Mixing and average mixing times for general Markov processes

Robert M. Anderson, Haosui Duanmu, and Aaron Smith

Abstract. Yuval Peres and Perla Sousi showed that the mixing times and average mixing times of reversible Markov chains on finite state spaces are equal up to some universal multiplicative constant. We use tools from nonstandard analysis to extend this result to reversible Markov chains on compact state spaces that satisfy the strong Feller property.

1 Introduction

Consider the simple random walk on $\mathbb{Z}_{2n}$, given by

$$(1.1) \qquad X_m = \sum_{i=1}^{m} \delta_i, \qquad \text{modulo } 2n$$

for an i.i.d. ~ sequence $\delta_1, \delta_2, \ldots$ of random variables with $\text{Unif}(\{-1, +1\})$ distribution. It is easy to see that X_{2m} is always even, while X_{2m+1} is always odd for $m \in \mathbb{N}$. This periodic behaviour means that the chain is not ergodic. On the other hand, there are various ways that this periodic behaviour seems to be essentially the *only* obstacle to mixing. For example, for $T \sim \text{Geom}(Cn^2)$ for a sufficiently large constant $C > 0$, one can easily check that the distribution of X_T is very close to uniform on $\mathbb{Z}_{2n}$ (see *e.g.*, [LPW09] for a coupling argument).

This sort of (near-)periodicity is often undesired, and a common way to "fix" the problem is to replace a chain with an ε-lazy version (in Example (1.1), the $\frac{1}{3}$-lazy chain can be obtained by sampling the driving sequence from $\delta_1, \delta_2, \ldots \overset{i.i.d.}{\sim} \text{Unif}(\{-1, 0, +1\})$). This chain is also close to uniform after $\Theta(n^2)$ steps, but it is natural to ask if smaller modifications can also eliminate periodic behaviour; for the above example, choosing $T \sim \text{Unif}(\{Cn^2, Cn^2 + 1\})$ spreads our random time over only two choices but still works well. This minimal modification turns out to work quite generally, and [PS15] shows that this gives an equivalent reduction in the time a discrete chain takes to mix (see also refinements in [HP17]). The modest goal of this paper is to give a quick proof of the analogous result for continuous chains.

Beyond providing a proof of this useful result, we were motivated to write this paper as a way to illustrate how the machinery developed in [ADS19], which shows that mixing times and hitting times are equal up to multiplicative constants for general

Markov processes satisfying regularity conditions, can be used to give fairly quick and simple translations of facts about discrete chains into facts about continuous chains.

2   Notation and Main Results

We fix a compact metric state space X endowed with Borel σ-algebra $\mathcal{B}[X]$ and let $\{P_x^{(t)}(\cdot)\}_{x \in X, t \in \mathbb{N}}$ denote the transition kernel of a Markov process with stationary measure π. Throughout this paper, all transition kernels are assumed to have a unique stationary distribution. We occasionally write $P^{(t)}(x, A)$ for $P_x^{(t)}(A)$. For any $x \in X$, let $P_x^{(0)}(\cdot)$ be the Dirac measure on x. Throughout the paper, we include 0 in $\mathbb{N}$. We write $P_x(A)$ and $P(x, A)$ as an abbreviation for $P_x^{(1)}(A)$ and $P^{(1)}(x, A)$, respectively.

For probability measures μ, ν on $(X, \mathcal{B}[X])$, we denote by

$$\| \mu - \nu \| = \sup_{A \in \mathcal{B}[X]} |\mu(A) - \nu(A)|$$

the usual *total variation distance* between μ and ν.

Definition 2.1   Let $\varepsilon \in \mathbb{R}_{>0}$. The mixing time $t_m(\varepsilon)$ of $\{P_x^{(t)}(\cdot)\}_{x \in X, t \in \mathbb{N}}$ is

$$\min\{t \geq 0 : d(t) \leq \varepsilon\},$$

where $d(t) = \sup_{x \in X} \| P^{(t)}(x, \cdot) - \pi(\cdot) \|$.

We use $\{P_L^{(t)}(x, \cdot)\}_{x \in X, t \in \mathbb{N}}$ to denote the *lazy transition kernel* of $\{P^{(t)}(x, \cdot)\}_{x \in X}$, given by the formula $P_L(x, \cdot) = \frac{1}{2}P(x, \cdot) + \frac{1}{2}\delta(x, \cdot)$ where $\delta(x, \cdot)$ denote the Dirac measure at x. Note that the lazy transition kernel $\{P_L^{(t)}(x, \cdot)\}_{x \in X, t \in \mathbb{N}}$ has the same stationary distribution as $\{P^{(t)}(x, \cdot)\}_{x \in X, t \in \mathbb{N}}$, and we denote its mixing time by $t_L(\varepsilon)$.

In general, it is possible to have $t_m \gg t_L$ due to (near)-periodicity. To avoid these issues, one could instead take an average over two successive steps. This suggests the following definition.

Definition 2.2   For $\varepsilon \in \mathbb{R}_{>0}$, the average mixing time is

$$t_a(\varepsilon) = \min\left\{t \geq 0 : \sup_{x \in X} \| \frac{1}{2}(P^{(t)}(x, \cdot) + P^{(t+1)}(x, \cdot)) - \pi(\cdot) \| \leq \varepsilon\right\}.$$

Recall that a transition kernel $\{P_x^{(t)}(\cdot)\}_{x \in X, t \in \mathbb{N}}$ is *reversible* if

$$\int_A P(x, B)\pi(\mathrm{d}x) = \int_B P(x, A)\pi(\mathrm{d}x).$$

for every $A, B \in \mathcal{B}[X]$. We can now quote the following result from [PS15].

Theorem 2.1 ([PS15, Thm. 1.4])   *For every $0 < \varepsilon \leq \frac{1}{4}$, there exist universal positive constants c_ε and c'_ε so that for every finite reversible Markov process,*

$$c_\varepsilon t_L(\varepsilon) \leq t_a(\varepsilon) \leq c'_\varepsilon t_L(\varepsilon).$$

We generalize Theorem 2.1 to transition kernels with compact metric state space satisfying the following continuity condition.

Definition 2.3   DSF The transition kernel $\{P_x^{(t)}(\cdot)\}_{x \in X, t \in \mathbb{N}}$ satisfies the *strong Feller property* if for every $x \in X$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$(\forall y \in X)(|y - x| < \delta \implies (\|P_x(\cdot) - P_y(\cdot)\| < \varepsilon)).$$

Throughout the paper, we use $\mathcal{C}$ to denote the collection of discrete time reversible transition kernels with compact metric state space satisfying Assumption 2.3. Our main theorem is the following.

Theorem 2.2   *For every $0 < \varepsilon < \frac{1}{4}$, there exist universal constants $d_\varepsilon, d'_\varepsilon$ such that, for every $\{P_x(\cdot)\}_{x \in X} \in \mathcal{C}$, we have*

$$d_\varepsilon t_L(\varepsilon) \le t_a(\varepsilon) \le d'_\varepsilon t_L(\varepsilon).$$

2.1 Equivalent Form of Mixing Times and Hitting times

In this section, we define a quantity that is asymptotically equivalent to the mixing times defined in the previous section. This equivalent form plays an important role throughout the entire paper. Let

$$\overline{d}(t) = \sup_{x, y \in X} \| P^{(t)}(x, \cdot) - P^{(t)}(y, \cdot) \| .$$

Similarly, for $\varepsilon \in \mathbb{R}_{>0}$, define the *standardized mixing time* to be

$$\overline{t}_m(\varepsilon) = \min\{t \ge 0 : \overline{d}(t) \le \varepsilon\},$$

and let $\overline{t}_L(\varepsilon)$ be the analogous quantity for the lazy kernel g_L.

2.2 Nonstandard Analysis and Nonstandard Probability Theory

In this paper, we use nonstandard analysis, powerful machinery derived from mathematical logic, as our main toolkit. For those who are not familiar with nonstandard analysis, [DRW18] and [DR16] provide reviews tailored to probabilists and statisticians. [ACH97, CNOSP95, WL00] provide thorough introductions.

For completeness, we give a brief introduction to nonstandard analysis as well as nonstandard probability theory. This section is taken from [ADS19, Section. 2.2] and [Kei84].

Given any set S, the *superstructure* $\mathbb{V}S$ over S is found by iterating the power set operation countably many times. That is

$$\mathbb{V}_1(S) = S,$$
$$\mathbb{V}_{n+1}(S) = \mathbb{V}_n(S) \sup \mathscr{P}(\mathbb{V}n(S)),$$
$$\mathbb{V}(S) = \bigcup_{n \in \mathbb{N}} \mathbb{V}_n(S).$$

We use $^*: \mathbb{V}(S) \to \mathbb{V}(^*S)$ to denote the nonstandard extension map taking elements, sets, functions, relations, etc., to their nonstandard counterparts. An *internal object* is an element of a set $^*b$ where $b \in \mathbb{V}(S)$. We assume that S contains $\mathbb{R}$ as a subset. In particular, $^*\mathbb{R}$ and $^*\mathbb{N}$ denote the nonstandard extensions of the reals and natural numbers, respectively. An element $r \in {}^*\mathbb{R}$ is *infinite* if $|r| > n$ for every $n \in \mathbb{N}$ and is *finite* otherwise. An element $r \in {}^*\mathbb{R}$ with $r > 0$ is *infinitesimal* if r^{-1} is infinite. For $r, s \in {}^*\mathbb{R}$, we use the notation $r \approx s$ as shorthand for the statement " $|r - s|$ is infinitesimal," and similarly, we use use $r \gtrapprox s$ as shorthand for the statement "either $r \geq s$ or $r \approx s$."

Given a topological space $(X, \mathcal{T})$, the monad of a point $x \in X$ is the set $\bigcap_{U \in \mathcal{T} : x \in U} {}^*U$. An element $x \in {}^*X$ is *near-standard* if it is in the monad of some $y \in X$. We say y is the standard part of x and write $y = \mathrm{st}(x)$. Note that such y is unique. We use $\mathrm{NS}^*(X)$ to denote the collection of near-standard elements of $^*X]$, and we say NS^*X is the *near-standard part* of $^*(X)$. The standard part map st is a function from $\mathrm{NS}^* * (X)$ to X, taking near-standard elements to their standard parts. In both cases, the notation elides the underlying space Y and the topology T, because the space and topology will always be clear from context. For a metric space (X, d), two elements $x, y \in {}^*X$ are *infinitely close* if $^*d(x, y) \approx 0$. An element $x \in {}^*X$ is near-standard if and only if it is infinitely close to some $y \in X$. An element $x \in {}^*X$ is *finite* if there exists $y \in X$ such that $^*d(x, y) < \infty$ and is infinite otherwise.

Let X be a topological space endowed with Borel σ-algebra $\mathcal{B}[X]$ and let $\mathcal{M}X$ denote the collection of all finitely additive probability measures on $(X, \mathcal{B}[X])$. An internal probability measure μ on $(^*X, {}^*\mathcal{B}[X])$ is an element of $^*\mathcal{M}(X)$. Specifically, an internal probability measure μ on $(^*X, {}^*\mathcal{B}[X])$ is an internal function from $^*\mathcal{B}[X] \to {}^*[0,1]$ such that

(1) $\mu(\varnothing) = 0$;
(2) $\mu(^*X) = 1$; and
(3) μ is hyperfinitely additive (that is, it satisfies the usual equality for an additive measure, but with the sum ranging from 1 to any element in $^*\mathbb{N}$).

The Loeb space of the internal probability space $(^*X, {}^*\mathcal{B}[X], \mu)$ is a countably additive probability space $(^*X, \overline{{}^*\mathcal{B}[X]}, \overline{\mu})$ such that

$$\overline{{}^*\mathcal{B}[X]} = \left\{ A \subset {}^*X \,\middle|\, (\forall \varepsilon > 0)(\exists A_i, A_o \in {}^*\mathcal{B}[X])(A_i \subset A \subset A_o \wedge \mu(A_o \smallsetminus A_i) < \varepsilon) \right\}$$

and

$$\overline{\mu}(A) = \sup \left\{ \mathrm{st}(\mu(A_i)) \,\middle|\, A_i \subset A, A_i \in {}^*\mathcal{B}[X] \right\}$$
$$= \inf \left\{ \mathrm{st}(\mu(A_o)) \,\middle|\, A_o \supset A, A_o \in {}^*\mathcal{B}[X] \right\}.$$

Every standard model is closely connected to its nonstandard extension via the *transfer principle*, which asserts that a first order statement is true in the standard model is true if and only if it is true in the nonstandard model. Finally, given a cardinal number κ, a nonstandard model is called κ-saturated if the following condition holds: let $\mathcal{F}$ be a family of internal sets, if $\mathcal{F}$ has cardinality less than κ and $\mathcal{F}$ has the finite intersection property, then the total intersection of $\mathcal{F}$ is non-empty. In this paper,

we assume our nonstandard model is as saturated as we need (see *e.g.,* [ACH97, Thm. 1.7.3] for the existence of κ-saturated nonstandard models for any uncountable cardinal κ).

3 Hyperfinite Representation of Compact Spaces

In this section, we give an overview of hyperfinite representation for compact metric spaces. Hyperfinite representation for more general metric spaces are discussed in [DRW18]. We use similar notation to [DRW18]. For the rest of the paper, we use the common notation $d(x, A) = \inf\{y \in X : d(x, y)\}$ for every $x \in X$ and every $A \subset X$. The formal definition of a hyperfinite representation of a compact metric space is given below.

Definition 3.1   Let (X, d) be a compact metric space. Let $\delta \in {}^*\mathbb{R}^+$ be an infinitesimal. A δ-hyperfinite representation of X is a tuple $(S, \{B(s)\}_{s \in S})$ such that

(i) S is a hyperfinite subset of ${}^*X$;
(ii) $s \in B(s) \in {}^*\mathcal{B}[X]$ for every $s \in S$;
(iii) for every $s \in S$, the diameter of $B(s)$ is no greater than δ;
(iv) $B(s_1) \cap B(s_2) = \varnothing$ for every $s_1 \neq s_2 \in S$;
(v) $\bigcup_{s \in S} B(s) = {}^*X$.

The set S is called the *base set* of the hyperfinite representation of X. For every $x \in \bigcup_{s \in S} B(s)$, we use s_x to denote the unique element in S such that $x \in B(s_x)$.

As discussed in [DRW18], hyperfinite representations exist for more general spaces. For simplicity, we focus on hyperfinite representations of compact metric spaces in this paper. Moreover, by [ADS19, Thm. 3.3], for every compact metric space X and every positive infinitesimal δ, there exists a δ-hyperfinite representation of X.

4 Hyperfinite Representation of Markov Processes

We give a brief introduction of hyperfinite representation of general Markov processes in this section. The construction of such hyperfinite representations is developed in [DRW18] and [ADS19].

Let $\{P_x^{(t)}(\cdot)\}_{x \in X, t \in \mathbb{N}}$ be the transition kernel of a discrete-time Markov process with state space X. We assume that X is a compact metric space for the remainder of the paper unless otherwise mentioned. The transition kernel can be viewed as a function $g : X \times \mathbb{N} \times \mathcal{B}[X] \to [0, 1]$ by letting $g(x, t, A) = P_x^{(t)}(A)$ for every $x \in X$, $t \in \mathbb{N}$ and $A \in \mathcal{B}[X]$. We will use $g(x, t, A)$ and $P_x^{(t)}(A)$ interchangeably. We will construct an internal transition kernel on S to represent the standard transition kernel g. We fix a set $M = \{1, 2, \ldots, K\}$ for some infinite $K \in {}^*\mathbb{N}$ throughout the paper. A *hyperfinite Markov process* is defined analogously to a finite Markov process. Namely, a hyperfinite Markov process is characterized by the following four ingredients:

(1)   a *state space S* that is a non-empty hyperfinite set;
(2)   a time line M;

(3)   a set $\{v_i : i \in S\} \subset {}^*\mathbb{R}$ where each $v_i \geq 0$ and $\sum_{i \in S} v_i = 1$;
(4)   a set $\{p_{ij}\}_{i,j \in S}$ of non-negative hyperreals with $\sum_{j \in S} p_{ij} = 1$ for every $i \in S$.

It was shown in [DRW18, Thm. 7.2] that one can always construct a hyperfinite Markov process given a fixed collection of $\{v_i : i \in S\}$ and $\{p_{ij}\}_{i,j \in S}$.

Let p be a standard probability measure on $(X, \mathcal{B}[X])$ and let $(S, \{B(s)\}_{s \in S})$ be a δ-hyperfinite representation of X for some infinitesimal δ. Define the associated hyperfinite probability measure P (with respect to p) to be the internal probability measure on S by letting $P(\{s\}) = {}^*p(B(s))$ for every $s \in S$. We now quote the following result from [ADS19].

Theorem 4.1 ([ADS19, Thm. 4.16])   *Suppose $g \in \mathcal{C}$ has a stationary measure π. Then there exist some hyperfinite representation $(S, \{B(s)\}_{s \in S})$ of X and some internal transition kernel H on S such that:*

(i)   *the associated internal probability measure Π (with respect to π) is a *stationary distribution of H;*
(ii)   *if g is reversible with respect to π, then H is *reversible with respect to Π;*
(iii)   *for every $t \in \mathbb{N}$, every $s \in \mathrm{NS}(S)$, and every $A \in {}^*\mathcal{B}[X]$, we have*

$$ {}^*g\left(s, t, \bigcup_{a \in A \cap S} B(a)\right) \approx H(s, t, A \cap S); $$

(iv)   *for every $s \in \mathrm{NS}(S)$, every $t \in \mathbb{N}$ and every $E \in \mathcal{B}[X]$, we have*

$$ g(\mathrm{st}(s), t, E) = \overline{H}(s, t, \mathrm{st}^{-1}(E) \cap S). $$

Note that $H(x, 1, \cdot)$ is generally *not* equal to the associated hyperfinite probability measure of $g(x, 1, \cdot)$, though they are closely related.

Lemma 4.2 ([DRW18, Lemma. 8.15]) $\overline{\Pi}(\mathrm{st}^{-1}(A) \cap S) = \pi(A)$ *for every $A \in \mathcal{B}[X]$.*

Define the lazy internal transition kernel $\{H_L(s, 1, \cdot)\}_{s \in S}$ associated with $\{H(s, 1, \cdot)\}_{s \in S}$ to be a collection of internal transition probabilities such that

$$ H_L(i, 1, A) \equiv \sum_{j \in A} H_L(i, 1, \{j\}) $$

$$ = \sum_{j \in A} \left(\frac{1}{2} H(i, 1, \{j\}) + \frac{1}{2}\Delta(i, \{j\})\right) = \frac{1}{2} H(i, 1, A) + \frac{1}{2}\Delta(i, A) $$

for every $i \in S$ and $A \in \mathcal{I}(S)$, where $\Delta(i, A) = 1$ if $i \in A$ and $\Delta(i, A) = 0$ if $i \notin A$.

Theorem 4.3 ([ADS19, Thm. 4.21])   *Suppose $g \in \mathcal{C}$ has a stationary measure π. Then:*

(i)   *For every $t \in \mathbb{N}$, every $x \in \mathrm{NS}({}^*X)$, and every $A \in {}^*\mathcal{B}[X]$, we have ${}^*g_L(x, t, \bigcup_{a \in A \cap S} B(a)) \approx H_L(s_x, t, A \cap S)$ where s_x is the unique element in S such that $x \in B(s_x)$.*
(ii)   *For every $x \in X$, every $t \in \mathbb{N}$, and every $E \in \mathcal{B}[X]$, we have $g_L(x, t, E) = \overline{H}_L(x, t, \mathrm{st}^{-1}(E) \cap S)$.*

5 Mixing Times and Average Mixing Times with Their Nonstandard Counterparts

In this section, we develop nonstandard notions of mixing and average mixing times for hyperfinite Markov processes and we show that the nonstandard notions agree with the standard notions. For the remainder of the paper, we fix $g \in \mathcal{C}$ with stationary measure π. We also fix a hyperfinite representation $(S, \{B(s)\}_{s \in S})$ of X and an internal transition kernel $\{H(s, 1, \cdot)\}_{s \in S}$ on S as in Theorem 4.1.

5.1 Agreement on Mixing Times

The following lemma shows that the mixing time of the lazy chain is no greater than the mixing time of the hyperfinite lazy chain.

Lemma 5.1 ([ADS19, Cor. 5.2])   *For every $\varepsilon \in \mathbb{R}_{>0}$, we have*

$$t_L(\varepsilon) \le T_L(\varepsilon),$$

where $T_L(\varepsilon) = {}^* \min\{t \in M : {}^* \sup_{i \in S} \| H_L(i, t, \cdot) - \Pi(\cdot) \| \le \varepsilon\}$ is the internal mixing time of the lazy version of the hyperfinite chain.

It is desirable to prove the reverse direction of Lemma 5.1. To do this, we introduce the following definition, which replaces a "soft" inequality $\le$ by a "strict" inequality $<$.

Definition 5.1   Let $\varepsilon \in \mathbb{R}_{>0}$. The strict mixing time $t_m^{(<)}(\varepsilon)$ of g is

$$\min\{t \ge 0 : d(t) < \varepsilon\},$$

where $d(t) = \sup_{x \in X} \| g(x, t, \cdot) - \pi(\cdot) \|$.

For every $\varepsilon > 0$, we write $t_L^{(<)}(\varepsilon)$ to denote the strict mixing time of the lazy chain.

Theorem 5.2   *For every $\varepsilon \in \mathbb{R}_{>0}$:*

$$t_L^{(<)}(\varepsilon) \ge T_L^{(<)}(\varepsilon),$$

where $T_L^{(<)}(\varepsilon) = {}^* \min\{t \in M : {}^* \sup_{i \in S} \| H_L(i, t, \cdot) - \Pi(\cdot) \| < \varepsilon\}$ is the internal strict mixing time of the lazy version of the hyperfinite chain.

Proof   Pick $\varepsilon \in \mathbb{R}_{>0}$. For every $t \in \mathbb{N}$,

$$
\begin{aligned}
{}^* \sup_{x \in {}^*X} \| {}^* g_L(x, t, \cdot) - {}^* \pi(\cdot) \| \\
\ge {}^* \sup_{i \in S} \| {}^* g_L(i, t, \cdot) - {}^* \pi(\cdot) \| \\
= {}^* \sup_{i \in S} {}^* \sup_{A \in {}^* \mathcal{B}[X]} |{}^* g_L(i, t, A) - {}^* \pi(A)| \\
\ge {}^* \sup_{i \in S} {}^* \sup_{A \in \mathcal{J}(S)} \left| {}^* g_L\left(i, t, \bigcup_{a \in A} B(a)\right) - {}^* \pi\left(\bigcup_{a \in A} B(a) \right) \right| \\
\overset{\text{Thm. 4.3}}{\approx} {}^* \sup_{i \in S} {}^* \sup_{A \in \mathcal{J}(S)} |H_L(i, t, A) - \Pi(A)| \\
= {}^* \sup_{i \in S} \| H_L(i, t, \cdot) - \Pi(\cdot) \| .
\end{aligned}
$$

Let t_0 be a natural number such that $^*\sup_{x\in{}^*X}\|\,{}^*g_L(x,t_0,\cdot)-{}^*\pi(\cdot)\,\|<\varepsilon$. Then

$$^*\sup_{i\in S}\|\,H_L(i,t_0,\cdot)-\Pi(\cdot)\,\|\stackrel{*}{\lesssim}{}^*\sup_{x\in{}^*X}\|\,{}^*g_L(x,t_0,\cdot)-{}^*\pi(\cdot)\,\|$$
$$=\sup_{x\in X}\|\,g_L(x,t_0,\cdot)-\pi(\cdot)\,\|<\varepsilon.$$

Hence, we have the desired result.   ∎

5.2 Agreement of Average Mixing Time

In this section, we show that the hyperfinite average mixing time is equivalent to the standard average mixing time. We begin by showing that standard average mixing times are no greater than hyperfinite average mixing times.

Theorem 5.3   *For every $\varepsilon\in\mathbb{R}_{>0}$:*

$$t_a(\varepsilon)\le\min\left\{t\in M:\sup_{i\in S}\operatorname{st}\left(\left\|\frac{1}{2}\big(H(i,t,\cdot)+H(i,t+1,\cdot)\big)-\Pi(\cdot)\right\|\right)\le\varepsilon\right\}.$$

Proof   Pick $\varepsilon\in\mathbb{R}_{>0}$. We have

$$\sup_{i\in S}\operatorname{st}\left(\left\|\frac{H(i,t,\cdot)+H(i,t+1,\cdot)}{2}-\Pi(\cdot)\right\|\right)$$
$$\ge\sup_{x\in X}\operatorname{st}\left(\sup_{A\in{}^*\mathcal{B}[X]}\left|\frac{H(x,t,A)+H(x,t+1,A)}{2}-\Pi(A)\right|\right).$$

By the construction of Loeb measure, for every $B\in\overline{{}^*\mathcal{B}[X]}$, there exists $B'\in{}^*\mathcal{B}[X]$ such that $B\triangle B'$ has Loeb measure 0. For every $A\in\mathcal{B}[X]$, the set $\operatorname{st}^{-1}(A)$ is Loeb measurable; thus,

$$\sup_{x\in X}\operatorname{st}\left(\sup_{A\in{}^*\mathcal{B}[X]}\left|\frac{H(x,t,A)+H(x,t+1,A)}{2}-\Pi(A)\right|\right)$$
$$\ge\sup_{x\in X}\sup_{A\in\mathcal{B}[X]}\left|\frac{\overline{H}(x,t,\operatorname{st}^{-1}(A)\cap S)+\overline{H}(x,t+1,\operatorname{st}^{-1}(A)\cap S)}{2}-\overline{\Pi}(\operatorname{st}^{-1}(A)\cap S)\right|.$$

By Theorem 4.1 and Lemma 4.2, we have

$$\sup_{x\in X}\sup_{A\in\mathcal{B}[X]}\left|\frac{\overline{H}(x,t,\operatorname{st}^{-1}(A)\cap S)+\overline{H}(x,t+1,\operatorname{st}^{-1}(A)\cap S)}{2}-\overline{\Pi}(\operatorname{st}^{-1}(A)\cap S)\right|$$
$$=\sup_{x\in X}\sup_{A\in\mathcal{B}[X]}\left|\frac{g(x,t,A)+g(x,t+1,A)}{2}-\pi(A)\right|$$
$$=\sup_{x\in X}\left\|\frac{g(x,t,\cdot)+g(x,t+1,\cdot)}{2}-\pi(\cdot)\right\|.$$

Hence, we know that

$$t_a(\varepsilon)\le\min\left\{t\in M:\sup_{i\in S}\operatorname{st}\left(\left\|\frac{H(i,t,\cdot)+H(i,t+1,\cdot)}{2}-\Pi(\cdot)\right\|\right)\le\varepsilon\right\},$$

completing the proof.   ∎

The following result is an immediate consequence of Lemma 5.3.

Corollary 5.4 *For every $\varepsilon \in \mathbb{R}_{>0}$, we have*

$$t_a(\varepsilon) \leq T_a(\varepsilon),$$

where

$$T_a(\varepsilon) = {}^*\min\left\{t \in M : {}^*\sup_{i \in S} \left\| \frac{H(i,t,\cdot) + H(i,t+1,\cdot)}{2} - \Pi(\cdot) \right\| \leq \varepsilon\right\}$$

is the internal average mixing time of the hyperfinite chain.

Proof Pick $\varepsilon \in \mathbb{R}_{>0}$. Note that if

$$ {}^*\sup_{i \in S} \left\| \frac{H(i,t,\cdot) + H(i,t+1,\cdot)}{2} - \Pi(\cdot) \right\| \leq \varepsilon,$$

then

$$\sup_{i \in S} \operatorname{st}\left(\left\| \frac{H(i,t,\cdot) + H(i,t+1,\cdot)}{2} - \Pi(\cdot) \right\| \right) \leq \varepsilon.$$

The result then follows from Lemma 5.3. ∎

It is desirable to prove the reverse direction of Corollary 5.4. To do this, we introduce the following definition.

Definition 5.2 For $\varepsilon \in \mathbb{R}_{>0}$, the strict average mixing time is

$$t_a^{(<)}(\varepsilon) = \min\left\{t \in M : \sup_{x \in X} \left\| \frac{g(x,t,\cdot) + g(x,t+1,\cdot)}{2} - \pi(\cdot) \right\| < \varepsilon\right\}.$$

Theorem 5.5 *For every $\varepsilon \in \mathbb{R}_{>0}$,*

$$t_a^{(<)}(\varepsilon) \geq T_a^{(<)}(\varepsilon),$$

where

$$T_a^{(<)}(\varepsilon) = {}^*\min\left\{t \in M : {}^*\sup_{i \in S} \left\| \frac{H(i,t,\cdot) + H(i,t+1,\cdot)}{2} - \Pi(\cdot) \right\| < \varepsilon\right\}$$

is the internal strict average mixing time of the hyperfinite chain.

Proof Pick $\varepsilon \in \mathbb{R}_{>0}$. For every $t \in \mathbb{N}$, by Theorem 4.1, we have

$$ {}^*\sup_{x \in {}^*X} \left\| \frac{{}^*g(x,t,\cdot) + {}^*g(x,t+1,\cdot)}{2} - {}^*\pi(\cdot) \right\|$$

$$\geq {}^*\sup_{i \in S} \left\| \frac{{}^*g(i,t,\cdot) + {}^*g(i,t+1,\cdot)}{2} - {}^*\pi(\cdot) \right\|$$

$$= {}^*\sup_{i \in S} {}^*\sup_{A \in {}^*\mathcal{B}[X]} \left| \frac{{}^*g(i,t,A) + {}^*g(i,t+1,A)}{2} - {}^*\pi(A) \right|$$

$$\geq {}^*\sup_{i \in S} {}^*\sup_{A \in \mathcal{J}(S)} \left| \frac{{}^*g(i,t,\bigcup_{a \in A} B(a)) + {}^*g(i,t+1,\bigcup_{a \in A} B(a))}{2} \right.$$

$$- {}^*\pi\Big(\bigcup_{a\in A} B(a)\Big)\Big|$$

$$\approx {}^*\sup_{i\in S} {}^*\sup_{A\in\mathcal{J}(S)}\Big|\frac{H(i,t,A)+H(i,t+1,A)}{2}-\Pi(A)\Big|$$

$$= {}^*\sup_{i\in S}\Big\|\frac{H(i,t,\cdot)+H(i,t+1,\cdot)}{2}-\Pi(\cdot)\Big\|.$$

Let t_0 be a natural number such that $t_a^{(<)}(\varepsilon)\le t$. Then

$${}^*\sup_{i\in S}\Big\|\frac{H(i,t_0,\cdot)+H(i,t_0+1,\cdot)}{2}-\Pi(\cdot)\Big\|$$

$$\lesssim {}^*\sup_{x\in{}^*X}\Big\|\frac{{}^*g(x,t_0,\cdot)+{}^*g(x,t_0+1,\cdot)}{2}-{}^*\pi(\cdot)\Big\|$$

$$= \sup_{x\in X}\Big\|\frac{g(x,t_0,\cdot)+g(x,t_0+1,\cdot)}{2}-\pi(\cdot)\Big\| < \varepsilon.$$

Hence, we have the desired result.   ∎

6 Mixing Times and Average Mixing Times on Compact Sets

In this section, we prove our main result, Theorem 2.2. The following well-known equivalence follows from submultiplicativity of $\overline{d}(t)$ and the fact that $d(t)\le\overline{d}(t)<2d(t)$ (see Lemmas 4.11 and 4.12 from [LPW09]). Recall that t_m denotes the mixing time of a Markov process (see Definition 2.1).

Lemma 6.1   *For every $0<\varepsilon_1<\varepsilon_2<\frac{1}{2}$, there exists a positive universal constant $c_{\varepsilon_1\varepsilon_2}$ such that*

$$t_m(\varepsilon_2)\le t_m(\varepsilon_1)\le c_{\varepsilon_1\varepsilon_2}t_m(\varepsilon_2)$$

for every Markov process with a unique stationary distribution.

We have the following result for strict mixing times and strict average mixing times.

Lemma 6.2   *For every $0<\varepsilon\le\frac{1}{4}$, there exist universal positive constants e_ε and e'_ε so that for every finite reversible Markov process*

$$e_\varepsilon t_L^{(<)}(\varepsilon)\le t_a^{(<)}(\varepsilon)\le e'_\varepsilon t_L^{(<)}(\varepsilon).$$

Proof   Pick some $0<\varepsilon<\frac{1}{4}$ and let $\varepsilon_0=\frac{\varepsilon}{2}$. By Theorem 2.1, we have

$$t_a^{(<)}(\varepsilon)\ge t_a(\varepsilon)\ge c_\varepsilon t_L(\varepsilon).$$

As the lazy chain of a reversible Markov process is reversible, by Lemma 6.1, we have

$$c_\varepsilon t_L(\varepsilon)\ge c_\varepsilon\frac{1}{c_{\varepsilon\varepsilon_0}}t_L(\varepsilon_0)\ge c_\varepsilon\frac{1}{c_{\varepsilon\varepsilon_0}}t_L^{(<)}(\varepsilon).$$

Let $e_\varepsilon = c_\varepsilon \frac{1}{c_{\varepsilon\varepsilon_0}}$. Then

$$e_\varepsilon t_L^{(<)}(\varepsilon) \le t_a^{(<)}(\varepsilon).$$

We now prove the other direction. By Theorem 2.1 and Lemma 6.1, we have

$$t_a^{(<)}(\varepsilon) \le t_a(\varepsilon_0) \le c'_{\varepsilon_0} t_L(\varepsilon_0) \le c'_{\varepsilon_0} c_{\varepsilon\varepsilon_0} t_L(\varepsilon) \le c'_{\varepsilon_0} c_{\varepsilon\varepsilon_0} t_L^{(<)}(\varepsilon).$$

By letting $e'_\varepsilon = c'_{\varepsilon_0} c_{\varepsilon\varepsilon_0}$, we have the desired result. ∎

We now prove Theorem 2.2 for strict mixing times and strict average mixing times.

Theorem 6.3 *For every $0 < \varepsilon \le \frac{1}{4}$, there exist universal constants $h_\varepsilon, h'_\varepsilon$ such that, for every $\{g(x,1,\cdot)\}_{x\in X} \in \mathcal{C}$, we have*

$$h_\varepsilon t_L^{(<)}(\varepsilon) \le t_a^{(<)}(\varepsilon) \le h'_\varepsilon t_L^{(<)}(\varepsilon).$$

Proof Pick some $0 < \varepsilon < \frac{1}{4}$. Let $\varepsilon_0 = \frac{\varepsilon}{2}$. By Theorem 5.5, we have $t_a^{(<)}(\varepsilon) \ge T_a^{(<)}(\varepsilon)$. By the transfer of Lemma 6.2, we have $T_a^{(<)}(\varepsilon) > e_\varepsilon T_L^{(<)}(\varepsilon)$. By the transfer of Lemmas 6.1 and 5.1, we have

$$
\begin{aligned}
T_L^{(<)}(\varepsilon) &= e_\varepsilon{}^* \min\big\{ t \in M : {}^* \sup_{i\in S} \| H_L(i,t,\cdot) - \Pi(\cdot) \| < \varepsilon \big\} \\
&\ge e_\varepsilon{}^* \min\big\{ t \in M : {}^* \sup_{i\in S} \| H_L(i,t,\cdot) - \Pi(\cdot) \| \le \varepsilon \big\} \\
&\ge e_\varepsilon \frac{1}{c_{\varepsilon\varepsilon_0}} {}^* \min\big\{ t \in M : {}^* \sup_{i\in S} \| H_L(i,t,\cdot) - \Pi(\cdot) \| \le \varepsilon_0 \big\} \\
&\ge e_\varepsilon \frac{1}{c_{\varepsilon\varepsilon_0}} t_L(\varepsilon_0) \ge e_\varepsilon \frac{1}{c_{\varepsilon\varepsilon_0}} t_L^{(<)}(\varepsilon).
\end{aligned}
$$

Let $h_\varepsilon = e_\varepsilon \frac{1}{c_{\varepsilon\varepsilon_0}}$. We then have $h_\varepsilon t_L^{(<)}(\varepsilon) \le t_a^{(<)}(\varepsilon)$.

We now prove the other direction. By Corollary 5.4, we have $t_a^{(<)}(\varepsilon) \le t_a(\varepsilon_0) \le T_a(\varepsilon_0)$. By the transfer of Theorem 2.1 and the transfer of Lemma 6.1, we have

$$T_a(\varepsilon_0) \le c'_{\varepsilon_0} T_L(\varepsilon_0) \le c'_{\varepsilon_0} c_{\varepsilon\varepsilon_0} T_L(\varepsilon).$$

By Lemma 5.2, we have

$$c'_{\varepsilon_0} c_{\varepsilon\varepsilon_0} T_L(\varepsilon) \le c'_{\varepsilon_0} c_{\varepsilon\varepsilon_0} T_L^{(<)}(\varepsilon) \le c'_{\varepsilon_0} c_{\varepsilon\varepsilon_0} t_L^{(<)}(\varepsilon).$$

By letting $h'_\varepsilon = c'_{\varepsilon_0} c_{\varepsilon\varepsilon_0}$, we have $t_a^{(<)}(\varepsilon) \le h'_\varepsilon t_L^{(<)}(\varepsilon)$, completing the proof. ∎

We now prove Theorem 2.2.

Proof Pick some $0 < \varepsilon \le \frac{1}{4}$. Let $\varepsilon_0 = \frac{\varepsilon}{2}$ and let $\varepsilon_1 = \varepsilon + \frac{\frac{1}{4}-\varepsilon}{2}$. By Theorem 6.3 and Lemma 6.1, we have

$$t_a(\varepsilon) \le t_a^{(<)}(\varepsilon) \le h'_\varepsilon t_L^{(<)}(\varepsilon) \le h'_\varepsilon t_L(\varepsilon_0) \le h'_\varepsilon c_{\varepsilon\varepsilon_0} t_L(\varepsilon).$$

Let $d'_\varepsilon = h'_\varepsilon c_{\varepsilon\varepsilon_0}$. We have $t_a(\varepsilon) \le d'_\varepsilon t_L(\varepsilon)$.

We now prove the other direction. By Theorem 6.3 and Lemma 6.1, we have

$$t_a(\varepsilon) \ge t_a^{(<)}(\varepsilon_1) \ge h_{\varepsilon_1} t_L^{(<)}(\varepsilon_1) \ge h_{\varepsilon_1} t_L(\varepsilon_1) \ge h_{\varepsilon_1} \frac{1}{c_{\varepsilon\varepsilon_1}} t_L(\varepsilon).$$

Let $d_\varepsilon = h_{\varepsilon_1} \frac{1}{c_{\varepsilon\varepsilon_1}}$. We have $t_a(\varepsilon) \ge d_\varepsilon t_L(\varepsilon)$, completing the proof.   ■

References

[ACH97]   L. O. Arkeryd, N. J. Cutland, and C. W. Henson (eds.), *Nonstandard analysis*. In: Theory and applications, NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences, 493, Kluwer Academic Publishers Group, Dordrecht, 1997. http://dx.doi.org/10.1007/978-94-011-5544-1

[ADS19]   R. M. Anderson, H. Duanmu, and A. Smith, *Mixing times and hitting times for general Markov processes*. Preprint, 2019.   arXiv:1810.06087

[CNOSP95]   N. J. Cutland, V. Neves, F. Oliveira, and J. Sousa-Pinto (eds.), *Developments in nonstandard mathematics*. Papers from the International Colloquium (CIMNS94) held in memory of Abraham Robinson at the University of Aveiro, Aveiro, July 18–22, 1994. Pitman Research Notes in Mathematics Series, 336, Longman, Harlow, 1995.

[DR16]   H. Duanmu and D. M. Roy, *On extended admissible procedures and their nonstandard Bayes risk*. Preprint, 2016.   arXiv:1612.09305

[DRW18]   H. Duanmu, J. S. Rosenthal, and W. Weiss, *Ergodicity of markov processes via non-standard analysis*. Mem. Amer. Math. Soc., to appear, 2018.

[HP17]   J. Hermon and Y. Peres, *The power of averaging at two consecutive time steps: Proof of a mixing conjecture by Aldous and Fill*. Ann. Inst. Henri Poincaré, Probab. Stat. 53(2017), 2030–2042.   http://dx.doi.org/10.1214/16-AIHP782

[Kei84]   H. J. Keisler, *An infinitesimal approach to stochastic analysis*. Mem. Amer. Math. Soc. 48(1984), no. 297.   http://dx.doi.org/10.1090/memo/0297

[LPW09]   D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*. With a chapter by James G. Propp and David B. Wilson. American Mathematical Society, Providence, RI, 2009.

[PS15]   Y. Peres and P. Sousi. *Mixing times are hitting times of large sets*. J. Theoret. Probab. 28(2015), 488–519.   http://dx.doi.org/10.1007/s10959-013-0497-9

[WL00]   M. Wolff and P. A. Loeb (eds.), *Nonstandard analysis for the working mathematician*. Mathematics and its Applications, 510, Kluwer Academic Publishers, Dordrecht, 2000. http://dx.doi.org/10.1007/978-94-4168-0

Department of Economics, University of California, Berkeley, CA
e-mail: robert.anderson@berkeley.edu   asmi28@uottawa.ca

Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada
e-mail: duanmuhaosui@berkeley.edu