 1$, let $s_n = t_n + c_0 b_n$. Then the family has an optimal (s_n, b_n) cutoff. By part (ii), we may choose $c < 0$ such that $\underline{G}(c + c_0) < M$. Also, by the definition of c_0, one has $\overline{G}(c/2 + c_0) > 0$ as desired.

Part (iii) is an immediate consequence of Lemma 2.1. □

Proof of Proposition 2.4. Observe that (2.9) implies

$$M = \limsup_{n\to\infty} f_n(0) \geq \delta_0 > 0.$$

As f_n vanishes at infinity, if $\eta > 0$ is small, then $T(f_n, \eta)$ is contained in $(0, \infty)$ for n large enough. This shows the limit of the ratios in (2.7) is well-defined for $0 < \eta < \delta < M$.

We first consider case (i) of Proposition 2.4. Assume that (2.7) holds for $0 < \eta < \delta$ with $C \geq 1$. Let $\delta_1 = \frac{1}{2}\min\{\delta_0, \delta\}$ and $t_n = T(f_n, \delta_1)$. Observe that, for $0 < \eta \leq \delta_1$, we may choose an integer $N(\eta)$ such that $T(f_n, \eta) < 2Ct_n$ for $n \geq N(\eta)$. This implies

$$\sup_{t > 2Ct_n} f_n(t) \leq \eta \quad \forall n \geq N(\eta),\ \eta \in (0, \delta_1).$$

The first condition in (c1) of Definition 2.1 follows. The definition of $T(f_n, \delta_1)$ gives

$$\inf_{t < t_n} f_n(t) \geq \delta_1$$

for n large enough. This proves the second condition in (c1) of Definition 2.1 with $a = 1$. Hence, $\mathcal{F}$ has a precutoff.

For the converse direction, let t_n, a, b be positive numbers as in (c1). Since $\sup_{t>bt_n} f_n(t) \to 0$, we must have $bt_n \geq T(f_n, \delta_0)$ for n large enough. By (2.9), this implies $t_n \to \infty$. Let $0 < \delta < \liminf_{n\to\infty} \inf_{t<t_n} f_n(t)$. For $\eta \in (0, \delta)$, we select $N(\eta) > 0$ such that

$$\forall n \geq N(\eta), \quad \sup_{t>bt_n} f_n(t) < \eta < \delta < \inf_{t<at_n} f_n(t).$$

For $n \geq N(\eta)$, let $r_n, s_n \in \mathbb{N}$ be such that $at_n - 1 \leq r_n < at_n$ and $bt_n < s_n \leq bt_n + 1$. Then the monotonicity of f_n implies

$$at_n - 1 \leq r_n \leq T(f_n, \delta) \leq T(f_n, \eta) \leq s_n \leq bt_n + 1,$$

which gives (2.7) with $C = b/a$.

Next, we consider case (ii) of Proposition 2.4. Assume that (2.7) holds for $0 < \eta < \delta < M$ with $C = 1$. Set $t_n = T(f_n, \delta_0/2)$. By assumption, we may choose, for each $\delta \in (0, M)$ and $\epsilon \in (0, 1)$, an integer $N(\delta, \epsilon)$ such that

$$(1 - \epsilon)t_n < T(f_n, \delta) < (1 + \epsilon)t_n \quad \forall n \geq N(\delta, \epsilon).$$

This implies, by the monotonicity of f_n,

$$\sup_{t>(1+\epsilon)t_n} f_n(t) \leq \delta \leq \inf_{t<(1-\epsilon)t_n} f_n(t) \quad \forall n \geq N(\delta, \epsilon).$$

The desired cutoff is proved by taking $n \to \infty$ and then letting $\delta \to 0$ and $\delta \to M$ respectively.

For the converse, assume that $\mathcal{F}$ presents a cutoff with cutoff sequence $(t_n)_1^\infty$. By definition, we may choose, for each $\delta \in (0, M)$ and $\epsilon \in (0, 1)$, an integer $N(\delta, \epsilon)$ such that

$$\sup_{t>(1+\epsilon)t_n} f_n(t) < \delta < \inf_{t<(1-\epsilon)t_n} f_n(t), \quad \forall n \geq N(\delta, \epsilon). \tag{A.1}$$

As in the proof of case (i), the monotonicity of f_n implies

$$(1 - \epsilon)t_n - 1 \leq T(f_n, \delta) \leq (1 + \epsilon)t_n + 1,$$

for $n \geq N(\delta, \epsilon)$ and $\epsilon \in (0, 1)$. Since (2.9) implies $t_n \to \infty$, we obtain $T(f_n, \delta) \sim t_n$ for all $0 < \delta < M$, which is equivalent to the desired property.

Finally, consider case (iii) of Proposition 2.4]. Assume that (2.8) holds for $\delta \in (0, M)$. This is equivalent to the existence of an integer $N(\delta, c_1)$, depending on $\delta \in (0, M)$ and $c_1 > 0$, such that

$$0 < t_n - cb_n \leq T(f_n, \delta) \leq t_n + cb_n \quad \forall n \geq N(\delta, c_1),\ c > c_1.$$

The above inequalities imply that

$$\sup_{t>t_n+cb_n} f_n(t) \leq \delta \leq \inf_{t<t_n-cb_n} f_n(t)$$

for all $n \geq N(\delta, c_1)$ and $c > c_1$. Letting $n \to \infty$, $c \to \infty$ and then respectively $\delta \to 0$ and $\delta \to M$ proves the (t_n, b_n) cutoff for $\mathcal{F}$

For the converse, assume that $\mathcal{F}$ has a (t_n, b_n) cutoff. By definition, we may choose, for any $\delta \in (0, M)$, a positive number $c(\delta) > 0$ and an integer $N(\delta)$ such that

$$\sup_{t > t_n + c(\delta) b_n} f_n(t) < \delta < \inf_{t < t_n - c(\delta) b_n} f_n(t)$$

for all $n \geq N(\delta)$. As before, this implies

$$t_n - c(\delta) b_n - 1 \leq T(f_n, \delta) \leq t_n + c(\delta) b_n + 1 \quad \forall n \geq N(\delta).$$

By (2.10), we may enlarge $N(\delta)$ so that $\inf_{n \geq N(\delta)} b_n = b > 0$. Then, the above inequalities become $|T(f_n, \delta) - t_n| \leq (c(\delta) + 1/b) b_n$ for $n \geq N(\delta)$. This proves (2.8). $\square$

Proof of Corollary 2.7. The direction (i) implies (ii) follows immediately from Proposition 2.4. For the other direction, assume that $\mathcal{F}$ has weakly optimal $(T(f_n, \delta), b_n)$ and $(T(f_n, \delta) - 1, b_n)$ cutoffs. Obviously, for any subsequence $(m_k)_1^\infty$, the subfamily $\mathcal{F}' = \{f_{m_k} : k = 1, 2, ...\} \subset \mathcal{F}$ has weakly optimal $(T(f_{m_k}, \delta), b_{m_k})$ and $(T(f_{m_k}, \delta) - 1, b_{m_k})$ cutoffs. By the weak optimality of the $(T(f_{m_k}, \delta), b_{m_k})$ cutoff and the positiveness of $\liminf\limits_{k \to \infty} b_{m_k}$, at least one of the following inequalities must hold.

$$\liminf_{k \to \infty} f_{m_k}(T(f_{m_k}, \delta) - 1) < M, \ \limsup_{k \to \infty} f_{m_k}(T(f_{m_k}, \delta) + 1) > 0. \tag{A.2}$$

Similarly, at least one of the following inequalities must hold.

$$\liminf_{k \to \infty} f_{m_k}(T(f_{m_k}, \delta) - 2) < M, \ \limsup_{k \to \infty} f_{m_k}(T(f_{m_k}, \delta)) > 0.$$

Suppose now that $\mathcal{F}$ has a (s_n, c_n) cutoff. It suffices to show that $b_n = O(c_n)$. If $\liminf_{n \to \infty} c_n > 0$, then, by Proposition 2.4, $\mathcal{F}$ has a $(T(f_n, \delta), c_n)$ cutoff. In this case, the weak optimality of the $(T(f_n, \delta), b_n)$ cutoff for $\mathcal{F}$ implies $b_n = O(c_n)$. It remains to prove that $\liminf_{n \to \infty} c_n > 0$. Assume the converse and let $(n_k)_1^\infty$ be a subsequence of $\mathbb{N}$ such that $c_{n_k} \to 0$ as $k \to \infty$. By the definition of the (s_n, c_n) cutoff, we may choose $C > 0$ and $N > 0$ such that

$$\sup_{t > s_n + C c_n} f_n(t) < \delta < \inf_{t < s_n - C c_n} f_n(t), \quad \forall n \geq N.$$

Since $c_{n_k} \to 0$, we may select $K > 0$ such that $C c_{n_k} < 1/2$ and $n_k \geq N$ for $k \geq K$. The monotonicity of f_n then implies that

$$s_{n_k} + C c_{n_k} \geq T(f_{n_k}, \delta) - 1, \quad T(f_{n_k}, \delta) \geq s_{n_k} - C c_{n_k}, \quad \forall k \geq K,$$

which gives

$$T(f_{n_k}, \delta) - 3/2 \leq s_{n_k} \leq T(f_{n_k}, \delta) + 1/2 \quad \forall k \geq K. \tag{A.3}$$

Since $\{f_{n_k} : k \geq 1\}$ has a (s_{n_k}, c_{n_k}) cutoff, we have

$$\lim_{k \to \infty} f_{n_k}(T(f_{n_k}, \delta) - 2) = M, \quad \lim_{k \to \infty} f_{n_k}(T(f_{n_k}, \delta) + 1) = 0.$$

By (A.3), there is no loss of generality in choosing a further subsequence n'_k of n_k such that $s_{n'_k} \geq T(f_{n'_k}, \delta) - 1/2$. Then $\lim\limits_{k \to \infty} f_{n'_k}(T(f_{n'_k}, \delta) - 1) = M$, which contradicts (A.2). Hence, $\inf_{n \geq 1} c_n > 0$. $\square$

Proof of Lemma 2.8. For (i), we assume that $\overline{F} < M$ (the case $\underline{F} > 0$ is similar). Fix $\delta \in (0, M)$ and let $s_n = T(f_n, \delta)$ for $n \geq 1$. By Corollary 2.6, it suffices to show that the (s_n, b_n) cutoff for $\mathcal{F}$ is weakly optimal. Observe that, by Corollary 2.5, we may choose positive numbers C_1, N_1 such that $|t_n - s_n| \leq C_1 b_n$ for $n \geq N_1$. Assume that $\mathcal{F}$ has a (s_n, c_n) cutoff. By Definition 2.1(c3), we may choose $C_2 > 0, N_2 > N_1$ such that

$$\inf_{t < s_n - C_2 c_n} f_n(t) > \sup_{t > t_n - 3C_1 b_n} f_n(t), \quad \forall n \geq N_2.$$

Let $(n_k)_1^\infty$ be a subsequence of $\mathbb{N}$ such that $b_{n_k} > 0$ for all $k \geq 1$ and $b_i = 0$ for $i \in \mathbb{N} \setminus \{n_k : k \geq 1\}$. On this subsequence, we have, for $k \geq 1$ such that $n_k \geq N_2$,

$$\begin{aligned} \inf_{t < s_{n_k} - C_2 c_{n_k}} f_{n_k}(t) &> \sup_{t > t_{n_k} - 3C_1 b_{n_k}} f_{n_k}(t) \\ &\geq \inf_{t < t_{n_k} - 2C_1 b_{n_k}} f_{n_k}(t) \geq \inf_{t < s_{n_k} - C_1 b_{n_k}} f_{n_k}(t). \end{aligned}$$

This implies $b_{n_k} = O(c_{n_k})$ and then $b_n = O(c_n)$. Hence, b_n is an optimal window.

For (ii), we consider the case $\overline{F} < M$ and $\underline{F}(C) = 0$ with $C > 0$ (the other case is similar). Assume the converse, that is, $\mathcal{F}$ has a strongly optimal cutoff, say (r_n, d_n). By part (i), we know that the (t_n, b_n) cutoff is optimal. Hence, we have $d_n = O(b_n)$ and $b_n = O(d_n)$. By Remark 2.5, the strong optimality of the (r_n, d_n) cutoff implies that $b_n > 0$ for n large enough. By Corollary 2.5, $|t_n - r_n| = O(b_n)$. Let $N_3 > 0, C_3 >$ be such that $C_3 > |C|$, $t_n \leq r_n + C_3 b_n$ and $b_n \leq C_3 d_n$ for $n \geq N_3$. By Lemma 2.1, we obtain

$$0 = \underline{F}(C) = \liminf_{n\to\infty} \inf_{t < t_n + C b_n} f_n(t) \geq \liminf_{n\to\infty} \inf_{t < r_n + C_3(C_3 + C) d_n} f_n(t) > 0,$$

a contradiction. ∎

Proof of Proposition 2.9. Note that if the family in Proposition 2.4 has a strongly optimal (t_n, b_n) cutoff, then $\liminf_{n\to\infty} b_n > 0$. By Corollary 2.5 and Definition 2.2(c3), the strong optimality of the (t_n, b_n) cutoff and the assumption that $|s_n - t_n| = O(b_n)$ imply that $\mathcal{F}$ has a strongly optimal (s_n, b_n) cutoff.

For case (i), Definition 2.1(c3) and Definition 2.2(w3) imply the existence of positive numbers C, C', N such that

$$\sup_{t > s_n + C' c_n} g_n(t) < \sup_{t > s_n + C b_n} f_n(t) \leq \sup_{t > s_n + C b_n} g_n(t), \quad \forall n \geq N.$$

By the monotonicity of g_n, we have $C' c_n > C b_n$ for $n \geq N$. This proves $b_n = O(c_n)$.

For (ii), we let n_k and n'_k be subsequences of $\mathbb{N}$ such that $f_{n_k} \geq g_{n_k}$ and $f_{n'_k} \leq g_{n'_k}$ for $k \geq 1$. By (i), we have $b_{n'_k} = O(c_{n'_k})$. For the subfamilies $\{f_{n_k} : k \geq 1\}$ and $\{g_{n_k} : k \geq 1\}$, Definition 2.1(c3) and Definition 2.2(w3) imply the existence of positive numbers C, C', K such that

$$\inf_{t < s_{n_k} - C b_{n_k}} g_{n_k}(t) \leq \inf_{t < s_{n_k} - C b_{n_k}} f_{n_k}(t) < \inf_{t < s_{n_k} - C' c_{n_k}} g_{n_k}(t), \quad \forall k \geq K.$$

This gives $C b_{n_k} \leq C' c_{n_k}$ for $k \geq K$. Hence, $b_n = O(c_n)$. ∎

Proof of Proposition 3.5. We first consider the spectral gaps λ_n^d and λ_n^c. Note that, for $t > 0$, $P_{n,t}^d = K_n^t$ and $P_{n,t}^c = e^{-t(I-K_n)}$. In the discrete time case, let K_n^* be the adjoint of K_n. Then the Markov kernel $K_n^*K_n$ is given by

$$K_n^*K_n(x,y) = \begin{cases} 1/2 & \text{if } y_i = x_i \text{ for } 1 \le i \le n-1 \\ 0 & \text{otherwise} \end{cases}$$

for all $x = x_n...x_1, y = y_n...y_1 \in \Omega_n$. This implies $\lambda_n^d = 0$. In the continuous time case, let $f(x)$ be the the number of 1's in x. Then $\mathrm{Var}_{\pi_n}(f) = n/4$ and

$$\langle (I-K_n)f, f\rangle = \frac{1}{2}\sum_{x,y\in\Omega_n} |f(x)-f(y)|^2 K_n(x,y)\pi_n(x) = \frac{1}{4}.$$

This implies $\lambda_n^c \le 1/n$.

To prove the desired cutoff, let $f_{n,p}^d(t) = D_{n,p}^d(0,t)$(resp. $f_{n,p}^c(t) = D_{n,p}^c(0,t)$), where $D_{n,p}^d(0,t)$(resp. $D_{n,p}^c(0,t)$) is the L^p distance between δ_0 and $\delta_0 P_{n,t}^d$(resp. $\delta_0 P_{n,t}^c$). We first consider (i). Obviously,

$$f_{n,p}^d(t) = \begin{cases} 2^{-n/p}\left((2^{n-t}-1)^p + 2^n - 2^t\right)^{1/p}\mathbf{1}_{[0,n]}(t) & \text{for } 1 \le p < \infty \\ (2^{n-t}-1)\mathbf{1}_{[0,n]}(t) & \text{for } p = \infty \end{cases}$$

Let $\underline{F}$ and $\overline{F}$ be functions in (2.2) w.r.t. $(t_n, b_n) = (n, 1)$. A few computations show that

$$\underline{F}(c) = \overline{F}(c) = \begin{cases} [(2^{-c}-1)^p 2^c + 1 - 2^c]^{1/p}\mathbf{1}_{[0,\infty)}(c) & \text{for } 1 \le p < \infty \\ (2^{-c}-1)\mathbf{1}_{[0,\infty)}(c) & \text{for } p = \infty \end{cases}$$

This proves the L^p $(n,1)$-cutoff for $1 \le p \le \infty$. The optimality of the window follows immediately from Lemma 2.8.

For (ii), observe that the transition function $p_n^c(t,\cdot,\cdot)$ can be expressed as follows.

$$p_n^c(t,0,y) = e^{-t}\sum_{j=i}^{n}\frac{t^j}{j!}2^{-j} + e^{-t}\sum_{j>n}\frac{t^j}{j!}2^{-n}$$

where $y = y_n\cdots y_1$ satisfies $y_n = y_{n-1} = \cdots = y_{i+1} = 0$ and $y_i = 1$ for $0 \le i \le n$. This implies, for $1 \le p < \infty$,

$$\left(f_{n,p}^c(t)\right)^p = e^{-pt}2^{-n}\left\{\sum_{i=1}^{n}\left|\sum_{j=i}^{n}\frac{t^j}{j!}(2^{n-j}-1) - \sum_{j=0}^{i-1}\frac{t^j}{j!}\right|^p 2^{i-1} + \left(\sum_{j=0}^{n}\frac{t^j}{j!}(2^{n-j}-1)\right)^p\right\}$$

and

$$f_{n,\infty}^c(t) = 2^n p_n^c(t,0,0) - 1 = e^{-t}\sum_{j=0}^{n}\frac{t^j}{j!}(2^{n-j}-1). \tag{A.4}$$

In the case of $1 < p < \infty$, the fact that

$$n^{1/p-1}\sum_{i=1}^{n} c_i \le \left(\sum_{i=1}^{n} c_i^p\right)^{1/p} \le \sum_{i=1}^{n} c_i, \quad \forall c_i \ge 0, 1 \le i \le n,$$

implies that, for $t > 0$,

$$2^{\frac{-1}{p}}(n+1)^{\frac{1-p}{p}}[G_p(n,t) - H_p(n,t)] \leq \|h^0_{n,t} - 1\|_p \leq G_p(n,t) + H_p(n,t) \tag{A.5}$$

where

$$\begin{aligned} G_p(n,t) &= e^{-t}2^{-n/p}\left(\sum_{i=1}^{n}\sum_{j=i}^{n}\frac{t^j}{j!}(2^{n-j}-1)2^{i/p} + \sum_{j=0}^{n}\frac{t^j}{j!}(2^{n-j}-1)\right) \\ &= e^{-t}2^{n(1-1/p)}\sum_{j=0}^{n-1}\frac{1}{j!}(t2^{(1-p)/p})^j\frac{(1-2^{j-n})(1-2^{-(j+1)/p})}{1-2^{-1/p}} \end{aligned}$$

and

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

Fix $1 < p < \infty$ and let $t_n = t_n(p)$ and $b_n = \log n$. Note that for $s > 1$, the map $s \mapsto \frac{\log s}{1-s^{-1}}$ is increasing and has limit 1 as $s \downarrow 1$. Using this observation, we may choose $\delta > 0$ and $N > 0$ such that $t_n(1-\delta) > n$ for $n \geq N$ and hence, Lemma A.1 implies

$$\lim_{n\to\infty} H_p(n, t_n + cb_n) \leq \lim_{n\to\infty} H_p(n, t_n(1-\delta/2)) = 0 \quad \forall c \in \mathbb{R}. \tag{A.6}$$

Hence it suffices to consider only the function G_p. Moreover, as

$$1/2 \leq \frac{(1-2^{j-n})(1-2^{-(j+1)/p})}{1-2^{-1/p}} \leq \frac{1}{1-2^{-1/p}}, \quad \forall 0 \leq j \leq n-1,$$

we can consider the function

$$g_p(n,t) = e^{-t}2^{n(1-1/p)}\sum_{j=0}^{n-1}\frac{1}{j!}(t2^{1/p-1})^j$$

instead of $G_p(n,t)$. A simple computation shows that

$$g_p(n, t_n + cb_n) = \exp\{-cb_n(1-2^{1/p-1})\}e^{-s_n}\sum_{j=0}^{n-1}\frac{s_n^j}{j!} \tag{A.7}$$

where $s_n = (t_n + cb_n)2^{1/p-1}$. Observe that, for fixed $c \in \mathbb{R}$,

$$n - s_n = n\left(1 - \frac{\log(2^{1-1/p})}{2^{1-1/p}-1}\right)(1+o(1)) \quad \text{as } n \to \infty.$$

Since the map $s \mapsto \frac{\log s}{s-1}$ for $s > 1$ is strictly decreasing and has limit 1 as $s \downarrow 1$, we may choose, for each $c \in \mathbb{R}$, positive numbers δ, N such that

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

By Lemma A.1, this implies

$$\lim_{n\to\infty} e^{-s_n}\sum_{j=0}^{n-1}\frac{s_n^j}{j!}=1 \quad \forall c\in\mathbb{R}. \tag{A.8}$$

Now combining (A.5), (A.6), (A.7) and (A.8), we get

$$\forall c>0,\quad \overline{F}(c)=\limsup_{n\to\infty} f^c_{n,p}(t_n+cb_n)\le \lim_{n\to\infty} n^{-c\left(1-2^{1/p-1}\right)}=0$$

and

$$\forall c<0,\quad \underline{F}(c)=\liminf_{n\to\infty} f^c_{n,p}(t_n+cb_n)\ge \lim_{n\to\infty} 2^{-1/p}n^{-c\left(1-2^{1/p-1}\right)+1/p-1}.$$

This proves the desired L^p-cutoff for $1<p<\infty$.

For the case $p=\infty$, we have, by using the identity in (A.4),

$$f^c_{n,\infty}(t)\begin{cases}\le 2^n e^{-t}\sum_{j=0}^{n-1}\frac{(t/2)^j}{j!}\\ \ge 2^{n-1}e^{-t}\sum_{j=0}^{n-1}\frac{(t/2)^j}{j!}\end{cases}.$$

Fix $c\in\mathbb{R}$. Observe that $t_n+c-(\log 2)n=\frac{t_n+c}{2}+\frac{c}{2}$. This implies

$$\frac{1}{2}e^{-c/2}c_n\le f^c_{n,\infty}(t_n+c)\le e^{-c/2}c_n,$$

where $c_n=e^{-(t_n+c)/2}\sum_{j=0}^{n-1}\frac{((t_n+c)/2)^j}{j!}$. Since $t_n/2=(\log 2)n<n$, by Lemma A.1, one has $c_n\to 1$ as $n\to\infty$. Hence, we have

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

This proves the strong optimality of the $((2\log 2)n,1)$ L^∞-cutoff, which is somewhat stronger than the statement in Proposition 3.5.

For the case $p=1$, note that the triangular inequality implies that

$$f^c_{n,1}(t)\le e^{-t}\sum_{i=0}^{n}\frac{t^i}{i!}f^d_{n,1}(t)\le 2e^{-t}\sum_{i=0}^{n}\frac{t^i}{i!}.$$

For $n\ge 1$, let $t_n=n$ and $b_n=\sqrt{n}$. By Lemma A.1, the above inequality implies that $\overline{F}(c)\le 2\Phi(-c)$ for all $c\in\mathbb{R}$, where $\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-s^2/2}ds$. For the lower, we set, for $\epsilon>0$,

$$A_n(\epsilon)=\{y=y_n\cdots y_1\in\Omega_n : y_n=y_{n-1}=\cdots=y_{n-[\epsilon\sqrt{n}]+1}=0\}.$$

Obviously, we have

$$f_{n,1}(t)\ge 2(p^c_n(t,0,A_n(\epsilon))-\pi_n(A_n(\epsilon)))\ge G(t)-H(t),$$

where

$$G(t)=2e^{-t}\sum_{j=0}^{n-[\epsilon\sqrt{n}]}\frac{t^j}{j!},\quad H(t)=2e^{-t}\sum_{j=0}^{n-[\epsilon\sqrt{n}]}\frac{(t/2)^j}{j!}+2^{1-[\epsilon\sqrt{n}]}.$$

Fix $c \in \mathbb{R}$. By Lemma A.1, we have

$$\lim_{n\to\infty} H(n+[c\sqrt{n}]) = 0, \;\; \lim_{n\to\infty} G(n+[c\sqrt{n}]) = 2\Phi(-\epsilon - c), \quad \forall \epsilon > 0.$$

This implies $\underline{F}(c) \geq 2\Phi(-c)$ for all $c \in \mathbb{R}$. Combining all above, we get the strong optimality of the $(n, \sqrt{n})$ L^1-cutoff. $\square$

Lemma A.1. *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*

$$\limsup_{n\to\infty} d_n = \Phi\left(\limsup_{n\to\infty} c_n\right), \quad \liminf_{n\to\infty} d_n = \Phi\left(\liminf_{n\to\infty} c_n\right), \tag{A.9}$$

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 of Lemma A.1. We prove the first identity in (A.9). The proof of the second identity is similar. Note that if (A.9) 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 A.1 under the assumption that the sequence $(a_n)_1^\infty$ either is bounded or tends to infinity. In the former case, one can easily prove it by Taylor expansion of the 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 $Y_1, Y_2, \ldots$ be i.i.d. Poisson(1) random variables and F_n the distribution function of $a_n^{-1/2}(Y_1 + Y_2 + \ldots + Y_{a_n} - a_n)$. 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

$$e^{-\lceil a_n\rceil}\sum_{i=0}^{b_n}\frac{\lceil a_n\rceil^i}{i!}\le d_n\le e^{-\lfloor a_n\rfloor}\sum_{i=0}^{b_n}\frac{\lfloor a_n\rfloor^i}{i!}. \tag{A.10}$$

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 $\limsup\limits_{n\to\infty}\frac{b_n-[a_n]}{\sqrt{[a_n]}}=\limsup\limits_{n\to\infty}c_n$. Hence, the first identity for nonnegative real-valued a_n is proved by applying (A.10) and the result in the case $a_n\in\mathbb{Z}_+$ for $n\ge 1$. $\square$

Acknowledgements

We thank the anonymous referee for his/her very careful reading of the manuscript. We thank Igor Pak for allowing us to includes a version of his example in Section 6. In addition to the seminal works by David Aldous and Persi Diaconis, Yuval Peres' suggestion of a simple criterion for the existence of a cutoff is at the origin of this work.

References

[1] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983. MR0770418

[2] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5):333–348, 1986. MR0841111

[3] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8(1):69–97, 1987. MR0876954

[4] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2(2):294–313, 1992. MR1161056

[5] Mark Brown and Y-S. Shao. Identifying coefficients in the spectral representation for first passage time distribution. *Probab. Engrg. Inform. Sci*, 1(1):69–74, 1987.

[6] Guan-Yu Chen. *The cut-off phenomenon for finite Markov chains.* PhD thesis, Cornell University, 2006.

[7] Guan-Yu Chen and Laurent Saloff-Coste. The cut-off phenomenon for randomized riffle shuffles. *Random Structures Algorithms*, ?(?):?–?, ?

[8] Persi Diaconis. *Group representations in probability and statistics.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988. MR0964069

[9] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.*, 93(4):1659–1664, 1996. MR1374011

[10] Persi Diaconis, James Allen Fill, and Jim Pitman. Analysis of top to random shuffles. *Combin. Probab. Comput.*, 1(2):135–155, 1992. MR1179244

[11] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *Ann. Probab.*, 21(4):2131–2156, 1993. MR1245303

[12] Persi Diaconis and Laurent Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006. MR2288715

[13] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981. MR0626813

[14] Nelson Dunford and Jacob T. Schwartz. *Linear operators. Part I.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1988.

[15] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968. MR0228020

[16] Mourad E. H. Ismail, David R. Masson, Jean Letessier, and Galliano Valent. Birth and death processes and orthogonal polynomials. In *Orthogonal polynomials (Columbus, OH, 1989)*, volume 294 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 229–255. Kluwer Acad. Publ., Dordrecht, 1990. MR1100296

[17] Samuel Karlin and James McGregor. Ehrenfest urn models. *J. Appl. Probability*, 2:352–376, 1965. MR0184284

[18] Samuel Karlin and James L. McGregor. The Hahn polynomials, formulas and an application. *Scripta Math.*, 26:33–46, 1961. MR0138806

[19] Julian Keilson. *Markov chain models—rarity and exponentiality*, volume 28 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1979. MR0528293

[20] Laurent Miclo. An example of application of discrete Hardy's inequalities. *Markov Process. Related Fields*, 5(3):319–330, 1999. MR1710983

[21] Yuval Peres and David Revelle. Mixing times for random walks on finite lamplighter groups. *Electron. J. Probab.*, 9:no. 26, 825–845 (electronic), 2004. MR2110019

[22] Ursula Porod. The cut-off phenomenon for random reflections. *Ann. Probab.*, 24(1):74–96, 1996. MR1387627

[23] Ursula Porod. The cut-off phenomenon for random reflections. II. Complex and quaternionic cases. *Probab. Theory Related Fields*, 104(2):181–209, 1996. MR1373375

[24] Jeffrey S. Rosenthal. Random rotations: characters and random walks on SO(N). *Ann. Probab.*, 22(1):398–423, 1994. MR1258882

[25] L. Saloff-Coste. Convergence to equilibrium and logarithmic Sobolev constant on manifolds with Ricci curvature bounded below. *Colloq. Math.*, 67(1):109–121, 1994. MR1292948

[26] L. Saloff-Coste. Precise estimates on the rate at which certain diffusions tend to equilibrium. *Math. Z.*, 217(4):641–677, 1994. MR1306030

[27] L. Saloff-Coste. On the convergence to equilibrium of Brownian motion on compact simple Lie groups. *J. Geom. Anal.*, 14(4):715–733, 2004. MR2111426

[28] Laurent Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 301–413. Springer, Berlin, 1997. MR1490046

[29] Laurent Saloff-Coste. Random walks on finite groups. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 263–346. Springer, Berlin, 2004. MR2023654

[30] Laurent Saloff-Coste and Wolfgang Woess. Transition operators on co-compact g-spaces. *Revista Matemática Iberoamericana*, 22(3):747–799, 2006. MR2320401

[31] W. Van Assche, P. R. Parthasarathy, and R. B. Lenin. Spectral representation of four finite birth and death processes. *Math. Sci.*, 24(2):105–112, 1999. MR1746330