ERGODICITY OF MARKOV PROCESSES VIA NONSTANDARD ANALYSIS

HAOSUI DUANMU, JEFFREY S. ROSENTHAL, AND WILLIAM.WEISS

Abstract. The Markov chain ergodic theorem is well-understood if either the time-line or the state space is discrete. However, there does not exist a very clear result for general state space continuous-time Markov processes. Using methods from mathematical logic and nonstandard analysis, we introduce a class of hyperfinite Markov processes-namely, general Markov processes which behave like finite state space discrete-time Markov processes. We show that, under moderate conditions, the transition probability of hyperfinite Markov processes align with the transition probability of standard Markov processes. The Markov chain ergodic theorem for hyperfinite Markov processes will then imply the Markov chain ergodic theorem for general state space continuous-time Markov processes.

Contents

1. Introduction

The transition probability of a time-homogeneous Markov process with a stationary probability distribution π will converge to π in an appropriate sense (i.e., will be "ergodic"), under suitable conditions (such as "irreducibility"). This phenomenon is well understood for processes in discrete time and space (see e.g. [Bil95; GS01]), and for processes in continuous time and discrete space (see e.g. [GS01]), and for processes in discrete time and continuous space (see e.g. [MT09] and [RR04]). However, for processes in continuous time and space, there are apparently no such clean results; the closest are apparently the results in [MT93a; MT93b; MT09] using awkward assumptions about skeleton chains together with drift conditions. Other existing results (see,e.g., [Ste94]) make extensive use of the techniques and results from [MT93a; MT93b].

Meanwhile, nonstandard analysis is a useful tool for providing intuitive new proofs as well as new results to all areas of mathematics, including probability and stochastic processes (see, eg.,[ACH97; Kei84; LW15]). One of the strengths of nonstandard analysis is to provide a direct passage to link discrete mathematical results to continuous mathematical results. This link is usually established by using "hyperfinite" sets which is an infinite set with the same basic logical properties as a finite set. Hence, they usually serve as a good approximation of general sets in nonstandard analysis.

In this new paper, we apply nonstandard analysis to general state space continuous time Markov processes. For a continuous Markov process $\{X_t\}_{t\geq 0}$ with general state space X, we will construct a nonstandard counterpart $\{X'_t\}_{t\in T}$ (which is called a hyperfinite Markov process). This nonstandard characterization $\{X'_t\}_{t\in T}$ will allow us to view every Markov chain as a "discrete" process. The time line $[0,\infty)$ of $\{X_t\}$ is approximated by a hyperfinite set $T=\{0,\delta t,2\delta t,\dots,K\}$ where δt is some positive infinitesimal and K is some infinite number. We then take the nonstandard extension of X and "cut" it into hyperfinitely pieces of mutually disjoint *Borel sets with infinitesimal diameters. For example, if the state space X is

$\mathbb{R}^n$ then we "cut" X into "rectangles" of the form $\{x \in {}^*\mathbb{R}^n : a \leq x_1 < a+\delta, a \leq x_2 < a+\delta, \ldots, a \leq x_n < a+\delta\}$ for some $a \in {}^*\mathbb{R}$ and some positive infinitesimal δ. We then pick one point from each of these *Borel sets to form a hyperfinite set S. The collection of these *Borel sets is usually denoted by $\{B(s) : s \in S\}$. The set S is called the "hyperfinite representation" of $^*X$ (the nonstandard extension of the state space X). The link between X and S are usually established by the standard part map st. The standard part of an element $x \in {}^*X$ is the unique element $\mathsf{st}(x) \in X$ such that x is infinitesimally close to $\mathsf{st}(x)$. The standard part map st maps points in $\{^*X\}$ to their standard part. Under moderate conditions, it can be shown that the measure of $E \subset X$ is the same as the corresponding measure of $\mathsf{st}^{-1}(E) \cap S$ (see, eg,.Lemma 6.8). There has been a rich literature on using hyperfinite measure spaces to represent standard measure spaces. (see, eg., [And82; Loe74]). In [And82], he gave a "hyperfinite representation" for every Hausdorff regular space with Borel σ-algebra. In this paper, we focus on σ-compact completely metric spaces hence obtaining a tighter control on our hyperfinite representation S.

The internal transition probability $\{G_{s_1}^{(\delta t)(\{s_2\})}\}_{s_1,s_2 \in S}$ of $\{X'_t\}_{t \in T}$ is defined to be $\{^*P_{s_1}^{(\delta t)}(B(s_2))\}_{s_1,s_2 \in S}$ with some minor modification. Roughly speaking, we obtain the internal transition probability of $\{X'_t\}_{t \in T}$ by considering the *transition probability from s_1 to $B(s_2)$ at time δt and collapse the mass to one point s_2. Hyperfinite Markov processes behave like Markov processes on finite state spaces with discrete time lines in many ways due to the close connection between finite sets and hyperfinite sets. Most of the concepts of hyperfinite Markov processes are naturally inherited from discrete Markov processes with finite state spaces. Meanwhile, $\{X'_t\}_{t \in T}$ also inherit most of the key properties of $\{X_t\}_{t \geq 0}$. Most importantly, the internal transition probability of $\{X'_t\}_{t \in T}$ agree with the transition probability of $\{X_t\}_{t \geq 0}$ via standard part mapping. Namely,for every Borel set E, every $x \in X$ and every $t > 0$ we have

$$(\forall s \approx x)(\forall t' \approx t)(P_x^{(t)}(E) = G_s^{(t')}(\mathsf{st}^{-1}(E))) \tag{1.1}$$

Thus, we refer $\{X'_t\}_{t\in T}$ as a "hyperfinite representation" of $\{X_t\}_{t\geq 0}$. Moreover, if π is a stationary distribution for $\{X_t\}_{t\geq 0}$, define $\pi'(\{s\}) = {}^*\pi(B(s))$ for every $s \in S$, it then follows that π' is "almost" a stationary distribution for $\{X'_t\}_{t\in T}$. Under moderate assumptions of $\{X_t\}_{t\geq 0}$, we can then show that $G_s^{(t)}(A) \approx \pi'(A)$ for all infinite t and all internal set A. Finally, we can push down this result to show that the transition probability of $\{X_t\}_{t\geq 0}$ converges to π in total variation distance hence establishing the Markov chain ergodic theorem for general state space continuous-time Markov processes.

The method used in this paper is an "up and down" argument. In probability theory and stochastic processes, it is usually easier to deal with discrete probability theory as well as discrete time stochastic processes. By using nonstandard analysis, we first "push up" the problem into the nonstandard universe and consider the hyperfinite counterpart of this problem. We can usually solve the hyperfinite counterpart of the problem by mimicking the method we used in solving the finite version of the problem. Once we solve the hyperfinite counterpart of the problem, we "push down" to obtain the desired result for our original problem. We believe that this method can be applied to many other areas in modern mathematics.

1.1. **Section Outline.** We conclude the introduction with a section-by-section summary, along the way mentioning some important results proved in the paper.

In Section 2 we give a short introduction to Markov processes. We start by introducing finite state space discrete time Markov processes and then move to more general Markov processes. We also give proofs to some basic facts of Markov processes. Most of the definitions as well as notations are adapted from [Ros06]. We state the main result of this paper at the end of section 1 (see Theorem 2.16).

In Section 3 we develop from the beginning the notions needed for nonstandard analysis, including the Extension, Transfer and Saturation Principles, internal sets and internal definition principles. The easiest way to visualize nonstandard analysis is to consider the $\mathbb{R}$ and its nonstandard extension ${}^*\mathbb{R}$. Hence, in Section 3.1, we introduce basic concepts in ${}^*\mathbb{R}$ including infinitesimals, infinite numbers, near-standard numbers,etc. For readers who are unfamiliar with nonstandard analysis, it is usually easy to make mistakes when it comes to identifying internal sets. In Example 3.18, we show that the set $\mathsf{st}^{-1}(\{0\})$ consisting of all infinitesimals is an external set. In Section 3.2, we generalized those concepts and notations developed in Section 3.1 to more general topological spaces.

In Section 4 we give an introduction to nonstandard measure theory. The nonstandard measure theory is formulated by Peter Loeb in his landmark paper [Loe75]. In [Loe75], Loeb constructed a standard countably additive probability space (called the Loeb space) which is the completion of some "nonstandard measure space" (called an internal probability space). We start Section 4 by introducing internal probability spaces followed by an explicit construction of Loeb spaces. A particular interesting class of internal probability spaces is the class consisting of hyperfinite probability spaces. Hyperfinite sets are infinite sets with the same first-order logic properties as finite sets. Hyperfinite probability spaces are simply internal probability spaces with hyperfinite sample space. Hyperfinite probability spaces can often serve as a "good representations" of standard probability spaces. We illustrate this idea in Example 4.5 and the remark after it. We also discuss nonstandard product measures and nonstandard integration theory in this section.

In Section 5, we discuss the measurability issue of the standard part map. The link between a standard probability space X and its hyperfinite representation S_X is usually established via the standard part map. Thus, it is natural to require that st (standard part map) to be a measurable function. In other words, we would like to find out conditions such that $\mathsf{st}^{-1}(E)$ is Loeb measurable for every Borel set E. In [LR87], it has been shown that this question largely depends on the Loeb measurability of $\mathrm{NS}({}^*X) = \{x \in {}^*X : (\exists y \in X)(y = \mathsf{st}(x))\}$. In Section 5, we investigate the conditions such that $\mathrm{NS}({}^*X)$ is Loeb measurable. In [ACH97, Exercise 4.19,1.20], $\mathrm{NS}({}^*X)$ is Loeb measurable if X is either a σ-compact, a locally compact Hausdorff or a complete metric space. We give a proof for the σ-compact case in Lemma 5.5. We are also able to further extend the result to merely Cech complete spaces (see Theorem 5.6).

In Section 6, we formally introduce the idea of hyperfinite representation. In Definition 6.3, we give the definition of hyperfinite representation of a metric spaces satisfying the Heine-Borel condition. The idea is to decompose X into hyperfinitely many *Borel sets with infinitesimal diameters and pick one representative from every such *Borel set. We usually denote the hyperfinite representation by S and the hyperfinite collection of *Borel sets by $\{B(s) : s \in S\}$. Note that it is generally impossible to decompose the space into hyperfinitely many *Borel sets with infinitesimal diameters. Thus, we only require our hyperfinite collection $\{B(s) : s \in S\}$ of *Borel sets to cover a "large enough" portion of ${}^*X$. A hyperfinite representation S has two parameters r and ϵ. The parameter r measures how large portion does $\{B(s) : s \in S\}$ cover while ϵ puts an upper bound on the diameters of $\{B(s) : s \in S\}$. Given an (ϵ, r)-hyperfinite representation S, in Theorem 6.11, we define an internal probability measure P' on $(S, \mathcal{I}[S])$ and establishes the link between $(X, \mathcal{B}[X], P)$ and $(S, \mathcal{I}(S), P')$. Theorem 6.11 is similar to [Cut+95, Theorem 3.5 page 159] which was proved in [And82].

In Section 7, we define hyperfinite Markov processes and investigate many of its properties. A hyperfinite Markov chain is characterized by four ingredients:

- a hyperfinite state space S.
- an initial distribution $\{\nu_i\}_{i \in S}$ consisting of non-negative hyperreals summing to 1.
- a hyperfinite time line $T = \{0, \delta t, \ldots, K\}$ for some infinitesimal δt and some infinite $K \in {}^*\mathbb{R}$.
- transition probabilities $\{p_{ij}\}_{i,j \in S}$ consisting of non-negative hyperreals with $\sum_{j \in S} p_{ij} = 1$ for all $i \in S$.

In other words, hyperfinite Markov processes behave much like discrete-time Markov processes with finite state spaces. The Markov chain ergodic theorem for discrete-time Markov processes with finite state spaces is proved using the "coupling" technique. Namely, for finite Markov processes, we can show that two i.i.d Markov chains starting at different points will eventually "couple" at the same point under moderate conditions. Similarly, for hyperfinite Markov processes, we can show that two i.i.d copies starting at different points will eventually get infinitesimally close. This infinitesimal coupling technique is illustrated in Lemma 7.8. In Theorems 7.19 and 7.26, we establish ergodic theorems for hyperfinite Markov processes.

In Section 8, we construct hyperfinite representations for discrete-time Markov processes. Given a discrete-time Markov process $\{X_t\}_{t\in\mathbb{N}}$, we construct a hyperfinite Markov process $\{X'_t\}_{t\in T}$ such that the internal transition probability of $\{X'_t\}_{t\in T}$ deviate from the transition probability of $\{X_t\}_{t\geq 0}$ by infinitesimal. $\{X'_t\}_{t\in T}$ is defined on some hyperfinite representation S of X. Note that the time line T of $\{X'_t\}_{t\in T}$ in this case will be $\{1,2,\dots,K\}$ for some infinite $K \in {}^*\mathbb{N}$. At each step, an infinitesimal difference between the internal transition probability of $\{X'_t\}_{t\in T}$ and the transition probability of $\{X_t\}$ is generated. As there are only countably many steps, the internal transition probability give a reasonably well approximation for the transition probability of of $\{X_t\}_{t\in\mathbb{N}}$. We illustrate such result in Theorem 8.16.

In Section 9, we apply similar ideas developed in Section 8 to continuous-time Markov processes with general state spaces. However, the construction of hyperfinite representation for a continuous-time Markov process $\{X_t\}_{t\geq 0}$ is much more complicated compared with the construction in Section 8. When the time-line is continuous, the time-line T for the hyperfinite representation is $\{0,\delta t,2\delta t,\dots,K\}$ where δt is some infinitesimal and K is some infinite number. As it takes hyperfinitely many infinitesimal steps to reach a non-infinitesimal time, we need to make sure that the difference between $\{X_t\}_{t\geq 0}$ and $\{X'_t\}_{t\in T}$ generated in every step is so small such that the accumulated difference will remain infinitesimal. We establish this by using internal induction principle (see Theorem 9.20). Unlike the construction of $\{X'_t\}_{t\in T}$ in Section 8, the construction of $\{X'_t\}_{t\in T}$ in Section 9 involves picking the underlying hyperfinite state space S carefully. Finally, we establish the connection between $\{X_t\}_{t\geq 0}$ and $\{X'_t\}_{t\in T}$ in Theorem 9.43.

In Section 10, we establish the Markov chain ergodic theorem for continuous-time general state space Markov processes. We show that the hyperfinite representation $\{X'_t\}_{t\in T}$ inherit many key properties from $\{X_t\}_{t\geq 0}$ (see Theorem 10.6 and Lemmas 10.8 and 10.15). By Theorem 7.26, we know that $\{X'_t\}_{t\in T}$ is ergodic. The ergodicity of $\{X_t\}_{t\in T}$ (Theorem 10.16) follows from pushing down Theorem 7.26.

One of the major assumptions on $\{X_t\}_{t\geq 0}$ is the strong Feller property which asserts that transition probability of $\{X_t\}_{t\geq 0}$ is a continuous function of the starting points with respect to the total variation distance. It is desirable to weaken this condition to only assert that the transition probability is a continuous function of the starting points for every Borel set (such condition is called the Feller condition). In Section 11, we establish how to construct a hyperfinite representation $\{X'_t\}_{t\in T}$ of $\{X_t\}_{t\geq 0}$ when $\{X_t\}_{t\geq 0}$ just satisfies the Feller condition. We also give a proof of a weaker Markov chain ergodic theorem under the Feller condition. It remains unclear to us whether the Markov chain ergodic theorem is true when $\{X_t\}_{t\geq 0}$ only satisfies the Feller condition.

In Section 12, we discuss how to construct standard Markov processes and stationary distributions from hyperfinite Markov processes and weakly stationary distributions ("stationary" distributions for hyperfinite Markov processes). This also gives rise to some new insights in establishing existence of stationary distributions for general Markov processes. A Markov process $\{X_t\}_{t\geq 0}$ satisfies the merging property if for all $x, y \in X$

$$\lim_{t\to\infty} \parallel P_x^{(t)}(\cdot) - P_y \parallel = 0. \tag{1.2}$$

Note that a Markov process with the merging property does not need to have a stationary distribution. In Section 13, we discuss conditions on $\{X_t\}_{t\geq 0}$ for it to have the merging property. In Section 14, we close with a few remarks, some open problems and a short literature review on existing Markov chain ergodic theorems.

2. Markov Processes and the Main Result

We start this paper by giving a brief introduction to Markov processes. Some of the notations and definitions are adapted from [Ros06]. Those who are familiar with Markov processes may skip to Definition 2.8.

In general, a continuous-time stochastic process is a collection $\{X_t\}_{t\geq 0}$ of random variables, defined jointly on some probability triple, taking values in some state space X with σ-algebra $\mathcal{F}$, and indexed by the non-negative real numbers $\{t \geq 0\}$. Usually we regard the variable t as representing time, so that X_t represents a random state at time t. Formally speaking, we have the following definition:

Definition 2.1. Given a probability space $(\Omega, \mathcal{F}, P)$ and a measurable space (X, Γ), a X-valued stochastic process is a collection of X-valued random variables $\{X_t\}_{t\in T}$, indexed by a totally ordered set T ("time line"). The space X is called the state space.

The "time line" is almost always taken to be either $\mathbb{R}^+ \cup \{0\}$ ("non-negative reals") or $\mathbb{N}$. When $T = \mathbb{N}$, the stochastic process is called a discrete-time stochastic process. Otherwise it is called a continuous-time stochastic process. In this paper, the σ-algebra on X is always taken to be the Borel σ-algebra $\mathcal{B}[X]$. The sample space Ω is usually taken to be the set of all measurable functions from T to X and $\mathcal{F}$ is taken to be the product σ-algebra. Every element in Ω is called a path.

We are now at the place to define Markov processes.

Definition 2.2. A stochastic process $\{X_t\}_{t\geq 0}$ on some measurable space $(X, \mathcal{B}[X])$ is a Markov process if there are transition probability measures $P_x^{(t)}(\cdot)$ on $(X, \mathcal{B}[X])$ for all $t \geq 0$ and all $x \in X$, and an initial distribution ν on $(X, \mathcal{B}[X])$, such that

(1) $P_x^{(t)}(\cdot)$ is a probability measure on $(X, \mathcal{B}[X])$ and $P_x^{(0)}(\cdot)$ is a point-mass at x.

(2) $P(X_0 \in A_0, X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n)$
$= \int_{x_0 \in A_0} \int_{x_{t_1} \in A_1} \ldots \int_{x_{t_n} \in A_n} \nu(\mathrm{d}x_0) P_{x_0}^{(t_1)}(\mathrm{d}x_{t_1}) P_{x_{t_1}}^{(t_2-t_1)}(\mathrm{d}x_{t_2}) \ldots P_{x_{t_{n-1}}}^{(t_n-t_{n-1})}(\mathrm{d}x_{t_n})$
for all $1 \leq t_0 < \ldots < t_n$ and all $A_1, \ldots, A_n \in \mathcal{B}[X]$.

(3) $P_x^{(s+t)}(A) = \int P_x^{(s)}(\mathrm{d}y) P_y^{(t)}(A)$ for all $s, t \geq 0$, all $x \in X$ and all $A \in \mathcal{B}[X]$.

where $\mathcal{B}[X]$ denote the Borel σ-algebra of X.

Intuitively, $P_x^{(t)}(A)$ refers to the probability of getting into set A at time t given that the chain starts at x. When the state space is countable, we write $p_{ij}^{(t)}$ to denote $P_i^{(t)}(\{j\})$.

The third property in Definition 2.2 is called the semigroup property. On countable state space, we have $p_{ij}^{(s+t)} = \sum_{k \in X} p_{ik}^{(s)} p_{kj}^{(t)}$. When the time line is discrete, it is easy to see that we can get all $P_x^{(t)}(\cdot)$ from $P_x^{(1)}(\cdot)$. Hence, when the time line is discrete, transition probabilities of a Markov process is uniquely determined by its "one-step" transition probability $P_x^{(1)}(\cdot)$ for all $x \in X$. In this case, we usually omit 1 and write $P_x(\cdot)$ instead.

Probably the most well-understood type of Markov chains are discrete time Markov chains on with discrete state spaces. By Definition 2.2, it is not difficult to see that such a Markov process is characterized by three ingredients:

(1) a state space S.
(2) an initial distribution $\{v_i : i \in S\}$ consisting of non-negative numbers summing to 1.
(3) one-step transition matrix $\{p_{ij}\}_{i,j \in S}$ consisting of non-negative numbers with $\sum_{j \in S} p_{ij} = 1$ for each $i \in S$.

Clearly we can generate n-th transition probability from one-step transition matrix $\{p_{ij}\}_{i,j \in S}$. We use $p_{ij}^{(n)}$ to denote the n-step transition probability from i to j.

Example 2.3. The simplest example of Markov process is the simple random walk. The state space is the set of all integers $\mathbb{Z}$. The initial distribution is the point mass at 0. The one-step transition matrix is given by $P_i(\{i+1\}) = \frac{1}{2}$ and $P_i(\{i-1\}) = \frac{1}{2}$.

We first explore some basic properties of discrete-time Markov processes with finite state spaces. Let $\{X_t\}_{t\geq 0}$ denote such a Markov process. By Definition 2.2, it is easy to see that $P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$. The following two properties can be established pretty easily for such Markov process:

(1) $P(X_{k+1} = j | X_k = i_0, X_{k-1} = i_1, \ldots, X_0 = i_k) = P(X_{k+1} = j | X_k = i_0)$. **Markov property**

(2) $P(X_{k+n} = j | X_k = i) = p_{ij}^{(n)}$ for all $i, j \in S$ and all $k, n \in \mathbb{N}$. **Time homogeneous**

Both of these properties can be generalized to more general Markov processes in a natural way.

As one can see, the discrete time Markov chain with discrete state space is easy to understand and work with. However, it is not the case for general Markov process. The level of complexity increases greatly when we analysis general Markov processes using standard method. In this paper, we will apply nonstandard analysis to turn every continuous time general state space Markov process into a "finite" Markov process. We will discuss these ideas in more details in later sections.

Before we discuss general Markov processes, we introduce the finite-dimensional distributions for a general stochastic process.

Definition 2.4. Given a stochastic process $\{X_t\}_{t\in T}$, and $k \in \mathbb{N}$, and a finite collection $t_1, t_2, \ldots, t_k \in T$ of distinct index values, we define the Borel probability measure $\mu_{t_1 \ldots t_k}$ on X^k by:

$$\mu_{t_1 \ldots t_k}(H) = P((X_{t_1}, \ldots, X_{t_k}) \in H), H \in \mathcal{B}[X^k] \tag{2.1}$$

The distribution $\{\mu_{t_1 \ldots t_k}; k \in \mathbb{N}, t_1, \ldots, t_k \in T \text{ distinct}\}$ are called the finite-dimensional distributions for the stochastic process $\{X_t : t \in T\}$.

Under suitable "consistency" conditions of the finite-dimensional distributions, we can determine a stochastic process from its finite-dimensional distributions. We first introduce the following definition.

Definition 2.5. A measure μ on a Hausdorff measure space $(X, \mathcal{F})$ is inner regular if

$$\mu(A) = \sup\{\mu(K) | \text{compact } K \subset A\}. \tag{2.2}$$

Theorem 2.6 (Kolmogorov Existence Theorem)**.** *Let T be any set. Let $\{(\Omega_t, \mathcal{F}_t)\}_{t\in T}$ be some collection of measurable spaces with Hausdorff topology on each Ω_t. For each $J \subset T$, let $\Omega_J = \prod_{t\in J} \Omega_t$. For subset $I \subset J \subset T$, let π_I^J be the projection map from $\prod_{t\in J} \Omega_t \to \prod_{t\in I} \Omega_t$. For each finite $F \subset T$, suppose we have a probability measure μ_F on Ω_F which is inner regular with respect to the product topology on Ω_F. Suppose that for finite sets $F \subset G \subset T$, we have that*

$$\mu_F(A) = \mu_G((\pi_F^G)^{-1}(A)). \tag{2.3}$$

for all measurable sets A. Then there exists an unique measure μ on Ω_T such that $\mu_F(A) = \mu((\pi_F^T)^{-1}(A))$ for all finite $F \subset T$ and all measurable sets A.

It is clear that $T = \{x \in \mathbb{R} : x \geq 0\}$ for a continuous Markov processes. For a detailed proof of this theorem, see e.g. [Bil95, 1995,Theorem 36.1].

We now turn our attention to general Markov processes. When $\nu(\{x\}) > 0$, we write $P_x(\cdot)$ for the probability of an event conditional on $X_0 = x$. In particular, we have $P_x^{(t)}(A) = P(X_t \in A | X_0 = x)$. However, we have $\nu(x) = 0$ for most of x when the state space is not discrete. Thus, we shall view P_x as a probability measure on the product space $(\{x\} \times X^{\mathbb{R}^+}, \mathcal{F})$ where $\mathcal{F}$ denote the product Borel σ-algebra with finite dimensional distribution:

$P_x(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n)$
$= \int_{x_{t_1}\in A_1} \dots \int_{x_{t_n}\in A_n} P_{x_0}^{(t_1)}(\mathrm{d}x_{t_1}) P_{x_{t_1}}^{(t_2-t_1)}(\mathrm{d}x_{t_2}) \dots P_{x_{t_{n-1}}}^{(t_n-t_{n-1})}(\mathrm{d}x_{t_n})$.

Then by Kolmogorov existence theorem, such probability measure P_x exists.

Definition 2.7. A probability distribution $\pi(\cdot)$ on $(X, \mathcal{B}[X])$ is a stationary distribution for the Markov process $\{X_t\}_{t\geq 0}$ if $\int_X P_x^{(t)}(A)\pi(\mathrm{d}x) = \pi(A)$ for all $t \geq 0$ and all $A \in \mathcal{B}[X]$.

The intuition behind stationarity is quite simple. It means that if we start the Markov chain in the distribution π then any time later the distribution will still be π. However, even if we start our process in some other distribution we would like to show that eventually the distribution will be π. This is the famous Markov chain Ergodic theorem. Before giving the formal statement of the Markov chain Ergodic theorem, we need to introduce some concepts for Markov processes. All these assumptions will be restated later in the paper.

Definition 2.8. Let $\mathcal{K}[X]$ denote the collection of compact subsets of X. The Markov chain $\{X_t\}_{t\geq 0}$ is said to be *vanishing in distance* if for all $t \geq 0$, all $K \in \mathcal{K}[X]$ and every $\epsilon > 0$, the set $\{x \in X : P_x^{(t)}(K) \geq \epsilon\}$ is contained in a compact subset of X.

Roughly speaking, if a Markov process $\{X_t\}$ is vanishing in distance, it means that the probability of $\{X_t\}$ "traveling far" within a fixed amount of time is small. This ensures that the Markov process is non-explosive. In Section 9, we give an equivalent formulation of Definition 2.8 using the metric d on the state space X. It will be easier to see the intuition behind Definition 2.8 there.

Definition 2.9. A Markov chain $\{X_t\}_{t\geq 0}$ is said to be *strong Feller* if for all $t > 0$, all $\epsilon > 0$, all $x \in X$, there exists $\delta > 0$ such that:

$$((\forall y \in X)(d(x,y) < \delta \implies (\forall A \in \mathcal{B}[X])|P_y^{(t)}(A) - P_x^{(t)}(A)| < \epsilon)). \tag{2.4}$$

Definition 2.10. Given two probability measures P_1, P_2 on some measurable space $(X, \mathcal{F})$. The total variation distance between P_1 and P_2 is

$$\| P_1 - P_2 \| = \sup_{A \in \mathcal{F}} |P_1(A) - P_2(A)|. \tag{2.5}$$

The strong Feller condition essentially says that, for any $t > 0$, the mapping $x \to P_x^{(t)}(\cdot)$ is continuous with respect to total variation norm. Given a strong Feller Markov process $\{X_t\}$, if we start at two points which are close to each other, after a fixed period of time, the probabilities of them reaching the same set is very close. This is certainly a reasonable assumption for most of the Markov processes. As a matter of fact, most of the diffusion and Gaussian processes satisfy this condition.

Definition 2.11. A Markov chain $\{X_t\}_{t\geq 0}$ is said to be *weakly continuous in time* if for any basic open set $A \subset X$, and any $x \in X$, we know that $P_x^{(t)}(A)$ is a right continuous function for $t > 0$. Moreover, for any $t_0 \in \mathbb{R}^+$, any $x \in X$ and any $E \in \mathcal{B}[X]$ we know that $\lim_{t\uparrow t_0} P_x^{(t)}(E)$ always exists although it not necessarily equals to $P_x^{(t_0)}(E)$

Definition 2.12. A Markov chain $\{X_t\}_{t\geq 0}$ with state space X is said to be open set irreducible on X if for every open ball $B \subseteq X$ and any $x \in X$, there exists $t \in \mathbb{R}^+$ such that $p_x^{(t)}(B) > 0$.

If a Markov process is open set irreducible, it means that it is possible to move from any point to any open set.

The classical proof of the Markov chain Ergodic theorem in the finite case uses the "coupling" idea. Roughly speaking, under moderate conditions, for two i.i.d Markov processes starting at two different points will eventually "couple" at some point. Thus it is worth to consider the product of two Markov processes.

Definition 2.13. Let $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ be two Markov processes on state spaces X and Y, respectively. Let $P_x^{(t)}(\cdot)$ denote the t-step transition measure of $\{X_t\}_{t\geq 0}$ had the chain started at x. Let $Q_y^{(t)}(\cdot)$ denote the t-step transition measure of $\{Y_t\}_{t\geq 0}$ had the chain started at y. The joint Markov chain $\{X_t \times Y_t\}_{t\geq 0}$ is a Markov process on the state space $X \times Y$ with transition probability:

$$F_{(x,y)}^{(t)}(A \times B) = P_x^{(t)}(A) Q_y^{(t)}(B). \tag{2.6}$$

for all $(x, y) \in X \times Y$, all $A \times B \in \mathcal{B}[X] \times \mathcal{B}[Y]$ and all $t \geq 0$.

The most common product Markov chain is the product between a Markov chain $\{X_t\}$ and itself. However, even if $\{X_t\}$ is open set irreducible this may not be the case for the product chain $\{X_t \times X_t\}$.

Example 2.14. Let $\{X_t\}_{t\in\mathbb{N}}$ be a Markov process with two-points state space $\{1,2\}$ with discrete topology. Let $P_1(\{2\}) = 1$ and $P_2(\{1\}) = 1$. It is clear that $\{X_t\}$ is open set irreducible. However, it is never possible to go from $(1,2)$ to $(1,1)$ since $P_1^{(t)}(\{1\})P_2^{(t)}(\{1\}) = 0$ for all $t \in \mathbb{N}$.

Thus we impose the following condition on $\{X_t\}_{t\geq 0}$ to eliminate such counter-example.

Definition 2.15. The Markov chain $\{X_t\}_{t\geq 0}$ is productively open set irreducible if the joint Markov chain $\{X_t \times Y_t\}_{t\geq 0}$ is open set irreducible on $X \times X$ where $\{Y_t\}_{t\geq 0}$ is an independent identical copy of $\{X_t\}_{t\geq 0}$.

We are now at the place to state the main result of this paper.

Theorem 2.16. *Let $\{X_t\}_{t\geq 0}$ be a general state space continuous-time Markov chain with separable locally compact metric state space (X,d). Suppose $\{X_t\}_{t\geq 0}$ is productively open set irreducible and has a stationary distribution π. Suppose $\{X_t\}_{t\geq 0}$ is vanishing in distance, strong Feller and weakly continuous. Then for π-almost surely $x \in X$ we have $\lim_{t\to\infty} \sup_{A\in\mathcal{B}[X]} |P_x^{(t)}(A) - \pi(A)| = 0$.*

A similar result holds for general-state-space discrete time Markov processes. We can drop weakly continuity in time and vanishing in distance in the discrete time case. To prove Theorem 2.16, we first establish a weaker Markov chain ergodic theorem. We start by introducing the following definition.

Definition 2.17. A metric space X is said to satisfy the Heine-Borel condition if the closure of every open ball is compact.

The proof of the following theorem will be delayed to Section 10, see Theorem 10.16

Theorem 2.18. *Let $\{X_t\}_{t\geq 0}$ be a general-state-space continuous in time Markov chain living on some metric space X satisfying the Heine-Borel condition. Suppose $\{X_t\}_{t\geq 0}$ is productively open set irreducible and has a stationary distribution π. Suppose $\{X_t\}_{t\geq 0}$ is strong Feller, weakly continuous in time and vanishes in distance. Then for π-almost surely $x \in X$ we have $\lim_{t\to\infty} \sup_{A\in\mathcal{B}[X]} |P_x^{(t)}(A) - \pi(A)| = 0$.*

Theorem 2.18 is interesting on its own. For example, Theorem 2.18 applies to all Markov processes with Euclidean state space. However, Theorem 2.18 does require that the state space X is a metric space satisfying the Heine-Borel property. Such an X is automatically a separable locally compact metric space. Hence Theorem 2.18 is an immediate consequence of Theorem 2.16. On one hand, the Heine-Borel condition is quite strong. For example, $(0,1)$ and $(0,\infty)$, while they are separable locally compact metric spaces, do not satisfy the Heine-Borel property. On the other hand, from the nonstandard perspective, the Heine-Borel condition is desirable because it guarantees that every finite nonstandard element is infinitely close to a standard element (see Theorem 6.2). Hence, it will be easier to establish Theorem 2.18 than Theorem 2.16.

It is easy to see that Theorem 2.18 follows from Theorem 2.16. For the remainder of this section, we establish Theorem 2.16 from Theorem 2.18. We start by proving the following theorem which shows that, for every separable locally compact metric space, there exists a Heine-Borel metric d_H on X that induces the same topology.

Theorem 2.19. *Let (X,d) be a separable locally compact metric space. There is a metric d_H on X inducing the same topology such that (X,d_H) satisfies the Heine-Borel property.*

Proof. It is a well-known topological fact that if (X,d) is a separable locally compact metric space then X is σ-compact. Let $X = \bigcup_{n\in\mathbb{N}} K_n$ where every K_n is a compact subset of X. We now define a non-decreasing of compact subsets of X as following:

- Let $V_1 = K_1$.

- Suppose we have defined V_n. As X is locally compact, there is a finite collection $\{U_1, \dots, U_k\}$ of open sets such that $\bigcup_{i \leq k} U_i \supset V_n$ and $\overline{U}_i$ is compact for every $i \leq k$. Let $V_{n+1} = (\bigcup_{i \leq k} \overline{U}_i) \cup K_{n+1}$.

Thus, $X = \bigcup_{n \in \mathbb{N}} V_n$ and $V_n \subset W_{n+1}$ where W_{n+1} is the interior of V_{n+1}. Define $f_n : X \mapsto \mathbb{R}$ by letting $f_n(x) = \frac{d(x,V_n)}{d(x,V_n)+d(x,X \setminus W_{n+1})}$. Let $f(x) = \sum_{n=1}^{\infty} f_n(x)$. Note that $\sum_{n=1}^{\infty} f_n(x)$ is always finite since each $x \in X$ is in some V_n. Moreover, as both V_n and $X \setminus W_{n+1}$ are closed, the function $f : X \mapsto \mathbb{R}$ is continuous. Define $d_H : X \times X \to \mathbb{R}$ by

$$d_H(x, y) = d(x, y) + |f(x) - f(y)|. \tag{2.7}$$

Then

$$d_H(x, z) = d(x, z) + |f(x) - f(z)| \leq d(x, y) + |f(x) - f(y)| + d(y, z) + |f(y) - f(z)| \tag{2.8}$$

hence d_H is a metric on X.

Claim 2.20. *d_H induces the same topology as d.*

Proof. Let $\{x_n : n \in \mathbb{N}\}$ be a subset of X and let $y \in X$. Suppose $\lim_{n\to\infty} d_H(x_n, y) = 0$. As $d(x_n, y) \leq d_H(x_n, y)$ for all $n \in \mathbb{N}$, we have $\lim_{n\to\infty} d(x_n, y) = 0$. Now suppose $\lim_{n\to\infty} d(x_n, y) = 0$. As f is continuous in the original metric, we have $\lim_{n\to\infty} f(x_n) = f(y)$ hence we have $\lim_{n\to\infty} d_H(x_n, y) = 0$. □

The metric space (X, d_H) satisfies the Heine-Borel condition since the following claim is true.

Claim 2.21. *For every $A \subset X$ bounded with respect to d_H, there is some V_n such that $A \subset V_n$.*

Proof. Suppose A is not a subset of any element in $\{V_n : n \in \mathbb{N}\}$. Fix some element $n \in \mathbb{N}$ and $r \in \mathbb{R}^+$. Pick $x \in V_{n+1} \setminus V_n$. By the construction of f, we know that $n + 1 \geq f(x) > n$. Thus, we can pick an element $a \in A$ such that $f(a) > f(x) + r$. Then $d_H(x, a) > r$. As n and r are arbitrary, this shows that A is not bounded. □

□

With the help of Theorem 2.19, we can prove Theorem 2.16 from Theorem 2.18.

Proof of Theorem 2.16. Let d_H be a Heine-Borel metric on X that induces the same topology as (X,d). By Theorem 2.18, it is sufficient to show that $\{X_t\}_{t\geq 0}$ is strong Feller and vanishes in distance under the metric d_H. Note that vanishing in distance (Definition 2.8) is a purely topological property. As d_H and d generate the same topology, we know that $\{X_t\}_{t\geq 0}$ vanishes in distance under the metric d_H.

We now show that $\{X_t\}_{t\geq 0}$ is strong Feller under the metric d_H. Pick $t>0$, $\epsilon>0$ and $x\in X$. As $\{X_t\}_{t\geq 0}$ is strong Feller under the metric d, there exists a $\delta>0$ such that

$$((\forall y\in X)(d(x,y)<\delta \implies (\forall A\in \mathcal{B}[X])|P_y^{(t)}(A)-P_x^{(t)}(A)|<\epsilon)). \tag{2.9}$$

Note that the set $\{y\in X: d(x,y)<\delta\}$ is an open subset of X. As the metric d_H generates the same topology as (X,d), there exists $\delta'>0$ such that $\{y\in X: d_H(x,y)<\delta'\}\subset\{y\in X: d(x,y)<\delta\}$. Thus, we can conclude that

$$((\forall y\in X)(d_H(x,y)<\delta' \implies (\forall A\in \mathcal{B}[X])|P_y^{(t)}(A)-P_x^{(t)}(A)|<\epsilon)). \tag{2.10}$$

Hence, $\{X_t\}_{t\geq 0}$ is strong Feller under the Heine-Borel metric d_H. □

To complete the proof of Theorem 2.16, it is sufficient to show Theorem 2.18. From this point on, we shall work with Markov processes with metric state space satisfying the Heine-Borel condition.

3. Preliminaries: Nonstandard Analysis

Those familiar with nonstandard methods may safely skip this section on their first reading. Nonstandard analysis is introduced by Abraham Robinson in [Rob66]. For modern applications of nonstandard analysis, interested readers can read [ACH97] or [Cut+95]. For those who are particularly interested in nonstandard measure theory, we recommend [LW15] which contains special measure-theoretic results obtained by nonstandard analysis that have no known classic analogues in various fields (see [LW15, Chapter. 8]). Our following introduction of nonstandard analysis owes much to [ACH97]. Some part of this section is adapted from [DD16].

For a set S, let $\mathscr{P}(S)$ denote its power set. Given any set S, define $\mathbb{V}_0(S) = S$ and $\mathbb{V}_{n+1}(S) = \mathbb{V}_n(S) \cup \mathscr{P}(\mathbb{V}_n(S))$ for all $n \in \mathbb{N}$. Then $\mathbb{V}(S) = \bigcup_{n\in\mathbb{N}} \mathbb{V}_n(S)$ is called the *superstructure* of S, and S is called the *ground set* of the superstructure $\mathbb{V}(S)$. We treat the elements in $\mathbb{V}(S)$ as indivisible atomics. The *rank* of an object $a \in \mathbb{V}(S)$ is the smallest k for which $a \in \mathbb{V}_k(S)$. The members of S have rank 0. The objects of rank no less than 1 in $\mathbb{V}(S)$ are precisely the sets in $\mathbb{V}(S)$. The empty set $\emptyset$ and S both have rank 1.

We now formally define the language $\mathcal{L}(\mathbb{V}(S))$ of $\mathbb{V}(S)$.

- *constants*: one for each element in $\mathbb{V}(S)$.
- *variables*: $x_1, x_2, x_3, \ldots$
- *relations*: $=$ and $\in$.
- *parentheses*:) and (
- *connectives*: $\wedge$ (and), $\vee$ (or) and $\neg$ (not).
- *quantifiers*: $\forall$ and $\exists$

The formulas in $\mathcal{L}(\mathbb{V}(S))$ are defined recursively:

- If x and y are variables and a and b are constants,
  $$(x = y), (x \in y), (a = x), (a \in x), (x \in a), (a = b), (a \in b) \text{ are formulas.}$$
- If ϕ and ψ are formulas, then $(\phi \wedge \psi), (\phi \vee \psi)$ and $(\neg\phi)$ are formulas.
- If ϕ is a formula, x is a variable and $A \in \mathbb{V}(S)$ then $(\forall x \in A)(\phi)$ and $(\exists x \in A)(\phi)$ are formulas.

A variable x is called a free variable if it is not within the scope of any quantifiers.

Let us agree to use the following abbreviations in constructing formulas in $\mathcal{L}(\mathbb{V}(S))$: We will write $(\phi \implies \psi)$ instead of $((\neg\phi) \vee (\psi))$ and $(\phi \iff \psi)$ instead of $(\phi \implies \psi) \wedge (\psi \implies \phi)$.

It may seem that we should include more relation symbols and function symbols in our language. For example, it is definitely natural to require $1 < 2$ to be a well-defined formula. However, every relation symbol and function syumbol can be viewed as an element in $\mathbb{V}(S)$ and we already have a constant symbol for that. Thus our language is powerful enough to describe all well-defined relation symbols and function symbols. In conclusion, there is no problem to include these symbols within our formula.

Definition 3.1. Let κ be an uncountable cardinal number. A *κ-saturated nonstandard extension* of a superstructure $\mathbb{V}(S)$ is a set ${}^*S$ and a rank-preserving map ${}^*: \mathbb{V}(S) \to \mathbb{V}({}^*S)$ satisfying the following three principles:

- *extension*: ${}^*S$ is a superset of S and ${}^*s = s$ for all $s \in S$.
- *transfer*: For every sentence ϕ in $\mathcal{L}(\mathbb{V}(S))$, ϕ is true in $\mathbb{V}(S)$ if and only if its $*$-transfer ${}^*\phi$ is true in $\mathbb{V}({}^*S)$.
- *κ-saturation*: For every family $\mathcal{F} = \{A_i : i \in I\}$ of internal sets indexed by a set I of cardinality less than κ, if $\mathcal{F}$ has the finite intersection property, i.e., if every finite intersection of elements in $\mathcal{F}$ is nonempty, then the total intersection of $\mathcal{F}$ is non-empty.

A $\aleph_1$ saturated model can be constructed via an ultrafilter, see [ACH97, Thm. 1.7.13].

The language of $\mathbb{V}({}^*S)$ is almost the same as $\mathcal{L}$ except that we enlarge the set of constants to include every element in $\mathbb{V}({}^*S)$. We denote the language of $\mathbb{V}({}^*(S))$ by $\mathcal{L}(\mathbb{V}({}^*S))$. If $\phi(x_1, \dots, x_n)$ is a formula in $\mathcal{L}(\mathbb{V}(S))$ with free variables $x_1, \dots, x_n$, then the $*$-transfer of ϕ is the formula in $\mathcal{L}(\mathbb{V}({}^*S))$ obtained by changing every constant a to ${}^*a$. Clearly, every constant in ${}^*\phi(x_1, \dots, x_n)$ is internal.

An important class of elements in $\mathbb{V}({}^*S)$ is the class of internal elements.

Definition 3.2. An element $a \in \mathbb{V}({}^*S)$ is *internal* when there exists $b \in \mathbb{V}(S)$ such that $a \in {}^*b$, and a is said to be *external* otherwise.

The next theorem shows that saturation to any uncountable cardinal number is possible:

Theorem 3.3 ([Lux69])**.** *For every superstructure $\mathbb{V}(S)$ and uncountable cardinal number κ, there exists a κ-saturated nonstandard extension of $\mathbb{V}(S)$.*

From this point on, we shall always assume that our nonstandard extension is always as saturated as we want.

As one can see, internal elements are those "well-behaved" elements which can be carried over via the transfer principle. It is natural to ask how to identify internal elements. By Definition 3.2, we know that an element $a \in \mathbb{V}({}^*S)$ is internal if and only if there exists a $k \in \mathbb{N}$ such that $a \in {}^*\mathbb{V}_k(S)$. It is then easy to see that every $a \in {}^*S$ is internal. The following lemma gives a characterization of internal elements in $\mathscr{P}({}^*S)$.

Lemma 3.4. *Consider a superstructure $\mathbb{V}(S)$ based on a set S with $\mathbb{N} \subset S$ and its nonstandard extension, for any standard set C from this superstructure, $\bigcup_{k<\omega} {}^*\mathbb{V}_k(S) \cap \mathscr{P}({}^*C) = {}^*\mathscr{P}(C)$.*

Proof. Let us assume that C has rank n for some $n \in \mathbb{N}$. $\mathscr{P}(C) \in \mathbb{V}_{n+1}(S)$ hence we have ${}^*\mathscr{P}(C) \in {}^*\mathbb{V}_{n+1}(S)$. Consider the following sentence $(\forall x \in \mathscr{P}(C))(\forall y \in x)(y \in C)$, the transfer of this sentence implies that ${}^*\mathscr{P}(C) \subset \mathscr{P}({}^*C)$. Hence we have ${}^*\mathscr{P}(C) \subset \bigcup_{k<\omega} {}^*\mathbb{V}_k(S) \cap \mathscr{P}({}^*C)$, completing the proof. □

Thus, we know that that $A \subset {}^*S$ is internal if and only if $A \in {}^*\mathscr{P}(S)$.

The following lemma shows a particularly useful fact of internal sets which will be used extensively in this paper.

Lemma 3.5. *Let a be an internal element in $\mathbb{V}({}^*S)$. Then the collection of all internal subsets of a is itself internal.*

Proof. As a is an internal element, there exists a $k \in \mathbb{N}$ such that $a \in {}^*\mathbb{V}_k(S)$. For any internal set $a' \subset a$, it is easy to see that $a' \in {}^*\mathbb{V}_k(S)$. Let b denote the collection of all internal subsets of a. The sentence $(\forall x \in y)(x \in \mathbb{V}_k(S)) \implies (Y \in \mathbb{V}_{k+1}(S))$ is true. Thus, by the transfer principle, we have that $b \in {}^*\mathbb{V}_{k+1}(S)$ hence b is an internal set. □

It takes practice to identify general internal sets. The main tool for constructing internal sets is the internal definition principle:

Lemma 3.6 (Internal Definition Principle). *Let $\phi(x)$ be a formula in $\mathcal{L}(\mathbb{V}(^*S))$ with free variable x. Suppose that all constants that occurs in ϕ are internal, then $\{x \in \mathbb{V}(^*S) : \phi(x)\}$ is internal in $\mathbb{V}(^*S)$.*

Saturation can be equivalently expressed in terms of the satisfiability of families of formulas. The role of the finite intersection property is played by finite satisfiability:

Definition 3.7. Let J be an index set and let $A \subseteq \mathbb{V}(^*S)$. A set of formulas $\{\phi_j(x) \mid j \in J\}$ over $\mathbb{V}(^*S)$ is said to be *finitely satisfiable in A* when, for every finite subset $\alpha \subset J$, there exists $c \in A$ such that $\phi_j(c)$ holds for all $j \in \alpha$.

We can now provide the following alternative expression of κ-saturation:

Theorem 3.8 ([ACH97, Thm. 1.7.2]). *Let $^*\mathbb{V}(S)$ be a κ-saturated nonstandard extension of the superstructure $\mathbb{V}(S)$, where κ is an uncountable cardinal number. Let J be an index set of cardinality less than κ. Let A be an internal set in $^*\mathbb{V}(S)$. For each $j \in J$, let $\phi_j(x)$ be a formula over $^*\mathbb{V}(S)$, so all objects mentioned in $\phi_j(x)$ are internal. Further, suppose that the set of formulas $\{\phi_j(x) \mid j \in J\}$ is finitely satisfied in A. Then there exists $c \in A$ such that $\phi_j(c)$ holds in $^*\mathbb{V}(S)$ simultaneously for all $j \in J$.*

Example 3.9. A particular interesting example of superstructure is $\mathbb{V}(\mathbb{R})$. The nonstandard extension of this superstructure is $\mathbb{V}(^*\mathbb{R})$. $\mathbb{V}(^*\mathbb{R})$ contains hyperreals, $^*\mathbb{N}$, etc. We will study this particular superstructure in detail in Section 3.1.

Through out this paper, we shall assume our ground set S always contain $\mathbb{R}$ as a subset.

We conclude this section by introducing a particularly useful class of sets in $\mathbb{V}(^*S)$: hyperfinite sets. A hyperfinite set A is an infinite set that has the basic logical properties of a finite set.

Definition 3.10. A set $A \in \mathbb{V}({}^*S)$ is hyperfinite if and only if there exists an internal bijection between A and $\{0, 1, ..., N-1\}$ for some $N \in {}^*\mathbb{N}$.

This N, if exists, is unique and this unique N is called the internal cardinality of A.

Just like finite sets, we can carry out all the basic arithmetics on a hyperfinite set. For example, we can sum over a hyperfinite set just like we did for finite set. Basic set theoretic operations are also preserved. For example, we can take hyperfinite unions and intersections just as taking finite unions and intersections.

We have rather nice characterization of internal subsets of a hyperfinite set.

Lemma 3.11 ([ACH97])**.** *A subset A of a hyperfinite set T is internal if and only if A is hyperfinite.*

An immediate consequence of Theorem 3.8 is:

Proposition 3.12 ([ACH97, Proposition. 1.7.4])**.** *Assume that the nonstandard extension is κ-saturated. Let a be an internal set in $\mathbb{V}({}^*S)$. Let A be a (possibly external) subset of a such that the cardinality of A is strictly less than κ. Then there exists a hyperfinite subset b of a such that b contains A as a subset.*

3.1. **The Hyperreals.** Probably the most well-known nonstandard extension is the nonstandard extension of $\mathbb{R}$. We investigate some basic properties and notations in ${}^*\mathbb{R}$.

Definition 3.13. The set ${}^*\mathbb{R}$ is called the set of hyperreals and every element in ${}^*\mathbb{R}$ is called a hyperreal number. An element $x \in {}^*\mathbb{R}$ is called an infinitesimal if $x < \frac{1}{n}$ for all $n \in \mathbb{N}$. An element $y \in {}^*\mathbb{R}$ is called an infinite number if $y > n$ for all $n \in \mathbb{N}$.

We write $x \approx 0$ when x is an infinitesimal.

Definition 3.14. Two elements $x, y \in {}^*\mathbb{R}$ are infinitesimally close if $|x - y| \approx 0$. In which case, we write $x \approx y$. An element $x \in {}^*\mathbb{R}$ is near-standard if x is infinitesimally close to some $a \in \mathbb{R}$. An element $x \in {}^*\mathbb{R}$ is finite if $|x|$ is bounded by some standard real number a.

It is easy to see that if $x \in {}^*\mathbb{R}$ is bounded then there exists some $a \in \mathbb{R}$ such that $|x - a|$ is finite.

Lemma 3.15. *An element $x \in {}^*\mathbb{R}$ is finite if and only if x is near-standard.*

Proof. It is clear that if x is near-standard then x is finite. Suppose there exists a $x \in {}^*\mathbb{R}$ such that x is finite but not near-standard. Then there exists a $a_0 \in \mathbb{R}$ such that $|x| \leq a_0$. This means that $x \in {}^*[-a_0, a_0]$. As x is not near-standard, for every standard $a \in [-a_0, a_0]$ we can find an open interval O_a centered at a with $x \notin {}^*O_a$. The family $\{O_a : a \in [-a_0, a_0]\}$ covers $[-a_0, a_0]$ and therefore has a finite subcover $\{O_1, ..., O_n\}$. As $[-a_0, a_0] \subset \bigcup_{i \leq n} O_i$, ${}^*[-a_0, a_0] \subset \bigcup_{i \leq n} {}^*O_i$. Since $x \notin \bigcup_{i \leq n} {}^*O_i$, $x \notin {}^*[-a_0, a_0]$ which is a contradiction. Hence $x \in {}^*\mathbb{R}$ is finite if and only if it is near-standard.

Pick an arbitrary near-standard $x \in {}^*\mathbb{R}$. Suppose there are two different $a_1, a_2 \in \mathbb{R}$ such that $x \approx a_1$ and $x \approx a_2$. This implies $a_1 \approx a_2$ which is impossible since $a_1, a_2 \in \mathbb{R}$. Hence there exists a unique $a \in \mathbb{R}$ such that $x \approx a$. □

This lemma would fail if we take some points from $\mathbb{R}$.

Example 3.16. Consider the set $\mathbb{R} \setminus \{0\}$. Then every infinitesimal element in ${}^*\mathbb{R}$ is finite since they are bounded by 1. However, they are not near-standard since 0 is excluded.

Definition 3.17. Let $\mathrm{NS}({}^*\mathbb{R})$ to denote the collection of all near-standard points in ${}^*\mathbb{R}$. For every near-standard point $x \in {}^*\mathbb{R}$, let $\mathsf{st}(x)$ denote the unique element in $a \in \mathbb{R}$ such that $|x - a| \approx 0$. $\mathsf{st}(x)$ is called the standard part of x. We call st the standard part map.

For $A \subset {}^*\mathbb{R}$, we write $\mathsf{st}(A)$ to mean $\{x \in \mathbb{R} : (\exists a \in A)(x \text{ is the standard part of } a)\}$. Similarly for every $B \subset \mathbb{R}$, we write $\mathsf{st}^{-1}(B)$ to mean $\{x \in {}^*\mathbb{R} : (\exists b \in B)(|x - b| \approx 0)\}$.

We now give an example of an external set. The example also shows that we have to be very careful when applying the transfer principle.

Example 3.18. The monad $\mu(0)$ of 0 is defined to be $\{a \in {}^*\mathbb{R} : a \approx 0\}$. We show that $\mu(0)$ is an external set. Consider the sentence: $\forall A \in \mathscr{P}(\mathbb{R})$ if A is bounded above then there is a least upper bound for A. By the transfer principle, we know that $(\forall A \in {}^*\mathscr{P}(\mathbb{R}))$(for all internal subsets of ${}^*\mathbb{R}$ if A is bounded above then there is a least upper bound for A). Suppose $\mu(0)$ is internal then there exists a $a_0 \in^* \mathbb{R}$ such that a_0 is an least upper bound for $\mu(0)$. Clearly $a_0 > 0$. Note that a_0 can not be infinitesimal since if a_0 is infinitesimal then $2a_0$ would also be infinitesimal and $2a_0 > a_0$. If a_0 is non-infinitesimal then so is $\frac{a_0}{2}$. But then $\frac{a_0}{2}$ is an upper bound for $\mu(0)$. This contradict with the fact that a_0 is the least upper bound. Hence $\mu(0)$ is not an internal set.

It is easy to make the following mistake: if we write the sentence as "$\forall A \subset \mathbb{R}$ if A is bounded above then there is a least upper bound for A" the transfer of it seems to give that "$\forall A \subset {}^*\mathbb{R}$ if A is bounded above then there is a least upper bound for A". As we have already seen, this is not correct. The reason is because $\subset$ is not in the language of set theory thus we have an "illegal" formation of a sentence. This shows that we have to be very careful when applying the transfer principle.

The following two principles derived from saturation are extremely useful in establishing the existence of certain nonstandard objects.

Theorem 3.19. *Let $A \subset {}^*\mathbb{R}$ be an internal set*

1. *(**Overflow**) If A contains arbitrarily large positive finite numbers, then it contains arbitrarily small positive infinite numbers.*
2. *(**Underflow**) If A contains arbitrarily small positive infinite numbers, then it contains arbitrarily large positive finite numbers.*

We conclude this section by the following lemma. This lemma will be used extensively in this paper.

Lemma 3.20. *Let N be an element in ${}^*\mathbb{N}$. Let $\{a_1,\dots,a_N\}$ be a set of non-negative hyperreals such that $\sum_{i=1}^{N} a_i = 1$. Let $\{b_1,\dots,b_N\}$ and $\{c_1,\dots,c_N\}$ be subsets of $\mathbb{R}$ such that $b_i \approx c_i$ for all $i \leq N$. Then $a_1b_1 + a_2b_2 + \cdots + a_Nb_N \approx a_1c_1 + a_2c_2 + \cdots + a_Nc_N$.*

Proof. By the transfer of convex combination theorem, we know that $(a_1b_1+a_2b_2+\cdots+a_Nb_N)-(a_1c_1+a_2c_2+\cdots+a_Nc_N) = a_1(b_1-c_1)+a_2(b_2-c_2)+\cdots+a_N(b_N-c_N) \leq \max\{a_i|b_i-c_i| : i \leq N\} \approx 0$. □

3.2. **Nonstandard Extensions of General Metric Spaces.** We generalize the concepts developed in Section 3.1 into generalized topological spaces. We especially emphasize on general metric spaces.

Let X be a topological space and let ${}^*X$ denote its nonstandard extension. For every $x \in X$, let $\mathcal{B}_x$ denote a local base at point x.

Definition 3.21. Given $x \in X$, the monad of x is

$$\mu(x) = \bigcap_{U \in \mathcal{B}_x} {}^*U. \tag{3.1}$$

The near-standard points in ${}^*X$ are the points in the monad of some standard points.

If X is a metric space with metric d, then ${}^*d$ is a metric for ${}^*X$. The monad of a point $x \in X$, in this case, is $\mu(x) = \bigcap_{n\in\mathbb{N}} {}^*U_n$ where each $U_n = \{y \in X : d(x,y) < \frac{1}{n}\}$. Thus we have the following definition:

Definition 3.22. Two elements $x, y \in {}^*X$ are infinitesimally close if ${}^*d(x,y) \approx 0$. An element $x \in {}^*X$ is near-standard if x is infinitesimally close to some $a \in X$. An element $x \in {}^*X$ is finite if ${}^*d(x,a)$ is finite for some $a \in X$.

If $x \in {}^*X$ is finite, then generally x is not near-standard. This is not even true for complete metric spaces.

Example 3.23. Consider the set of natural numbers $\mathbb{N}$. Define the metric d on $\mathbb{N}$ to be $d(x,y)=1$ if $x\neq y$ and equals to 0 otherwise. Then $(\mathbb{N},d)$ is a complete metric space. Every element in ${}^*\mathbb{N}$ is finite. But those elements in ${}^*\mathbb{N}\setminus\mathbb{N}$ are not near-standard.

Just as in ${}^*\mathbb{R}$, we have the following definition.

Definition 3.24. Let $\mathrm{NS}({}^*X)$ to denote the collection of all near-standard points in ${}^*X$. For every near-standard point $x\in {}^*X$, let $\mathsf{st}(x)$ denote the unique element in $a\in X$ such that ${}^*d(x,a)\approx 0$. $\mathsf{st}(x)$ is called the standard part of x. We call st the standard part map.

In general, $\mathrm{NS}({}^*X)$ is a proper subset of ${}^*X$. However, when X is compact, we have $\mathrm{NS}({}^*X)={}^*X$. This is the nonstandard way to characterize a compact space.

Theorem 3.25 ([ACH97, Theorem 3.5.1])**.** *A set $A\subset X$ is compact if and only if ${}^*A=\mathrm{NS}({}^*A)$.*

Proof. Assume A is compact but there exists $y\in A$ such that y is not near-standard. Then for every $x\in A$, there exists an open set O_x containing x with $y\notin {}^*O_x$. The family $\{O_x : x\in A\}$ forms an open cover of A. As A is compact, there exists a finite subcover $\{O_1,\dots,O_n\}$ for some $n\in\mathbb{N}$. As $A\subset\bigcup_{i=1}^n O_i$, by the transfer principle, we have ${}^*A\subset\bigcup_{i=1}^n {}^*O_i$. However, $y\notin O_i$ for all $i\leq n$. This implies that $y\notin A$, a contradiction.

We now show the reverse direction. Let $\mathcal{U}=\{O_\alpha : \alpha\in\mathcal{A}\}$ be an open cover of A with no finite subcover. By Proposition 3.12, let $\mathcal{B}$ be a hyperfinite collection of ${}^*\mathcal{U}$ containing ${}^*O_\alpha$ for all $\alpha\in\mathcal{A}$. By the transfer principle, there exists a $y\in {}^*A$ such that $y\notin U$ for all $U\in\mathcal{B}$. Thus, $y\notin {}^*O_\alpha$ for all $\alpha\in\mathcal{A}$. Hence y can not be near-standard, completing the proof. □

This relationship breaks down for non-compact spaces as is shown by the following example.

Example 3.26. Consider ${}^*[0,1] = \{x \in {}^*\mathbb{R} : 0 \leq x \leq 1\}$, as $[0,1]$ is compact we have ${}^*[0,1] = \mathrm{NS}({}^*[0,1])$. $(0,1)$ is not compact and this implies that ${}^*(0,1) \neq \mathrm{NS}({}^*(0,1))$. Indeed, consider any positive infinitesimal $\epsilon \in {}^*\mathbb{R}$. Then $\epsilon \in {}^*(0,1)$ but $\epsilon \notin \mathrm{NS}({}^*(0,1))$.

However, under enough saturation, the standard part map st maps internal sets to compact sets.

Theorem 3.27 ([Lux69])**.** *Let $(X, \mathcal{T})$ be a regular Hausdorff space. Suppose the nonstandard extension is more saturated than the cardinality of $\mathcal{T}$. Let A be a near-standard internal set. Then $E = \mathsf{st}(A) = \{x \in X : (\exists a \in A)(a \in \mu(x))\}$ is compact.*

Proof. Fix $y \in {}^*E$. If U is a standard open set with $y \in {}^*U$, then $U \cap E \neq \emptyset$. Let $x \in E \cap U$. By the definition of E, there exists an $a \in A$ such that $a \in \mu(x) \subset {}^*U$. Thus, for every open set U with $y \in {}^*U$, there exists $a \in A \cap {}^*U$. By saturation, there exists an $a_0 \in A$ such that $a_0 \in A \cap {}^*U$ for all standard open set U with $y \in {}^*U$.

Let $x_0 = \mathsf{st}(a_0)$. In order to finish the proof, by Theorem 3.25, it is sufficient to show that $y \in \mu(x_0)$. Suppose not, then there exists an open set V such that $x_0 \in V$ and $y \notin {}^*V$. By regularity of X, there exists an open set V' such that $x_0 \in V' \subset \overline{V'} \subset V$. Then $x \in V'$ and $y \in {}^*X \setminus {}^*\overline{V'}$. It then follows that $a_0 \in {}^*V'$ and $a_0 \in {}^*X \setminus {}^*\overline{V'}$. This is a contradiction. $\square$

Moreover, for σ-compact locally compact spaces, we have the following result.

Theorem 3.28. *Let X be a Hausdorff space. Suppose X is σ-compact and locally compact. Then there exists a non-decreasing sequence of compact sets K_n with $\bigcup_{n \in \mathbb{N}} K_n = X$ such that $\bigcup_{n \in \mathbb{N}} {}^*K_n = \mathrm{NS}({}^*X)$.*

Proof. As X is σ-compact, there exists a sequence of non-decreasing compact sets G_n such that $X = \bigcup_{n \in \mathbb{N}} G_n$. Let $K_0 = G_0$. By locally compactness of X, for every $x \in K_0 \cup G_1$, let C_x denote a compact subset of X containing a neighborhood U_x

of x. The collection $\{U_x : x \in K_0 \cup G_1\}$ is a cover of $K_0 \cup G_1$ hence there is a finite subcover $\{U_{x_1}, \dots, U_{x_n}\}$. Let $K_1 = \bigcup_{i \leq n} C_{x_i}$. It is easy to see that K_1 is a compact and $K_0 \subset {K_1}^{\mathrm{o}}$ where ${K_1}^{\mathrm{o}}$ denotes the interior of K_1. For any $n \in \mathbb{N}$, we can construct K_n based on $K_{n-1} \cup G_n$ in exactly the same way as we constructs K_1. Hence we have a sequence of compact sets K_n such that $\bigcup_{n\in\mathbb{N}} K_n = X$ and $K_n \subset {K_{n+1}}^{\mathrm{o}}$ for all $n \in \mathbb{N}$.

We now show that $\bigcup_{n\in\mathbb{N}} {}^*K_n = \mathrm{NS}({}^*X)$. As every K_n is compact, by **??**, we know that $\bigcup_{n\in\mathbb{N}} {}^*K_n \subset \mathrm{NS}({}^*X)$. Now pick any element $x \in \mathrm{NS}({}^*X)$. Then $\mathsf{st}(x) \in {}^*K_n$ for some n. As $K_n \subset {K_{n+1}}^{\mathrm{o}}$, we know that $\mu(\mathsf{st}(x)) \subset {}^*K_{n+1}$ hence we have $x \in {}^*K_{n+1}$. Thus, we know that $\mathrm{NS}({}^*X) \subset \bigcup_{n\in\mathbb{N}} {}^*K_n$, completing the proof. □

A merely Hausdorff σ-compact space may not have this property. For a σ-compact, locally compact and Hausdorff space X, the sequence $\{K_n : n \in \mathbb{N}\}$ has to be chosen carefully.

Example 3.29. The set of rational numbers $\mathbb{Q}$ is a Hausdorff σ-compact space. Every compact subset of $\mathbb{Q}$ is finite. Thus, for any collection $\{K_n : n \in \mathbb{N}\}$ of $\mathbb{Q}$ that covers $\mathbb{Q}$, we have $\bigcup_{n\in\mathbb{N}} {}^*K_n = \mathbb{Q}$. That is, any near-standard hyperrational is not in any of the ${}^*K_n$.

Now consider the real line $\mathbb{R}$. Let $K_n = [-n, -\frac{1}{n}] \cup [\frac{1}{n}, n] \cup \{0\}$ for $n \geq 1$. It is easy to see that $\bigcup_{n\in\mathbb{N}} K_n = \mathbb{R}$. However, an infinitesimal is not an element of any ${}^*K_n$.

4. Internal Probability Theory

In this section, we give a brief introduction to nonstandard probability theory. The interested reader can consult [Kei84] and [ACH97, Section 4] for more details. The expert may safely skip this section on first reading.

Let Ω be an internal set. An *internal algebra* $\mathcal{A} \subset \mathscr{P}(\Omega)$ is an internal set containing Ω and closed under complementation and hyperfinite unions/intersections.

A set function $P\colon \mathcal{A} \to {}^*\mathbb{R}$ is *hyperfinitely additive* when, for every $n \in {}^*\mathbb{N}$ and mutually disjoint internal family $\{A_1, \dots, A_n\} \subset \mathcal{A}$, we have $P(\bigcup_{i\leq n} A_i) = \sum_{i\leq n} P(A_i)$.

We are now at the place to introduce the definition of internal probability spaces.

Definition 4.1. An internal finitely-additive probability space is a triple $(\Omega, \mathcal{A}, P)$ where:

(1) Ω is an internal set.
(2) $\mathcal{A}$ is an internal subalgebra of $\mathscr{P}(\Omega)$
(3) $P : \mathcal{A} \to {}^*\mathbb{R}$ is a non-negative hyperfinitely additive internal function such that $P(\Omega) = 1$ and $P(\emptyset) = 0$.

Example 4.2. Let $(X, \mathcal{A}, P)$ be a standard probability space. Then $({}^*X, {}^*\mathcal{A}, {}^*P)$ is an internal probability space. Although $\mathcal{A}$ is a σ-algebra and P is countably additive, $\mathcal{A}$ is just an internal algebra and ${}^*P$ is only hyperfinitely additive. This is because "countable" is not an element of the superstructure.

A special class of an internal probability spaces are hyperfinite probability spaces. Hyperfinite probability spaces behave like finite probability spaces but can be good "approximation" of standard probability space as we will see in future sections.

Definition 4.3. A hyperfinite probability space is an internal probability space $(\Omega, \mathcal{A}, P)$ where:

(1) Ω is a hyperfinite set.
(2) $\mathcal{A} = \mathcal{I}(\Omega)$ where $\mathcal{I}(\Omega)$ denote the collection of all internal subsets of Ω.

Like finite probability spaces, we can specify the internal probability measure P by defining its mass at each $\omega \in \Omega$.

Peter Loeb in [Loe75] showed that any internal probability space can be extended to a standard countably additive probability space. The extension is called the Loeb space of the original internal probability space. The central theorem in modern nonstandard measure theory is the following:

Theorem 4.4 ([Loe75]). *Let $(\Omega, \mathcal{A}, P)$ be an internal finitely additive probability space; then there is a standard (σ-additive) probability space $(\Omega, \overline{\mathcal{A}}, \overline{P})$ such that:*

(1) $\overline{\mathcal{A}}$ is a σ-algebra with $\mathcal{A} \subset \overline{\mathcal{A}} \subset \mathscr{P}(\Omega)$.

(2) $\overline{P}(A) = \mathsf{st}(P(A))$ for any $A \in \mathcal{A}$.

(3) For every $A \in \overline{\mathcal{A}}$ and standard $\epsilon > 0$ there are $A_i, A_o \in \mathcal{A}$ such that $A_i \subset A \subset A_o$ and $P(A_o \setminus A_i) < \epsilon$.

(4) For every $A \in \overline{\mathcal{A}}$ there is a $B \in \mathcal{A}$ such that $\overline{P}(A \triangle B) = 0$.

The probability triple $(\Omega, \overline{\mathcal{A}}, \overline{P})$ is called the Loeb space of $(\Omega, \mathcal{A}, P)$. It is a σ-additive standard probability space. From Loeb's original proof, we can give the explicit form of $\overline{\mathcal{A}}$ and $\overline{P}$:

(1) $\overline{\mathcal{A}}$ equals to:

$$\{A \subset \Omega | \forall \epsilon \in \mathbb{R}^{+} \exists A_i, A_o \in \mathcal{A} \text{ such that } A_i \subset A \subset A_o \text{ and } P(A_o \setminus A_i) < \epsilon\}. \tag{4.1}$$

(2) For all $A \in \overline{\mathcal{A}}$ we have:

$$\overline{P}(A) = \inf\{\overline{P}(A_o) | A \subset A_o, A_o \in \mathcal{A}\} = \sup\{\overline{P}(A_i) | A_i \subset A, A_i \in \mathcal{A}\}. \tag{4.2}$$

In fact, the Loeb σ-algebra can be taken to be the $\overline{P}$-completion of the smallest σ-algebra generated by $\mathcal{A}$. In this paper, we shall assume that our Loeb space is always complete.

The following example of hyperfinite probability space motivates the idea of hyperfinite representation.

Example 4.5. Let $(\Omega, \mathcal{A}, P)$ be a hyperfinite probability space. Pick any $N \in {}^*\mathbb{N} \setminus \mathbb{N}$ and let $\delta t = \frac{1}{N}$. Then δt is an infinitesimal. Let $\Omega = \{\delta t, 2\delta t, ..., 1\}$ and $\mathcal{A} = \mathcal{I}(\Omega)$(Recall that $\mathcal{I}(\Omega)$ is the collection of all internal subsets of Ω). Define P on $\mathcal{A}$ by letting $P(\omega) = \delta t$ for all $\omega \in \Omega$. This is called the uniform hyperfinite Loeb measure.

Claim 4.6. $\mathsf{st}^{-1}(0) \cap \Omega \in \overline{\mathcal{A}}$

Proof. $\mathsf{st}^{-1}(0) \cap \Omega$ consists of elements from Ω that are infinitesimally close to 0. Let $A_n = \{\omega \in \Omega : \omega \leq \frac{1}{n}\}$. By the internal definition principle, A_n is internal for all $n \in \mathbb{N}$. Thus $A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$. Hence $\bigcap_{n\in\mathbb{N}} A_n \in \overline{\mathcal{A}}$. Thus $\mathsf{st}^{-1}(0) \cap \Omega = \bigcap_{n\in\mathbb{N}} A_n \in \overline{\mathcal{A}}$, completing the proof. □

Let ν denote the Lebesgue measure on $[0, 1]$. In Section 6, we will show that $\nu(A) = \overline{P}(\mathsf{st}^{-1}(A) \cap \Omega)$ for every Lebesgue measurable set A. This shows that we can use $(\Omega, \mathcal{A}, P)$ to represent the Lebesgue measure on $[0, 1]$. $(\Omega, \mathcal{A}, P)$ is called a "hyperfinite representation" of the Lebesgue measure space on $[0, 1]$. We will investigate such hyperfinite representation space in more detail in Section 6.

As $\mathsf{st}^{-1}(0)$ is an external set, Example 4.5 shows that the Loeb σ-algebra contains external sets.

4.1. **Product Measures.** In this section, we introduce internal product measures. This would be useful when we are dealing with the product of two hyperfinite Markov chains in later sections.

In this section, let $(\Omega, \mathcal{A}, P_1)$ and $(\Gamma, \mathcal{D}, P_2)$ be two internal probability spaces. Let $(\Omega, \overline{\mathcal{A}}, \overline{P}_1)$ and $(\Gamma, \overline{\mathcal{D}}, \overline{P}_2)$ be the Loeb spaces of $(\Omega, \mathcal{A}, P_1)$ and $(\Gamma, \mathcal{D}, P_2)$, respectively.

Definition 4.7. The product Loeb measure $\overline{P}_1 \times \overline{P}_2$ is defined to be the probability measure on $(\Omega \times \Gamma, \overline{\mathcal{A}} \otimes \overline{\mathcal{D}})$ satisfying:

$$(\overline{P}_1 \times \overline{P}_2)(A \times B) = \overline{P}_1(A) \cdot \overline{P}_2(B). \tag{4.3}$$

for all $A \times B \in \overline{\mathcal{A}} \times \overline{\mathcal{D}}$, where $\overline{\mathcal{A}} \otimes \overline{\mathcal{D}}$ denotes the σ-algebra generated by sets from $\overline{\mathcal{A}} \times \overline{\mathcal{D}}$.

Note that this is nothing more than the standard definition of product measures. Thus $(\Omega \times \Gamma, \overline{\mathcal{A}} \otimes \overline{\mathcal{D}}, \overline{P}_1 \times \overline{P}_2)$ is a standard σ-additive probability space.

It is sometimes more natural to consider the product internal measure $P_1 \times P_2$.

Definition 4.8. The product internal measure $P_1 \times P_2$ is defined to be the internal probability measure on $(\Omega \times \Gamma, \mathcal{A} \otimes \mathcal{D})$ satisfying:

$$(P_1 \times P_2)(A \times B) = P_1(A) \cdot P_2(B). \tag{4.4}$$

for all $A \times B \in \mathcal{A} \times \mathcal{D}$, where $\mathcal{A} \otimes \mathcal{D}$ denote the internal algebra generated by sets from $\mathcal{A} \times \mathcal{D}$.

In this case, we form a product internal probability space $(\Omega \times \Gamma, \mathcal{A} \otimes \mathcal{D}, P_1 \times P_2)$.

Example 4.9. Suppose both $(\Omega, \mathcal{A}, P_1)$ and $(\Gamma, \mathcal{D}, P_2)$ are hyperfinite probability spaces. Recall from Definition 4.3 that $\mathcal{A} = \mathcal{I}(\Omega)$ and $\mathcal{D} = \mathcal{I}(\Gamma)$ where $\mathcal{I}(\Omega)$ and $\mathcal{I}(\Gamma)$ denote the collection of all internal sets of Ω and Γ,respectively. Then the product internal measure $P_1 \times P_2$ is defined on $\mathcal{I}(\Omega \times \Gamma)$. To see this, it is enough to note that every internal subset of $\Omega \times \Gamma$ is hyperfinite hence is a hyperfinite union of singletons.

Once we have the product internal probability space $(\Omega \times \Gamma, \mathcal{A} \otimes \mathcal{D}, P_1 \times P_2)$, the Loeb construction can be applied to give a Loeb probability space $(\Omega \times \Gamma, \overline{(\mathcal{A} \otimes \mathcal{D})}, \overline{(P_1 \times P_2)})$. The Loeb probability space $(\Omega \times \Gamma, \overline{(\mathcal{A} \otimes \mathcal{D})}, \overline{(P_1 \times P_2)})$ is called *the Loeb product space*. It is shown in [KS04] that the Loeb product space $(\Omega \times \Gamma, \overline{(\mathcal{A} \otimes \mathcal{D})}, \overline{(P_1 \times P_2)})$ is uniquely determined by the factor Loeb spaces $(\Omega, \overline{\mathcal{A}}, \overline{P_1})$ and $(\Gamma, \overline{\mathcal{D}}, \overline{P_2})$, and not on the internal probability spaces $(\Omega, \mathcal{A}, P_1)$ and $(\Gamma, \mathcal{D}, P_2)$ that generate $(\Omega, \overline{\mathcal{A}}, \overline{P_1})$ and $(\Gamma, \overline{\mathcal{D}}, \overline{P_2})$. It is natural to seek for relation between $(\Omega \times \Gamma, \overline{(\mathcal{A} \otimes \mathcal{D})}, \overline{(P_1 \times P_2)})$ and $(\Omega \times \Gamma, \overline{\mathcal{A}} \otimes \overline{\mathcal{D}}, \overline{P}_1 \times \overline{P}_2)$.

Theorem 4.10 ([Kei84])**.** *Consider two Loeb probability spaces* $(\Omega, \overline{\mathcal{A}}, \overline{P}_1)$ *and* $(\Gamma, \overline{\mathcal{D}}, \overline{P}_2)$. *We have* $\overline{(P_1 \times P_2)} = \overline{P}_1 \times \overline{P}_2$ *on* $\overline{\mathcal{A}} \otimes \overline{\mathcal{D}}$.

Proof. We first show that $\overline{\mathcal{A}} \otimes \overline{\mathcal{D}} \subset \overline{(\mathcal{A} \otimes \mathcal{D})}$. It is enough to show that for any $A \times B \in \overline{\mathcal{A}} \times \overline{\mathcal{D}}$ we have $A \times B \in \overline{(\mathcal{A} \otimes \mathcal{D})}$. Fix an $\epsilon \in (0, 1)$. As $A \subset \overline{\mathcal{A}}$, by Loeb's construction, there exists $A_i, A_o \in \mathcal{A}$ with $A_i \subset A \subset A_o$ such that $P_1(A_o \setminus A_i) < \epsilon$.

Similarly, there exist such $B_i, B_o \in \mathcal{D}$ for B. Then we have

$$(P_1 \times P_2)((A_o \times B_o) \setminus (A_i \times B_i)) = (P_1 \times P_2)((A_o \setminus A_o) \times (A_i \setminus B_i)) = \epsilon^2 < \epsilon. \tag{4.5}$$

As our choice of ϵ is arbitrary, we have $A \times B \in \overline{(\mathcal{A} \otimes \mathcal{D})}$.

We now show that $\overline{(P_1 \times P_2)} = \overline{P}_1 \times \overline{P}_2$ on $\overline{\mathcal{A}} \otimes \overline{\mathcal{D}}$. Again it is enough to just consider $A \times B \in \overline{\mathcal{A}} \times \overline{\mathcal{D}}$. We then have:

$$\overline{P}_1 \times \overline{P}_2(A \times B) \tag{4.6}$$

$$= \sup\{\mathsf{st}(P_1(A_i))|A_i \subset A, A_i \in \mathcal{A}\} \times \sup\{\mathsf{st}(P_2(B_i))|B_i \subset A, B_i \in \mathcal{D}\} \tag{4.7}$$

$$= \sup\{\mathsf{st}(P_1(A_i))\mathsf{st}(P_2(B_i))|A_i \subset A, A_i \in \mathcal{A}, B_i \subset A, B_i \in \mathcal{D}\} \tag{4.8}$$

$$= \overline{(P_1 \times P_2)}(A \times B), \tag{4.9}$$

completing the proof. □

However, $\overline{\mathcal{A}} \otimes \overline{\mathcal{D}}$ will generally be a smaller σ-algebra than $\overline{(\mathcal{A} \otimes \mathcal{D})}$ as is shown by the following example which is due to Doug Hoover.

Example 4.11. [Kei84] Let Ω be an infinite hyperfinite set. Let $\Gamma = \mathcal{I}(\Omega)$. Let $(\Omega, \mathcal{I}(\Omega), P)$ and $(\Gamma, \mathcal{I}(\Gamma), Q)$ be two uniform hyperfinite probability spaces over the respective sets. Let $E = \{(\omega, \lambda) : \omega \in \lambda \in \Gamma\}$. It can be shown that $E \in \overline{(\mathcal{I}(\Omega) \otimes \mathcal{I}(\Gamma))}$ but $E \notin \overline{\mathcal{I}(\Omega)} \otimes \overline{\mathcal{I}(\Gamma)}$. It can be shown that $(\overline{P \times Q})(E) > 0$ while $\overline{P}(A)\overline{Q}(B) = 0$ for every $A \in \mathcal{I}[\Omega]$ and every $B \in \mathcal{I}[\Gamma]$.

In [Sun98], the author gave a complete characterization of the relationship between the two types of product spaces for Loeb spaces. Before we quote the result from [Sun98], we first recall the following common definition from measure theory.

Definition 4.12. Let $(X, \mathcal{F}, P)$ be a probability space. An atom is a set $A \in \mathcal{F}$ with $P(A) > 0$ and the property that for each measurable subset $B \subset A$, either $P(B) = 0$ or $P(B) = P(A)$. $(X, \mathcal{F}, P)$ is called purely atomic if every measurable set with positive measure contains an atom.

We quote the following result from [Sun98] which gives a complete characterization between two types of product spaces for Loeb spaces.

Theorem 4.13 ([Sun98, Proposition. 6.6]). *Let $(\Omega, \mathcal{A}, P_1)$ and $(\Gamma, \mathcal{D}, P_2)$ be two internal probability spaces. The completion of the product Loeb σ-algebra $\overline{\mathcal{A}} \otimes \overline{\mathcal{D}}$ is strictly contained in the Loeb product σ-algebra $\overline{\mathcal{A} \otimes \mathcal{D}}$ if and only if both $\overline{P_1}$ and $\overline{P_2}$ are not purely atomic.*

It is natural to ask whether the product Loeb σ-algebra is a "rich" subset of the Loeb product σ-algebra. In particular, let $(\Omega, \mathcal{A}, P)$ be an internal probability space and suppose $(\overline{P \times P})(B) > 0$ for some $B \in \overline{\mathcal{A} \otimes \mathcal{A}}$, does there exists $C \in \overline{\mathcal{A}} \otimes \overline{\mathcal{A}}$ such that $C \subset B$ and $(\overline{P \times P})(C) > 0$? The answer of this question is generally negative. It is shown in [Ber+02, Thm. 5.1] that for any two atomless Loeb spaces and any number $s \in [0, 1]$, there is a measurable set E in the corresponding Loeb product space with Loeb product measure s such that its inner and outer measure, based on the usual product of the factor Loeb spaces, are zero and one respectively.

4.2. **Nonstandard Integration Theory.** In this section we establish the nonstandard integration theory on Loeb spaces. Fix an internal probability space (Ω, Γ, P) and let $(\Omega, \overline{\Gamma}, \overline{P})$ denote the corresponding Loeb space. If Γ is ${}^*\sigma$-algebra then we have the notion of "P-integrability" which is nothing more than the usual integrability "copied" from the standard measure theory. Note that the Loeb space $(\Omega, \overline{\Gamma}, \overline{P})$ is a standard countably additive probability space. The Loeb integrability is the same as the integrability with respect to the probability measure $\overline{P}$. We mainly focus on discussing the relationship between "P-integrability" and Loeb integrability in this section.

Corollary 4.14 ([ACH97, Corollary 4.6.1]). *Suppose (Ω, Γ, P) is an internal probability space, and $F : \Omega \to {}^*\mathbb{R}$ is an internal measurable function such that $\mathsf{st}(F)$ exists everywhere. Then $\mathsf{st}(F)$ is Loeb integrable and $\int F \mathrm{d}P \approx \int \mathsf{st}(F) \mathrm{d}\overline{P}$.*

The situation is more difficult when $\mathsf{st}(F)$ exists almost surely. We present the following well-known result.

Theorem 4.15 ([ACH97, Theorem 4.6.2])**.** *Suppose* (Ω, Γ, P) *is an internal probability space, and* $F : \Omega \to {}^*\mathbb{R}$ *is an internally integrable function such that* $\mathsf{st}(F)$ *exists* $\overline{P}$*-almost surely. Then the following are equivalent:*

1. $\mathsf{st}(\int |F| \mathrm{d}P)$ *exists and it equals to* $\lim_{n\to\infty} \mathsf{st}(\int |F_n| \mathrm{d}P)$ *where for* $n \in \mathbb{N}$, $F_n = \min\{F, n\}$ *when* $F \geq 0$ *and* $F_n = \max\{F, -n\}$ *when* $F \leq 0$.
2. *For every infinite* $K > 0$, $\int_{|F|>K} |F| \mathrm{d}P \approx 0$.
3. $\mathsf{st}(\int |F| \mathrm{d}P)$ *exists, and for every* B *with* $P(B) \approx 0$*, we have* $\int_B |F| \mathrm{d}P \approx 0$.
4. $\mathsf{st}(F)$ *is* $\overline{P}$*-integrable, and* ${}^*\int F \mathrm{d}P \approx \int \mathsf{st}(F) \mathrm{d}\overline{P}$.

Definition 4.16. Suppose (Ω, Γ, P) is an internal probability space, and $F : \Omega \to {}^*\mathbb{R}$ is an internally integrable function such that $\mathsf{st}(F)$ exists $\overline{P}$-almost surely. If F satisfies any of the conditions (1)-(4) in Theorem 4.15, then F is called a S-integrable function.

Up to now, we have been discussing the internal integrability as well as the Loeb integrability of internal functions. An external function is never internally integrable. However, it is possible that some external functions are Loeb integrable. We start by introducing the following definition.

Definition 4.17. Suppose that $(\Omega, \overline{\Gamma}, \overline{P})$ is a Loeb space, that X is a Hausdorff space, and that f is a measurable (possibly external) function from Ω to X. An internal function $F : \Omega \to {}^*X$ is a lifting of f provided that $f = \mathsf{st}(F)$ almost surely with respect to $\overline{P}$.

We conclude this section by the following Loeb integrability theory.

Theorem 4.18 ([ACH97, Theorem 4.6.4])**.** *Let* $(\Omega, \overline{\Gamma}, \overline{P})$ *be a Loeb space, and let* $f : \Omega \to \mathbb{R}$ *be a measurable function. Then* f *is Loeb integrable if and only if it has a S-integrable lifting.*

5. Measurability of Standard Part Map

When we apply nonstandard analysis to attack measure theory questions, the standard part map st plays an essential role since $\mathsf{st}^{-1}(E)$ for $E \in \mathcal{B}[X]$ is usually considered to be the nonstandard counterpart for E. Thus a natural question to ask is: when is the standard part map st a measurable function? There are quite a few answers to this question in the literature (see, eg,. [ACH97, Section 4.3]) and they should cover most of the interesting cases. It turns out that, in most interesting cases, the measurability of st depends on the Loeb measurability of $\mathrm{NS}({}^*X)$. Such results are mentioned in [ACH97, Exescise 4.19,4.20]. However, we give a proof for more general topological spaces in this section.

The following theorem of Ward Henson in [Hen79] is a key result regarding the measurability of st.

Theorem 5.1 ([ACH97, Theorems 4.3.1 and 4.3.2])**.** *Let X be a regular topological space, let P be an internal, finitely additive probability measure on $({}^*X, {}^*\mathcal{B}[X])$ and suppose $\mathrm{NS}({}^*X) \in \overline{{}^*\mathcal{B}[X]}$; then st is Borel measurable from $({}^*X, \overline{{}^*\mathcal{B}[X]})$ to $(X, \mathcal{B}[X])$.*

Thus we only need to figure out what conditions on X will guarantee that $\mathrm{NS}({}^*X) \in \overline{{}^*\mathcal{B}[X]}$. In the literature, people have shown that, for σ-compact, locally compact or completely metrizable spaces X, we have $\mathrm{NS}({}^*X) \in \overline{{}^*\mathcal{B}[X]}$. In this section we will generalize such results to more general topological spaces.

We first recall the following definitions from general topology.

Definition 5.2. Let X be a topological space. A subset A is a G_δ set if A is a countable intersection of open sets. A subset is a F_σ set if its complement is a G_δ set.

Definition 5.3. For a Tychonoff space X, it is Cech complete if there exist a compactification Y such that X is a G_δ subset of Y.

The following lemma is due to Landers and Rogge. We provide a proof here since it is closely related to our main result of this section.

Lemma 5.4 ([LR87])**.** *Suppose that $(\Omega, \mathcal{A}, P)$ is an internal finitely additive probability space with corresponding Loeb space $(\Omega, \mathcal{A}_L, \overline{P})$ and suppose that $\mathcal{C}$ is a subset* *of $\mathcal{A}$ such that the nonstandard model is more saturated than the external cardinality of $\mathcal{C}$. Then $\bigcap \mathcal{C} \in \mathcal{A}_L$. Furthermore, if $P(A) = 1$ for all $A \in \mathcal{C}$, then $\overline{P}(\bigcap \mathcal{C}) = 1$*

Proof. Without loss of generality we can assume that $\mathcal{C}$ is closed under finite intersections. Let $r = \inf\{\overline{P}(C) : C \in \mathcal{C}\}$. Fix a standard $\epsilon > 0$. We can find $C_o \in \mathcal{C} \subset \mathcal{A}$ such that $P(C_o) < r + \epsilon$. Denote $\mathcal{C} = \{C_\alpha : \alpha \in J\}$ where J is some index set. Consider the set of formulas $\{\phi_\alpha(A) | \alpha \in J\}$ where $\phi_\alpha(A)$ is $(A \in \mathcal{A}) \wedge (P(A) > r - \epsilon) \wedge ((\forall a \in A)(a \in C_\alpha))$. As $\mathcal{C}$ is closed under finite intersection and $r = \inf\{\overline{P}(C) : C \in \mathcal{C}\}$, we have $\{\phi_\alpha(A) : \alpha \in J\}$ is finitely satisfiable. By saturation, we can find a set $A_i \in \mathcal{A}$ such that $P(A_i) > r - \epsilon$ and $A_i \subset \bigcap \mathcal{C}$. So $\bigcap \mathcal{C} \in \mathcal{A}_L$.

If $\forall C \in \mathcal{C}$ we have $P(C) = 1$, by the same construction in the last paragraph, we have $1 - \epsilon \leq \overline{P}(A_i) \leq \overline{P}(\bigcap \mathcal{C}) \leq \overline{P}(A_o) = 1$ for every positive $\epsilon \in \mathbb{R}$. Thus we have the desired result. □

In the context of Lemma 5.4, by considering the complement, it is easy to see that $\bigcup \mathcal{C} \in \overline{\mathcal{A}}$. Similarly, if we have $P(A) = 0$ for all $A \in \mathcal{C}$ then $\overline{P}(\bigcup \mathcal{C}) = 0$.

We quote the next lemma which establishes the Loeb measurability of $\mathrm{NS}({}^*X)$ for σ-compact spaces.

Lemma 5.5 ([LR87])**.** *Let X be a σ-compact space with Borel σ-algebra $\mathcal{B}[X]$ and let $({}^*X, {}^*\mathcal{B}[X]_L, \overline{P})$ be a Loeb space. Then $\mathrm{NS}({}^*X) \in \overline{{}^*\mathcal{B}[X]}$.*

We are now at the place to prove the measurability of $\mathrm{NS}({}^*X)$ for Cech complete spaces.

Theorem 5.6. *If the Tychnoff space X is Cech complete then $\mathrm{NS}({}^*X) \in {}^*\mathcal{B}[X]_L$.*

Proof. Let Y be a compactification of X such that X is a G_δ subset of Y. We use S to denote $Y \setminus X$. Then S is a F_σ subset of T hence is a σ-compact subset of Y. Let $S = \bigcup_{i \in \omega} S_i$ where each S_i is a compact subset of Y. Note that

$$ {}^*Y = {}^*X \cup {}^*S = \mathrm{NS}({}^*X) \cup {}^*S \cup Z. \tag{5.1} $$

where $Z = {}^*X \setminus \mathrm{NS}({}^*X)$. As Y is compact, we know that $Z = \{x \in {}^*X : (\exists s \in S)(x \in \mu(s))\}$. Note that $\mathrm{NS}({}^*X), {}^*S, Z$ are mutually disjoint sets. Let $N_i = \{y \in {}^*Y : (\exists x \in S_i)(y \in \mu(x))\}$.

Claim 5.7. *For any $i \in \omega$, $N_i \in \overline{{}^*\mathcal{B}[X]}$.*

Proof. : Without loss of generality, it is enough to prove the claim for N_1. Let $\mathcal{U} = \{U \subset X\text{: U is open and } S_1 \subset U\}$. We claim that $N_1 = \bigcap\{{}^*U : U \in \mathcal{U}\}$. To see this, we first consider any $u \in \bigcap\{{}^*U : U \in \mathcal{U}\}$. Suppose $u \notin N_1$, this means that for any $y \in S_1$ there exists ${}^*U_y$ such that U_y is open and $u \notin {}^*U_y$. As S_1 is compact, we can pick finitely many $y_1, \ldots, y_n$ such that $S_1 \subset \bigcup_{i\leq n} U_{y_i}$. Thus we have ${}^*\bigcup_{i\leq n} U_{y_i} = \bigcup_{i\leq n} {}^*U_{y_i} \subset \bigcup_{y\in S_1} {}^*U_y$. Note that $u \notin \bigcup_{y\in S_1} {}^*U_y$ implies that $u \notin {}^*\bigcup_{i\leq n} U_{y_i}$. But $\bigcup_{i\leq n} U_{y_i}$ is an element of $\mathcal{U}$. Hence we have a contradiction. Conversely, it is easy to see that $N_1 \subset \bigcap\{{}^*U : U \in \mathcal{U}\}$. We also know that each ${}^*U \in {}^*\mathcal{B}[X]$. Assume that we are working on a nonstandard extension which is more saturated than the cardinality of the topology of X, then for any $i \in \omega$ $N_i \in \overline{{}^*\mathcal{B}[X]}$ by Lemma 5.4. □

It is also easy to see that $\bigcup_{i<\omega} N_i = \mathrm{NS}({}^*S) \cup Z$. By Lemma 5.5, we know that both $\bigcup_{i<\omega} N_i$ and $\mathrm{NS}({}^*S)$ belong to $\overline{{}^*\mathcal{B}[Y]}$. Hence $Z \in \overline{{}^*\mathcal{B}[Y]}$.

As S is σ-compact in Y, we know that $S \in \mathcal{B}[Y]$. By the transfer principle, we know that ${}^*S \in {}^*\mathcal{B}[Y] \subset \overline{{}^*\mathcal{B}[Y]}$. As both ${}^*S$ and Z belong to $\overline{{}^*\mathcal{B}[Y]}$, it follows that $\mathrm{NS}({}^*X) \in \overline{{}^*\mathcal{B}[Y]}$.

We now show that $\mathrm{NS}({}^*X) \in \overline{{}^*\mathcal{B}[X]}$. Fix an arbitrary internal probability measure P on $({}^*X, {}^*\mathcal{B}[X])$. Let P' be the extension of P to $({}^*Y, {}^*\mathcal{B}[Y])$ defined by $P'(A) = P'(A \cap X)$. We already know that $\mathrm{NS}({}^*X) \in \overline{{}^*\mathcal{B}[Y]}$. By definition, this means that for every positive $\epsilon \in \mathbb{R}$ there exist $A_i, A_o \in {}^*\mathcal{B}[Y]$ such that $A_i \subset \mathrm{NS}({}^*X) \subset A_o$ and $P'(A_o \setminus A_i) < \epsilon$. Let $B_i = A_i \cap {}^*X$ and $B_o = A_o \cap {}^*X$. By the construction of P and P', it is clear that $B_i \subset \mathrm{NS}({}^*X) \subset B_o$ and $P(B_o \setminus B_i) < \epsilon$. It remains to show that B_i and B_o both lie in ${}^*\mathcal{B}[X]$. The transfer of $(\forall A \in \mathcal{B}[Y])(A \cap X \in \mathcal{B}[X])$ gives us the final result. □

Thus, by Theorem 5.1, we know that st is measurable for Cech-complete spaces. For regular spaces, either locally compact spaces or completely metrizable spaces are Cech-complete. Thus we have established the measurability of st for more general topological spaces. However, note that σ-compact metric spaces need not be Cech complete.

We now introduce the concept of universally Loeb measurable sets.

Recall from Section 4 that given an internal algebra $\mathcal{A}$ its Loeb extension $\overline{\mathcal{A}}$ is actually the $\overline{P}$-completion of the σ-algebra generated by $\mathcal{A}$. So $\mathcal{A}_L$ could differ for different internal probability measures. We use $\overline{\mathcal{A}}^P$ to denote the Loeb extension of $\mathcal{A}$ with respect to the internal probability measure P.

Definition 5.8. A set $A \subset {}^*X$ is called universally Loeb-measurable if $A \in \overline{\mathcal{A}}^P$ for every internal probability measure P on $({}^*X, \mathcal{A})$.

We denote the collection of all universally-Loeb measurable sets by $\mathcal{L}(\mathcal{A})$. By Theorem 5.6, $\mathrm{NS}({}^*X)$ is universally Loeb measurable if X is Cech complete. Moreover, Theorem 5.1 can be restated as following:

Theorem 5.9 ([LR87])**.** *Let X be a Hausdorff regular space equipped with Borel σ-algebra $\mathcal{B}[X]$. If $B \in \mathcal{B}[X]$ then $\mathsf{st}^{-1}(B) \in \{A \cap \mathrm{NS}({}^*X) : A \in \mathcal{L}(\mathcal{B}[X])\}$.*

Thus, by Theorem 5.6, $\mathsf{st}^{-1}(B)$ is universally measurable for every $B \in \mathcal{B}[X]$ if X is Cech complete.

We conclude this section by giving an example of a relatively nice space where $\mathrm{NS}({}^*X)$ is not measurable.

Theorem 5.10. *[ACH97, Example 4.1] There is a separable metric space X and a Loeb space $({}^*X, {}^*\mathcal{B}[X], \overline{P})$ such that $\mathrm{NS}({}^*X)$ is not measurable.*

Proof. Let X be the Bernstein set of $[0,1]$; for every uncountable closed subset A of $[0,1]$, both $A \cap X$ and $A \cap ([0,1] \setminus X)$ are nonempty. The topology on X is the natural subspace topology inherited from standard topology on $[0,1]$. Clearly $B \subset X$ is Borel if and only if $B = X \cap B'$ for some Borel subset B' of $[0,1]$. Let μ

denote the Lebesgue measure on $([0,1],\mathcal{B}[[0,1]])$. Let $\mathcal{A}$ be the σ-algebra generated from $\mathcal{B}[[0,1]]\cup\{X\}$. Let m be the extension of μ to $\mathcal{A}$ by letting $m(X)=1$.

Claim 5.11. *m is a probability measure on* $([0,1],\mathcal{A})$.

Proof. It is sufficient to show that, for any $A,B\in\mathcal{B}[[0,1]]$, we have

$$m(A\cap X)=m(B\cap X)\rightarrow m(A)=m(B). \tag{5.2}$$

Suppose not. Then $m(A\triangle B)>0$. As $m(A\cap X)=m(B\cap X)$, we have $m((A\triangle B)\cap X)=0$. But we already know that $m([0,1]\setminus X)=0$ □

Let P be the restriction of ${}^*m$ to ${}^*\mathcal{B}[X]$. Consider the internal probability space $({}^*X,{}^*\mathcal{B}[X],P)$. Let $A\in \mathrm{NS}({}^*X)\cap{}^*\mathcal{B}[X]$ and let $A'=\mathsf{st}_X(A)$ where $\mathsf{st}_X(A)=\{x\in X:(\exists a\in A)(a\approx x)\}$. By Theorem 3.27, we know that A' is a compact subset of X. Thus A' is a closed subset of $[0,1]$. As X does not contain any uncountable closed subset of $[0,1]$, we conclude that A' must be countable. Thus, for any $\epsilon>0$, there exists an open set $U_\epsilon\subset[0,1]$ of Lebesgue measure less than ϵ that contains A'. As $A'=\mathsf{st}_X(A)$, we know that $A\subset{}^*X\cap{}^*U_\epsilon\subset{}^*U_\epsilon$. Then $P(A)\leq{}^*m({}^*U_\epsilon)<\epsilon$. Thus the $\overline{P}$-inner measure of $\mathrm{NS}({}^*X)$ is 0. By applying the same technique to $[0,1]\setminus X$, we can show that the $\overline{P}$-outer measure of $\mathrm{NS}({}^*X)$ is 1. Thus $\mathrm{NS}({}^*X)$ can not be Loeb measurable. □

This is slightly different from [ACH97, Example 4.1]. In [ACH97, Example 4.1], the author let m be a finitely-additive extension of Lebesgue measure to all subsets of $[0,1]$. In this paper, we let m to be a countably-additive extension of the Lebesgue measure to include the Bernstein set.

6. Hyperfinite Representation of a Probability Space

In the literature of nonstandard measure theory, there exist quite a few results to represent standard measure spaces using hyperfinite measure spaces. For example, see [BW69; Loe74; Hen72; And82]. In this section, we establish a hyperfinite representation theorem for Heine-Borel metric spaces with Radon probability measures. Although we restrict ourselves to a smaller class of spaces, we believe that we provide a more intuitive and simple construction. Moreover, such a construction will be used extensively in later sections.

Let X be a σ-compact metric space. Let d denote the metric in X. Then ${}^*d$ will denote the metric on ${}^*X$. We impose the following definition on our space X.

Definition 6.1. A metric space is said to satisfy the Heine-Borel condition if the closure of every open ball is compact.

Note that the Heine-Borel condition is equivalent to that every closed bounded set is compact.

As we mentioned in Section 3.2, finite elements of complete metric spaces need not be near-standard. However, finite elements are near-standard for metric spaces satisfying the Heine-Borel condition.

Theorem 6.2. *A metric space X satisfies the Heine-Borel condition if and only if every finite element in ${}^*X$ is near-standard.*

Proof. Let X be a metric space with metric d. Suppose X satisfies the Heine-Borel condition. Let $y \in {}^*X$ be a finite element. Then there exists $x \in X$ and $k \in \mathbb{N}$ such that ${}^*d(x, y) < k$. Let U_y^k denote the open ball centered at y with radius k. Clearly we know that $y \in {}^*U_y^k \subset {}^*(\overline{U_y^k})$. As X satisfies the Heine-Borel condition, we know that $\overline{U_y^k}$ is a compact set. By Theorem 3.25, there exists an element $x_0 \in \overline{U_y^k}$ such that $y \in \mu(x_0)$.

We now prove the reverse direction. Suppose X does not satisfy the Heine-Borel condition. Then there exists an open ball U such that $\overline{U}$ is not compact. By Theorem 3.25, there exists an element $y \in {}^*(\overline{U})$ such that y is not in the monad of any element $x \in \overline{U}$. As $y \in {}^*(\overline{U})$, y is finite hence is near-standard. Thus there exists a $x_0 \in X \setminus \overline{U}$ such that $y \in \mu(x_0)$. Thus there exists an open ball V centered at x_0 such that $V \cap \overline{U} = \emptyset$. Then we have $y \in {}^*V$ and $y \in {}^*\overline{U}$, which is a contradiction. Thus the closure of every open ball of X must be compact, completing the proof. $\square$

We shall assume our state space X is a metric space satisfying the Heine-Borel condition in the remainder of this paper unless otherwise mentioned. Note that metric spaces satisfying the Heine-Borel condition are complete and σ-compact.

We are now at the place to introduce the hyperfinite representation of a topological space. The idea behind hyperfinite representation is quite simple: For a metric space X, we partition an "initial segment" of ${}^*X$ into hyperfinitely pieces of sets with infinitesimal diameters. We then pick exactly one element from each element of the partition to form our hyperfinite representation. The formal definition is stated below.

Definition 6.3. Let X be a metric space satisfying the Heine-Borel condition. Let $\epsilon \in {}^*\mathbb{R}^+$ be an infinitesimal and r be an infinite nonstandard real number. A hyperfinite set $S \subset {}^*X$ is said to be an (ϵ, r)-hyperfinite representation of ${}^*X$ if the following three conditions hold:

(1) For each $s \in S$, there exists a $B(s) \in {}^*\mathcal{B}[X]$ with diameter no greater than ϵ containing s such that $B(s_1) \cap B(s_2) = \emptyset$ for any two different $s_1, s_2 \in S$.
(2) For any $x \in \mathrm{NS}({}^*X)$, ${}^*d(x, {}^*X \setminus \bigcup_{s\in S} B(s)) > r$.
(3) There exists $a_0 \in X$ and some infinite r_0 such that

$$\mathrm{NS}({}^*X) \subset \bigcup_{s\in S} B(s) = \overline{U(a_0, r_0)} \tag{6.1}$$

where $\overline{U(a_0, r_0)} = \{x \in {}^*X : {}^*d(x, a_0) \leq r_0\}$.

If X is compact, then $\bigcup_{s\in S} B(s) = {}^*X$. In this case, the second parameter of an (ϵ, r)-hyperfinite representation is redundant. Thus, we have ϵ-hyperfinite representation for compact space X.

Definition 6.4. Let $\mathcal{T}$ denote the topology of X and $\mathcal{K}$ denote the collection of compact sets of X. A *open set is an element of ${}^*\mathcal{T}$ and a *compact set is an element of ${}^*\mathcal{K}$.

By the transfer principle, a set A is a *compact set if for every *open cover of A there is a hyperfinite subcover. By the Heine-Borel condition, the closure of every open ball is a compact subset of X. By the transfer principle, we know that $\overline{U(a_0, r_0)}$ in Definition 6.3 is *compact.

Example 6.5. Consider the real line $\mathbb{R}$ with standard metric. Fix $N_1, N_2 \in {}^*\mathbb{N} \setminus \mathbb{N}$. Let $\epsilon = \frac{1}{N_1}$ and let $r = 2N_2$. It then follows that

$$S = \{-2N_2, -2N_2 + \frac{1}{N_1}, \ldots, -\frac{1}{N_1}, 0, \frac{1}{N_1}, \ldots, 2N_2\} \tag{6.2}$$

is a (ϵ, r)-hyperfinite representation of ${}^*\mathbb{R}$.

To see this, we need to check the three conditions in Definition 6.3. For $s = 2N_2$, let $B(s) = \{2N_2\}$. For other $s \in S$, let $B(s) = [s, s + \frac{1}{N_1})$. Clearly $\{B(s) : s \in S\}$ is a mutually disjoint collection of *Borel sets with diameter no greater than $\frac{1}{N_1}$. Moreover, it is easy to see that $\bigcup_{s \in S} B(s) = [-2N_2, 2N_2] \supset \mathrm{NS}({}^*\mathbb{R})$. For every element $y \in {}^*\mathbb{R} \setminus [-2N_2, 2N_2]$, we have ${}^*d(y, 0) > 2N_2$. Then the distance between y and any near-standard element is greater than N_2. Finally, by the transfer principle, we know that $\bigcup_{s \in S} B(s) = [-2N_2, 2N_2]$ is a *compact set.

Theorem 6.6. *Let X be a metric space satisfying the Heine-Borel condition. Then for every positive infinitesimal ϵ and every positive infinite r there exists a (ϵ, r)-hyperfinite representation S_ϵ^r of ${}^*X$.*

Proof. Let us start by assuming X is non-compact. Since X satisfies the Heine-Borel condition, X must be unbounded. Fix an infinitesimal $\epsilon_0 \in {}^*\mathbb{R}^+$ and an infinite r_0. Pick any standard $x_0 \in X$ and consider the open ball

$$U(x_0, 2r_0) = \{x \in {}^*X : {}^*d(x, x_0) < 2r_0\}. \tag{6.3}$$

As X is unbounded, $U(x_0, 2r_0)$ is a proper subset of ${}^*X$. Moreover, as X satisfies the Heine-Borel condition, $\overline{U}(x_0, 2r_0)$ is a *compact proper subset of ${}^*X$. The following sentence is true for X:

$(\forall r, \epsilon \in \mathbb{R}^+)(\exists N \in \mathbb{N})(\exists \mathcal{A} \in \mathscr{P}(\mathcal{B}[X]))(\mathcal{A}$ has cardinality N and $\mathcal{A}$ is a collection of mutually disjoint sets with diameters no greater than ϵ and $\mathcal{A}$ covers $\overline{U}(x_0, r))$

By the transfer principle, we have:

$(\exists K \in {}^*\mathbb{N})(\exists \mathcal{A} \in {}^*\mathscr{P}(\mathcal{B}[X]))(\mathcal{A}$ has internal cardinality K and $\mathcal{A}$ is a collection of mutually disjoint sets with diameters no greater than ϵ_0 and $\mathcal{A}$ covers $\overline{U}(x_0, 2r_0))$

Let $\mathcal{A} = \{U_i : i \leq K\}$. Without loss of generality, we can assume that U_i is a subset of $\overline{U}(x_0, 2r_0)$ for all $i \leq K$. It follows that $\bigcup_{i\leq K} U_i = \overline{U}(x_0, 2r_0)$ which implies that $\mathrm{NS}({}^*X) \subset \bigcup_{i\leq K} U_i$. For any $x \in \mathrm{NS}({}^*X)$ and any $y \in {}^*X \setminus \overline{U}(x_0, 2r_0)$, we have ${}^*d(x, y) > r_0$. By the axiom of choice, we can pick one element $s_i \in U_i$ for $i \leq K$. Let $S_{\epsilon_0}^{r_0} = \{s_i : i \leq K\}$ and it is easy to check that this $S_{\epsilon_0}^{r_0}$ satisfies all the conditions in Definition 6.3.

It is easy to see that an essentially same but much simpler proof would work when X is compact. □

For an (ϵ, r)-hyperfinite representation S_ϵ^r, it is possible for S_ϵ^r to contain every element of X.

Lemma 6.7. *Suppose our nonstandard model is more saturated than the cardinality of X, then we can construct S_ϵ^r so that $X \subset S_\epsilon^r$.*

Proof. Let $\mathcal{A} = \{U_i : i \leq K\}$ be the same object as in Theorem 6.6 and let $S_\epsilon^r = \{s_i : i \leq K\}$ be a hyperfinite representation constructed from $\mathcal{A}$. Let a=$\{S$: S is a hyperfinite subset of ${}^*X$ with internal cardinality $K\}$. Note that a is itself an internal set. Pick $x \in X$ and let $\phi_x(S)$ be the formula

$$(S \in a) \wedge ((\forall s \in S)(\exists! U \in \mathcal{A})(s \in U)) \wedge (x \in S). \tag{6.4}$$

Consider the family $\mathcal{F} = \{\phi_x(S) | x \in X\}$, we now show that this family is finitely satisfiable. Fix finitely many elements $x_1, ...x_k$ from X, we define a function f from S_ϵ^r to ${}^*X$ as follows: For each $i \leq N$, if $\{x_1, ...x_k\} \cap U_i = \emptyset$ then $f(s_i) = s_i$. If the intersection is nonempty, then $\{x_1, ...x_k\} \cap C_i = \{x\}$ for some $x \in \{x_1, ...x_k\}$. In this case, we let $f(s_i) = x$. By the internal definition principle, such f is an internal function and $f(S_{\epsilon}^{r})$ is the realization of the formula $\phi_{x_1}(S) \cap \cdots \cap \phi_{x_k}(S)$. By saturation, there would be a $S_0 \in a$ satisfies all the formulas in $\mathcal{F}$ simultaneously. This S_0 is the desired set. □

Let $(X, \mathcal{B}[X], P)$ be a Borel probability space satisfying the conditions of Theorem 6.6 and let S be an (ϵ, r)-hyperfinite representation of ${}^*X$. We now show that we can define an internal measure on $(S, \mathcal{I}(S))$ such that the resulting internal probability space is a good representation of $(X, \mathcal{B}[X], P)$. Similar theorems have been given assuming that X is merely Hausdorff [And82]. Here we assume X is a metric space satisfying Heine-Borel conditions and as a consequence we will obtain tighter control on the representation of $(X, \mathcal{B}[X], P)$.

Before we introduce the main theorem of this section, we first quote the following useful lemma by Anderson.

Lemma 6.8 ([ACH97, Thm 4.1])**.** *Let $(X, \mathcal{B}[X], \mu)$ be a σ-compact Borel probability space. Then st is measure preserving from $({}^*X, \overline{{}^*\mathcal{B}[X]}, \overline{{}^*\mu})$ to $(X, \mathcal{B}[X], \mu)$. That is, we have $\mu(E) = \overline{{}^*\mu}(\mathsf{st}^{-1}(E))$ for all $E \in \mathcal{B}[X]$.*

Proof. Let $E \in \mathcal{B}[X]$, $\epsilon \in \mathbb{R}^+$ and choose K compact, U open with $K \subset E \subset U$ and $\mu(U) - \mu(K) < \epsilon$. Note that ${}^*K \subset \mathsf{st}^{-1}(K) \subset \mathsf{st}^{-1}(E) \subset \mathsf{st}^{-1}(U) \subset {}^*U$, and ${}^*\overline{\mu}({}^*U) - {}^*\overline{\mu}({}^*K) < \epsilon$. By Theorem 5.9, we have $\mathsf{st}^{-1}(E) \in \overline{{}^*\mathcal{B}[X]}$. Since ϵ is arbitrary, we have $\mu(E) = \overline{{}^*\mu}(\mathsf{st}^{-1}(E))$. □

The following two lemmas are crucial in the proof of the main theorem of this section.

Lemma 6.9. *Consider any (ϵ, r)-hyperfinite representation S of ${}^*X$. Let F denote $\bigcup\{B(s) : s \in \mathsf{st}^{-1}(E) \cap S\}$. Then for any $E \in \mathcal{B}[X]$, we have $\mathsf{st}^{-1}(E) = F$*

Proof. First we show that $F \subset \mathsf{st}^{-1}(E)$. Let $x \in F$ then x must lie in $B(s_0)$ for some $s_0 \in \mathsf{st}^{-1}(E) \cap S$. Since $s_0 \in \mathsf{st}^{-1}(E)$, there exists a $y \in E$ such that $s_0 \in \mu(y)$. As $B(s_0)$ has infinitesimal radius, $B(s_0) \subset \mu(y)$. Hence $x \in B(s_0) \subset \mu(y) \subset \mathsf{st}^{-1}(E)$. Hence, $F \subset \mathsf{st}^{-1}(E)$.

Now we show the reverse direction. Let $x \in \mathsf{st}^{-1}(E)$. Since $\bigcup_{s\in S} B(s) \supset \mathrm{NS}({}^*X)$, $x \in B(s_0)$ for some $s_0 \in S$. As $x \in \mathsf{st}^{-1}(E)$, there exists a $y \in E$ such that $x \in \mu(y)$. This shows that $s_0 \in \mathsf{st}^{-1}(E) \cap S$ which implies that $x \in F$, completing the proof. □

Before proving the next lemma, recall that $\mathcal{L}(\mathcal{A})$ denote the collection of universally Loeb measurable sets of the internal algebra $\mathcal{A}$.

Lemma 6.10. *Let X be a metric space satisfying the Heine-Borel condition equipped with Borel σ-algebra $\mathcal{B}[X]$. Let S be a (ϵ, r)-hyperfinite representation of ${}^*X$ for some positive infinitesimal ϵ. Then for any $E \in \mathcal{B}[X]$ we have*

$$\mathsf{st}^{-1}(E) \in \mathcal{L}({}^*\mathcal{B}[X]) \textit{ and } \mathsf{st}^{-1}(E) \cap S \in \mathcal{L}(\mathcal{I}(S)). \tag{6.5}$$

Proof. By Theorem 5.9, $\mathsf{st}^{-1}(E) \in \{A \cap \mathrm{NS}({}^*X) : A \in \mathcal{L}({}^*\mathcal{B}[X])\}$. As X is σ-compact, by Lemma 5.5, we have $\mathrm{NS}({}^*X) \in \mathcal{L}({}^*\mathcal{B}[X])$ hence $\mathsf{st}^{-1}(E) \in \mathcal{L}({}^*\mathcal{B}[X])$. Let P be any internal probability measure on $(S, \mathcal{I}(S))$ Let P' be an internal probability measure on $({}^*X, {}^*\mathcal{B}[X])$ with $P'(B) = P(B \cap S)$. As S is internal and $\mathsf{st}^{-1}(E)$ is universally Loeb measurable, we know that $\mathsf{st}^{-1}(E) \cap S \in \overline{{}^*\mathcal{B}[X]}^{P'}$ where $\overline{{}^*\mathcal{B}[X]}^{P'}$ denotes the Loeb σ-algebra of ${}^*\mathcal{B}[X]$ under P'. Fix any $\epsilon > 0$. We can then find $A_i, A_o \in {}^*\mathcal{B}[X]$ such that $A_i \subset \mathsf{st}^{-1}(E) \cap S \subset A_o$ and $P'(A_o \setminus A_i) < \epsilon$. We thus have

$$P'(A_o \setminus A_i) = P((A_o \setminus A_i) \cap S) = P((A_o \cap S) \setminus (A_i \cap S)) < \epsilon. \tag{6.6}$$

As both $A_i, A_o \in {}^*\mathcal{B}[X]$, we know that $A_i \cap S, A_o \cap S \in \mathcal{I}(S)$. Moreover, we have $A_i \cap S \subset \mathsf{st}^{-1}(E) \cap S \subset A_o \cap S$. Hence, by the construction of Loeb measure, $\mathsf{st}^{-1}(E) \cap S$ is Loeb measurable with respect to P. As P is arbitrary, we know that $\mathsf{st}^{-1}(E) \cap S \in \mathcal{L}(\mathcal{I}(S))$. □

We are now at the place to prove the main theorem of this section.

Theorem 6.11. *Let $(X,\mathcal{B}[X],P)$ be a Borel probability space where X is a metric space satisfying the Heine-Borel condition, and let $({}^*X,{}^*\mathcal{B}[X],{}^*P)$ be its nonstandard extension. Then for every positive infinitesimal ϵ,every positive infinite r and every (ϵ,r)-hyperfinite representation S of ${}^*X$ there exists an internal probability measure P' on $(S,\mathcal{I}(S))$*

(1) $P'(\{s\}) \approx {}^*P(B(s))$.
(2) $P(E) = \overline{P'}(\mathsf{st}^{-1}(E)\cap S)$ *for every* $E\in\mathcal{B}[X]$.

where $\overline{P'}$ denotes the Loeb measure of P'.

Proof. Fix an infinitesimal $\epsilon \in {}^*\mathbb{R}^+$ and an positive infinite number r. Let S be a (ϵ,r)-hyperfinite representation of ${}^*X$ and consider the hyperfinite measurable space $(S,\mathcal{I}(S))$. Let $P'(\{s\}) = \frac{{}^*P(B(s))}{{}^*P(\bigcup_{s\in S}B(s))}$ for every $s\in S$. It follows that P' is internal because the map $s\mapsto P'(\{s\})$ is internal. For any $A\in\mathcal{I}(S)$, let $P'(A)=\sum_{s\in A}P'(\{s\})$. Since $\sum_{s\in A}{}^*P(B(s)) = {}^*P(\bigcup_{s\in S}B(s))$ by the hyperfinite additivity of ${}^*P$, it is easy to see that P' is an internal probability measure on $(S,\mathcal{I}(S))$.

As $\bigcup_{s\in S}B(s)\supset \mathrm{NS}({}^*X)$, by Lemma 6.8, we know that ${}^*P(\bigcup_{s\in S}B(s))\approx 1$. Hence we have $P'(\{s\})\approx {}^*P(B(s))$.

It remains to show that $P(E)=\overline{P'}(\mathsf{st}^{-1}(E)\cap S)$ for every $E\in\mathcal{B}[X]$. As X is a σ-compact Borel probability space, by Lemma 6.8 and Lemma 6.10, we have $P(E)=\overline{{}^*P}(\mathsf{st}^{-1}(E))$. By Lemma 6.9, we then have

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) = \overline{{}^*P}(\bigcup\{B(s): s\in \mathsf{st}^{-1}(E)\cap S\}). \tag{6.7}$$

Consider any set $A_o\supset \mathsf{st}^{-1}(E)\cap S, A_o\in\mathcal{I}(S)$, then A_o is an internal subset of S hence is hyperfinite. This means that $\bigcup_{s\in A_o}B(s)$ is a hyperfinite union of *Borel sets hence is *Borel. Because $\mathsf{st}^{-1}(E)\cap S\subset A_o$, we have

$$\overline{{}^*P}(\bigcup\{B(s): s\in\mathsf{st}^{-1}(E)\cap S\}) \leq \overline{{}^*P}(\bigcup_{s\in A_o}B(s)) = \mathsf{st}({}^*P(\bigcup_{s\in A_o}B(s))). \tag{6.8}$$

As $\bigcup_{s\in S} B(s) \supset \mathrm{NS}({}^*X)$, by Lemma 6.8, we have ${}^*P(\bigcup_{s\in S} B(s)) \approx 1$. Thus we have

$$\mathsf{st}({}^*P(\bigcup_{s\in A_o} B(s))) = \mathsf{st}(\frac{{}^*P(\bigcup_{s\in A_o} B(s))}{{}^*P(\bigcup_{s\in S} B(s))}) = \mathsf{st}(P'(A_o)) = \overline{P'}(A_o). \tag{6.9}$$

Hence, for every set $A_o \in \mathcal{I}(S)$ such that $A_o \supset \mathsf{st}^{-1}(E) \cap S$, we have

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) = \overline{{}^*P}(\bigcup\{B(s) : s \in \mathsf{st}^{-1}(E) \cap S\}) \leq \overline{P'}(A_o). \tag{6.10}$$

This means that

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) \leq \inf\{\overline{P'}(A_o) : A_o \supset \mathsf{st}^{-1}(E) \cap S, A_o \in \mathcal{I}(S)\}. \tag{6.11}$$

By a similar argument, we have

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) \geq \sup\{\overline{P'}(A_i) : A_i \subset \mathsf{st}^{-1}(E) \cap S, A_i \in \mathcal{I}(S)\}. \tag{6.12}$$

By Lemma 6.10, we have $\mathsf{st}^{-1}(E) \cap S \in \mathcal{I}(S)_L$. Thus by the construction of Loeb measure, we have

$$\inf\{\overline{P'}(A_o) : A_o \supset \mathsf{st}^{-1}(E) \cap S, A_o \in \mathcal{I}(S)\} \tag{6.13}$$

$$= \sup\{\overline{P'}(A_i) : A_i \subset \mathsf{st}^{-1}(E) \cap S, A_i \in \mathcal{I}(S)\} \tag{6.14}$$

$$= \overline{P'}(\mathsf{st}^{-1}(E) \cap S). \tag{6.15}$$

Hence $P(E) = \overline{{}^*P}(\mathsf{st}^{-1}(E)) = \overline{P'}(\mathsf{st}^{-1}(E) \cap S)$ finishing the proof. □

From the above proof, we see that

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) = \inf\{\overline{P'}(A_o) : A_o \supset \mathsf{st}^{-1}(E) \cap S, A_o \in \mathcal{I}(S)\} \tag{6.16}$$

$$= \inf\{\mathsf{st}(\frac{{}^*P(\bigcup_{s\in A_o} B(s))}{{}^*P(\bigcup_{s\in S} B(s))}) : A_o \supset \mathsf{st}^{-1}(E) \cap S, A_o \in \mathcal{I}(S)\} \tag{6.17}$$

$$= \inf\{\overline{{}^*P}(\bigcup_{s\in A_o} B(s)) : A_o \supset \mathsf{st}^{-1}(E) \cap S, A_o \in \mathcal{I}(S)\}. \tag{6.18}$$

Similarly we have:

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) = \sup\{\overline{{}^*P}(\bigcup_{s\in A_i} B(s)) : A_i \subset \mathsf{st}^{-1}(E)\cap S, A_i \in \mathcal{I}(S)\} \tag{6.19}$$

Note that if X is compact, then ${}^*P(\bigcup_{s\in S} B(s)) = {}^*P({}^*X) = 1$. Hence $P'(\{s\}) = {}^*P(B(s))$ in Theorem 6.11. We no longer need to normalize the probability space when X is compact.

We conclude this section by giving an explicit application of Theorem 6.11 to Example 4.5.

Example 6.12. Let μ be the Lebesgue measure on the unit interval $[0,1]$ restricted to the Borel σ-algebra on $[0,1]$. Let N be an infinite element in ${}^*\mathbb{N}$ and let $\Omega = \{\frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N}, 1\}$. Let μ' be an internal probability measure on $(\Omega, \mathcal{I}(\Omega))$ such that $\mu'(\{\omega\}) = \frac{1}{N}$ for every $\omega \in \Omega$.

Theorem 6.13. *For every Borel measurable set A, we have*

$$\mu(A) = \overline{\mu'}(\mathsf{st}^{-1}(A)\cap\Omega). \tag{6.20}$$

Proof. Clearly Ω is a $(\frac{1}{N})$-hyperfinite representation of ${}^*[0,1]$. For every $\omega \in \Omega$, we have $B(\omega) = (\omega - \frac{1}{N}, \omega]$ for $\omega \neq \frac{1}{N}$ and $B(\frac{1}{N}) = [0, \frac{1}{N}]$. It is easy to see that $\{B(\omega) : \omega \in \Omega\}$ covers ${}^*[0,1]$ and $\mu'(\{\omega\}) = {}^*\mu(B(\omega))$ for every $\omega \in \Omega$. Thus, by Theorem 6.11, we have $\mu(A) = \overline{\mu'}(\mathsf{st}^{-1}(A)\cap\Omega)$, completing the proof. □

7. General Hyperfinite Markov Processes

In this section, we introduce the concept of general hyperfinite Markov processes. Intuitively, hyperfinite Markov processes behaves like finite Markov processes but can be used to represent standard continuous time Markov processes under certain conditions.

Definition 7.1. A general hyperfinite Markov chain is characterized by the following four ingredients:

(1) A hyperfinite state space $S \subset {}^*X$ where X is a metric space satisfying the Heine-Borel condition.
(2) A hyperfinite time line $T = \{0, \delta t, ..., K\}$ where $\delta t = \frac{1}{N!}$ for some $N \in {}^*\mathbb{N} \setminus \mathbb{N}$ and $K \in {}^*\mathbb{N} \setminus \mathbb{N}$.
(3) A set $\{v_i : i \in S\}$ where each $v_i \geq 0$ and $\sum_{i \in S} v_i = 1$.
(4) A set $\{p_{ij}\}_{i,j \in S}$ consisting of non-negative hyperreals with $\sum_{j \in S} p_{ij} = 1$ for each $i \in S$

Thus the state space S naturally inherits the *metric of $^*X$. An element $s \in S$ is near-standard if it is near-standard in $^*X$. The near-standard part of S, $\mathrm{NS}(S)$, is defined to be $\mathrm{NS}(S) = \mathrm{NS}({}^*X) \cap S$.

Note that the time line T contains all the standard rational numbers but contains no standard irrational number. However, for every standard irrational number r there exists $t_r \in T$ such that $t_r \approx r$.

Intuitively, the $\{p_{ij}\}_{i,j \in S}$ refers to the internal probability of going from i to j at time δt.

The following theorem shows the existence of hyperfinite Markov Processes.

Theorem 7.2. *Given a non-empty hyperfinite state space S, a hyperfinite time line $T = \{0, \delta t, ..., K\}$, $\{v_i\}_{i \in S}$ and $\{p_{ij}\}_{i,j \in S}$ as in Definition 7.1. Then there exists internal probability triple $(\Omega, \mathcal{A}, P)$ with an internal stochastic process $\{X_t\}_{t \in T}$ defined on $(\Omega, \mathcal{A}, P)$ such that*

$$P(X_0 = i_0, X_{\delta t} = i_{\delta t}, ...X_t = i_t) = v_{i_0} p_{i_0 i_{\delta t}} ... p_{i_{t-\delta t} i_t} \tag{7.1}$$

for all $t \in T$ and $i_0, ...i_t \in S$.

Note that $v_{i_0} p_{i_0 i_{\delta t}} ... p_{i_{t-\delta t} i_t}$ is a product of hyperfinitely many hyperreal numbers. It is well-defined by the transfer principle.

Proof. Let Ω=$\{\omega \in S^T : \omega \text{ is internal}\}$ which is the set of internal functions from T to S. As both S and T are hyperfinite, Ω is hyperfinite. Let $\mathcal{A}$ be the set consisting of all internal subsets of Ω. We now define the internal measure P on $(\Omega,\mathcal{A})$. For every $\omega\in\Omega$, let

$$P(\omega)=v_{i_{\omega(0)}}p_{i_{\omega(0)}i_{\omega(1)}}\cdots p_{i_{\omega(K-\delta t)}i_{\omega(K)}}. \tag{7.2}$$

For every $A\in\mathcal{A}$, let $P(A)=\sum_{\omega\in A}P(\omega)$. Let $X_t(\omega)=\omega(t)$. It is easy to check that $(\Omega,\mathcal{A},P)$ and $\{X_t\}_{t\in T}$ satisfy the conditions of this theorem. $\square$

We use $(\Omega,\overline{\mathcal{A}},\overline{P})$ to denote the Loeb extension of the internal probability triple $(\Omega,\mathcal{A},P)$ in Theorem 7.2. The construction of hyperfinite Markov processes is similar to the construction of finite state space discrete time Markov processes. Unlike the construction of general Markov processes, we do not need to use the Kolmogorov extension theorem.

We introduce the following definition.

Definition 7.3. For any $i,j\in S$ and any $t\in T$, we define:

$$p_{ij}^{(t)}=\sum_{\omega\in\mathcal{M}}P(\{\omega\}|X_0=i) \tag{7.3}$$

where $\mathcal{M}=\{\omega\in\Omega:\omega(0)=i\wedge\omega(t)=j\}$.

It is easy to see that $p_{ij}^{(\delta t)}=p_{ij}$. For general $t\in T$, $p_{ij}^{(t)}$ is the sum of $p_{ii_{\delta t}}p_{i_{\delta t}i_{2\delta t}}\cdots p_{i_{t-\delta t}j}$ over all possible $i_{\delta t},i_{2\delta t},\ldots,i_{t-\delta t}$ in S. Intuitively, $p_{ij}^{(t)}$ is the internal probability of the chain reaches state j at time t provided that the chain started at i. For any set $A\in\mathcal{I}(S)$, any $i\in S$ and any $t\in T$, the internal transition probability from x to A at time t is denoted by $p_i^{(t)}(A)$ or $p^{(t)}(i,A)$. In both cases, they are defined to be $\sum_{j\in A}p_{ij}^{(t)}$.

We are now at the place to show that the hyperfinite Markov chain is time-homogeneous.

Lemma 7.4. *For any $t,k\in T$ and any $i,j\in S$, we have $P(X_{k+t}=j|X_k=i)=p_{ij}^{(t)}$ provided that $P(X_k=i)>0$.*

Proof. It is sufficient to show that $P(X_{k+\delta t}=j|X_k=i)=p_{ij}$ since the general case follows from a similar calculation.

$$P(X_{k+\delta t}=j|X_k=i)=\frac{P(X_{k+\delta t}=j,X_k=i)}{P(X_k=i)} \tag{7.4}$$

$$=\frac{\sum_{i_0,i_{\delta t},\dots,i_{k-\delta t}} v_{i_0}p_{i_0 i_{\delta t}}\cdots p_{i_{k-\delta t}i}p_{ij}}{\sum_{i_0,i_{\delta t},\dots,i_{k-\delta t}} v_{i_0}p_{i_0 i_{\delta t}}\cdots p_{i_{k-\delta t}i}} \tag{7.5}$$

$$=p_{ij}. \tag{7.6}$$

Hence we have the desired result. □

We write $P_i(X_t\in A)$ for $P(X_t\in A|X_0=i)$. It is easy to see that $p_i^{(t)}(A)=P_i(X_t\in A)$. Note that for every $i\in S$ and every $t\in T$, $p_i^{(t)}(\cdot)$ is an internal probability measure on $(S,\mathcal{I}(S))$. We use $\overline{p}_i^{(t)}$ to denote the Loeb extension of this internal probability measure. For every $A\in\overline{\mathcal{I}(S)}$, it is easy to see that $\overline{p}_i^{(t)}(A)=\overline{P}_i(X_t\in A)$.

We are now at the place to define some basic concepts for Hyperfinite Markov processes.

Definition 7.5. Let π be an internal probability measure on $(S,\mathcal{I}(S))$. We call π a weakly stationary if there exists an infinite $t_0\in T$ such that for any $t\leq t_0$ and any $A\in\mathcal{I}(S)$ we have $\pi(A)\approx\sum_{i\in S}\pi(\{i\})p^{(t)}(i,A)$.

The definition of weakly stationary distribution is similar to the definition of stationary distribution for discrete time finite Markov processes. However, we only require $\pi(A)\approx\sum_{i\in S}\pi(\{i\})p^{(t)}(i,A)$ for t no greater than some infinite t_0 for weakly stationary distributions. We use $\overline{\pi}$ to denote the Loeb extension of π.

Definition 7.6. A hyperfinite Markov chain is said to be strong regular if for any $A\in\mathcal{I}(S)$, any $i,j\in\mathrm{NS}(S)$ and any non-infinitesimal $t\in T$ we have

$$(i\approx j)\implies(P_i(X_t\in A)\approx P_j(X_t\in A)). \tag{7.7}$$

One might wonder whether $P_i(X_t\in A)\approx P_j(X_t\in A)$ for infinitesimal $t\in T$. This is generally not true.

Example 7.7. Let the time line $T = \{0, \delta t, 2\delta t, \ldots, K\}$ for some infinitesimal δt and some infinite K. Let the state space $S = \{\frac{-K}{\sqrt{\delta t}}, \ldots, -\sqrt{\delta t}, 0, \sqrt{\delta t}, \ldots, \frac{K}{\sqrt{\delta t}}\}$. For any $i \in S$, we have $p^{(\delta t)}(i, i+\sqrt{\delta t}) = \frac{1}{2}$ and $p_i i - \sqrt{\delta t} = \frac{1}{2}$. This is Anderson's construction of Brownian motion which motivates the study of infinitesimal stochastic processes (see [And76]). It can also be viewed as a hyperfinite Markov process. As the normal distributions with different means converge in total variational distance, the hyperfinite Brownian motion is strong Feller. However, we have $p^{(\delta t)}(0, \sqrt{\delta t}) = \frac{1}{2}$ and $p^{(\delta t)}(\sqrt{\delta t}, \sqrt{\delta t}) = 0$.

For a general state space Markov processes, the transition probability to a specific point is usually 0. For hyperfinite Markov process, under some conditions, we can get infinitesimally close to a specific point with probability 1.

Lemma 7.8. *Consider a hyperfinite Markov chain on a state space S and two states $i, j \in S$, let $\{U_j^{\frac{1}{n}} : n \in \mathbb{N}\}$ be the collection of balls with radius $\frac{1}{n}$ around j. Suppose $\forall n \in \mathbb{N}$, we have $\overline{P}_i(\{\omega : (\exists t \in \mathrm{NS}(T))(X_t(\omega) \in U_j^{\frac{1}{n}})\}) = 1$. Then for any infinite $s_0 \in T$, we have $\overline{P}_i(\{\omega : \exists t < s_0 X_t(\omega) \approx j\}) = 1$.*

Proof. Pick any infinite $s_0 \in T$ and from the hypothesis we know that $\forall n \in \mathbb{N}$, $P_i(\{\omega : (\exists t \leq s_0)(X_t \in U_j^{\frac{1}{n}})\}) > 1 - \frac{1}{n}$.

Consider the set $B = \{n \in {}^*\mathbb{N} : P_i(\{\omega : (\exists t \leq s_0)(X_t \in U_j^{\frac{1}{n}})\}) > 1 - \frac{1}{n}\}$, by the internal definition principle, B is an internal set and contains $\mathbb{N}$. By overspill, B contains an infinite number in ${}^*\mathbb{N}$ and we denote it by n_0. Thus we have $P_i(\{\omega : (\exists t \leq s_0)(X_t \in U_j^{\frac{1}{n_0}})\}) > 1 - \frac{1}{n_0}$. Hence $\overline{P}_i(\{\omega : (\exists t \leq s_0)(X_t \in U_j^{\frac{1}{n_0}})\}) = 1$. The set $\{\omega : (\exists t \leq s_0)(X_t \approx j)\}$ is a superset of $\{\omega : (\exists t \leq s_0)(X_t \in U_j^{\frac{1}{n_0}})\}$. Since the Loeb measure is complete we know that $\overline{P}_i(\{\omega : (\exists t \leq s_0)(X_t \approx j)\}) = 1$. □

.

In the study of standard Markov processes, it is sometimes useful to consider the product of two i.i.d Markov processes. The similar idea can be applied to hyperfinite Markov processes.

Definition 7.9. Let $\{X_t\}_{t\in T}$ be a hyperfinite Markov chain with internal transition probability $\{p_{ij}\}_{i,j\in S}$. Let $\{Y_t\}_{t\in T}$ be a i.i.d copy of $\{X_t\}$. Then product chain Z_t is defined on the state space $S\times S$ with transition probability

$$\{q_{(i,j),(k,l)} = p_{ik}p_{jl}\}_{i,j,k,l\in S}. \tag{7.8}$$

Similarly $q_{(i,j),(k,l)}$ refers to the internal probability of going from point (i,j) to point (k,l). The following lemma is an immediate consequence of this definition.

Lemma 7.10. *Let $\{X_t\}_{t\in T}$, $\{Y_t\}_{t\in T}$ and $\{Z_t\}_{t\in T}$ be the same as in Definition 7.9. Then for any $t\in T$,any $i,j\in S$ and any $A,B\in\mathcal{I}(S)$ we have $q^{(t)}_{(i,j)}(A\times B) = p^{(t)}_i(A)p^{(t)}_j(B)$.*

Proof. We prove this lemma by internal induction on T.

Fix any $i,j\in S$ and any $A,B\in\mathcal{I}(S)$. We have

$$p^{(\delta t)}_i(A)p^{(\delta t)}_j(B) \tag{7.9}$$

$$= \sum_{(a,b)\in A\times B} p^{(\delta t)}_i(\{a\})\times p^{(\delta t)}_j(\{b\}) \tag{7.10}$$

$$= \sum_{(a,b)\in A\times B} q^{(\delta t)}_{(i,j)}(\{(a,b)\}) \tag{7.11}$$

$$= q^{(\delta t)}_{(i,j)}(A\times B). \tag{7.12}$$

Hence we have shown the base case.

Suppose we know that the lemma is true for $t=k$. We now prove the lemma for $k+\delta t$. Fix any $i,j\in S$ and any $A,B\in\mathcal{I}(S)$. We have

$$p^{(k+\delta t)}_i(A)\times p^{(k+\delta t)}_j(B) \tag{7.13}$$

$$= \sum_{s\in S} p^{(\delta t)}_i(\{s\})p^{(k)}_s(A)\times \sum_{s'\in S} p^{(\delta t)}_j(\{s'\})p^{(k)}_{s'}(B) \tag{7.14}$$

$$= \sum_{(s,s')\in S\times S} p^{(\delta t)}_i(\{s\})p^{(\delta t)}_j(\{s'\})p^{(k)}_s(A)p^{(k)}_{s'}(B) \tag{7.15}$$

By induction hypothesis, this equals to:

$$\sum_{(s,s')\in S\times S} q^{(\delta t)}_{(i,j)}(\{(s,s')\})q^{(k)}_{(s,s')}(A\times B) = q^{(k+\delta t)}_{(i,j)}(A\times B). \tag{7.16}$$

As all the parameters are internal, by internal induction principle we have shown the result. □

Definition 7.11. Consider a hyperfinite Markov chain $\{X_t\}_{t\in T}$ and two near-standard $i,j\in S$. A near-standard $(x,y)\in S\times S$ is called a near-standard absorbing point with respect to i,j if $\overline{P'}_{(i,j)}((\exists t\in \mathrm{NS}(T))(Z_t\in U_x^{\frac{1}{n}}\times U_y^{\frac{1}{n}}))=1$ for all $n\in\mathbb{N}$ where P' denotes the internal probability measure of the product chain $\{Z_t\}_{t\in T}$ and $U_x^{\frac{1}{n}}, U_y^{\frac{1}{n}}$ denote the open ball centered at x,y with radius $\frac{1}{n}$,respectively.

It is a natural to ask when a hyperfinite Markov chain has a near-standard absorbing point. We start by introducing the following definitions.

Definition 7.12. For any $A\in\mathcal{I}(S)$, the stopping time $\tau(A)$ with respect to a hyperfinite Markov chain $\{X_t\}_{t\in T}$ is defined to be $\tau(A)=\min\{t\in T: X_t\in A\}$.

Definition 7.13. A hyperfinite Markov chain $\{X_t\}_{t\in T}$ is productively near-standard open set irreducible if for any $i,j\in\mathrm{NS}(S)$ and any near-standard open ball B with non-infinitesimal radius we have $\overline{P'}_{(i,j)}(\tau(B\times B)<\infty)>0$ where P' denotes the internal probability measure of the product chain $\{Z_t\}_{t\in T}$ as in Definition 7.9.

Recall that the state space of $\{X_t\}_{t\in T}$ is a hyperfinite set $S\subset {}^*X$ where X is a metric space satisfying the Heine-Borel condition. Let d denote the metric on X. A near-standard open ball of S is an internal set taking the form $\{s\in S: {}^*d(s,s_0)<r\}$ for some near-standard point $s_0\in S$ and some near-standard $r\in{}^*\mathbb{R}$.

Theorem 7.14. *Let $\{X_t\}_{t\in T}$ be a hyperfinite Markov chain with weakly stationary distribution π such that $\overline{\pi}(\mathrm{NS}(S))=1$. Suppose $\pi\times\pi$ is a weakly stationary distribution for the product Markov process $\{Z_t\}_{t\in T}$. If $\{X_t\}_{t\in T}$ is productively near-standard open set irreducible then for $\overline{\pi\times\pi}$ almost all $(i,j)\in S\times S$ there exists an near-standard absorbing point (i_0,i_0) for (i,j) as in Definition 7.11.*

Before we prove this theorem, we first establish the following technical lemma. Although this lemma takes place in the non-standard universe, the proof of this lemma is similar to the proof of a similar standard result in [RR04].

Lemma 7.15 ([RR04, Lemma. 20])**.** *Consider a general hyperfinite Markov chain on a state space S, having a weakly stationary distribution $\pi(\cdot)$ such that $\overline{\pi}(\mathrm{NS}(S)) = 1$. Suppose that for some internal $A \subset S$, we have $\overline{P}_x(\tau(A) < \infty) > 0$ for $\overline{\pi}$ almost all $x \in S$. Then for $\overline{\pi}$-almost-all $x \in S$, $\overline{P}_x(\tau(A) < \infty) = 1$.*

Proof. Suppose to the contrary that the conclusion does not hold. That means $\overline{\pi}(x \in S : \overline{P}_x(\tau(A) < \infty) < 1) > 0$.

Claim 7.16. *There exist $l \in \mathbb{N}$, $\delta \in \mathbb{R}^+$ and internal set $B \subset S$ with $\overline{\pi}(B) > 0$ such that $\overline{P}_x(\tau(A) = \infty, \max\{k \in T : X_k \in B\} < l) \geq \delta$ for all $x \in B$.*

Proof. As $\overline{\pi}(x \in S : \overline{P}_x(\tau(A) = \infty) > 0) > 0$, This implies that there exist $\delta_1 \in \mathbb{R}^+$ and $B_1 \in \mathcal{F}$ with $\overline{\pi}(B_1) > 0$ such that $\overline{P}_x(\tau(A) < \infty) \leq 1 - \delta_1$ for all $x \in B_1$ where $\mathcal{F}$ denote the Loeb extension of the internal algebra $\mathcal{I}(S)$ with respect to π. By the construction of Loeb measure, we can assume that B_1 is internal. On the other hand, as $\overline{P}_x(\tau(A) < \infty) > 0$ for $\overline{\pi}$ almost surely $x \in S$, by countable additivity, we can find $l_0 \in \mathbb{N}$ and $\delta_2 \in \mathbb{R}^+$ and internal $B_2 \subset B_1$(again by the construction of Loeb measure) with $\overline{\pi}(B_2) > 0$ such that $\forall x \in B_2$, $\overline{P}_x((\exists t \leq l_0 \wedge t \in T)(X_t \in A)) \geq \delta_2$. Let $\eta = |\{k \in \mathbb{N} \cup \{0\} : (\exists t \in T \cap [k, k+1))(X_k \in B_2)\}|$. Then for any $r \in \mathbb{N}$ and $x \in S$, we have $\overline{P}_x(\tau(A) = \infty, \eta > r(l_0 + 1)) \leq \overline{P}_x(\tau(A) = \infty | \eta > r(l_0 + 1)) \leq (1 - \delta_2)^r$. In particular, $\overline{P}_x(\tau(A) = \infty, \eta = \infty) = 0$.

Hence for $x \in B_2$, we have

$$\overline{P}_x(\tau(A) = \infty, \eta < \infty) \tag{7.17}$$

$$= 1 - \overline{P}_x(\tau(A) = \infty, \eta = \infty) - \overline{P}_x(\tau(A) < \infty) \tag{7.18}$$

$$\geq 1 - 0 - (1 - \delta_1) = \delta_1. \tag{7.19}$$

By countable additivity again there exist $l \in \mathbb{N}$, $\delta \in \mathbb{R}^+$ and $B \subset B_2$ (again pick B to be internal) with $\overline{\pi}(B) > 0$ such that $\overline{P}_x(\tau(A) = \infty, \max\{t \in T : X_t \in B_2\} < l) \geq \delta$ for all $x \in B$. Finally as $B \subset B_2$, we have

$$\max\{t \in T : X_t \in B_2\} \geq \max\{t \in T : X_t \in B\} \tag{7.20}$$

establishing the claim. □

Claim 7.17. *Let B, l, δ be as in Claim 7.16. Let K' be the biggest hyperinteger such that $K'l \leq K$ where K is the last element in T. Let*

$$s = \max\{k \in {}^*\mathbb{N} : (1 \leq k \leq K') \wedge (X_{kl} \in B)\} \tag{7.21}$$

and $s = 0$ if the set is empty. Then for all $1 \leq r \leq j \in \mathbb{N}$ we have

$$\sum_{x \in S} \pi(\{x\}) P_x(s = r, X_{jl} \notin A) \gtrapprox \mathsf{st}(\pi(B)\delta). \tag{7.22}$$

Proof. Pick any $j \in \mathbb{N}$. we have

$$\sum_{x \in S} \pi(\{x\}) P_x(s = r, X_{jl} \notin A) = \sum_{x \in S} \pi(\{x\}) \sum_{y \in B} P_x(X_{rl} = y) P_y(s = 0, X_{(j-r)l} \notin A) \tag{7.23}$$

Note that $\tau(A) = \infty$ implies $X_{(j-r)l} \notin A$ and $\max\{k \in T : X_k \in B\} < l$ implies that $s = 0$. As $r, l \in \mathbb{N}$ and π is a weakly stationary distribution, we have

$$\sum_{x \in S} \pi(\{x\}) \sum_{y \in B} P_x(X_{rl} = y) P_y(s = 0, X_{(j-r)l} \notin A) \tag{7.24}$$

$$\geq \sum_{x \in S} \pi(\{x\}) \sum_{y \in B} P_x(X_{rl} = y)\delta \tag{7.25}$$

$$\approx \pi(B)\delta. \tag{7.26}$$

By the definition of standard part, it is easy to see that this claim holds. □

Now we are at the position to prove the theorem. For all $j \in \mathbb{N}$, by Claim 7.16, we have

$$\pi(A^c) \approx \sum_{x \in S} \pi(\{x\}) P_x(X_{jl} \in A^c) \tag{7.27}$$

$$= \sum_{x \in S} \pi(\{x\}) P_x(X_{jl} \notin A) \tag{7.28}$$

$$\geq \sum_{r=1}^{j} \sum_{x \in S} \pi(\{x\}) P_x(s = r, X_{jl} \notin A) \tag{7.29}$$

$$\geq \sum_{r=1}^{j} \mathsf{st}(\pi(B)\delta). \tag{7.30}$$

As $\overline{\pi}(B) > 0$, so we can pick $j \in \mathbb{N}$ such that $j > \frac{1}{\mathsf{st}(\pi(B)\delta)}$. This gives that $\pi(A^c) > 1$ which is a contradiction, proving the result. □

.

We are now at the place to prove Theorem 7.14.

proof of Theorem 7.14. Pick any near-standard $i_0 \in S$. Recall that $U_{i_0}^{\frac{1}{n}}$ denote the open ball around i_0 with radius $\frac{1}{n}$. It is clear that $U_{i_0}^{\frac{1}{n}} \times U_{i_0}^{\frac{1}{n}} \in \mathcal{I}(S) \times \mathcal{I}(S)$. By Definition 7.13, we have $\overline{P'}_{(i,j)}(\tau(U_{i_0}^{\frac{1}{n}} \times U_{i_0}^{\frac{1}{n}}) < \infty) > 0$ for all $n \in \mathbb{N}$ and $\overline{\pi \times \pi}$ almost all $(i,j) \in S \times S$. As $\pi \times \pi$ is a weakly stationary distribution, by Lemma 7.15, we have $\overline{P'}_{(i,j)}(\tau(U_{i_0}^{\frac{1}{n}} \times U_{i_0}^{\frac{1}{n}}) < \infty) = 1$ for π' almost surely $(i,j) \in S \times S$ and every $n \in \mathbb{N}$. By Definition 7.11, we know that (i_0, i_0) is a near-standard absorbing point for π' almost all $(i,j) \in S \times S$. □

Note that this proof shows that every near-standard point (i,j) is a near-standard absorbing point for $\overline{\pi \times \pi}$ almost all $(x,y) \in S \times S$.

In the statement of Theorem 7.14, we require $\pi \times \pi$ to be a weakly stationary distribution of the product hyperfinite Markov chain $\{Z_t\}_{t \in T}$. Recall that t_0 is an infinite element in T such that $\pi(A) \approx \sum_{i \in S} \pi(\{i\}) p_i^{(t)}(A)$ for all $A \in \mathcal{I}(S)$ and all $t \leq t_0$.

Lemma 7.18. *Let $\pi' = \pi \times \pi$. For any $A, B \in \mathcal{I}(S)$ and any $t \leq t_0$, we have $\pi'(A \times B) \approx \sum_{(i,j)\in S\times S} \pi'_{(i,j)} q^{(t)}_{(i,j)}(A \times B)$ where $q^{(t)}_{(i,j)}(A \times B)$ denotes the t-step transition probability from (i, j) to the set $A \times B$.*

Proof. Pick $A, B \in \mathcal{I}(S)$ and $t \leq t_0$. Then, by Definition 7.5 and Lemma 7.10, we have

$$\sum_{(i,j)\in S\times S} \pi'(\{(i,j)\}) q^{(t)}_{(i,j)}(A \times B) \tag{7.31}$$

$$= \sum_{(i,j)\in S\times S} \pi(\{i\})\pi(\{j\}) p^{(t)}_i(A) p^{(t)}_j(B) \tag{7.32}$$

$$= (\sum_{i\in S} \pi(\{i\}) p^{(t)}_i(A))(\sum_{j\in S} \pi(\{j\}) p^{(t)}_j(B)) \tag{7.33}$$

$$\approx \pi(A)\pi(B) \tag{7.34}$$

$$= \pi(A \times B). \tag{7.35}$$

□

However, we do not know whether π' would always be a weakly stationary distribution since $\overline{\mathcal{I}(S) \times \mathcal{I}(S)}$ is a bigger σ-algebra than $\overline{\mathcal{I}(S)} \times \overline{\mathcal{I}(S)}$. This gives rise to the following open questions.

Open Problem 1. *Does there exists a π' that fails to be a weakly stationary distribution of the product hyperfinite Markov process $\{Z_t\}_{t\in T}$?*

It is natural to ask whether the product of two weakly stationary distributions is a weakly stationary distribution for the product chain. More generally, suppose P_1, P_2 are two internal probability measures on $(\Omega, \mathcal{A})$ with $P_1(A) \approx P_2(A)$ for all $A \in \mathcal{A}$, is it true that $(P_1 \times P_1)(B) \approx (P_2 \times P_2)(B)$ for all $B \in \mathcal{A} \otimes \mathcal{A}$? The answer to this question is affirmative by the result in [KS04] that the Loeb product space is uniquely determined by its factor Loeb spaces.

We are now at the place to prove the hyperfinite Markov chain Ergodic theorem.

Theorem 7.19. *Consider a strongly regular hyperfinite Markov chain having a weakly stationary distribution π such that $\overline{\pi}(\mathrm{NS}(S))=1$. Suppose for $\overline{\pi\times\pi}$ almost surely $(i,j)\in S\times S$ there exists a near-standard absorbing point (i_0,i_0) for (i,j). Then there exists an infinite $t_0\in T$ such that for $\overline{\pi}$-almost every $x\in S$, any internal set A, any infinite $t\leq t_0$ we have $P_x(X_t\in A)\approx\pi(A)$.*

Proof. Let $\{X_t\}_{t\in T}$ be such a hyperfinite Markov chain with internal transition probability $\{p_{ij}^{(t)}\}_{i,j\in S,t\in T}$. Let $\{Y_t\}_{t\in T}$ be a i.i.d copy of $\{X_t\}_{t\in T}$ and let $\{Z_t\}_{t\in T}$ denote the product hyperfinite Markov chain. We use P' and $\overline{P'}$ to denote the internal probability and Loeb probability of $\{Z_t\}_{t\in T}$. Let $\pi'(\{(i,j)\})=\pi(\{i\})\pi(\{j\})$.

By the assumption of the theorem, we know that for $\overline{\pi'}$ almost surely $(i,j)\in S\times S$ there exists a near-standard absorbing point (i_0,i_0) for (i,j). As $\overline{\pi}(\mathrm{NS}(S))=1$, both i,j can be taken to be near-standard points. Pick an infinite $t_0\in T$ such that $\pi(A)\approx\sum_{i\in S}\pi(\{i\})P_i(X_t\in A)$ for all $t\leq t_0$ and all internal sets $A\subset S$. Now fix some internal set A and some infinite time $t_1\leq t_0$. Let M denote the set $\{\omega:\exists t<t_1-1, X_s(\omega)\approx Y_s(\omega)\approx i_0\}$. By Definition 7.11, we know that for $\overline{\pi'}$ almost surely $(i,j)\in S\times S$ and any $n\in\mathbb{N}$ we have

$$\overline{P'}_{(i,j)}((\exists t\in\mathrm{NS}(T))(Z_t\in U_{i_0}^{\frac{1}{n}}\times U_{i_0}^{\frac{1}{n}}))=1. \tag{7.36}$$

By Lemma 7.8, we know that for $\overline{\pi'}$ almost surely $(i,j)\in S\times S$ we have $\overline{P'}_{(i,j)}(M)=1$. Thus by strongly regularity of the chain, we know that for $\overline{\pi'}$ almost surely $(i,j)\in S\times S$:

$$|\overline{P}_i(X_{t_1}\in A)-\overline{P}_j(X_{t_1}\in A)| \tag{7.37}$$

$$=|\overline{P'}_{(i,j)}(X_{t_1}\in A)-\overline{P'}_{(i,j)}(Y_{t_1}\in A)| \tag{7.38}$$

$$=|\overline{P'}_{(i,j)}((X_{t_1}\in A)\cap M^c)-\overline{P'}_{(i,j)}((Y_{t_1}\in A)\cap M^c)| \tag{7.39}$$

$$\leq\overline{P'}_{(i,j)}(M^c)=0 \tag{7.40}$$

To see Eq. (7.39), note that $|\overline{P'}_{(i,j)}((X_{t_1} \in A) \cap M) - \overline{P'}_{(i,j)}((Y_{t_1} \in A) \cap M)| = 0$ since $\{X_t\}_{t \in T}$ is strong regular. Hence we know that for $\overline{\pi'}$ almost surely $(i,j) \in S \times S$ we have $|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \approx 0$.

Let the set $F = \{(i,j) \in S \times S : |P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \approx 0\}$. We know that $\overline{\pi'}(F) = 1$. For each $i \in S$, define $F_i = \{j \in S : (i,j) \in F\}$.

Claim 7.20. *For $\overline{\pi}$ almost surely $i \in S$, $\overline{\pi}(F_i) = 1$.*

Proof. Note that $\pi' = \pi \times \pi$ and is defined on all $\mathcal{I}(S \times S)$. Fix some $n \in \mathbb{N}$. Let

$$F^n = \{(i,j) \in S \times S : |P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \leq \frac{1}{n}\}. \tag{7.41}$$

For each $i \in S$, let $F_i^n = \{j \in S : (i,j) \in F^n\}$. Note that both F^n and F_i^n are internal sets. Moreover, as $F^n \supset F$, we know that $\overline{\pi'}(F^n) = 1$. We will show that, for $\overline{\pi}$ almost surely $i \in S$, F_i^n has $\overline{\pi}$ measure 1. Let $E^n = \{i \in S : (\exists j \in S)((i,j) \in F^n)\}$. By the internal definition principle, E^n is an internal set. We first show that $\overline{\pi}(E^n) = 1$. Suppose not, then there exist a positive $\epsilon \in \mathbb{R}$ such that $\mathsf{st}(\pi(E^n)) \leq 1 - \epsilon$. As $F^n \subset E^n \times S$, we have

$$\pi'(F^n) = \pi(E^n) \times \pi(S) \leq 1 - \epsilon \tag{7.42}$$

Contradicting the fact that $\overline{\pi'}(F^n) = 1$.

Now suppose that there exists a set with positive π measure such that $\overline{\pi}(F_i^n) < 1$ for every i from this set. By countable additivity and the fact that $\overline{\pi}(E^n) = 1$, there exist positive $\epsilon_1, \epsilon_2 \in \mathbb{R}$ and an internal set $D^n \subset E^n$ such that $\overline{\pi}(D^n) = \epsilon_1$ and $\overline{\pi}(F_i^n) < 1 - \epsilon_2$ for all $i \in D^n$. As each F_i^n is internal, the collection $\{F_i^n : i \in D^n\}$ is internal. Then the set $A = \bigcup_{i \in D^n} \{i\} \times F_i^n$ is internal. Thus we have

$$\overline{\pi'}(F^n) \leq \overline{\pi'}(F^n \cup A) = \overline{\pi}'(F^n \setminus A) + \overline{\pi'}(A). \tag{7.43}$$

Note that

$$\overline{\pi'}(F^n \setminus A) \leq \overline{\pi'}((E^n \setminus D^n) \times S) \leq \overline{\pi'}((S \setminus D^n) \times S) \leq 1 - \epsilon_1 \tag{7.44}$$

$$\overline{\pi'}(A) = \mathsf{st}(\pi(A)) = \mathsf{st}(\sum_{i \in D^n} \pi(\{i\})\pi(F_i^n)) \leq \mathsf{st}(\sum_{i \in D^n} \pi(\{i\})(1-\epsilon_2)) = \epsilon_1(1-\epsilon_2). \tag{7.45}$$

In conclusion, $\overline{\pi'}(F^n) = \overline{\pi'}(F^n \setminus A) + \overline{\pi'}(A) \leq (1-\epsilon_1) + \epsilon_1(1-\epsilon_2) < 1$. A contradiction. Hence, for every $n \in \mathbb{N}$, there exists a B_n with $\overline{\pi}(B_n) = 1$ such that $\overline{\pi}(F_i^n) = 1$ for every $i \in B_n$. Without loss of generality, we can assume $\{B_n\}_{n \in \mathbb{N}}$ is a decreasing sequence of sets. Thus, we have $\overline{\pi}(\bigcap_{n\in\mathbb{N}} B_n) = 1$. For every $i \in \bigcap_{n\in\mathbb{N}} B_n$, we know that $\overline{\pi}(\bigcap_{n\in\mathbb{N}} F_i^n) = 1$. As $\bigcap_{n\in\mathbb{N}} F_i^n = F_i$, we have the desired result. □

Thus we have

$$|P_i(X_{t_1} \in A) - \pi(A)| \approx |\sum_{j \in S} \pi(\{j\})(P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A))| \tag{7.46}$$

$$\leq \sum_{j \in S} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)|. \tag{7.47}$$

Recall that $F_i = \{j \in S : |P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \approx 0\}$. By the previous claim, for $\overline{\pi}$ almost all i we have $\overline{\pi}(F_i) = 1$. Pick some arbitrary positive $\epsilon \in \mathbb{R}^+$, we can find an internal $F_i' \subset F_i$ such that $\pi(F_i') > 1 - \epsilon$. Now for $\overline{\pi}$ almost all i we have

$$\sum_{j \in S} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \tag{7.48}$$

$$= \sum_{j \in S \setminus F_i'} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| + \sum_{j \in F_i'} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \tag{7.49}$$

The first part of the last equation is less than ϵ and the second part is infinitesimal. Thus we have $\sum_{j \in S} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \lessapprox \epsilon$. As ϵ is arbitrary, we know that $\sum_{j \in S} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)|$ is infinitesimal. Hence we know that for $\overline{\pi}$ almost all $i \in S$ we have $|P_i(X_{t_1} \in A) - \pi(A)| \approx 0$. As t_1 is arbitrary, we have the desired result. □

An immediate consequence of this theorem is the following result.

Corollary 7.21. *Consider a strongly regular hyperfinite Markov chain having a weakly stationary distribution π such that $\overline{\pi}(\mathrm{NS}(S)) = 1$. Suppose $\{X_t\}_{t\in T}$ is productively near-standard open set irreducible and $\pi \times \pi$ is a weakly stationary distribution of the product hyperfinite Markov chain $\{Z_t\}_{t\in T}$. Then there exists an infinite $t_0 \in T$ such that for $\overline{\pi}$-almost every $x \in S$, any internal set A, any infinite $t \leq t_0$ we have $P_x(X_t \in A) \approx \pi(A)$.*

Proof. The proof follows immediately from Theorems 7.14 and 7.19. □

It follows immediately from the construction of Loeb measure that for any internal $A \subset S$, we have $\overline{P}_x(X_t \in A) = \overline{\pi}(A)$ for any infinite $t \leq t_0$. We can extend this result to all universally Loeb measurable sets.

Lemma 7.22. *Let $\mathcal{L}(\mathcal{I}(S))$ denote the collection of all universally Loeb measurable sets (see Definition 5.8). Under the same assumptions of Theorem 7.19. For every $B \in \mathcal{L}(\mathcal{I}(S))$, every infinite $t \leq t_0$ we have $\overline{P}_x(X_t \in B) = \overline{\pi}(B)$ for $\overline{\pi}$-almost every $x \in S$.*

Proof. The proof follows directly from the construction of Loeb measures. □

As X is a metric space satisfying the Heine-Borel condition, we always have $\mathsf{st}^{-1}(E) \in \mathcal{L}(\mathcal{I}(S))$ for every $E \in \mathcal{B}[X]$.

We now show that we can actually obtain a stronger type of convergence than in Theorem 7.19 and Corollary 7.21.

Definition 7.23. Given two hyperfinite probability spaces $(S, \mathcal{I}(S), P_1)$ and $(S, \mathcal{I}(S), P_2)$, the total variation distance is defined to be

$$\| P_1(\cdot) - P_2(\cdot) \| = \sup_{A \in \mathcal{I}(S)} |P_1(A) - P_2(A)|. \tag{7.50}$$

Lemma 7.24. *We have*

$$\| P_1(\cdot) - P_2(\cdot) \| \geq \sup_{f:S\to^*[0,1]} |\sum_{i\in S} P_1(\{i\})f(i) - \sum_{i\in S} P_2(\{i\})f(i)|. \tag{7.51}$$

The sup *is taken over all internal functions.*

Proof. $|\sum_{i\in S} P_1(i)f(i) - \sum_{i\in S} P_2(i)f(i)| = |\sum_{i\in S} f(i)(P_1(i) - P_2(i))|$. This is maximized at $f(i) = 1$ for $P_1 > P_2$ and $f(i) = 0$ for $P_1 \leq P_2$ (or vice versa). Note that such f is an internal function. Thus we have $|\sum_{i\in S} f(i)(P_1(i) - P_2(i))| \leq |P_1(A) - P_2(A)|$ for $A = \{i \in S : P_1(i) > P_2(i)\}$ (or $\{i \in S : P_1(i) \leq P_2(i)\}$). This establishes the desired result. □

Consider the general hyperfinite Markov chain, for any fixed $x \in S$ and any $t \in T$ it is natural to consider the total variation distance $\| p_x^{(t)}(\cdot) - \pi(\cdot) \|$. Just as standard Markov chains, we can show that the total variation distance is non-increasing.

Lemma 7.25. *Consider a general hyperfinite Markov chain with weakly stationary distribution π. Then for any $x \in S$ and any $t_1, t_2 \in T$ such that $t_1 + t_2 \in T$, we have $\| p_x^{(t_1)}(\cdot) - \pi(\cdot) \| \gtrapprox \| p_x^{(t_1+t_2)}(\cdot) - \pi(\cdot) \|$*

Proof. Pick $t_1, t_2 \in T$ such that $t_1 + t_2 \in T$ and any internal set $A \subset S$. Then we have $|p_x^{(t_1+t_2)}(A) - \pi(A)| \approx |\sum_{y\in S} p_{xy}^{(t_1)} p_y^{(t_2)}(A) - \sum_{y\in S} \pi(y) p_y^{(t_2)}(A)|$. Let $f(y) = p_y^{(t_2)}(A)$. By the internal definition principle, we know that $p_y^{(t_2)}(A)$ is an internal function. By the previous lemma we know that

$$|p_x^{(t_1+t_2)}(A) - \pi(A)| \lessapprox \| p_x^{(t_1)}(\cdot) - \pi(\cdot) \| . \tag{7.52}$$

Since this is true for all internal A, we have shown the lemma. □

We conclude this section by introducing the following theorem which gives a stronger convergence result compared with Theorem 7.19 and Corollary 7.21.

Theorem 7.26. *Under the same hypotheses in Theorem 7.19. For $\overline{\pi}$ almost every $s \in \mathrm{NS}(S)$, the sequence $\{\sup_{B\in\mathcal{L}(\mathcal{I}(S))} |\overline{P}_s(X_t \in B) - \overline{\pi}(B)| : t \in \mathrm{NS}(T)\}$ converges to 0.*

Proof. We need to show that for any positive $\epsilon \in \mathbb{R}$ there exists a $t_1 \in \mathrm{NS}(T)$ such that for every $t \geq t_1$ we have

$$\sup_{B\in\mathcal{L}(\mathcal{I}(S))} |\overline{P}_s(X_t \in B) - \overline{\pi}(B)| < \epsilon. \tag{7.53}$$

Pick any real $\epsilon > 0$, by Theorem 7.19, we know that for any infinite $t \leq t_0$ we have $\parallel p_s^{(t)}(\cdot) - \pi(\cdot) \parallel < \frac{\epsilon}{2}$. By underspill, there exist a $t_1 \in \mathrm{NS}(T)$ such that $\parallel p_s^{(t_1)}(\cdot) - \pi(\cdot) \parallel < \frac{\epsilon}{2}$. Fix any $t_2 \geq t_1$. Then by Lemma 7.25 we have $\parallel p_s^{(t_2)}(\cdot) - \pi(\cdot) \parallel < \epsilon$. Now fix any internal set $A \subset S$. By the definition of total variation distance, we have $|P_s(X_{t_2} \in A) - \pi(A)| < \epsilon$. This implies that $|\overline{P}_s(X_{t_2} \in A) - \overline{\pi}(A)| \leq \epsilon$ for all $A \in \mathcal{I}(S)$. For external $B \in \mathcal{L}(\mathcal{I}(S))$, we have

$$\overline{P}_s(X_{t_2} \in B) = \sup\{\overline{P}_s(X_{t_2} \in A_i) : A_i \subset B, A_i \in \mathcal{I}(S)\} \tag{7.54}$$

$$\overline{\pi}(B) = \sup\{\overline{\pi}(A_i) : A_i \subset B, A_i \in \mathcal{I}(S)\} \tag{7.55}$$

hence we have $|\overline{P}_s^{(t_2)}(B) - \overline{\pi}(B)| \leq \epsilon$ for all $B \in \mathcal{L}(\mathcal{I}(S))$. Thus we have the desired result. $\square$

As $\mathsf{st}^{-1}(E) \in \mathcal{L}(\mathcal{I}(S))$ for all $E \in \mathcal{B}[X]$, we have

$$\lim_{t\to\infty} \{ \sup_{E\in\mathcal{B}[X]} |\overline{P}_x^{(t)}(\mathsf{st}^{-1}(E)) - \overline{\pi}(\mathsf{st}^{-1}(E))| : t \in \mathrm{NS}(T)\} = 0. \tag{7.56}$$

Note that the statement of Theorem 7.26 is very similar to the statement of the standard Markov chain Ergodic theorem. We will use this theorem in later sections to establish the standard Markov chain Ergodic theorem.

8. Hyperfinite Representation for Discrete-time Markov Processes

As one can see from Section 7, hyperfinite Markov processes behave like discrete-time finite state space Markov processes in many ways. Discrete-time finite state space Markov processes are well-understood and easy to work with. This makes hyperfinite Markov processes easy to work with. Thus it is desirable to construct a hyperfinite Markov process for every standard Markov process. In this section, we illustrate this idea by constructing a hyperfinite Markov process for every discrete-time general state space Markov process. Such hyperfinite Markov process is called a hyperfinite representation of the standard Markov process. For continuous-time general state space Markov processes, such construction will be done in the next section.

We start by establishing some basic properties of general Markov processes. Note that we establish these properties for general state space continuous time Markov processes. It is easy to see that these properties also hold for discrete-time general state space Markov processes.

8.1. **General properties of the transition probability.** Consider a Markov chain $\{X_t\}_{t\geq 0}$ on $(X, \mathcal{B}[X])$ where X is a metric space satisfying the Heine-Borel condition. Note that X is then a σ-compact complete metric space. We shall denote the transition probability of $\{X_t\}_{t\geq 0}$ by

$$\{P_x^{(t)}(A) : x \in X, t \in \mathbb{R}^+, A \in \mathcal{B}[X]\}. \tag{8.1}$$

Once again $P_x^{(t)}(A)$ refers to the probability of going from x to set A at time t. For each fixed $x \in X, t \geq 0$, we know that $P_x^{(t)}(\cdot)$ is a probability measure on $(X, \mathcal{B}[X])$. It is sometimes desirable to treat the transition probability as a function of three variables. Namely, we define a function $g : X \times \mathbb{R}^+ \times \mathcal{B}[X] \mapsto [0, 1]$ by $g(x, t, A) = P_x^{(t)}(A)$. We will use these to notations of transition probability interchangeably.

The nonstandard extension of g is then a function from ${}^*X \times {}^*\mathbb{R}^+ \times {}^*\mathcal{B}[X]$ to ${}^*[0, 1]$.

Lemma 8.1. *For any given $x \in {}^*X$, any $t \in {}^*\mathbb{R}^+$, ${}^*g(x, t, .)$ is an internal finitely-additive probability measure on $({}^*X, {}^*\mathcal{B}[X])$.*

Proof. : Clearly ${}^*X$ is internal and ${}^*\mathcal{B}[X]$ is an internal algebra. The following sentence is clearly true:

$(\forall x \in X)(\forall t \in \mathbb{R})(g(x,t,\emptyset) = 0 \wedge g(x,t,X) = 1 \wedge ((\forall A, B \in \mathcal{B}[X])(g(x,t,A \cup B) = g(x,t,A) + g(x,t,B) - g(x,t,A \cap B))))$.

By the transfer principle and the definition of internal probability space, we have the desired result. □

Recall that for every fixed $A \in \mathcal{B}[X]$ and any $t \geq 0$, we require that $P_x^{(t)}(A)$ is a measurable function from X to $[0,1]$. This gives rise to the following lemma.

Lemma 8.2. *For each fixed $A \in {}^*\mathcal{B}[X]$ and time point $t \in {}^*\mathbb{R}^+$, ${}^*g(x,t,A)$ is a *-Borel measurable function from ${}^*X$ to ${}^*[0,1]$.*

Proof. We know that $\forall A \in \mathcal{B}[X]\ \forall t \in \mathbb{R}^+\ \forall B \in \mathcal{B}[[0,1]]\ \{x : g(x,t,A) \in B\} \in \mathcal{B}[X]$. By the transfer principle, we get the desired result. □

For every $x \in {}^*X$ and $t \in {}^*\mathbb{R}^+$, we use $\overline{{}^*P_x^{(t)}}(\cdot)$ or $\overline{{}^*g}(x,t,\cdot)$ to denote the Loeb measure with respect to the internal probability measure ${}^*g(x,t,.)$.

We now investigate some properties of the internal function ${}^*g$. We first introduce the following definition.

Definition 8.3. For any $A, B \in \mathcal{B}[X]$, any $k_1, k_2 \in \mathbb{R}^+$ and any $x \in X$, let $f_x^{(k_1,k_2)}(A,B)$ be $P_x(X_{k_1+k_2} \in B | X_{k_1} \in A)$ when $P_x^{(k_1)}(A) > 0$ and let $f_x^{(k_1,k_2)}(A,B) = 1$ otherwise.

Intuitively, $f_x^{(k_1,k_2)}(A,B)$ denotes the probability that $\{X_t\}_{t \geq 0}$ reaches set B at time k_1+k_2 conditioned on the chain reaching set A at time k_1 had the chain started at x. For every $x \in X$, every $k_1, k_2 \in \mathbb{R}^+$ and every $A \in \mathcal{B}[X]$ it is easy to see that $f_x^{(k_1,k_2)}(A,.)$ is a probability measure on $(X, \mathcal{B}[X])$ provided that $P_x^{(k_1)}(A) > 0$. For those A such that $P_x^{(k_1)}(A) > 0$, by the definition of conditional probability, we know that $f_x^{(k_1,k_2)}(A,B) = \frac{P_x(X_{k_1+k_2} \in B \wedge X_{k_1} \in A)}{P_x^{(k_1)}(A)}$. We can view f as a function from $X \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathcal{B}[X] \times \mathcal{B}[X]$ to $[0,1]$. By the transfer principle, we know that ${}^*f$ is an internal function from ${}^*X \times {}^*\mathbb{R}^+ \times {}^*\mathbb{R}^+ \times {}^*\mathcal{B}[X] \times {}^*\mathcal{B}[X]$ to ${}^*[0,1]$.

Moreover, ${}^*f_x^{(k_1,k_2)}(A,.)$ is an internal probability measure on $({}^*X, {}^*\mathcal{B}[X])$ provided that ${}^*g(x,k_1,A) > 0$.

We first establish the following standard result of the functions g and f.

Lemma 8.4. *Consider any $k_1, k_2 \in \mathbb{R}^+$, any $x \in X$ and any two sets $A, B \in \mathcal{B}[X]$ such that $g(x,k_1,A) > 0$. If there exists an $\epsilon > 0$ such that for any two points $x_1, x_2 \in A$ we have $|g(x_1,k_2,B) - g(x_2,k_2,B)| \leq \epsilon$, then for any point $y \in A$ we have $|g(y,k_2,B) - f_x^{(k_1,k_2)}(A,B)| \leq \epsilon$.*

Proof. Since $g(x,k_1,A) > 0$, we have

$$f_x^{(k_1,k_2)}(A,B) = \frac{P_x(X_{k_1+k_2} \in B, X_{k_1} \in A)}{P_x(X_{k_1} \in A)} = \frac{\int_A g(s,k_2,B)g(x,k_1,\mathrm{d}s)}{g(x,k_1,A)}. \tag{8.2}$$

For any $y \in A$, we have

$$|g(y,k_2,B) - f_x^{(k_1,k_2)}(A,B)| = \frac{\int_A |g(y,k_2,B) - g(s,k_2,B)|g(x,k_1.ds)}{g(x,k_1,A)}. \tag{8.3}$$

As $|g(x_1,k_2,B) - g(x_2,k_2,B)| \leq \epsilon$ for any $x_1, x_2 \in A$, we have

$$\frac{\int_A |g(y,k_2,B) - g(s,k_2,B)|g(x,k_1.ds)}{g(x,k_1,A)} \leq \frac{\epsilon \cdot g(x,k_1,A)}{g(x,k_1,A)} = \epsilon. \tag{8.4}$$

□

Intuitively, this lemma means that if the probability of going from any two different points from A to B are similar then it does not matter much which point in A do we start.

Transferring Lemma 8.4, we obtain the following lemma

Lemma 8.5. *Consider any $k_1, k_2 \in {}^*\mathbb{R}^+$, any $x \in {}^*X$ and any two internal sets $A, B \in {}^*\mathcal{B}[X]$ such that $g(x,k_1,A) > 0$. If there exists a positive $\epsilon \in {}^*\mathbb{R}$ such that for any two points $x_1, x_2 \in A$ we have $|{}^*g(x_1,k_2,B) - {}^*g(x_2,k_2,B)| \leq \epsilon$, then for any point $y \in A$ we have $|{}^*g(y,k_2,B) - {}^*f_x^{(k_1,k_2)}(A,B)| \leq \epsilon$.*

In particular, if $|{}^*g(x_1,k_2,B) - {}^*g(x_2,k_2,B)| \approx 0$ for all x_1, x_2 in some A then we have $|{}^*g(y,k_2,B) - {}^*f_x^{(k_1,k_2)}(A,B)| \approx 0$ for all $y \in A$. It is easy to see that

Lemmas 8.4 and 8.5 hold for discrete-time Markov processes. We simply restrict to k_1, k_2 in $\mathbb{N}$ or ${}^*\mathbb{N}$,respectively. When $k_1 = 1$ and the context is clear, we write $f_x^{(k_2)}(A, B)$ instead of $f_x^{(k_1,k_2)}(A, B)$.

8.2. **Hyperfinite Representation for Discrete-time Markov Processes.** In this section, we consider a discrete-time general state space Markov process $\{X_t\}_{t\in\mathbb{N}}$ with a metric state space X satisfying the Heine-Borel condition. Let $\{P_x(.)\}_{x\in X}$ denote the one-step transition probability of $\{X_t\}_{t\in\mathbb{N}}$. The probability $P_x(A)$ refers to the probability of going from x to A in one step. For general n-step transition probability $P_x^{(n)}(A)$, we view it as a function $g : X \times \mathbb{N} \times \mathcal{B}[X] \mapsto [0, 1]$ in a same way as in last section. The nonstandard extension ${}^*g$ is an internal function from ${}^*X \times {}^*\mathbb{N} \times {}^*\mathcal{B}[X]$ to ${}^*[0, 1]$. We start by making the following assumption on $\{X_t\}_{t\in\mathbb{N}}$.

Condition DSF. A discrete-time Markov process $\{X_t\}_{t\in\mathbb{N}}$ is called strong Feller if for every $x \in X$ and every $\epsilon > 0$ there exists $\delta > 0$ such that

$$(\forall x_1 \in X)(|x_1 - x| < \delta \implies ((\forall A \in \mathcal{B}[X])|P_{x_1}(A) - P_x(A)| < \epsilon)). \tag{8.5}$$

We quote the following lemma regarding total variation distance. This lemma is the "standard counterpart" of Lemma 7.24.

Lemma 8.6 ([RR04])**.** *Let ν_1, ν_2 be two different probability measures on some space $(X, \mathcal{F})$ and let $\parallel \nu_1 - \nu_2 \parallel$ denote the total variation distance between ν_1, ν_2. Then $\parallel \nu_1 - \nu_2 \parallel = \sup_{f:X\to[0,1]} |\int f d\nu_1 - \int f d\nu_2|$ where f is measurable.*

An immediate consequence of Lemma 8.6 is the following result which can be viewed as a discrete-time counterpart of Lemma 7.25.

Lemma 8.7. *Consider the discrete-time Markov process $\{X_t\}_{t\in\mathbb{N}}$ with state space X. For every $\epsilon > 0$, every $x_1, x_2 \in X$ and every positive $k \in \mathbb{N}$ we have*

$$((\forall A \in \mathcal{B}[X])(|P_{x_1}^{(k)}(A) - P_{x_2}^{(k)}(A)| \leq \epsilon)) \implies ((\forall A \in \mathcal{B}[X])(|P_{x_1}^{(k+1)}(A) - P_{x_2}^{(k+1)}(A)| \leq \epsilon)). \tag{8.6}$$

Proof. : Pick any arbitrary $\epsilon > 0$, any $x_1, x_2 \in X$ and any $k \in \mathbb{N}$. We have

$$\sup_{A \in \mathcal{B}[X]} \{|P_{x_1}^{(k+1)}(A) - P_{x_2}^{(k+1)}(A)|\} \tag{8.7}$$

$$= \sup_{A \in \mathcal{B}[X]} \{|\int_{y \in X} P_y(A) P_{x_1}^{(k)}(dy) - \int_{y \in X} P_y(A) P_{x_2}^{(k)}(dy)|\} \tag{8.8}$$

$$\leq \| P_{x_1}^{(k)}(\cdot) - P_{x_2}^{(k)}(\cdot) \| \leq \epsilon. \tag{8.9}$$

Thus we have proved the result. □

By the transfer principle and (DSF), we have the following result.

Lemma 8.8. *Suppose $\{X_t\}_{t \in \mathbb{N}}$ satisfies (DSF). Let $x_1 \approx x_2 \in \mathrm{NS}({}^*X)$. Then for every positive $k \in \mathbb{N}$ and every $A \in {}^*\mathcal{B}[X]$ we have ${}^*g(x_1, k, A) \approx {}^*g(x_2, k, A)$.*

Proof. Fix $x_1, x_2 \in \mathrm{NS}({}^*X)$. We first prove the result for $k = 1$. Let $x_0 = \mathsf{st}(x_1) = \mathsf{st}(x_2)$ and let ϵ be any positive real number. By (DSF) and the transfer principle, we know that there exists $\delta \in \mathbb{R}^+$ such that

$$(\forall x \in {}^*X)(|x - x_0| < \delta \implies ((\forall A \in {}^*\mathcal{B}[X])|{}^*g(x, 1, A) - {}^*g(x_0, 1, A)| < \epsilon)) \tag{8.10}$$

As $x_1 \approx x_2 \approx x_0$ and ϵ is arbitrary, we know that ${}^*g(x_1, 1, A) \approx {}^*g(x_0, 1, A) \approx {}^*g(x_2, 1, A)$ for all $A \in {}^*\mathcal{B}[X]$.

We now prove the lemma for all $k \in \mathbb{N}$. Again fix some $\epsilon \in \mathbb{R}^+$. We know that

$$(\forall A \in {}^*\mathcal{B}[X])(|{}^*g(x_1, 1, A) - {}^*g(x_2, 1, A)| < \epsilon). \tag{8.11}$$

By the transfer of Lemma 8.7, we know that for every $k \in \mathbb{N}$ we have

$$(\forall A \in {}^*\mathcal{B}[X])(|{}^*g(x_1, k, A) - {}^*g(x_2, k, A)| < \epsilon). \tag{8.12}$$

As ϵ is arbitrary, we have the desired result. □

We are now at the place to construct a hyperfinite Markov process $\{X'_t\}_{t \in \mathbb{N}}$ which represents our standard Markov process $\{X_t\}_{t \in \mathbb{N}}$. Our first task is to specify the state space of $\{X'_t\}_{t\in\mathbb{N}}$. Pick any positive infinitesimal δ and any positive infinite number r. Our state space S for $\{X'_t\}_{t\in\mathbb{N}}$ is simply a (δ, r)-hyperfinite representation of ${}^*X$. The following properties of S will be used later.

(1) For each $s \in S$, there exists a $B(s) \in {}^*\mathcal{B}[X]$ with diameter no greater than δ containing s such that $B(s_1) \cap B(s_2) = \emptyset$ for any two different $s_1, s_2 \in S$.
(2) $\mathrm{NS}({}^*X) \subset \bigcup_{s\in S} B(s)$.

For every $x \in {}^*X$, we know that ${}^*g(x,1,.)$ is an internal probability measure on $({}^*X, {}^*\mathcal{B}[X])$. When X is non-compact, $\bigcup_{s\in S} B(s) \neq {}^*X$. We can truncate ${}^*g$ to an internal probability measure on $\bigcup_{s\in S} B(s)$.

Definition 8.9. For $i \in \{0,1\}$, let $g'(x,i,A) : \bigcup_{s\in S} B(s) \times {}^*\mathcal{B}[X] \to {}^*[0,1]$ be given by:

$$g'(x,i,A) = {}^*g(x,i,A \cap \bigcup_{s\in S} B(s)) + \delta_x(A){}^*g(x,i,{}^*X \setminus \bigcup_{s\in S} B(s)). \tag{8.13}$$

where $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if otherwise.

Intuitively, this means that if our *Markov chain is trying to reach ${}^*X \setminus \bigcup_{s\in S} B(s)$ then we would force it to stay at where it is. For any $x \in \bigcup_{s\in S} B(s)$ and any $A \in {}^*\mathcal{B}[X]$, it is easy to see that $g'(x,0,A) = 1$ if $x \in A$ and equals to 0 otherwise. Clearly, $g'(x,0,.)$ is an internal probability measure for every $x \in \bigcup_{s\in S} B(s)$.

We first show that g' is a valid internal probability measure.

Lemma 8.10. *Let* $\mathcal{B}[\bigcup_{s\in S} B(s)] = \{A \cap \bigcup_{s\in S} B(s) : A \in {}^*\mathcal{B}[X]\}$. *Then for any* $x \in \bigcup_{s\in S} B(s)$*, the triple* $(\bigcup_{s\in S} B(s), \mathcal{B}[\bigcup_{s\in S} B(s)], g'(x,1,.))$ *is an internal probability space.*

Proof. Fix $x \in \bigcup_{s\in S} B(s)$. We only need to show that $g'(x,1,.)$ is an internal probability measure on $(\bigcup_{s\in S} B(s), \mathcal{B}[\bigcup_{s\in S} B(s)])$.

By definition, it is clear that $g'(x,1,\emptyset)=0$ and $g'(x,1,\bigcup_{s\in S}B(s))=1$. Consider two disjoint $A,B\in\mathcal{B}[\bigcup_{s\in S}B(s)]$, we have:

$$g'(x,1,A\cup B) \tag{8.14}$$

$$={}^*g(x,1,A\cup B)+\delta_x(A\cup B){}^*g(x,1,{}^*X\setminus\bigcup_{s\in S}B(s)) \tag{8.15}$$

$$={}^*g(x,1,A)+\delta_x(A){}^*g(x,1,{}^*X\setminus\bigcup_{s\in S}B(s))+{}^*g(x,1,B)+\delta_x(B){}^*g(x,1,{}^*X\setminus\bigcup_{s\in S}B(s)) \tag{8.16}$$

$$=g'(x,1,A)+g'(x,1,B). \tag{8.17}$$

Thus we have the desired result. □

In fact, for $x\in \mathrm{NS}({}^*X)=\mathsf{st}^{-1}(X)$, the probability of escaping to infinity is always infinitesimal.

Lemma 8.11. *Suppose $\{X_t\}_{t\in\mathbb{N}}$ satisfies (DSF). Then for any $x\in\mathrm{NS}({}^*X)$ and any $t\in\mathbb{N}$, we have $\overline{{}^*g}(x,t,\mathsf{st}^{-1}(X))=1$.*

Proof. Pick a $x\in\mathrm{NS}({}^*X)$ and some $t\in\mathbb{N}$. Let $x_0=\mathsf{st}(x)$. By Lemma 8.8, we know that ${}^*g(x,t,A)\approx{}^*g(x_0,t,A)$ for every $A\in{}^*\mathcal{B}[X]$. Thus we have $\overline{{}^*g}(x,t,\mathsf{st}^{-1}(X))=\overline{{}^*g}(x_0,t,\mathsf{st}^{-1}(X))=1$, completing the proof. □

We now define the hyperfinite Markov chain $\{X'_t\}_{t\in\mathbb{N}}$ on $(S,\mathcal{I}(S))$ from $\{X_t\}_{t\in\mathbb{N}}$ by specifying its "one-step" transition probability. For $i,j\in S$ let $G^{(0)}_{ij}=g'(i,0,B(j))$ and $G_{ij}=g'(i,1,B(j))$. Intuitively, G_{ij} refers to the probability of going from i to j in one step. For any internal set $A\subset S$ and any $i\in S$, $G_i(A)=\sum_{j\in A}G_{ij}$. Then $\{X'_t\}_{t\in\mathbb{N}}$ is the hyperfinite Markov chain on $(S,\mathcal{I}(S))$ with "one-step" transition probability $\{G_{ij}\}_{i,j\in S}$. We first verify that $G_i(\cdot)$ is an internal probability measure on $(S,\mathcal{I}(S))$ for every $i\in S$.

Lemma 8.12. *For every $i\in S$, $G_i(\cdot)$ and $G^{(0)}_i(\cdot)$ are internal probability measure on $(S,\mathcal{I}(S))$.*

Proof. Clearly $G_i^{(0)}(A) = 1$ if $i \in A$ and $G_i^{(0)}(A) = 0$ otherwise. Thus $G_i^{(0)}(\cdot)$ is an internal probability measure on $(S, \mathcal{S})$.

Now consider $G_i(\cdot)$. By definition, it is clear that

$$G_i(\emptyset) = g'(i, 1, \emptyset) = 0 \tag{8.18}$$

$$G_i(S) = g'(i, 1, \bigcup_{s \in S} B(s)) = {}^*g(i, 1, \bigcup_{s \in S} B(s)) + \delta_i(\bigcup_{s \in S} B(s))^* g(i, 1, {}^*X \setminus \bigcup_{s \in S} B(s)) = 1. \tag{8.19}$$

For hyperfinite additivity, it is sufficient to note that for any two internal sets $A, B \subset S$ and any $i \in S$ we have $G_i(A \cup B) = \sum_{j \in A \cup B} G_{ij} = G_i(A) + G_i(B)$. $\square$

We use $G_i^{(t)}(\cdot)$ to denote the t-step transition probability of $\{X'_t\}_{t \in \mathbb{N}}$. Note that $G_i^{(t)}(\cdot)$ is purely determined from the "one-step" transition matrix $\{G_{ij}\}_{i,j \in S}$. We now show that $G_i^{(t)}(\cdot)$ is an internal probability measure on $(S, \mathcal{I}(S))$.

Lemma 8.13. *For any $i \in S$ and any $t \in \mathbb{N}$, $G_i^{(t)}(\cdot)$ is an internal probability measure on $(S, \mathcal{I}(S))$.*

Proof. We will prove this by internal induction on t.

For t equals to 0 or 1, we already have the results by Lemma 8.12.

Suppose the result is true for $t = t_0$. We now show that it is true for $t = t_0 + 1$. Fix any $i \in S$. For all $A \in \mathcal{I}(S)$ we have $G_i^{(t_0+1)}(A) = \sum_{j \in S} G_{ij} G_j^{(t_0)}(A)$. Thus we have $G_i^{(t_0+1)}(\emptyset) = \sum_{j \in S} G_{ij} G_j^{(t_0)}(\emptyset) = 0$. Similarly we have $G_i^{(t_0+1)}(S) = \sum_{j \in S} G_{ij} G_j^{(t_0)}(S) = 1$. Pick any two disjoint sets $A, B \in \mathcal{I}(S)$. We have:

$$G_i^{(t_0+1)}(A \cup B) = \sum_{j \in S} G_{ij}(G_j^{(t_0)}(A) + G_j^{(t_0)}(B)) = G_j^{(t_0+1)}(A) + G_j^{(t_0+1)}(B). \tag{8.20}$$

Hence $G_i^{(t_0+1)}(\cdot)$ is an internal probability measure on $(S, \mathcal{I}(S))$. Thus by internal induction, we have the desired result. $\square$

The following lemma establishes the link between *transition probability and the internal transition probability of $\{X'_t\}_{t \in \mathbb{N}}$.

Theorem 8.14. *Suppose $\{X_t\}_{t\in\mathbb{N}}$ satisfies (DSF). Then for any $n\in\mathbb{N}$, any $x\in \mathrm{NS}(S)$ and any $A\in {}^*\mathcal{B}[X]$, ${}^*g(x,n,\bigcup_{s\in A\cap S}B(s))\approx G_x^{(n)}(A\cap S)$.*

Proof. We prove the theorem by induction on $n\in\mathbb{N}$.

Let $n=1$. Fix any $x\in \mathrm{NS}({}^*X)\cap S$ and any $A\in {}^*\mathcal{B}[X]$. We have

$$G_x(A\cap S) \tag{8.21}$$

$$= g'(x,1,\bigcup_{s\in A\cap S}B(s)) \tag{8.22}$$

$$= {}^*g(x,1,\bigcup_{s\in A\cap S}B(s))+\delta_x(\bigcup_{s\in A\cap S}B(s))^*g(x,1,{}^*X\setminus\bigcup_{s\in S}B(s)) \tag{8.23}$$

$$\approx {}^*g(x,1,\bigcup_{s\in A\cap S}B(s)) \tag{8.24}$$

where the last $\approx$ follows from Lemma 8.11.

We now prove the general case. Fix any $x\in \mathrm{NS}({}^*X)\cap S$ and any $A\in {}^*\mathcal{B}[X]$. Assume the theorem is true for $t=k$ and we will show the result holds for $t=k+1$. We have

$$^*g(x,k+1,\bigcup_{s'\in A\cap S}B(s')) \tag{8.25}$$

$$=(\sum_{s\in S}{}^*g(x,1,B(s))^*f_x^{(k)}(B(s),\bigcup_{s'\in A\cap S}B(s'))+{}^*g(x,1,{}^*X\setminus\bigcup_{s\in S}B(s))^*f_x^{(k)}({}^*X\setminus\bigcup_{s\in S}B(s),\bigcup_{s'\in A\cap S}B(s')) \tag{8.26}$$

$$\approx\sum_{s\in S}{}^*g(x,1,B(s))^*f_x^{(k)}(B(s),\bigcup_{s'\in A\cap S}B(s')). \tag{8.27}$$

where the last $\approx$ follows from Lemma 8.11.

By Lemmas 8.5 and 8.8, we have ${}^*f_x^{(k)}(B(s),\bigcup_{s'\in A\cap S}B(s'))\approx {}^*g(s,k,\bigcup_{s'\in A\cap S}B(s'))$. Thus we have

$$\sum_{s\in S}{}^*g(x,1,B(s))^*f_x^{(k)}(B(s),\bigcup_{s'\in A\cap S}B(s'))\approx\sum_{s\in S}{}^*g(x,1,B(s))^*g(s,k,\bigcup_{s'\in A\cap S}B(s')). \tag{8.28}$$

It remains to show that $\sum_{s\in S} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) \approx G_x^{(k+1)}(A\cap S)$. Fix any positive $\epsilon \in \mathbb{R}$. By Lemma 8.11, we can pick an internal set $M \subset \mathrm{NS}(S)$ such that ${}^*g(x,1,\bigcup_{s\in M} B(s)) > 1-\epsilon$. We then have

$$\sum_{s\in S} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) \tag{8.29}$$

$$= \sum_{s\in M} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) + \sum_{s\in S\setminus M} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')). \tag{8.30}$$

By induction hypothesis, we have ${}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) \approx G_s^{(k)}(A\cap S)$ for all $s\in M$. By Lemma 3.20 we have

$$\sum_{s\in M} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) \approx \sum_{s\in M} {}^*g(x,1,B(s))G_s^{(k)}(A\cap S). \tag{8.31}$$

As all $B(s)$ are mutually disjoint, x lies in at most one element of the collection $\{B(s) : s\in M\}$. Suppose $x\in B(s_0)$ for some $s_0\in M$. Then we have

$$|\sum_{s\in M} {}^*g(x,1,B(s))G_s^{(k)}(A\cap S) - \sum_{s\in M} g'(x,1,B(s))G_s^{(k)}(A\cap S)| \tag{8.32}$$

$$= |({}^*g(x,1,B(s_0)) - g'(x,1,B(s_0)))G_{s_0}^{(k)}(A\cap S)| \tag{8.33}$$

$$= |{}^*g(x,1,{}^*X\setminus \bigcup_{s\in S} B(s))G_{s_0}^{(k)}(A\cap S)| \approx 0 \tag{8.34}$$

where the last $\approx$ follows from Lemma 8.11. Thus, by Eq. (8.31), we have

$$\sum_{s\in M} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) \tag{8.35}$$

$$\approx \sum_{s\in M} g'(x,1,B(s))G_s^{(k)}(A\cap S) \tag{8.36}$$

$$= \sum_{s\in M} G_x(\{s\})G_s^{(k)}(A\cap S). \tag{8.37}$$

As ${}^*g(x,1,\bigcup_{s\in M} B(s)) > 1-\epsilon$, we know that

$$\sum_{s\in S\setminus M} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) < \epsilon. \tag{8.38}$$

On the other hand, we have

$$\sum_{s\in S\setminus M} G_x(\{s\})G_s^{(k)}(A\cap S) \tag{8.39}$$

$$= \sum_{s\in S\setminus M} g'(x,1,B(s))G_s^{(k)}(A\cap S) \tag{8.40}$$

$$\leq \sum_{s\in S\setminus M} g'(x,1,B(s)) \tag{8.41}$$

$$\leq {}^*g(x,1,\bigcup_{s\in S\setminus M} B(s)) + {}^*g(x,1,{}^*X\setminus\bigcup_{s\in S} B(s)) \tag{8.42}$$

$$\approx {}^*g(x,1,\bigcup_{s\in S\setminus M} B(s)) < \epsilon \tag{8.43}$$

where the second last $\approx$ follows from Lemma 8.11.

Thus the difference between

$\sum_{s\in M} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s'))+\sum_{s\in S\setminus M} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s'))$ and $\sum_{s\in M} G_x(\{s\})G_s^{(k)}(A\cap S)+\sum_{s\in S\setminus M} G_x(\{s\})G_s^{(k)}(A\cap S)$ is less or approximately to ϵ. Hence we have

$$|{}^*g(x,k+1,\bigcup_{s'\in A\cap S} B(s')) - G_x^{(k+1)}(A\cap S)| \lessapprox \epsilon \tag{8.44}$$

As our choice of ϵ is arbitrary, we have ${}^*g(x,k+1,\bigcup_{s'\in A\cap S} B(s')) \approx G_x^{(k+1)}(A\cap S)$, completing the proof. $\square$

The following lemma is a slight generalization of [ACH97, Thm 4.1].

Lemma 8.15. *Suppose $\{X_t\}_{t\in\mathbb{N}}$ satisfies (DSF). Then for any Borel set E, any $x\in \mathrm{NS}({}^*X)$ and any $n\in\mathbb{N}$, we have ${}^*g(x,n,{}^*E)\approx \overline{{}^*g}(x,n,\mathsf{st}^{-1}(E))$.*

Proof. Fix $x\in \mathrm{NS}({}^*X)$ and $n\in\mathbb{N}$. Let $x_0=\mathsf{st}(x)$. Fix any positive $\epsilon\in\mathbb{R}$, as $g(x_0,n,.)$ is a Radon measure, we can find K compact, U open with $K\subset E\subset U$ such that $g(x_0,n,U)-g(x_0,n,K)<\frac{\epsilon}{2}$. By the transfer principle, we know that ${}^*g(x_0,n,{}^*U)-{}^*g(x_0,n,{}^*K)<\epsilon/2$. By (DSF) we know that ${}^*g(x_0,n,{}^*U)\approx {}^*g(x,n,{}^*U)$ and ${}^*g(x_0,n,{}^*K)\approx {}^*g(x,n,{}^*K)$. Hence we know that ${}^*g(x,n,{}^*U)-{}^*g(x,n,{}^*K)<\epsilon$. Note that ${}^*K\subset \mathsf{st}^{-1}(K)\subset \mathsf{st}^{-1}(E)\subset \mathsf{st}^{-1}(U)\subset {}^*U$. Both

${}^*g(x,n,{}^*E)$ and $\overline{{}^*g}(x,n,\mathsf{st}^{-1}(E))$ lie between ${}^*g(x,n,{}^*U)$ and ${}^*g(x,n,{}^*K)$. So $|{}^*g(x,n,{}^*E)-\overline{{}^*g}(x,n,\mathsf{st}^{-1}(E))|<\epsilon$. This is true for any ϵ and hence ${}^*g(x,n,{}^*E)\approx\overline{{}^*g}(x,n,\mathsf{st}^{-1}(E))$. □

We are now at the place to establish the link between the transition probability of $\{X_t\}_{t\in\mathbb{N}}$ and the internal transition probability of $\{X'_t\}_{t\in\mathbb{N}}$.

Theorem 8.16. *Suppose $\{X_t\}_{t\in\mathbb{N}}$ satisfies (DSF). Then for any $s\in\mathrm{NS}(S)$, any $n\in\mathbb{N}$ and any $E\in\mathcal{B}[X]$, $P^{(n)}_{\mathsf{st}(s)}(E)=\overline{G}^{(n)}_s(\mathsf{st}^{-1}(E)\cap S)$.*

Proof. Fix any $s\in\mathrm{NS}(S)$, any $n\in\mathbb{N}$ and any Borel set E. By Lemma 8.15, we have $P^{(n)}_{\mathsf{st}(s)}(E)={}^*g(\mathsf{st}(s),n,{}^*E)\approx{}^*g(s,n,{}^*E)\approx\overline{{}^*g}(s,n,\mathsf{st}^{-1}(E))$. By Eq. (6.19), we have

$$\overline{{}^*g}(s,n,st^{-1}(E))=\sup\{\overline{{}^*g}(s,n,\bigcup_{s\in A_i}B(s)):A_i\subset\mathsf{st}^{-1}(E)\cap S,A_i\in\mathcal{I}(S)\}. \tag{8.45}$$

By Theorem 8.14, we have $\overline{{}^*g}(s,n,\bigcup_{s\in A_i}B(s))=\overline{G}^{(n)}_s(A_i)$. Thus we have

$$\overline{{}^*g}(s,n,\mathsf{st}^{-1}(E))=\sup\{\overline{G}^{(n)}_s(A_i):A_i\subset\mathsf{st}^{-1}(E)\cap S,A_i\in\mathcal{I}(S)\}=\overline{G}^{(n)}_s(\mathsf{st}^{-1}(E)\cap S). \tag{8.46}$$

Hence we have the desired result. □

Thus the transition probability of $\{X_t\}_{t\in\mathbb{N}}$ agrees with the Loeb probability of $\{X'_t\}_{t\in\mathbb{N}}$ via standard part map.

9. Hyperfinite Representation for Continuous-time Markov Processes

In Section 8.2, for every standard discrete-time Markov process, we construct a hyperfinite Markov process that represents it. In this section, we extend the results developed in Section 8 to continuous-time Markov processes. Let $\{X_t\}_{t\geq 0}$ be a continuous-time Markov process on a metric state space X satisfying the Heine-Borel condition. The transition probability of $\{X_t\}_{t\geq 0}$ is given by

$$\{P_x^{(t)}(A) : x \in X, t \in \mathbb{R}^+, A \in \mathcal{B}[X]\}. \tag{9.1}$$

When we view the transition probability as a function of three variables, we again use $g(x,t,A)$ to denote the transition probability $P_x^{(t)}(A)$. We have already established some general properties regarding the transition probability $g(x,t,A)$ in Section 8.1. We recall some important definitions are results here.

Definition 9.1. For any $A, B \in \mathcal{B}[X]$, any $k_1, k_2 \in \mathbb{R}^+$ and any $x \in X$, let $f_x^{(k_1,k_2)}(A,B)$ be $P_x(X_{k_1+k_2} \in B | X_{k_1} \in A)$ when $P_x^{(k_1)}(A) > 0$ and let $f_x^{(k_1,k_2)}(A,B) = 1$ otherwise.

Again, f can be viewed as a function of five variables. Let $\{A_n : n \in \mathbb{N}\}$ be a partition of X consisting of Borel sets and let $k_1, k_2 \in \mathbb{R}^+$. For any $x \in X$ and any $A \in \mathcal{B}[X]$, we have

$$g(x, k_1 + k_2, A) = \sum_{n \in \mathbb{N}} g(x, k_1, A_n) f_x^{(k_1,k_2)}(A_n, A). \tag{9.2}$$

Intuitively, this means that the Markov chain first go to one of the A_n's at time k_1 and then go from that A_n to A in time k_2.

As in Section 8.1, we are interested in the relation between the nonstandard extensions of g and f. Recall Lemma 8.5 from Section 8.1.

Lemma 9.2. *Consider any $k_1, k_2 \in {}^*\mathbb{R}^+$, any $x \in {}^*X$ and any two sets $A, B \in {}^*\mathcal{B}[X]$ such that $g(x, k_1, A) > 0$. If there exists a positive $\epsilon \in {}^*\mathbb{R}$ such that for any two points $x_1, x_2 \in A$ we have $|{}^*g(x_1, k_2, B) - {}^*g(x_2, k_2, B)| \leq \epsilon$, then for any point $y \in A$ we have $|{}^*g(y, k_2, B) - {}^*f_x^{(k_1,k_2)}(A, B)| \leq \epsilon$.*

Let the hyperfinite time line $T = \{\delta t, \ldots, K\}$ as in Section 7. When $k_1 = \delta t$ and the context is clear, we write $f_x^{(k_2)}(A,B)$ instead of $f_x^{(k_1,k_2)}(A,B)$.

In Section 8.2, we constructed a hyperfinite Markov chain $\{X'_t\}_{t\in\mathbb{N}}$ which represents our standard Markov chain $\{X_t\}_{t\in\mathbb{N}}$. The idea was that the difference between the transition probability of $\{X_t\}_{t\in\mathbb{N}}$ and the internal transition probability $\{X'_t\}_{t\in\mathbb{N}}$ generated from each step is infinitesimal. Since the time-line was discrete, this implies that the transition probability of $\{X_t\}_{t\in\mathbb{N}}$ and $\{X'_t\}_{t\in\mathbb{N}}$ agree with each other. However, for continuous-time Markov process, we need to make sure that if we add up the errors up to any near-standard time t_0 the sum is still infinitesimal. Thus, instead of taking any hyperfinite representation of ${}^*X$ to be our state space we need to carefully choose our state space for our hyperfinite Markov process.

9.1. **Construction of Hyperfinite State Space.** In this section, we will carefully pick a hyperfinite set $S \subset {}^*X$ to be the hyperfinite state space for our hyperfinite Markov chain. The set S will be a (δ_0, r)-hyperfinite representation of ${}^*X$ for some infinitesimal δ_0 and some positive infinite r. Intuitively, δ_0 measures the closeness between the points in S and r measures the portion of ${}^*X$ to be covered by S. We first pick ϵ_0 such that $\epsilon_0 \frac{t}{\delta t} \approx 0$ for all $t \in T$. This ϵ_0 will be fixed for the remainder of this section. We first choose r according to this ϵ_0. We first recall the following definitions from Section 2.

Definition 9.3 (Definition 2.8)**.** Let $\mathcal{K}[X]$ denote the collection of compact subsets of X. The Markov chain $\{X_t\}_{t\geq 0}$ is said to be *vanishing in distance* if for all $t \geq 0$, all $K \in \mathcal{K}[X]$ and every $\epsilon > 0$, the set $\{x \in X : P^{(t)}_x(K) \geq \epsilon\}$ is contained in a compact subset of X.

Definition 9.4 (Definition 2.9)**.** A Markov chain $\{X_t\}_{t\geq 0}$ is said to be *strong Feller* if for all $t > 0$, all $\epsilon > 0$, all $x \in X$, there exists $\delta > 0$ such that:

$$((\forall y \in X)(d(x,y) < \delta \implies (\forall A \in \mathcal{B}[X])|P^{(t)}_y(A) - P^{(t)}_x(A)| < \epsilon)). \tag{9.3}$$

We first introduce the following definition from general topology.

Definition 9.5. Let X be a metric space. For every $x \in X$ and every $A \subset X$, the *distance between x and A* is defined by $d(x,A) = \inf\{d(x,a) : a \in A\}$.

As promised in Section 2, we now give an equivalent formulation of Definition 2.8 using the metric. It is easier to see the intuition behind Definition 2.8 using the following formulation of "diminishing in transition probability".

Condition DT. For all $t \geq 0$ and all compact $K \subset X$ we have:

(1) $(\forall \epsilon > 0)(\exists r > 0)(\forall x \in K)(\forall A \in \mathcal{B}[X])(d(x, A) > r \implies g(x, t, A) < \epsilon)$.

(2) $(\forall \epsilon > 0)(\exists r > 0)(\forall x \in X)(d(x, K) > r \implies g(x, t, K) < \epsilon)$.

Note that every Markov process with bounded state space satisfies (DT) automatically. We now show that (DT) can be derived from Definitions 2.8 and 2.9.

Theorem 9.6. *Suppose a Markov process $\{X_t\}_{t \geq 0}$ has the strong Feller property, then it also has property (1) of (DT). Property (2) of (DT) is equivalent to vanishing in distance whenever the metric on X has the Heine-Borel property.*

Proof. We first show that property (1) of (DT) follows automatically from the strong Feller property. Fix $t \geq 0$, $K \in \mathcal{K}[X]$ and $\epsilon > 0$. Let $F_n = \{x \in X : d(x, K) > n\}$. By countable additivity, for every $x \in X$, there exists $n(x) \in \mathbb{N}$ such that $P_x^{(t)}(F_{n(x)}) < \epsilon$. By the strong Feller property, for every $x \in X$, there exists an open ball $U(x)$ such that $P_y^{(t)}(F_{n(x)}) < \epsilon$ for all $y \in U(x)$. By compactness of K, there exists a fixed $m \in \mathbb{N}$ such that $P_y^{(t)}(F_m) < \epsilon$ for all $y \in K$. Let $r = m + s$ where s is the diameter of K. By the triangle inequality, if $x \in K$ and $A \in \mathcal{B}[X]$ with $d(x, A) > r$ then $A \subset F_m$. Hence, if $d(x, A) > r$ then $P_x^{(t)}(A) < \epsilon$.

We now show that property (2) of (DT) is equivalent to vanishing in distance whenever the metric on X has the Heine-Borel property. Logically, property (2) of (DT) is equivalent to: for all $t \geq 0$, all $K \in \mathcal{K}[X]$ and all $\epsilon > 0$, there exists $r > 0$ such that

$$\{x \in X : P_x^{(t)}(K) \geq \epsilon\} \subset \{x \in X : d(x, K) \leq r\}. \tag{9.4}$$

Suppose $\{X_t\}_{t \geq 0}$ has property (2) of (DT) and the metric on X has the Heine-Borel property. As the metric on X has the Heine-Borel property, the set $\{x \in X : d(x, K) \leq r\}$ is compact so $\{X_t\}_{t \geq 0}$ is vanishing in distance.

Suppose $\{X_t\}_{t\geq 0}$ is vanishing in distance. Fix $t \geq 0$, $K \in \mathcal{K}[X]$ and $\epsilon > 0$. The set $\{x \in X : P_x^{(t)}(K) \geq \epsilon\}$ is contained in some compact set $T \subset X$. As T is compact, there exists some $r > 0$ such that $\{x \in X : P_x^{(t)}(K) \geq \epsilon\} \subset T \subset \{x \in X : d(x, K) \leq r\}$. Hence $\{X_t\}_{t\geq 0}$ has property (2) of (DT). □

As we will always assume that $\{X_t\}_{t\geq 0}$ is vanishing in distance and strong Feller, we shall use (DT) instead of Definition 2.8 from now on. The following condition is an alternative condition to (DT), but stronger.

Condition SDT. For all $t \geq 0$ we have

$$(\forall \epsilon > 0)(\exists r > 0)(\forall x \in X)(\forall A \in \mathcal{B}[X])(d(x, A) > r \implies g(x, t, A) < \epsilon). \qquad (9.5)$$

It is easy to see that (SDT) implies (DT).

Example 9.7. The Ornstein-Uhlenbeck is a continuous time stochastic process $\{X_t\}_{t\geq 0}$ satisfies the stochastic differential equation:

$$\mathrm{d}X_t = \theta(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t. \qquad (9.6)$$

where $\theta > 0$, $\mu > 0$ and $\sigma > 0$ are parameters and W_t denote the Wiener process. The Ornstein-Uhlenbeck process is a stationary Gauss-Markov process. Note that the Ornstein-Uhlenbeck process satisfies (DT) but not (SDT). As the state space of the Ornstein-Uhlenbeck process satisfies the Heine-Borel condition, by Theorem 9.6, the Ornstein-Uhlenbeck process also satisfies Definition 2.8.

An open ball centered at some $x_0 \in {}^*X$ with radius r is simply the set

$$\{x \in {}^*X : {}^*d(x, x_0) \leq r\}. \qquad (9.7)$$

We usually use $U(x_0, r)$ to denote such set.

Theorem 9.8. *Suppose (DT) holds. For every positive $\epsilon \in {}^*\mathbb{R}$, there exists an open ball $U(a, r)$ centered at some standard point a with radius r such that:*

(1) ${}^*g(x, \delta t, {}^*X \setminus \overline{U}(a, r)) < \epsilon$ for all $x \in \mathrm{NS}({}^*X)$.

(2) ${}^*g(y,t,A)<\epsilon$ *for all* $y\in {}^*X\setminus \overline{U}(a,r)$, *all near-standard* $A\in {}^*\mathcal{B}[X]$ *and all* $t\in T$.

where $\overline{U}(a,r)=\{x\in {}^*X: {}^*d(x,a)\leq r\}$.

Proof. : Fix a positive $\epsilon\in {}^*\mathbb{R}$. Let $X=\bigcup_{n\in\mathbb{N}}K_n$. For every $n\in\mathbb{N}$, by the transfer of condition 1 of (DT), there exists $r\in {}^*\mathbb{R}^+$ such that the following formula $\psi_n(r)$ holds:

$$(\forall x\in {}^*K_n)(\forall A\in {}^*\mathcal{B}[X])({}^*d(x,A)>r\implies {}^*g(x,\delta t,A)<\epsilon). \tag{9.8}$$

It is easy to see that $\{\psi_n(r):n\in\mathbb{N}\}$ is a family of finitely satisfiable internal formulas. By the saturation principle, there is a $r_{\delta t}$ such that

$$(\forall x\in \bigcup_{n\in\mathbb{N}}{}^*K_n)(\forall A\in {}^*\mathcal{B}[X])({}^*d(x,A)>r_{\delta t}\implies {}^*g(x,\delta t,A)<\epsilon). \tag{9.9}$$

Claim 9.9. *For every* $n\in\mathbb{N}$, *the formula* $\phi_n(r)$

$$(\forall x\in {}^*X)({}^*d(x,{}^*K_n)>r\implies ((\forall t\in T)({}^*g(x,t,{}^*K_n)<\epsilon))). \tag{9.10}$$

is satisfiable.

Proof. Fix some $n\in\mathbb{N}$. For every $t\in T$, by the transfer of condition 2 of (DT), there exists $r\in {}^*\mathbb{R}^+$ such that the following formula holds:

$$(\forall x\in {}^*X)({}^*d(x,{}^*K_n)>r\implies {}^*g(x,t,{}^*K_n)<\epsilon). \tag{9.11}$$

Define $h:T\to {}^*\mathbb{R}^+$ by

$$h(t)=\min\{r\in {}^*\mathbb{R}^+:(\forall x\in {}^*X)({}^*d(x,{}^*K_n)>r\implies {}^*g(x,t,{}^*K_n)<\epsilon)\} \tag{9.12}$$

By the internal definition principle, h is an internal function thus $h(T)$ is a hyperfinite set. Let $r_n=\max\{r:r\in h(T)\}$. Then r_n witnesses the satisfiability of the formula $\phi_n(r)$. □

For any $k \in \mathbb{N}$, it is easy to see that $\max\{r_{n_i} : i \leq k\}$ witnesses the satisfiability of $\{\phi_{n_i}(r) : i \leq k\}$. Hence the family $\{\phi_n(r) : n \in \mathbb{N}\}$ is finitely satisfiable. By the saturation principle, there exists a r' satisfies all $\phi_n(r)$ simultaneously. This means

$$(\forall x \in {}^*X)(\forall n \in \mathbb{N})({}^*d(x, {}^*K_n) > r' \implies ((\forall t \in T)({}^*g(x, t, {}^*K_n) < \epsilon))). \tag{9.13}$$

Consider any near-standard internal set A.

Claim 9.10. *There exists $n \in \mathbb{N}$ such that $A \subset {}^*K_n$.*

Proof. Suppose not. Then $\mathcal{M}_n = \{a \in A : a \not\in {}^*K_n\}$ is non-empty for every $n \in \mathbb{N}$. It is easy to see that any finite intersection of these is non-empty. By saturation, we know that $\bigcap_{n\in\mathbb{N}} \mathcal{M}_n \neq \emptyset$. Hence there exists $a \in A$ such that $a \not\in \bigcup_{n\in\mathbb{N}} {}^*K_n$. By Theorem 3.28, we know that $\bigcup_{n\in\mathbb{N}} {}^*K_n = \mathrm{NS}({}^*X)$. This contradicts with the fact that A is near-standard. □

Thus, we know that for every $x \in {}^*X$ and every near-standard $A \in {}^*\mathcal{B}[X]$ we have

$$((\forall n \in \mathbb{N})({}^*d(x, {}^*K_n) > r')) \implies ((\forall t \in T)({}^*g(x, t, A) < \epsilon)). \tag{9.14}$$

Pick an infinite $r_\infty \in {}^*\mathbb{R}_{>0}$. Let a be any standard element in X and let $r = 2\max\{r_{\delta t}, r', r_\infty\}$. We claim that $\overline{U}(a, r)$ satisfies the two conditions of this lemma. By the choice of r, we know that ${}^*d(x, {}^*X \setminus \overline{U}(a, r)) > r_{\delta t}$ for all $x \in \bigcup_{n\in\mathbb{N}} {}^*K_n$. As $\bigcup_{n\in\mathbb{N}} {}^*K_n = \mathrm{NS}({}^*X)$, by Eq. (9.9), we have

$$(\forall x \in \mathrm{NS}({}^*X))({}^*g(x, \delta t, {}^*X \setminus \overline{U}(a, r)) < \epsilon). \tag{9.15}$$

Fix any $y \in {}^*X \setminus \overline{U}(a, r)$ and any near-standard $A \in {}^*\mathcal{B}[X]$. By the choice of r, we know that ${}^*d(y, {}^*K_n) > r'$ for all $n \in \mathbb{N}$. Thus, by Eq. (9.14) we have ${}^*g(y, t, A) < \epsilon$ for all $t \in T$. As our choices of y and A are arbitrary, we have the desired result. □

For the particular ϵ_0 fixed above, we can find a standard $a_0 \in {}^*X$ and some positive infinite $r_1 \in {}^*\mathbb{R}$ such that the open ball $U(a_0, r_1)$ satisfies the conditions in Theorem 9.8. We fix a_0 and r_1 for the remainder of this section.

Lemma 9.11. *Suppose (DT) holds. There exists a positive infinite $r_0 > 2r_1$ such that*

$$(\forall y \in \overline{U}(a_0, 2r_1))({}^*g(y, \delta t, {}^*X \setminus \overline{U}(a_0, r_0)) < \epsilon_0). \tag{9.16}$$

Proof. By the transfer of the Heine-Borel condition, $\overline{U}(a_0, 2r_1)$ is a *compact set. Then the proof follows easily from the transfer of condition 1 of (DT). Note that we can always pick r_0 to be bigger than $2r_1$. □

We will see how do we use Lemma 9.11 in Theorem 9.20. We now fix r_0 for the remainder of this section. An immediate consequence of Theorem 9.8 and Lemma 9.11 is:

Lemma 9.12. *Suppose (DT) holds. For any $x \in X$, any $t \in T$, any near-standard internal set $A \subset {}^*X$ we have ${}^*f_x^{(t)}({}^*X \setminus \overline{U}(a_0, 2r_0), A) < 2\epsilon_0$.*

Proof. Fix a $x \in {}^*X$, a near-standard internal set A and some $t \in T$. By Theorem 9.8, we know that $(\forall y \in {}^*X \setminus \overline{U}(a_0, 2r_0))({}^*g(y, t, A) < \epsilon_0)$. This means that for any $y_1, y_2 \in {}^*X \setminus \overline{U}(a_0, 2r_0)$ we have $|{}^*g(y_1, t, A) - {}^*g(y_2, t, A)| < \epsilon_0$. By Lemma 8.5, we know that for any $y \in {}^*X \setminus \overline{U}(a_0, 2r_0)$ we have

$$|{}^*g(y, t, A) - {}^*f_x^{(t)}({}^*X \setminus \overline{U}(a_0, 2r_0), A)| < \epsilon_0 \tag{9.17}$$

which then implies that ${}^*f_x^{(t)}({}^*X \setminus \overline{U}(a_0, 2r_0), A) < 2\epsilon_0$. □

Thus, our hyperfinite state space S is a $(\delta_0, 2r_0)$-hyperfinite representation of ${}^*X$ such that $\bigcup_{s \in S} B(s) = \overline{U}(a_0, 2r_0)$. We now choose an appropriate δ_0 to partition $\overline{U}(a_0, 2r_0)$ into hyperfinitely pieces. We use the strong Feller condition to control the diameter of each $B(s)$ for $s \in S$. For reader's convenience, we restate the strong Feller condition below:

Condition SF. The Markov chain$\{X_t\}_{t\geq 0}$ is said to be *strong Feller* if for every $t>0$, every $x\in X$ and every $\epsilon>0$ there exists $\delta>0$ such that

$$(\forall x\in X)(\exists \delta>0)((\forall y\in X)(d(x,y)<\delta \implies (\forall A\in \mathcal{B}[X])|P_y^{(t)}(A)-P_x^{(t)}(A)|<\epsilon)). \tag{9.18}$$

Note that this δ depends on ϵ, t and x. View the transition probability as the function g and by the transfer principle, we have for every $t\in T\setminus\{0\}$,every $\epsilon\in {}^*\mathbb{R}^+$ and every $x\in {}^*X$ there exists $\delta\in {}^*\mathbb{R}^+$ such that:

$$((\forall y\in {}^*X)(d(x,y)<\delta \implies (\forall A\in {}^*\mathcal{B}[X])|{}^*g(y,t,A)-{}^*g(x,t,A)|<\epsilon)). \tag{9.19}$$

We can then show that the total variation distance between transition probabilities for Markov processes is non-increasing. The following lemma is a "standard counterpart" of Lemma 7.25. The proof is identical to Lemma 8.7 hence omitted.

Lemma 9.13. *Consider a standard Markov process with transition probability measure $P_x^{(t)}(\cdot)$, then for every $\epsilon\in\mathbb{R}^+$, every $x_1,x_2\in X$, every $t_1,t_2\in\mathbb{R}^+$ and every $A\in\mathcal{B}[X]$ we have*

$$(|P_{x_1}^{(t_1)}(A)-P_{x_2}^{(t_1)}(A)|\leq\epsilon \implies |P_{x_1}^{(t_1+t_2)}(A)-P_{x_2}^{(t_1+t_2)}(A)|\leq\epsilon). \tag{9.20}$$

Apply the transfer principle to the above lemma and restrict out time line to T, we know that for every $\epsilon\in {}^*\mathbb{R}^+$, every $x_1,x_2\in {}^*X$,every $t_1,t_2\in T^+$ and every $A\in {}^*\mathcal{B}[X]$ we have:

$$((|{}^*P_{x_1}^{(t_1)}(A)-{}^*P_{x_2}^{(t_1)}(A)|\leq\epsilon) \implies (|{}^*P_{x_1}^{(t_1+t_2)}(A)-{}^*P_{x_2}^{(t_1+t_2)}(A)|\leq\epsilon)). \tag{9.21}$$

where ${}^*P_x^{(t)}(A)={}^*g(x,t,A)$.

(SF) ensures the uniform continuity of the transition probability $g(x,t,A)$ with respect to x as is shown by the following lemma.

Lemma 9.14. *Suppose (SF) holds. There exists $\delta_0 \in {}^*\mathbb{R}^+$ such that for any $x_1, x_2 \in \overline{U}(a_0, 2r_0)$ with $|x_1 - x_2| < \delta_0$ we have $|{}^*g(x_1, t, A) - {}^*g(x_2, t, A)| < \epsilon_0$ for all $A \in {}^*\mathcal{B}[X]$ and all $t \in T^+$.*

Proof. By the transfer of strong Feller, for every $x \in \overline{U}(a_0, 2r_0)$ there exists $\delta_x \in {}^*\mathbb{R}^+$ such that:

$$(\forall y \in {}^*X)(d(x,y) < \delta_x \implies (\forall A \in {}^*\mathcal{B}[X])|{}^*g(x, \delta t, A) - {}^*g(y, \delta t, A)| < \frac{\epsilon_0}{2}). \tag{9.22}$$

The internal collection $\mathcal{L} = \{U(x, \frac{\delta_x}{2}) : x \in \overline{U}(a_0, 2r_0)\}$ of open balls forms an open cover of $\overline{U}(a_0, 2r_0)$. By the transfer of Heine-Borel condition, we know that $\overline{U}(a_0, 2r_0)$ is *compact hence there exists a hyperfinite subset of the cover $\mathcal{L}$ that covers $\overline{U}(a_0, 2r_0)$. Denote this hyperfinite subcover by $\mathcal{F} = \{B(x_i, \frac{\delta_{x_i}}{2}) : i \leq N\}$ for some $N \in {}^*\mathbb{N}$. The set $\Delta = \{\frac{\delta_{x_i}}{2} : i \leq N\}$ is a hyperfinite set thus there exists a minimum element of Δ. Let $\delta_0 = \min\{\frac{\delta_{x_i}}{2} : i \leq N\}$.

Pick any $x, y \in \overline{U}(a_0, 2r_0)$ with $d(x, y) < \delta_0$. We have $x \in U(x_i, \frac{\delta_{x_i}}{2})$ for some $i \leq N$. Then we have ${}^*d(y, x_i) \leq {}^*d(y, x) + {}^*d(x, x_i) \leq \delta_{x_i}$. Thus both x, y are in $U(x_i, \delta_{x_i})$. This means that $(\forall A \in {}^*\mathcal{B}[X])(|{}^*g(x, \delta t, A) - {}^*g(y, \delta t, A)| < \epsilon_0)$. By Eq. (9.21), we know that $(\forall A \in {}^*\mathcal{B}[X])(\forall t \in T|{}^*g(x, t, A) - {}^*g(y, t, A)| < \epsilon_0)$, completing the proof. □

Now we have determined a_0,r_0 and δ_0. We now construct a $(\delta_0, 2r_0)$-hyperfinite representation set S with $\bigcup_{s \in S} B(s) = \overline{U}(a_0, 2r_0)$. The following lemma is an immediate consequence.

Theorem 9.15. *Suppose (SF) holds. Let S be a $(\delta_0, 2r_0)$-hyperfinite representation with $\bigcup_{s \in S} B(s) = \overline{U}(a_0, 2r_0)$. For any $s \in S$, any $x_1, x_2 \in B(s)$, any $A \in {}^*\mathcal{B}[X]$ and any $t \in T^+$ we have $|{}^*g(x_1, t, A) - {}^*g(x_2, t, A)| < \epsilon_0$*

An immediate consequence of the above lemma is:

Lemma 9.16. *Suppose (SF) holds. Let S be a $(\delta_0, 2r_0)$-hyperfinite representation with $\bigcup_{s\in S} B(s) = \overline{U}(a_0, 2r_0)$. For for any $s \in S$, any $y \in B(s)$, any $x \in {}^*X$, any $A \in {}^*\mathcal{B}[X]$ and any $t \in T^+$ we have $|{}^*g(y,t,A) - {}^*f_x^{(t)}(B(s),A)| < \epsilon_0$.*

Proof. First recall that we use ${}^*f_x^{(t)}(B(s),A)$ to denote ${}^*f_x^{(\delta t,t)}(B(s),A)$. This lemma then follows easily by applying Lemma 8.4 to Theorem 9.15. □

For the remainder of this paper we shall fix our hyperfinite state space S to be a $(\delta_0, 2r_0)$-hyperfinite representation of ${}^*X$ with $\bigcup_{s\in S} B(s) = \overline{U}(a_0, 2r_0)$. That is:

(1) $\bigcup_{s\in S} B(s) = \overline{U}(a_0, 2r_0)$.
(2) $\{B(s) : s \in S\}$ is a mutually disjoint collection of *Borel sets with diameters no greater than δ_0.

This S will be the state space of our hyperfinite Markov process which is a hyperfinite representation of our standard Markov process $\{X_t\}_{t\geq 0}$.

9.2. **Construction of Hyperfinite Markov Processess.** In the last section, we have constructed the hyperfinite state space S to be a $(\delta_0, 2r_0)$-hyperfinite representation of ${}^*X$. In this section, we will construct a hyperfinite Markov $\{X'_t\}_{t\in T}$ process on S which is hyperfinite representation of our standard Markov process $\{X_t\}_{t\geq 0}$.

The following definition is very similar to Definition 8.9.

Definition 9.17. Let $g'(x,\delta t,A) : \bigcup_{s\in S} B(s) \times {}^*\mathcal{B}[X] \to {}^*[0,1]$ be given by:

$$g'(x,\delta t,A) = {}^*g(x,\delta t, A \cap \bigcup_{s\in S} B(s)) + \delta_x(A){}^*g(x,\delta t, {}^*X \setminus \bigcup_{s\in S} B(s)). \tag{9.23}$$

where $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if otherwise.

For any $i,j \in S$, let $G_i^{(\delta t)}(\{j\}) = g'(i,\delta t, B(j))$ and let $G_i^{(\delta t)}(A) = \sum_{j\in A} G_i^{(\delta t)}(\{j\})$ for all internal $A \subset S$. For any internal $A \subset S$, $G_i^{(0)}(A) = 1$ if $i \in A$ and $G_i^{0}(A) = 0$ otherwise.

The following two lemmas are identical to Lemmas 8.10, 8.12 and 8.13 after substituting δt for 1. Likewise, $G_i^{(t)}(\cdot)$ denotes the t-step transition probability of $\{X'_t\}_{t\in T}$ which is purely generated from $\{G_i^{(\delta t)}(\cdot)\}_{i\in S}$.

Lemma 9.18. *Let $\mathcal{B}[\bigcup_{s\in S} B(s)] = \{A\cap \bigcup_{s\in S} B(s) : A \in {}^*\mathcal{B}[X]\}$. Then for any $x \in \bigcup_{s\in S} B(s)$ we have $(\bigcup_{s\in S} B(s), \mathcal{B}[\bigcup_{s\in S} B(s)], g'(x,\delta t,.))$ is an internal probability space.*

Lemma 9.19. *For any $i \in S$ and any $t \in T$, we know that $G_i^{(t)}(\cdot)$ is an internal probability measure on $(S,\mathcal{I}(S))$.*

For any $i \in S$ and any $t \in T$ we shall use $\overline{G}_i^{(t)}(\cdot)$ to denote the Loeb extension of the internal probability measure $G_i^{(t)}(\cdot)$ on $(S,\mathcal{I}(S))$.

In order for the hyperfinite Markov chain $\{X'_t\}_{t\in T}$ to be a good representation of $\{X_t\}_{t\geq 0}$, one of the key properties which needs to be shown is that the internal transition probability of $\{X'_t\}_{t\in T}$ agrees with the transition probability of $\{X_t\}_{t\geq 0}$ up to an infinitesimal. The following technical result is a key step towards showing this property (recall that ϵ_0 is a positive infinitesimal such that $\epsilon_0\frac{t}{\delta t}\approx 0$ for all $t\in T$). This result is similar to Theorem 8.14 but is more complicated.

Theorem 9.20. *Suppose (DT) and (SF) hold. Then for any $t \in T$, any $x \in S$ and any near-standard $A \in {}^*\mathcal{B}[X]$, we have*

$$|{}^*g(x,t,\bigcup_{s'\in A\cap S} B(s')) - G_x^{(t)}(A\cap S)| \leq \epsilon_0 + 5\epsilon_0\frac{t-\delta t}{\delta t}. \tag{9.24}$$

In particular, we have $|{}^*g(x,t,\bigcup_{s'\in A\cap S} B(s')) - G_x^{(t)}(A\cap S)| \approx 0$ for all $t \in T$, all $x \in S$ and all near-standard $A \in {}^*\mathcal{B}[X]$.

Proof. We will prove the result by internal induction on $t \in T$.

We first prove the theorem for $t = 0$. As $x \in S$, it is easy to see that $x \in \bigcup_{s'\in A\cap S} B(s')$ if and only if $x \in A\cap S$. Hence ${}^*g(x,0,\bigcup_{s'\in A\cap S} B(s')) = G_x^{(0)}(A\cap S)$

We now show the case where $t = \delta t$. Pick any near-standard set $A \in {}^*\mathcal{B}[X]$ and any $x \in S$. By definition, we have:

$$G_x^{(\delta t)}(A \cap S) = g'(x, \delta t, \bigcup_{s' \in A \cap S} B(s')) \tag{9.25}$$

$$= {}^*g(x, \delta t, \bigcup_{s' \in A \cap S} B(s')) + \delta_x(\bigcup_{s' \in A \cap S} B(s')) {}^*g(x, \delta t, {}^*X \setminus \bigcup_{s \in S} B(s)). \tag{9.26}$$

For any $x \in \bigcup_{s' \in A \cap S} B(s')$, by Theorem 9.8 and the fact that $\bigcup_{s' \in A \cap S} B(s')$ is near-standard, we have ${}^*g(x, \delta t, {}^*X \setminus \bigcup_{s \in S} B(s)) < \epsilon_0$ since ${}^*d(x, {}^*X \setminus \bigcup_{s \in S} B(s)) > r_0$. Thus we have $|{}^*g(x, \delta t, \bigcup_{s' \in A \cap S} B(s')) - G_x^{(\delta t)}(A \cap S)| < \epsilon_0$.

We now prove the induction case. Assume the statement is true for some $t \in T$. We now show that it is true for $t + \delta t$. Fix a near-standard $A \in {}^*\mathcal{B}[X]$ and any $x \in S$. We know that:

${}^*g(x, t + \delta t, \bigcup_{s' \in A \cap S} B(s')) = \sum_{s \in S} {}^*g(x, \delta t, B(s)) {}^*f_x^{(t)}(B(s), \bigcup_{s' \in A \cap S} B(s')) + {}^*g(x, \delta t, {}^*X \setminus \bigcup_{s \in S} B(s)) {}^*f_x^{(t)}({}^*X \setminus \bigcup_{s \in S} B(s), \bigcup_{s' \in A \cap S} B(s'))$.

Consider ${}^*g(x, \delta t, {}^*X \setminus \bigcup_{s \in S} B(s)) {}^*f_x^{(t)}({}^*X \setminus \bigcup_{s \in S} B(s), \bigcup_{s' \in A \cap S} B(s'))$. By Lemma 9.12, we have ${}^*f_x^{(t)}({}^*X \setminus \bigcup_{s \in S} B(s), \bigcup_{s' \in A \cap S} B(s')) < 2\epsilon_0$. Thus we conclude that:

$$|{}^*g(x, t + \delta t, \bigcup_{s' \in A \cap S} B(s')) - \sum_{s \in S} {}^*g(x, \delta t, B(s)) {}^*f_x^{(t)}(B(s), \bigcup_{s' \in A \cap S} B(s'))| < 2\epsilon_0. \tag{9.27}$$

By the construction of our hyperfinite representation S and Lemma 9.16, we know that for any $s \in S$ we have $|{}^*g(s, t, \bigcup_{s' \in A \cap S} B(s')) - {}^*f_x^{(t)}(B(s), \bigcup_{s' \in A \cap S} B(s'))| < \epsilon_0$. By the transfer of Lemma 3.20, we have that:

$$|\sum_{s \in S} {}^*g(x, \delta t, B(s)) {}^*f_x^{(t)}(B(s), \bigcup_{s' \in A \cap S} B(s')) - \sum_{s \in S} {}^*g(x, \delta t, B(s)) {}^*g(s, t, \bigcup_{s' \in A \cap S} B(s'))| < \epsilon_0. \tag{9.28}$$

Let us now consider the formulas $\sum_{s\in S} {}^*g(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S} B(s'))$ and $\sum_{s\in S} g'(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S} B(s'))$. There exists an unique $s_0 \in S$ such that $x \in B(s_0)$. This means that ${}^*g(x,\delta t,B(s))$ is the same as $g'(x,\delta t,B(s))$ for all $s \neq s_0$. Thus we have:

$$|\sum_{s\in S} {}^*g(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S} B(s')) - \sum_{s\in S} g'(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S} B(s'))| \tag{9.29}$$

$$= |{}^*g(x,\delta t,B(s_0)) - g'(x,\delta t,B(s_0))|{}^*g(s_0,t,\bigcup_{s'\in A\cap S} B(s')). \tag{9.30}$$

Recall the properties of r_1 constructed after Theorem 9.8. If ${}^*d(s_0,y) > r_1$ for all near-standard $y \in \mathrm{NS}({}^*X)$, by Theorem 9.8, we have ${}^*g(s_0,t,\bigcup_{s'\in A\cap S} B(s')) < \epsilon_0$. This implies that

$$|{}^*g(s_0,\delta t,B(s)) - g'(s_0,\delta t,B(s))|{}^*g(s_0,t,\bigcup_{s'\in A\cap S} B(s')) < \epsilon_0. \tag{9.31}$$

If there exists some $x_0 \in \mathrm{NS}({}^*X)$ such that ${}^*d(s_0,x_0) < r_1$ then $s_0 \in \overline{U}(a_0,2r_1)$. By the definition of g' and Lemma 9.11, we know that ${}^*g(s_0,\delta t,{}^*X \setminus \bigcup_{s\in S} B(s)) < \epsilon_0$. As $x \in B(s_0)$, by Theorem 9.15, we know that

$$|{}^*g(x,\delta t,B(s_0)) - g'(x,\delta t,B(s_0))| = |{}^*g(x,\delta t,{}^*X \setminus \bigcup_{s\in S} B(s))| < 2\epsilon_0. \tag{9.32}$$

To conclude we have:

$$|\sum_{s\in S} {}^*g(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S} B(s')) - \sum_{s\in S} g'(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S} B(s'))| < 2\epsilon_0. \tag{9.33}$$

Finally by induction hypothesis and the transfer of Lemma 3.20 we know that:

$$|\sum_{s\in S} g'(x,\delta t,B(s))^{*}g(s,t,\bigcup_{s'\in A\cap S} B(s')) - G_x^{(t+\delta t)}(A\cap S)| \tag{9.34}$$

$$= |\sum_{s\in S} g'(x,\delta t,B(s))^{*}g(s,t,\bigcup_{s'\in A\cap S} B(s')) - \sum_{s\in S} g'(x,\delta t,B(s))G_s^{(t)}(A\cap S)| \tag{9.35}$$

$$\leq |^{*}g(s,t,\bigcup_{s'\in A\cap S} B(s')) - G_s^{(t)}(A\cap S)| \leq \epsilon_0 + 5\epsilon_0\frac{t-\delta t}{\delta t}. \tag{9.36}$$

Thus by Eq. (9.27),Eq. (9.28),Eq. (9.33) and Eq. (9.36) we conclude that

$$|^{*}g(x,t+\delta t,\bigcup_{s'\in A\cap S} B(s')) - G_x^{(t+\delta t)}(A\cap S)| \tag{9.37}$$

$$\leq \epsilon_0 + 4\epsilon_0\frac{t-\delta t}{\delta t} + 5\epsilon_0 = \epsilon_0 + 5\epsilon_0\frac{t}{\delta t}. \tag{9.38}$$

As all the parameters in this statement are internal, by internal induction principle, we have shown the statement. As $\epsilon_0\frac{t}{\delta t} \approx 0$ for all $t \in T$, in particular, we have $|^{*}g(x,t,\bigcup_{s'\in A\cap S} B(s')) - G_x^{(t)}(A\cap S)| \approx 0$ for all $t \in T$, all $x \in S$ and all near-standard $A \in {}^{*}\mathcal{B}[X]$. $\square$

As the state space X is σ-compact, by Lemma 5.5 and Theorem 5.9, we know that $\mathsf{st}^{-1}(A)$ is universally Loeb measurable for $A \in \mathcal{B}[X]$. We now extend Theorem 9.20 to establish the relationship between $\overline{^{*}g}$ and $\overline{G}$.

Theorem 9.21. *For any $x \in \bigcup_{s\in S} B(s)$ let s_x denote the unique element in S such that $x \in B(s_x)$. Then, under (DT) and (SF), for any $E \in \mathcal{B}[X]$ and any $t \in T$, we have $\overline{^{*}g}(x,t,\mathsf{st}^{-1}(E)) = \overline{G}_{s_x}^{(t)}(\mathsf{st}^{-1}(E)\cap S)$ for any $x \in {}^{*}X$.*

Proof. When $t = 0$, $\overline{^{*}g}(x,0,\mathsf{st}^{-1}(E))$ is 1 if $x \in \mathsf{st}^{-1}(E)$ and is 0 otherwise. Note that $x \in \mathsf{st}^{-1}(E)$ if and only if $s_x \in \mathsf{st}^{-1}(E)\cap S$. Hence $\overline{^{*}g}(x,t,\mathsf{st}^{-1}(E)) = \overline{G}_{s_x}^{(t)}(\mathsf{st}^{-1}(E)\cap S)$.

We now prove the case for $t > 0$. By the transfer principle, we know that for any $x \in {}^{*}X$ and any $t \in T$ we have ${}^{*}g(x,t,.)$ is an internal probability measure. By the construction of Loeb measures (Eq. (6.19)), for $t > 0$ we have

$$\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E)) = \sup\{\overline{{}^*g}(x,t,\bigcup_{s\in A_i} B(s)) : A_i \subset \mathsf{st}^{-1}(E)\cap S, A_i \in \mathcal{I}(S)\}. \tag{9.39}$$

As the distance between x and s_x is less than δ_0, by Theorem 9.15 we know that $|{}^*g(x,t,\bigcup_{s\in A_i} B(s)) - {}^*g(s_x,t,\bigcup_{s\in A_i} B(s))| < \epsilon_0$. By Theorem 9.20, we know that $|{}^*g(s_x,t,\bigcup_{s\in A_i} B(s)) - G^{(t)}_{s_x}(A_i)| \approx 0$ as A_i is a near-standard internal set. Thus we know that $\overline{{}^*g}(x,t,\bigcup_{s\in A_i} B(s)) = \overline{G}^{(t)}_{s_x}(A_i)$. Thus we know that

$$\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E)) = \sup\{\overline{G}_{s_x}(A_i) : A_i \subset \mathsf{st}^{-1}(E)\cap S, A_i \in \mathcal{I}(S)\} = \overline{G}^{(t)}_{s_x}(\mathsf{st}^{-1}(E)\cap S) \tag{9.40}$$

finishing the proof. □

One of the desired properties for a hyperfinite Markov chain is strong regularity. Recall from Definition 7.6 that a hyperfinite Markov chain is strong regular if for any $A \in \mathcal{I}(S)$, any non-infinitesimal $t \in T$ and any $i \approx j \in \mathrm{NS}(S)$ we have $G^{(t)}_i(A) \approx G^{(t)}_j(A)$. We now show that $\{X'_t\}$ satisfies strong regularity. We first prove the following "locally continuous" property for ${}^*g$.

Lemma 9.22. *Suppose (SF) holds. For any two near-standard $x_1 \approx x_2$ from ${}^*X$,any $t \in {}^*\mathbb{R}^+$ that is not infinitesimal and any $A \in {}^*\mathcal{B}[X]$ we have ${}^*g(x_1,t,A) \approx {}^*g(x_2,t,A)$.*

Proof. Fix two near-standard x_1, x_2 from ${}^*X$. Let $x_0 = \mathsf{st}(x_1) = \mathsf{st}(x_2)$. Fix some $t_0 \in {}^*\mathbb{R}^+$ that is not infinitesimal and also fix some positive $\epsilon \in \mathbb{R}$. Pick some standard $t' \in \mathbb{R}^+$ with $t' \leq t_0$. By strong Feller we can pick a $\delta \in \mathbb{R}^+$ such that $(\forall y \in X)(|y - x_0| < \delta \implies ((\forall A \in \mathcal{B}[X])|g(y,t',A) - g(x_0,t',A)| < \epsilon))$. By the transfer principle and the fact that $x_1 \approx x_2 \approx x_0$ we know that

$$(\forall A \in {}^*\mathcal{B}[X])(|{}^*g(x_1,t',A) - {}^*g(x_2,t',A)| < \epsilon). \tag{9.41}$$

As $t' \leq t_0$, by Eq. (9.21), we know that $|{}^*g(x_1,t_0,A) - {}^*g(x_2,t_0,A)| < \epsilon$ for all $A \in {}^*\mathcal{B}[X]$. Since our choice of ϵ is arbitrary, we can conclude that ${}^*g(x_1,t_0,A) \approx {}^*g(x_2,t_0,A)$ for all $A \in {}^*\mathcal{B}[X]$. □

An immediate consequence of this lemma is the following:

Lemma 9.23. *Suppose (SF) holds. For any two near-standard $x_1 \approx x_2$ from ${}^*X$,any $t \in {}^*\mathbb{R}^+$ that is not infinitesimal and any universally Loeb measurable set A we have $\overline{{}^*g}(x_1,t,A) = \overline{{}^*g}(x_2,t,A)$.*

Next we show that the internal measure ${}^*g(x,t,.)$ concentrates on the near-standard part of ${}^*X$ for near-standard x and standard t.

Lemma 9.24. *Suppose (SF) holds. For any Borel set E, any $x \in \mathrm{NS}({}^*X)$ and any $t \in \mathbb{R}^+$ we have ${}^*g(x,t,{}^*E) \approx \overline{{}^*g}(x,t,\mathsf{st}^{-1}(E))$.*

Proof. Fix any $x \in \mathrm{NS}({}^*X)$ and any $t \in \mathbb{R}^+$. Let $x_0 = \mathsf{st}(x)$. Fix any ϵ, as the probability measure $P^{(t)}_{x_0}(\cdot)$ is Radon, we can find K compact, U open with $K \subset E \subset U$ and $P^{(t)}_{x_0}(U) - P^{(t)}_{x_0}(K) < \epsilon/2$. By the transfer principle, we know that ${}^*g(x_0,t,{}^*U) - {}^*g(x_0,t,{}^*K) < \epsilon/2$. By Lemma 9.22, we know that ${}^*g(x_0,t,{}^*U) \approx {}^*g(x,t,{}^*U)$ and ${}^*g(x_0,t,{}^*K) \approx {}^*g(x,t,{}^*K)$. Hence we know that ${}^*g(x,t,{}^*U) - {}^*g(x,t,{}^*K) < \epsilon$. Note that ${}^*K \subset \mathsf{st}^{-1}(K) \subset \mathsf{st}^{-1}(E) \subset \mathsf{st}^{-1}(U) \subset {}^*U$. Both ${}^*g(x,t,{}^*E)$ and $\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E))$ lie between ${}^*g(x,t,{}^*U)$ and ${}^*g(x,t,{}^*K)$. So $|{}^*g(x,t,{}^*E) - \overline{{}^*g}(x,t,\mathsf{st}^{-1}(E))| < \epsilon$. This is true for any ϵ and hence ${}^*g(x,t,{}^*E) \approx \overline{{}^*g}(x,t,\mathsf{st}^{-1}(E))$. □

We are now at the place to establish that $\{X'_t\}$ is strong regular. Note that the time line $T = \{0, \delta t, ..., K\}$ contains all the rational numbers but none of the irrational numbers.

Theorem 9.25. *Suppose (DT) and (SF) hold. For any two near-standard $s_1 \approx s_2$ from S, any $t \in T$ that is not infinitesimal and any $A \in \mathcal{I}(S)$ we have $G^{(t)}_{s_1}(A) \approx G^{(t)}_{s_2}(A)$.*

Proof. Fix any two near-standard $s_1 \approx s_2 \in S$ and any non-infinitesimal $t \in T$. Pick a non-zero $t' \in \mathbb{Q}$ such that $t' \leq t$. By Theorem 9.21, we know that $\overline{{}^*g}(x, t, \mathsf{st}^{-1}(E)) = \overline{G}^{(t)}_{s_x}(\mathsf{st}^{-1}(E) \cap S)$. Fix any $\epsilon \in \mathbb{R}^+$ and any $A \in \mathcal{I}(S)$, we now consider $G^{t'}_{s_1}(A)$ and $G^{t'}_{s_2}(A)$. By Lemma 9.24, we can find a near-standard $A_i \in \mathcal{I}(S)$ such that $|G^{(t')}_{s_1}(A) - G^{(t')}_{s_1}(A_i)| < \frac{\epsilon}{3}$ and $|G^{(t')}_{s_2}(A) - G^{(t')}_{s_2}(A_i)| < \frac{\epsilon}{3}$. As A_i is near-standard, by Theorem 9.20, we know that $G^{(t')}_{s_1}(A_i) \approx {}^*g(s_1, t', \bigcup_{s \in A_i \cap S} B(s))$ and $G^{(t')}_{s_2}(A_i) \approx {}^*g(s_2, t', \bigcup_{s \in A_i \cap S} B(s))$. Moreover,by Lemma 9.22, we know that $|{}^*g(s_1, t', \bigcup_{s \in A_i \cap S} B(s)) - {}^*g(s_2, t', \bigcup_{s \in A_i \cap S} B(s))| \approx 0$. Hence we know that $|G^{(t')}_{s_1}(A_i) - G^{(t')}_{s_2}(A_i)| \approx 0$. Thus we have $|G^{(t')}_{s_1}(A) - G^{(t')}_{s_2}(A)| < \epsilon$. As our choice ϵ is arbitrary, we know that $|G^{(t')}_{s_1}(A) - G^{(t')}_{s_2}(A)| \approx 0$. Hence we know that $\| G^{(t')}_{s_1}(\cdot) - G^{(t')}_{s_1}(\cdot) \| \approx 0$ where $\| G^{(t')}_{s_1}(\cdot) - G^{(t')}_{s_1}(\cdot) \|$ denotes the total variation distance between $G^{(t')}_{s_1}$ and $G^{(t')}_{s_1}$. By Lemma 7.25, we know that $\| G^{(t)}_{s_1}(\cdot) - G^{(t)}_{s_1}(\cdot) \| \approx 0$ hence finishes the proof. $\square$

We are now able to establish to following theorem which is an immediate consequence of Theorem 9.25.

Lemma 9.26. *Suppose (DT) and (SF) hold. For any two near-standard $s_1 \approx s_2$ from S, any $t \in T$ that is not infinitesimal and any universally Loeb measurable set A we have $\overline{G}^{(t)}_{s_1}(A) = \overline{G}^{(t)}_{s_2}(A)$.*

There exists a natural link between the transition probability g of $\{X_t\}$ and its nonstandard extension ${}^*g$. We have already established a strong link between ${}^*g$ and the internal transition probability G of $\{X'_t\}$. We have also established the " local continuity" of ${}^*g$. We are now at the place to establish the relationship between the internal transition probability of $\{X'_t\}$ and the transition probability of $\{X_t\}$.

Theorem 9.27. *Suppose (DT) and (SF) hold. For any $s \in \mathrm{NS}(S)$, any non-negative $t \in \mathbb{Q}$ and any $E \in \mathcal{B}[X]$, we have $P^{(t)}_{\mathsf{st}(s)}(E) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(E) \cap S)$.*

Proof. We first prove the theorem when $t = 0$. Fix any $s \in \mathrm{NS}(S)$ and any $E \in \mathcal{B}[X]$. We know that $P^{(t)}_{\mathsf{st}(s)}(E) = 1$ if $\mathsf{st}(s) \in E$ and $P^{(t)}_{\mathsf{st}(s)}(E) = 0$ otherwise. For any

$x \in S$ and $A \in \mathcal{I}(S)$, note that $G_x^{(0)}(A) = 1$ if and only if $x \in A$ and $G_x^{(0)}(A) = 0$ otherwise. This establishes the theorem for $t = 0$.

We now prove the result for positive $t \in \mathbb{Q}$. Fix any $s \in \mathrm{NS}(S)$, any positive $t \in \mathbb{Q}$ and any $E \in \mathcal{B}[X]$. By Lemmas 9.22 and 9.24 and Theorem 9.21, we know that

$$g(\mathsf{st}(s), t, E) = {}^*g(\mathsf{st}(s), t, {}^*E) \approx {}^*g(s, t, {}^*E) \approx \overline{{}^*g}(s, t, \mathsf{st}^{-1}(E)) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S). \tag{9.42}$$

Thus we have for any $s \in \mathrm{NS}(S)$, any non-zero $t \in \mathbb{Q}^+$ and any $E \in \mathcal{B}[X]$: $P_{\mathsf{st}(s)}^{(t))}(E) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S)$. $\square$

It is desirable to extend Theorem 9.27 to all non-negative $t \in \mathbb{R}$. In order to do this, we need some continuity condition of the transition probability with respect to time.

Condition OC. The Markov chain $\{X_t\}$ is said to be *continuous in time* if there exists a basis $\mathscr{B}_0$ such that $g(x, t, U)$ is a continuous function of $t > 0$ for every $x \in X$ and every U which is a finite intersection of elements from $\mathscr{B}_0$.

It is easy to see that $g(x, t, U)$ is continuous function of $t > 0$ for every $x \in X$ and every U which is a finite union of elements from $\mathscr{B}_0$. Note that (OC) is weaker than assuming $g(x, t, U)$ is a continuous function of $t > 0$ for every $x \in X$ and every open set U. We establish this by the following counterexample.

Example 9.28. Let μ_n be the uniform probability measure on the set $\{\frac{1}{n}, \dots, 1\}$ for every $n \geq 1$. Let μ be the Lebesgue measure on $[0, 1]$. It is easy to see that $\mu_n(I)$ converges to $\mu(A)$ for every open interval I. However, it is not the case that $\mu_n(U)$ converges to $\mu(U)$ for every open set. To see this, let U be an open set containing the the set of rational numbers $\mathbb{Q}$ such that $\mu(Q) \leq \frac{1}{2}$. We can find such U since $\mu(\mathbb{Q}) = 0$. We know $\lim_{n\to\infty} \mu_n(U) = 1$ which does not equal to $\mu(U) = \frac{1}{2}$.

Let us fix a basis $\mathscr{B}_0$ satisfying the conditions in (OC) for the remainder of this section.

Lemma 9.29. *Suppose (SF) and (OC) hold. For any near-standard $x_1 \approx x_2$, any non-infinitesimal $t_1, t_2 \in \mathrm{NS}({}^*\mathbb{R}^+)$ such that $t_1 \approx t_2$ and any U which is a finite intersection of elements in $\mathscr{B}_0$, we have ${}^*g(x_1, t_1, {}^*U) \approx {}^*g(x_2, t_2, {}^*U)$.*

Proof. Fix near-standard $x_1 \approx x_2 \in {}^*X$, some $U \subset X$ which is a finite intersection of elements in $\mathscr{B}_0$ and some $\epsilon \in \mathbb{R}^+$. Also fix two non-infinitesimal $t_1, t_2 \in \mathrm{NS}({}^*\mathbb{R}^+)$ such that $t_1 \approx t_2$. Let $x_0 \in X$ and $t_0 \in \mathbb{R}^+$ denote the standard parts of x_1, x_2 and t_1, t_2,respectively. Note that $t_0 > 0$.

As U is a finite intersection of elements from $\mathscr{B}_0$, by (OC), there exists $\delta \in \mathbb{R}^+$ such that

$$(\forall t \in \mathbb{R}^+)((|t - t_0| < \delta) \implies (|g(x_0, t, U) - g(x_0, t_0, U)| < \epsilon)). \tag{9.43}$$

By the transfer principle, we know that

$$(\forall t \in {}^*\mathbb{R}^+)((|t - t_0| < \delta) \implies (|{}^*g(x_0, t, {}^*U) - {}^*g(x_0, t_0, {}^*U)| < \epsilon)). \tag{9.44}$$

Since ϵ is arbitrary and $\mathsf{st}(t_1) = \mathsf{st}(t_2) = t_0$, we have

$${}^*g(x_0, t_1, {}^*U) \approx {}^*g(x_0, t_0, {}^*U) \approx {}^*g(x_0, t_2, {}^*U). \tag{9.45}$$

By Lemma 9.22, we then have

$${}^*g(x_1, t_1, {}^*U) \approx {}^*g(x_0, t_1, {}^*U) \approx {}^*g(x_0, t_2, {}^*U) \approx {}^*g(x_2, t_2, {}^*U), \tag{9.46}$$

completing the proof. □

The next lemma establishes the relation between U and $\mathsf{st}^{-1}(U)$.

Lemma 9.30. *Suppose (SF) and (OC) hold. For any U which is a finite intersection of elements from $\mathscr{B}_0$, any $x \in \mathrm{NS}({}^*X)$ and any $t \in \mathrm{NS}({}^*\mathbb{R}^+)$ we have ${}^*g(x, t, {}^*U) \approx \overline{{}^*g}(x, t, \mathsf{st}^{-1}(U))$.*

Proof. Fix some U which is a finite intersection of elements from $\mathscr{B}_0$, some $x \in \mathrm{NS}({}^*X)$ and some $t \in \mathrm{NS}({}^*\mathbb{R}^+)$. As $\mathsf{st}^{-1}(U) \subset {}^*U$, it is sufficient to show that

${}^*g(x,t,{}^*U) - \overline{{}^*g}(x,t,\mathsf{st}^{-1}(U)) < \epsilon$ for every $\epsilon \in \mathbb{R}^+$. Fix some $\epsilon_1 \in \mathbb{R}^+$. By Lemma 9.29, we know that

$$ {}^*g(x,t,{}^*U) \approx {}^*g(\mathsf{st}(x),\mathsf{st}(t),{}^*U). \tag{9.47} $$

Let $U = \bigcup_{n\in\mathbb{N}} U_n$ where $U_n \in \mathscr{B}_0$ for all $n \in \mathbb{N}$. As X is a metric space satisfying the Heine-Borel condition, X is locally compact. Thus, without loss of generality, we can assume that $\overline{U}_n \subset U$ for all $n \in \mathbb{N}$. By the continuity of probability and the transfer principle, there exists a $N \in \mathbb{N}$ such that

$$ {}^*g(\mathsf{st}(x),\mathsf{st}(t),{}^*U) - {}^*g(\mathsf{st}(x),\mathsf{st}(t),{}^*(\bigcup_{n\leq N})U_n) < \epsilon_1. \tag{9.48} $$

By Lemma 9.29 again, we know that ${}^*g(x,t,{}^*U) - {}^*g(x,t,{}^*(\bigcup_{n\leq N})U_n) < \epsilon_1$. As $\overline{\bigcup_{n\leq N} U_n} \subset U$, we know that ${}^*(\bigcup_{n\leq N})U_n \subset \mathsf{st}^{-1}(U)$. Hence we know that ${}^*g(x,t,{}^*U) - \overline{{}^*g}(x,t,\mathsf{st}^{-1}(U)) < \epsilon_1$. As the choice of ϵ_1 is arbitrary, we have the desired result. $\square$

Before we extend Theorem 9.27 to all non-negative $t \in \mathbb{R}$, we introduce the following concept.

Definition 9.31. A class $\mathcal{C}$ of subsets of some space X is called a π-system if it is closed under finite intersections.

π-system can be used to determine the uniqueness of measures.

Lemma 9.32 ([Kal02, Lemma 1.17])**.** *Let μ and ν be bounded measures on some measurable space $(\Omega,\mathcal{A})$, and let $\mathcal{C}$ be a π-system in Ω such that $\Omega \in \mathcal{C}$ and $\sigma(\mathcal{C}) = \mathcal{A}$ where $\sigma(\mathcal{C})$ denote the σ-algebra generated by $\mathcal{C}$. Then $\mu = \nu$ if and only if $\mu(A) = \nu(A)$ for all $A \in \mathcal{C}$.*

Lemma 9.32 allows us to obtain slightly stronger results than Lemmas 9.29 and 9.30.

Lemma 9.33. *Suppose (SF) and (OC) hold. For any near-standard $x_1 \approx x_2$, any non-infinitesimal $t_1, t_2 \in \mathrm{NS}({}^*\mathbb{R}^+)$ such that $t_1 \approx t_2$ and any $E \in \mathcal{B}[X]$, we have ${}^*g(x_1, t_1, {}^*E) \approx {}^*g(x_2, t_2, {}^*E)$.*

Proof. Fix two near-standard $x_1 \approx x_2$ and two near-standard $t_1 \approx t_2$. Let $\mu_1(A) = \overline{{}^*g}(x_1, t_1, {}^*A)$ and $\mu_2(A) = \overline{{}^*g}(x_2, t_2, {}^*A)$ for all $A \in \mathcal{B}[X]$. It is easy to see that both μ_1 and μ_2 are probability measures on X. By Lemma 9.29, we know that $\mu_1(U) = \mu_2(U)$ for any U which is a finite intersection of elements in $\mathscr{B}_0$. By Lemma 9.32, we have the desired result. □

By using essentially the same argument, we have

Lemma 9.34. *Suppose (SF) and (OC) hold. For any $E \in \mathcal{B}[X]$, any $x \in \mathrm{NS}({}^*X)$ and any $t \in \mathrm{NS}({}^*\mathbb{R}^+)$ we have ${}^*g(x, t, {}^*E) \approx \overline{{}^*g}(x, t, \mathsf{st}^{-1}(E))$.*

We are now at the place to extend Theorem 9.27 to all non-negative $t \in \mathbb{R}$.

Theorem 9.35. *Suppose (DT), (SF) and (OC) hold. For any $s \in \mathrm{NS}(S)$, any non-infinitesimal $t \in \mathrm{NS}(T)$ and any $E \in \mathcal{B}[X]$, we have $P^{(\mathsf{st}(t))}_{\mathsf{st}(s)}(E) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(E) \cap S)$.*

Proof. Fix any $s \in \mathrm{NS}(S)$, any non-infinitesimal $t \in \mathrm{NS}(T)$ and any $E \in \mathcal{B}[X]$. By Lemmas 9.33 and 9.34, we know that

$$g(\mathsf{st}(s), \mathsf{st}(t), E) = {}^*g(\mathsf{st}(s), \mathsf{st}(t), {}^*E) \approx {}^*g(s, t, {}^*E) \approx \overline{{}^*g}(s, t, \mathsf{st}^{-1}(E)). \tag{9.49}$$

By Theorem 9.21, we know that $\overline{{}^*g}(s, t, \mathsf{st}^{-1}(E)) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(E) \cap S)$. Thus we know that $g(\mathsf{st}(s), \mathsf{st}(t), E) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(E) \cap S)$, completing the proof. □

It is possible to weaken (OC) to: $g(x, t, U)$ is a continuous function of $t > 0$ for $x \in X$ and $U \in \mathscr{B}_0$. From the proof of Theorem 9.35, we can show that $g(\mathsf{st}(s), \mathsf{st}(t), U) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(U) \cap S)$ for all $U \in \mathscr{B}_0$. Then the question is: if two finite Borel measures on some metric space agree on all open balls, do they agree on all Borel sets? Unfortunately, this is not true even for compact metric spaces.

Theorem 9.36 ([Dav71, Theorem .2]). *There exists a compact metric space Ω, and two distinct probability Borel measures μ_1, μ_2 on Ω, such that $\mu_1(U) = \mu_2(U)$ for every open ball $U \subset \Omega$.*

However, we do have an affirmative answer of the above question for metric spaces we normally encounter.

Theorem 9.37 ([PT91]). *Whenever finite Borel measures μ and ν over a separable Banach space agree on all open balls, then $\mu = \nu$.*

The following definition of "continuous in time" is weaker than (OC).

Condition WC. The Markov chain $\{X_t\}$ is said to be *weakly continuous in time* if for any open ball $A \subset X$, and any $x \in X$, the function $t \mapsto P_x^{(t)}(A)$ is a right continuous function for $t > 0$. Moreover, for any $t_0 \in \mathbb{R}^+$, any $x \in X$ and any $E \in \mathcal{B}[X]$ we have $\lim_{t \uparrow t_0} P_x^{(t)}(E)$ always exists although it not necessarily equals to $P_x^{(t_0)}(E)$.

This condition is usually assumed for all the continuous time Markov processes. An immediate implication of this definition is the following lemma:

Lemma 9.38. *Suppose (SF) and (WC) hold. For any near-standard $x_1 \approx x_2$, any non-infinitesimal $t_1, t_2 \in \mathrm{NS}({}^*\mathbb{R}^+)$ such that $t_1 \approx t_2$ and $t_1, t_2 \geq \mathsf{st}(t_1)$ and any open ball A we have ${}^*g(x_1, t_1, {}^*A) \approx {}^*g(x_2, t_2, {}^*A)$.*

Proof. The proof is similar to the proof of Lemma 9.29. □

This lemma, just like Lemma 9.29, is stronger than Lemma 9.22 since t_1 and t_2 need not be standard positive real numbers. We now generalize Lemma 9.24 to all $t \in \mathrm{NS}({}^*\mathbb{R})$. Before proving it, we first recall the following theorem.

Theorem 9.39 (Vitali-Hahn-Saks Theorem). *Let μ_n be a sequence of countably additive functions defined on some fixed σ-algebra Σ, with values in a given Banach* *space B such that*

$$\lim_{n\to\infty}\mu_n(X)=\mu(X). \tag{9.50}$$

exists for every $X\in\Sigma$, then μ is countably additive.

An immediate consequence of Theorem 9.39 is that the limit of probability measures remain a probability measure. The following lemma generalizes Lemma 9.24 to all $t\in\mathrm{NS}({}^*\mathbb{R})$.

Lemma 9.40. *Suppose (SF) and (WC) hold. For any $x\in\mathrm{NS}({}^*X)$ and for any non-infinitesimal $t\in\mathrm{NS}({}^*\mathbb{R})$ we have ${}^*g(x,t,{}^*E)\approx\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E))$ for all $E\in\mathcal{B}[X]$. Moreover, $\overline{{}^*g}(x,t,\mathsf{st}^{-1}(X))=1$ for all $x\in\mathrm{NS}({}^*X)$ and all $t\in\mathrm{NS}({}^*\mathbb{R})$.*

Proof. Pick any $x\in\mathrm{NS}({}^*X)$, any $t\in\mathrm{NS}({}^*\mathbb{R})$ and any $E\in\mathcal{B}[X]$. Let $x_0=\mathsf{st}(x)$ and $t_0=\mathsf{st}(t)$. We first show the result for $t<t_0$. For any $B\in\mathcal{B}[X]$, let $h(x_0,t_0,B)$ denote $\lim_{s\uparrow t_0}g(x_0,s,B)$. By Vitali-Hahn-Saks theorem, h is a probability measure on $(X,\mathcal{B}[X])$. Since X is a Polish space, h is a Radon measure. By Lemma 6.8, we know that $\overline{{}^*h}(x_0,t_0,\mathsf{st}^{-1}(X))=1$. As $t\approx t_0$, we know that ${}^*g(x_0,t,{}^*B)\approx{}^*h(x_0,t_0,{}^*B)$ for all $B\in\mathcal{B}[X]$. Pick some $\epsilon\in\mathbb{R}^+$ and choose K compact, U open with $K\subset E\subset U$ and $h(x_0,t_0,U)-h(x_0,t_0,K)<\frac{\epsilon}{2}$. We have

$$|\overline{{}^*g}(x_0,t,\mathsf{st}^{-1}(E))-\overline{{}^*h}(x_0,t_0,\mathsf{st}^{-1}(E))| \tag{9.51}$$

$$\lessapprox|\overline{{}^*g}(x_0,t,\mathsf{st}^{-1}(E))-\overline{{}^*g}(x_0,t,{}^*K)|+|\overline{{}^*h}(x_0,t_0,{}^*K)-\overline{{}^*h}(x_0,t_0,\mathsf{st}^{-1}(E))|\lessapprox\epsilon \tag{9.52}$$

As ϵ is arbitrary, we have $\overline{{}^*g}(x_0,t,\mathsf{st}^{-1}(E))=\overline{{}^*h}(x_0,t_0,\mathsf{st}^{-1}(E))$. Hence we have $\overline{{}^*g}(x_0,t,\mathsf{st}^{-1}(E))={}^*g(x_0,t,{}^*E)$. By Lemma 9.22, we know that ${}^*g(x_0,t,D)\approx{}^*g(x,t,D)$ for all $D\in{}^*\mathcal{B}[X]$. Thus, we have $\overline{{}^*g}(x_0,t,\mathsf{st}^{-1}(E))=\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E))$ and ${}^*g(x_0,t,{}^*E)\approx{}^*g(x,t,{}^*E)$. Hence we have $\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E))={}^*g(x,t,{}^*E)$.

For $t\geq t_0$, we can simply take $h(x_0,t_0,B)$ to be $g(x_0,t_0,B)$ for every $B\in\mathcal{B}[X]$.

Suppose there exist some $x_0\in\mathrm{NS}({}^*X)$ and some infinitesimal t_0 such that $\overline{{}^*g}(x_0,t_0,\mathsf{st}^{-1}(X))<1$. This implies that there exist $n\in\mathbb{N}$ and $A\in{}^*\mathcal{B}[X]$ such that

$$(A \cap \mathsf{st}^{-1}(X) = \emptyset) \wedge ({}^*g(x_0, t_0, A) > \frac{1}{n}). \tag{9.53}$$

Pick some positive $t_1 \in \mathbb{R}$.

Claim 9.41. *${}^*f_{x_0}^{(t_0,t_1)}(A, {}^*K) \approx 0$ for all compact $K \subset X$.*

Proof. Pick some compact subset K and some positive $\epsilon \in \mathbb{R}$. By condition (2) of (DT), there exists positive $r \in \mathbb{R}$ such that

$$(\forall x \in X)(d(x, K) > r \implies g(x, t_1, K) < \epsilon). \tag{9.54}$$

By the transfer principle, we know that ${}^*g(x, t_1, {}^*K) \approx 0$ for all $x \in A$. By Lemma 8.5, we have ${}^*f_{x_0}^{(t_0,t_1)}(A, {}^*K) \approx 0$. □

Fix some compact $K \subset X$. Note that

$${}^*g(x_0, t_0 + t_1, K) = {}^*g(x_0, t_0, A){}^*f_{x_0}^{(t_0,t_1)}(A, {}^*K) + {}^*g(x_0, t_0, {}^*X \setminus A){}^*f_{x_0}^{(t_0,t_1)}({}^*X \setminus A, {}^*K). \tag{9.55}$$

Hence ${}^*g(x_0, t_0 + t_1, K) \lessapprox 1 - \frac{1}{n}$. As this is true for all compact $K \subset X$, we know that $\overline{{}^*g}(x_0, t_0 + t_1, \mathsf{st}^{-1}(X)) \leq 1 - \frac{1}{n}$. This is a contradiction hence we have the desired result. □

A consequence of this lemma is the following result:

Lemma 9.42. *Suppose (SF) and (WC) hold. For any $s \in \mathrm{NS}(S)$ and any $t \in \mathrm{NS}(T)$ we have $G_s^{(t)}(S) = \overline{G}_s^{(t)}(\mathrm{NS}(S)) = 1$.*

Proof. Fix any $s \in \mathrm{NS}(S)$ and any $t \in \mathrm{NS}(T)$. By Theorem 9.21 and Lemma 9.40, we know that

$$\overline{G}_s^{(t)}(\mathsf{st}^{-1}(X) \cap S) = {}^*g(s, t, \mathsf{st}^{-1}(X)) = 1. \tag{9.56}$$

□

Assuming (WC) instead of (OC), we have the following result which is similar to Theorem 9.35.

Theorem 9.43. *Suppose (DT), (SF) and (WC) hold. Suppose the state space X of $\{X_t\}_{t\geq 0}$ is a separable Banach space. Then for any $s \in \mathrm{NS}(S)$, any $t \in \mathrm{NS}(T)$ with $t > \mathsf{st}(t)$ and any $E \in \mathcal{B}[X]$, we have $P^{(\mathsf{st}(t))}_{\mathsf{st}(s)}(E) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(E) \cap S)$.*

Proof. We require X to be a separable Banach space to apply Theorem 9.37. The proof is similar to the proof of Theorem 9.35 hence omitted. □

10. Markov Chain Ergodic Theorem

In the last section, we established the relationship between the transition probability of $\{X_t\}_{t\geq 0}$ and $\{X'_t\}_{t\in T}$. In this section, we will show that $\{X'_t\}_{t\in T}$ inherits some other key properties from $\{X_t\}_{t\geq 0}$. Most importantly, we show that if π is a stationary distribution then its nonstandard counterpart is a weakly stationary distribution. Finally we will establish the Markov chain Ergodic theorem for $\{X_t\}_{t\geq 0}$ by showing that $\{X'_t\}_{t\in T}$ converges.

Let π be a stationary distribution for our standard Markov process $\{X_t\}_{t\geq 0}$. We now show that π', the hyperfinite representation measure of π, is a weakly stationary distribution for $\{X'_t\}_{t\in T}$.

Since X is a Polish space equipped with Borel σ-algebra, the stationary distribution π for $\{X_t\}$ must be a Radon measure. We first establish the following fact of stationary distributions.

Lemma 10.1. *For any $t \in \mathbb{R}^+$, any finite partition of X with Borel sets $A_1, ..., A_n, B$ of X and any $A \in \mathcal{B}[X]$ such that:*

1. *for each $A_i \in \{A_1, ..., A_n\}$ there exists an $\epsilon_i \in \mathbb{R}^+$ such that for any $x, y \in A_i$ we have $|P^{(t)}_x(A) - P^{(t)}_y(A)| < \epsilon_i$.*
2. *there exists an $\epsilon \in \mathbb{R}^+$ such that $\pi(B) < \epsilon$.*

We have $|\pi(A) - \sum_{i\leq n} \pi(A_i)P^{(t)}_{x_i}(A)| \leq \sum_{i\leq n} \pi(A_i)\epsilon_i + \epsilon$ for any $x_i \in A_i$.

Proof. Fix a $t \in \mathbb{R}^+$ and suppose there exists such a finite partition $A_1, ..., A_n, B$ of X satisfying the two conditions in the lemma. Pick any $A \in \mathcal{B}[X]$ and any $x_i \in A_i$. We then have:

$$|\pi(A) - \sum_{i \leq n} \pi(A_i) P_{x_i}^{(t)}(A)| \tag{10.1}$$

$$= |\int_X P_x^{(t)}(A)\pi(dx) - \sum_{i \leq n}(\int_{A_i} \pi(dx)) P_{x_i}^{(t)}(A)| \tag{10.2}$$

$$= |\sum_{i \leq n} \int_{A_i} P_x^{(t)}(A)\pi(dx) + \int_B P_x^{(t)}(A)\pi(dx) - \sum_{i \leq n} \int_{A_i} P_{x_i}^{(t)}(A)\pi(dx)| \tag{10.3}$$

$$\leq |\sum_{i \leq n}(\int_{A_i} (P_x^{(t)}(A) - P_{x_i}^{(t)}(A))\pi(dx))| + \epsilon \tag{10.4}$$

$$\leq \sum_{i \leq n}(\int_{A_i} \epsilon_i \pi(dx)) + \epsilon \tag{10.5}$$

$$= \sum_{i \leq n} \pi(A_i)\epsilon_i + \epsilon. \tag{10.6}$$

□

Write $P_x^{(t)}(A)$ as $g(x, t, A)$ and then apply the transfer principle, we have the following lemma:

Lemma 10.2. *For any $t \in {}^*\mathbb{R}^+$, for any hyperfinite partition of ${}^*X$ with *Borel sets $A_1, ..., A_N, B$ of ${}^*X$ and any $A \in {}^*\mathcal{B}[X]$ such that:*

1. *for each $A_i \in \{A_1, ..., A_N\}$ there exists an $\epsilon_i \in {}^*\mathbb{R}^+$ such that for any $x, y \in A_i$ $|{}^*g(x, t, A) - {}^*g(x, t, A)| < \epsilon_i$.*
2. *there exists an $\epsilon \in {}^*\mathbb{R}^+$ such that $\pi(B) < \epsilon$.*

We have

$$|{}^*\pi(A) - \sum_{i \leq N} {}^*\pi(A_i) {}^*g(x_i, t, A)| \leq \sum_{i \leq N} {}^*\pi(A_i)\epsilon_i + \epsilon. \tag{10.7}$$

for any $x_i \in A_i$

Recall the definition of weakly stationary distribution:

Definition 10.3. An internal distribution π' on $(S, \mathcal{I}(S))$ is called weakly stationary with respect to the Markov chain $\{X'_t\}_{t\in T}$ if there exists an infinite $t_0 \in T$ such that for every $t \leq t_0$ and every $A \in \mathcal{I}(S)$ we have $\pi'(A) \approx \sum_{s\in S} \pi'(\{s\}) G_s^{(t)}(A)$.

We now construct a weak-stationary distribution for $\{X'_t\}_{t\in T}$ from the stationary distribution π of $\{X_t\}_{t\geq 0}$.

Definition 10.4. Define an internal probability measure π' on $(S, \mathcal{I}(S))$ as following:

(1) For all $s \in S$ let $\pi'(\{s\}) = \frac{{}^*\pi(B(s))}{{}^*\pi(\bigcup_{s'\in S} B(s'))}$.
(2) For all internal sets $A \subset S$ let $\pi'(A) = \sum_{s\in A} \pi'(\{s\})$.

The following lemma is a direct consequence of Definition 10.4.

Lemma 10.5. *π' is an internal probability measure on $(S, \mathcal{I}(S))$. Moreover, for any $A \in \mathcal{B}[X]$, we have $\pi(A) = \overline{\pi'}(\mathsf{st}^{-1}(A) \cap S)$.*

Proof. Clearly π' is an internal measure on $(S, \mathcal{I}(S))$. The second part of the lemma follows directly from Theorem 6.11. □

We now show that π' is a weakly stationary distribution for $\{X'_t\}$.

Theorem 10.6. *Suppose (DT), (SF) and (WC) hold. Then π' is a weakly stationary distribution for $\{X'_t\}_{t\in T}$.*

Proof. Fix an internal set $A \in S$ and some near-standard $t \in T$. Consider the hyperfinite partition $\mathcal{F} = \{B(s_1), ..., B(s_N), {}^*X \setminus \bigcup_{s\in S} B(s)\}$ of ${}^*X$ where $S = \{s_1, s_2, ..., s_N\}$ is the state space of $\{X'_t\}$. Note that every member of $\mathcal{F}$ is an member of ${}^*\mathcal{B}[X]$. By Theorem 9.15 and Eq. (9.21), we know that

$$(\forall i \leq N)(\forall x, y \in B(s_i))(\forall C \in {}^*\mathcal{B}[X])(|{}^*g(x,t,C) - {}^*g(y,t,C)| < \epsilon_0). \tag{10.8}$$

Let $B = \bigcup_{s\in A} B(s)$ then $B \in {}^*\mathcal{B}[X]$ since it is a hyperfinite union of *Borel sets. As π is a Radon measure, we know that there exists an infinitesimal ϵ_1 such that ${}^*\pi({}^*X \setminus \bigcup_{s\in S} B(s)) = \epsilon_1$.

By Lemma 10.2 , we have

$$|{}^*\pi(B) - \sum_{i\leq N} {}^*\pi(B(s_i)){}^*g(s_i,t,B)| \leq \sum_{i\leq N} {}^*\pi(B(s_i))\epsilon_0 + \epsilon_1 \leq \epsilon_0 + \epsilon_1 \approx 0. \tag{10.9}$$

By Definition 10.4, we know that $\pi'(A) = \frac{{}^*\pi(B)}{{}^*\pi(\bigcup_{s'\in S} B(s'))}$ and $\pi'(s_i) = \frac{{}^*\pi(B(s_i))}{{}^*\pi(\bigcup_{s'\in S} B(s'))}$. Thus, we have

$$|\pi'(A) - \sum_{i\leq N} \pi'(s_i){}^*g(s_i,t,B)| \approx 0. \tag{10.10}$$

Fix positive $\epsilon \in \mathbb{R}$. As $\overline{\pi'}$ concentrates on $\mathrm{NS}(S)$, there is a near-standard internal set C with $\pi'(C) > 1-\epsilon$. Thus we have

$$|\sum_{s\in S} \pi'(\{s\}){}^*g(s_i,t,B) - \sum_{s\in C} \pi'(\{s\}){}^*g(s_i,t,B)| < \epsilon \tag{10.11}$$

Claim 10.7. *Suppose (DT), (SF) and (WC) hold. Then ${}^*g(s,t,B) \approx G_s^{(t)}(A)$ for all $s \in \mathrm{NS}(S)$ and $t \in \mathrm{NS}(T)$.*

Proof. Fix $n_0 \in \mathbb{N}$, $s \in \mathrm{NS}(S)$ and $t \in \mathrm{NS}(T)$. By Lemma 9.42, there exist near-standard $A_i \in \mathcal{I}(S)$ with $A_i \subset A$ such that $G_s^{(t)}(A) - G_s^{(t)}(A_i) < \frac{1}{n_0}$. By Lemma 9.40, there exist near-standard $C_i \in {}^*\mathcal{B}[X]$ with $C_i \subset B$ such that ${}^*g(s,t,B) - {}^*g(s,t,C_i) < \frac{1}{n_0}$. As X is σ-compact, let $X = \bigcup_{n\in\mathbb{N}} K_n$ where $\{K_n : n \in \mathbb{N}\}$ is a sequence of non-decreasing compact sets. Without loss of generality, we can assume $C_i \subset {}^*K_m$ for some $m \in \mathbb{N}$. As K_m is compact, there exists a near-standard $B_i \in \mathcal{I}(S)$ such that ${}^*K_m \in \bigcup_{s\in B_i} B(s)$. Thus, we have $C_i \subset \bigcup_{s\in B_i} B(s) \subset B$. By the construction of B, it is easy to see that $B_i \subset A$. Note that, by Theorem 9.20, we have

$${}^*g(s,t,\bigcup_{s'\in A_i\cup B_i} B(s')) \approx G_s^{(t)}(A_i \cup B_i) \tag{10.12}$$

Thus we have

$$|^*g(s,t,B) - G_s^{(t)}(A)| \tag{10.13}$$

$$\approx |^*g(s,t,B) - {}^*g(s,t,\bigcup_{s'\in A_i\cup B_i} B(s')) + G_s^{(t)}(A_i\cup B_i) - G_s^{(t)}(A)| \tag{10.14}$$

$$\leq |^*g(s,t,B) - {}^*g(s,t,\bigcup_{s'\in A_i\cup B_i} B(s'))| + |G_s^{(t)}(A_i\cup B_i) - G_s^{(t)}(A)| < \frac{2}{n_0} \tag{10.15}$$

As the choice of n_0 is arbitrary, we have the desired result. □

As C is near-standard, by Lemma 3.20, we have

$$|\sum_{s\in C}\pi'(\{s\})^*g(s_i,t,B) - \sum_{s\in C}\pi'(\{s\})G_s^{(t)}(A)| \approx 0. \tag{10.16}$$

By the construction of C again, we have

$$|\sum_{s\in C}\pi'(\{s\})G_s^{(t)}(A) - \sum_{s\in S}\pi'(\{s\})G_s^{(t)}(A)| < \epsilon. \tag{10.17}$$

By Eqs. (10.10), (10.11), (10.16) and (10.17), we have

$$|\pi'(A) - \sum_{s\in S}\pi'(\{s\})G_s^{(t)}(A)| < 2\epsilon. \tag{10.18}$$

As the choice of ϵ is arbitrary, we have $\pi'(A) \approx \sum_{s\in S}\pi'(\{s\})G_s^{(t)}(A)$ for all $t \in \mathrm{NS}(T)$.

Consider the set $\mathcal{D} = \{t \in T : (\forall A \in \mathcal{I}(S))(|\pi'(A) - \sum_{s\in S}\pi'(\{s\})G_s^{(t)}(A)| < \frac{1}{t})\}$. This is an internal set and contains all $t \in \mathrm{NS}(T)$. Suppose there is no infinite t_0 such that $\mathcal{D}$ contains all the infinite t no greater than t_0. This implies $T \setminus \mathcal{D}$ contains arbitrarily small infinite element hence, by underspill, $T\setminus\mathcal{D}$ contains some $t_0 \in \mathrm{NS}(T)$. This contradicts with the fact that $\mathcal{D}$ contains all $t \in \mathrm{NS}(T)$. Thus π' is a weakly stationary distribution of $\{X'_t\}_{t\in T}$.

□

Note that if π is a stationary distribution of $\{X_t\}_{t\geq 0}$ then $\pi\times\pi$ is a stationary distribution of $\{X_t \times X_t\}_{t\geq 0}$. Thus, we have the following lemma.

Lemma 10.8. *Suppose (DT) and (SF) hold. Then $\pi' \times \pi'$ is a weakly stationary distribution of $\{X'_t \times X'_t\}_{t\in T}$.*

Proof. It is straightforward to verify that $S\times S$ is a (δ_0, r)-hyperfinite representation of ${}^*X \times {}^*X$. Since $\pi \times \pi$ is a stationary distribution, by Theorem 10.6, $(\pi \times \pi)'$ is a weakly stationary distribution of $\{X'_t \times X'_t\}_{t\in T}$. In order to finish the proof, it is sufficient to show that $(\pi \times \pi)' = \pi' \times \pi'$.

Pick any $(s_1, s_2) \in S \times S$. As $\{B(s) : s \in S\}$ is a collection of mutually disjoint sets, we have

$$(\pi \times \pi)'(\{(s_1, s_2)\}) = \frac{{}^*(\pi \times \pi)(B(s_1) \times B(s_2))}{{}^*(\pi \times \pi)(\bigcup_{s\in S} B(s) \times \bigcup_{s\in S} B(s))} \tag{10.19}$$

$$= \frac{{}^*\pi(B(s_1))}{{}^*\pi(\bigcup_{s\in S} B(s))} \cdot \frac{{}^*\pi(B(s_2))}{{}^*\pi(\bigcup_{s\in S} B(s))} \tag{10.20}$$

$$= \pi'(s_1)\pi'(s_2). \tag{10.21}$$

Hence we have $(\pi \times \pi)' = \pi' \times \pi'$, completing the proof. □

In order to show that $\{X'_t\}_{t\in T}$ converges to π', by Theorem 7.19, it remains to show that for $\overline{\pi' \times \pi'}$-almost surely $(i, j) \in S\times S$ there exists a near-standard absorbing point i_0. By Theorem 7.14, it is enough to show that $\{X'_t\}_{t\in T}$ is productively near-standard open set irreducible. We first recall the definition of productively near-standard open set irreducible. We now impose some conditions on $\{X_t\}_{t\geq 0}$ to show that $\{X'_t\}_{t\in T}$ is productively near-standard open set irreducible. We first recall the following definitions.

Definition 10.9. A Markov chain $\{X_t\}_{t\geq 0}$ with state space X is said to be open set irreducible on X if for every open ball $B \subseteq X$ and any $x \in X$, there exists $t \in \mathbb{R}^+$ such that $P^{(t)}_x(B) > 0$.

An internal set $B \subset S$ is an open ball if $B = \{s \in S : {}^*d(s, s_0) < r\}$ for some $s_0 \in S$ and $r \in {}^*\mathbb{R}$. An open ball is near-standard if it contains only near-standard elements.

Definition 10.10. A hyperfinite Markov chain $\{Y_t\}_{t\in T}$ is called near-standard open set irreducible if for any near-standard $s \in S$, any near-standard open ball $B \subset {}^*X$ with non-infinitesimal radius we have $\overline{P}_i(\tau(B) < \infty) > 0$

We first establish the connection between open set irreducibility of $\{X_t\}_{t\geq 0}$ and $\{X'_t\}_{t\in T}$. Note that the consequence of the following theorem implies the near-standard open-set irreducibility of $\{X'_t\}_{t\in T}$.

Theorem 10.11. *Suppose (DT), (SF) and (WC) hold. If $\{X_t\}_{t\geq 0}$ is open set irreducible, then for any near-standard $s \in S$, any near-standard open ball B with non-infinitesimal radius there is a positive $t \in \mathrm{NS}(T)$ such that $\overline{G}^{(t)}_s(B) > 0$.*

Proof. Consider any near-standard open ball $B \subset S$ with non-infinitesimal radius k. Without loss of generality let $B = \{s \in S : {}^*d(s, s_0) < r\}$ for some near-standard $s_0 \in S$ and some near-standard $r \in {}^*\mathbb{R}^+$. Let A be the ball in X centered at $\mathsf{st}(s_0)$ with radius $\frac{\mathsf{st}(r)}{2}$.

Claim 10.12. $\mathsf{st}^{-1}(A) \cap S \subset B$.

Proof. Pick any point $x \in \mathsf{st}^{-1}(A) \cap S$. There exists $a \in A$ such that $x \in \mu(a)$. We then have ${}^*d(x, s_0) \leq {}^*d(x, a) + {}^*d(a, \mathsf{st}(s_0)) + {}^*d(\mathsf{st}(s_0), s_0) \lessapprox \frac{\mathsf{st}(r)}{2}$. Thus ${}^*d(x, s_0) \lessapprox \frac{\mathsf{st}(r)}{2} < r$. This implies that $\mathsf{st}^{-1}(A) \cap S \subset B$. □

Consider any near-standard $s \in S$, there exists a $x \in X$ such that $x = \mathsf{st}(s)$. As $\{X_t\}_{t\geq 0}$ is open set irreducible, there exists a $t \in \mathbb{R}^+$ such that $P^{(t)}_x(A) > 0$. Pick $t' \in T$ such that $t' \approx t$ and $t' \geq t$. By Lemma 9.38, Lemma 9.40 and Theorem 9.21, we know that

$$P^{(t)}_x(A) = g(x, t, A) = {}^*g(x, t, {}^*A) \approx {}^*g(s, t', {}^*A) \approx \overline{{}^*g}(s, t', \mathsf{st}^{-1}(A)) = \overline{G}^{(t')}_s(\mathsf{st}^{-1}(A) \cap S). \tag{10.22}$$

Then we have $\mathsf{st}((G^{(t)}_s(B))) > 0$. □

Let $\{X'_t\}_{t\in T}$ and $\{Y'_t\}_{t\in T}$ be two i.i.d hyperfinite Markov chains on $(S, \mathcal{I}(S))$ both with internal transition probability $\{G^{(\delta t)}_i j)\}_{i,j\in S}$. Let $\{Z'_t\}_{t\in T}$ be the product hyperfinite Markov chain live on $(S \times S, \mathcal{I}(S \times S))$ with respect to $\{X'_t\}_{t\in T}$ and $\{Y'_t\}_{t\in T}$. Recall that the internal transition probability of $\{Z'_t\}_{t\in T}$ is then defined to be

$$F^{(\delta t)}_{(i,j)}(\{(a,b)\}) = G^{(\delta t)}_{i}(\{a\}) \times G^{(\delta t)}_{j}(\{b\}). \tag{10.23}$$

where $(F^{(\delta t)}_{(i,j)}(\{(a,b)\})$ denote the internal probability of Z'_t starts at (i,j) and reach (a,b) at δt.

Before we prove that $\{Z'_t\}_{t\in T}$ is near-standard open set irreducible, we impose the following condition on the standard joint Markov chain.

Definition 10.13. The Markov chain $\{X_t\}_{t\geq 0}$ is productively open set irreducible if the joint Markov chain $\{X_t \times Y_t\}_{t\geq 0}$ is open set irreducible on $X \times X$ where $\{Y_t\}_{t\geq 0}$ is an independent identical copy of $\{X_t\}_{t\geq 0}$.

The following lemma gives a sufficient condition for a Markov process being productively open set irreducible.

Lemma 10.14. *Let $\{X_t\}_{t\geq 0}$ be an open set irreducible Markov process. If there exists $t_0 \in \mathbb{R}^+$ such that for any open set A and any $x \in A$, we have $P^{(t)}_x(A) > 0$ for all $t \geq t_0$. Then $\{X_t\}_{t\in\mathbb{R}}$ is productively open set irreducible.*

Proof. Consider a basic open set $A \times B$. Suppose $\{X_t\}$ reaches A first. Then $\{X_t\}$ will wait for $\{Y_t\}$ to reach B. □

Most of the diffusion processes satisfy the condition of this lemma.

Recall that $\{X'_t\}_{t\in T}$ is productively near-standard open set irreducible if $\{Z'_t\}_{t\in T}$ is near-standard open set irreducible.

Lemma 10.15. *Suppose (DT), (SF) and (WC) hold. If $\{X_t\}_{t\geq 0}$ is productively open set irreducible, then $\{X'_t\}_{t\in T}$ is productively near-standard open set irreducible.*

Proof. Let $\{Y_t\}_{t\geq 0}$ denote an independent identical copy of $\{X_t\}_{t\geq 0}$. We use P to denote the transition probability of X_t and Y_t. Let $\{Z_t\}_{t\in\mathbb{R}}$ be the product chain of $\{X_t\}$ and $\{Y_t\}$. We use Q to denote the transition probability of the joint chain Z_t. Let $\{Y'_t\}_{t\in T}$ denote an independent identical copy of $\{X'_t\}_{t\in T}$. We use G to denote the internal transition probability of X'_t and Y'_t and use F to denote the internal transition probability of the product hyperifnite chain Z'_t. It is sufficient to show that $\{Z'_t\}_{t\in T}$ is near-standard open set irreducible.

Pick any near-standard open ball B with non-infinitesimal radius from $S\times S$ and fix some near-standard $(i,j)\in S\times S$. Then there exists $(x,y)\in X\times X$ such that $(i,j)\in \mu((x,y))$. We can find two open balls $B_1,B_2\in S$ with non-infinitesimal radius such that $B_1\times B_2\subset B$. As in Theorem 10.11, we can find two open balls A_1,A_2 such that $\mathsf{st}^{-1}(A_1)\cap S\subset B_1$ and $\mathsf{st}^{-1}(A_2)\cap S\subset B_2$,respectively. Thus in conclusion we have $(\mathsf{st}^{-1}(A_1)\cap S)\times(\mathsf{st}^{-1}(A_2)\cap S)=(\mathsf{st}^{-1}(A_1\times A_2))\cap(S\times S)\subset B$. As $\{X_t\}_{t\geq 0}$ is productively open set irreducible, there exists $t\in\mathbb{R}^+$ such that $Q^{(t)}_{(x,y)}(A_1\times A_2)>0$. By (WC), we can pick t to be a rational number. By the definition of $\{Z_t\}_{t\geq 0}$ and Theorem 9.27, we have

$$Q^{(t)}_{(x,y)}(A_1\times A_2)=P^{(t)}_x(A_1)\times P^{(t)}_y(A_2)=\overline{G}^{(t)}_i(\mathsf{st}^{-1}(A_1)\cap S)\times \overline{G}^{(t)}_j(\mathsf{st}^{-1}(A_2)\cap S). \tag{10.24}$$

By Lemma 7.10 and the construction of Loeb measure, we know that

$$\overline{G}^{(t)}_i(\mathsf{st}^{-1}(A_1)\cap S)\times \overline{G}^{(t)}_j(\mathsf{st}^{-1}(A_2)\cap S)=\overline{F}^{(t)}_{(i,j)}(\mathsf{st}^{-1}(A_1\times A_2))\cap(S\times S)). \tag{10.25}$$

Thus $\overline{F}^{(t)}_{(i,j)}(\mathsf{st}^{-1}(A_1\times A_2))\cap(S\times S))>0$. As $(\mathsf{st}^{-1}(A_1\times A_2))\cap(S\times S)\subset B$ we have that $\overline{F}^{(t)}_{(i,j)}(B)>0$, completing the proof. $\square$

Now we are at the place to prove the main theorem of this paper.

Theorem 10.16. *Let $\{X_t\}_{t\geq 0}$ be a general-state-space continuous in time Markov chain on some metric space X satisfying the Heine-Borel condition. Suppose $\{X_t\}_{t\geq 0}$ is productively open set irreducible and has a stationary distribution π. Suppose $\{X_t\}_{t\geq 0}$ is vanishing in distance and also satisfies (SF) and (WC). Then for π-almost surely $x\in X$ we have $\lim_{t\to\infty}\sup_{A\in\mathcal{B}[X]}|P^{(t)}_x(A)-\pi(A)|=0$.*

Proof. As $\{X_t\}$ is vanishing in distance and satisfies (SF), by Theorem 9.6, we know that $\{X_t\}_{t\geq 0}$ satisfies (DT). Let $\{X'_t\}_{t\in T}$ denote the corresponding hyperfinite Markov chain on the hyperfinite set S. We use P to denote the transition probability of $\{X_t\}_{t\geq 0}$ and use G to denote the internal transition probability for $\{X'_t\}_{t\in T}$. Let π' be defined as in Definition 10.4. By Theorem 10.6, we know that π' is a weakly stationary distribution for $\{X'_t\}_{t\in T}$. We first show that the internal transition probability of $\{X'_t\}_{t\in T}$ converges to π'. As $\{X_t\}_{t\geq 0}$ is productively open set irreducible, by Lemma 10.15, we know that $\{X'_t\}_{t\in T}$ is productively near-standard open set irreducible. By Theorem 9.25, we know that $\{X'_t\}_{t\in T}$ is strong regular. Thus by Theorems 7.19 and 7.26, we know that for $\overline{\pi'}$ almost surely $s\in S$ and any $A\in\mathcal{L}(\mathcal{I}(S))$, $\lim_{t\to\infty}\sup_{B\in\mathcal{L}(\mathcal{I}(S))}|\overline{G}_s^{(t)}(B)-\overline{\pi'}(B)|=0$.

Now fix any $A\in\mathcal{B}[X]$. Then by Theorem 5.9, we know that $\mathsf{st}^{-1}(A)\in\mathcal{L}(\mathcal{I}(S))$. Consider any $x\in X$ and any $s\in\mathsf{st}^{-1}(\{x\})\cap S$. By Theorem 9.27, we know that for any $t\in\mathbb{Q}^+$ we have $P_x^{(t)}(A)=\overline{G}_s^{(t)}(\mathsf{st}^{-1}(A)\cap S)$. By Lemma 10.5, we know that $\pi(A)=\overline{\pi'}(\mathsf{st}^{-1}(A)\cap S)$. Suppose that there exists a set $B\in\mathcal{B}[X]$ with $\pi(B)>0$ such that, for any $x\in B$, $P_x^{(t)}(\cdot)$ does not converge to $\pi(\cdot)$ in total variation distance. This means that for any $s\in\mathsf{st}^{-1}(B)\cap S$ we have

$$\sup_{A\in\mathcal{B}[X]}|\overline{G}_s^{(t)}(\mathsf{st}^{-1}(A)\cap S)-\overline{\pi'}(\mathsf{st}^{-1}(A)\cap S)|\nrightarrow 0. \tag{10.26}$$

where we can restrict t to $\mathbb{Q}^+\subset T$ since total variation distance is non-increasing. However, as $\pi(B)>0$, we know that $\overline{\pi'}(\mathsf{st}^{-1}(B)\cap S)>0$. This contradict the fact that for $\overline{\pi'}$ almost surely s, $\lim_{t\to\infty}\sup_{B\in\mathcal{L}(\mathcal{I}(S))}|\overline{G}_s^{(t)}(B)-\overline{\pi'}(B)|=0$. Hence we have the desired result. □

Using Theorems 9.6 and 10.16 and results in Section 2, we can obtain Theorem 2.16. We restate it here.

Theorem 10.17. *Let $\{X_t\}_{t\geq 0}$ be a general state space continuous-time Markov chain with separable locally compact metric state space (X,d). Suppose $\{X_t\}_{t\geq 0}$ is productively open set irreducible and has a stationary distribution π. Suppose*

$\{X_t\}_{t\geq 0}$ is vanishing in distance, strong Feller and weakly continuous. Then for π-almost surely $x \in X$ we have $\lim_{t\to\infty} \sup_{A\in\mathcal{B}[X]} |P_x^{(t)}(A) - \pi(A)| = 0$.

11. The Feller Condition

In Sections 8 and 10, our analysis depend on the strong Feller condition ((SF)). In the literature, however, it is sometimes more desirable to replace strong Feller condition with a weaker condition which we call Feller condition. In this section, we will discuss the difference between strong Feller and Feller conditions. Moreover, we will construct a hyperfinite representation $\{X'_t\}_{t\in T}$ of $\{X_t\}_{t\geq 0}$ under Feller condition. Finally, we will establish some of the key properties of $\{X'_t\}_{t\in T}$ inherited from $\{X_t\}_{t\geq 0}$.

We first recall the definition of strong Feller.

Remark 11.1. (SF) The Markov chain$\{X_t\}_{t\geq 0}$ is said to be *strong Feller* if for any $t > 0$ and any $\epsilon > 0$ we have:

$$(\forall x \in X)(\exists \delta > 0)((\forall y \in X)(d(x, y) < \delta \implies (\forall A \in \mathcal{B}[X])|P_y^{(t)}(A) - P_x^{(t)}(A)| < \epsilon)). \tag{11.1}$$

We then introduce the Feller condition.

Condition WF. The Markov chain$\{X_t\}_{t\geq 0}$ is said to be *Feller* if for all $t > 0$ and all $\epsilon > 0$ we have:

$$(\forall A \in \mathcal{B}[X])(\forall x \in X)(\exists \delta > 0)((\forall y \in X)(d(x, y) < \delta \implies |P_y^{(t)}(A) - P_x^{(t)}(A)| < \epsilon)). \tag{11.2}$$

As one can see, the choice of δ in (WF) depends on the Borel set A. We present the following Feller Markov process which is not strong Feller.

Example 11.2 (suggested by Neal Madras)**.** [MS10, Page. 889] Let $\{X_t\}_{t\in\mathbb{N}}$ be a discrete-time Markov processes with state space $[-\pi, \pi]$. For every $n \in \mathbb{N}$, let $\frac{1+\sin(ny)}{2\pi}$ be the density of $P_{\frac{1}{n}}(\mathrm{d}y)$. Let μ be the Lebesgue measure on $[-\pi, \pi]$ divided by 2π and let $\mu(A) = P_0(A)$ for all Borel sets A.

Claim 11.3. $\lim_{n\to\infty} P_{\frac{1}{n}}(A) = \mu(A)$ *for all Borel sets* A.

Proof. Let A be an internal with end points a and b.

$$\lim_{n\to\infty} P_{\frac{1}{n}}(A) = \lim_{n\to\infty} \int_a^b \frac{1+\sin(ny)}{2\pi} \mathrm{d}y \tag{11.3}$$

$$= \lim_{n\to\infty} \left(\frac{b-a}{2\pi} - \frac{\cos(nb)-\cos(na)}{2n\pi}\right) \tag{11.4}$$

$$= \frac{b-a}{2\pi} = \mu(A) \tag{11.5}$$

By Theorem 9.37, we have the desired result. □

Claim 11.4. $\sup_{A\in\mathcal{B}[[-\pi,\pi]]} |P_{\frac{1}{n}}(A) - \mu(A)| \geq \frac{1}{\pi}$ *for all* $n \in \mathbb{N}$

Proof. Let A be an internal with end points a and b. Then we have $|P_{\frac{1}{n}}(A)-\mu(A)| = |\frac{\cos(nb)-\cos(na)}{2n\pi}|$. For any $m \in \mathbb{N}$, we can find an open set U_m which is a union of m open intervals $(a_1, b_1), \ldots, (a_m, b_m)$ such that $\cos(nb_n) - \cos(na_n) = 2$ for all $n \leq m$. Then $|P_{\frac{1}{m}}(U_m) - \mu(U_m)| = \frac{1}{\pi}$, completing the proof. □

11.1. **Hyperfinite Representation under the Feller Condition.** In this section, we will show that, by carefully picking a hyperfinite representation, we can construct a hyperfinite Markov process $\{X'_t\}_{t\in T}$ which is a hyperfinite representation of $\{X_t\}_{t\geq 0}$. We use $P^{(t)}_x(A)$ to denote the transition probability of $\{X_t\}_{t\geq 0}$. When we view the transition probability as a function of three variables, we denote it by $g(x, t, A)$.

The state space of $\{X'_t\}_{t\in T}$ is a hyperfinite representation S of ${}^*X$. By Definition 6.3, the hyperfinite set S should be a (δ_0, r_0)-hyperfinite representation of ${}^*X$ for some positive infinitesimal δ_0 and some positive infinite number r_0. We need to pick δ_0 and r_0 carefully. Recall that the time line $T = \{0, \delta t, \ldots, K\}$. Let ϵ_0 be a positive infinitesimal such that $\epsilon_0 \frac{t}{\delta t} \approx 0$ for all $t \in T$. We can pick r_0 the same way as we did in Section 8. Recall (DT) and Theorem 9.8 from Section 8.

Remark 11.5 ((DT)). The Markov chain $\{X_t\}_{t\geq 0}$ is said to be *vanish in distance* if for all $t \geq 0$ and all $K \in \mathcal{K}[X]$ we have:

(1) $(\forall \epsilon > 0)(\exists r > 0)(\forall x \in K)(\forall A \in \mathcal{B}[X])(d(x,A) > r \implies g(x,t,A) < \epsilon)$.

(2) $(\forall \epsilon > 0)(\exists r > 0)(\forall x \in X)(d(x,K) > r \implies g(x,t,K) < \epsilon)$.

where $\mathcal{K}$ denote the collection of all compact sets of X.

By mimicking the proof of Theorem 9.6, we immediately obtain the following result.

Lemma 11.6. *If a Markov process has the weak Feller property, then it also satisfies property (1) from (DT).*

From (DT), we have the following lemma.

Lemma 11.7 (Theorem 9.8)**.** *Suppose (DT) holds. For every positive $\epsilon \in {}^*\mathbb{R}$, there exists an open ball centered at some standard point a with radius r such that:*

(1) ${}^*g(x, \delta t, {}^*X \setminus \overline{U}(a,r)) < \epsilon$ *for all* $x \in \mathrm{NS}({}^*X)$.

(2) ${}^*g(y,t,A) < \epsilon$ *for all* $y \in {}^*X \setminus \overline{U}(a,r)$*, all near-standard* $A \in {}^*\mathcal{B}[X]$ *and all* $t \in T$.

where $\overline{U}(a,r) = \{x \in {}^*X : {}^*d(x,a) \leq r\}$.

Fix a standard $a_0 \in X$. For the particular ϵ_0, we can find a r_1 such that the ball $U(a_0, r_1)$ satisfies the conditions in Lemma 11.7.

Recall the following results from Section 8

Lemma 11.8 (Lemma 9.11)**.** *Suppose (DT) holds. There exists a positive infinite $r_0 > 2r_1$ such that*

$$(\forall y \in \overline{U}(a_0, 2r_1))({}^*g(y, \delta t, {}^*X \setminus \overline{U}(a_0, r_0)) < \epsilon_0). \tag{11.6}$$

Just as in Section 8,we fix a_0, r_1 and r_0 for the remainder of this section.

Lemma 11.9 (Lemma 9.12)**.** *Suppose (DT) holds. For any $x \in X$, any $t \in T$, any near-standard internal set $A \subset {}^*X$ we have* ${}^*f_x^{(t)}({}^*X \setminus \overline{U}(a_0, r_0), A) < 2\epsilon_0$.

Just as in Section 8, our hyperfinite state space will cover $\overline{U}(a_0, 2r_0)$. We will choose δ_0 to partition $\overline{U}(a_0, 2r_0)$ into *Borel sets with diameters no greater than δ_0.

We start by picking an arbitrary positive infinitesimal δ_1 and let S_1 be a $(\delta_1, 2r_0)$-hyperfinite representation of ${}^*X$ such that $\{B_1(s) : s \in S_1\} = \overline{U}(a_0, 2r_0)$. We fix S_1 for the remainder of this section.

Lemma 11.10. *Suppose (DT) and (WF) hold. There exists a positive infinitesimal δ_0 such that for any $x_1, x_2 \in \overline{U}(a_0, 2r_0)$ with $|x_1 - x_2| < \delta_0$ we have for all $A \in \mathcal{I}(S_1)$ and all $t \in T^+$:*

$$|{}^*g(x_1, t, \bigcup_{s\in A} B_1(s)) - {}^*g(x_2, t, \bigcup_{s\in A} B_1(s))| < \epsilon_0 \tag{11.7}$$

Proof. Fix a $A \in \mathcal{I}(S_1)$. By the transfer of (WF), for every $x \in \overline{U}(a_0, 2r_0)$ there exists $\delta_x \in {}^*\mathbb{R}^+$ such that $\forall y \in {}^*X$ we have

$$d(x, y) < \delta_x \implies |{}^*g(x, \delta t, \bigcup_{s\in A} B_1(s)) - {}^*g(y, \delta t, \bigcup_{s\in A} B_1(s))| < \frac{\epsilon_0}{2}. \tag{11.8}$$

The collection $\{U(x, \frac{\delta_x}{2}) : x \in \overline{U}(a_0, 2r_0)\}$ forms an open cover of $\overline{U}(a_0, 2r_0)$. By the transfer of Heine-Borel condition, $\overline{U}(a_0, 2r_0)$ is *compact hence there exists a hyperfinite subset of the cover $\{U(x, \frac{\delta_x}{2}) : x \in \overline{U}(a_0, 2r_0)\}$ that covers $\overline{U}(a_0, 2r_0)$. Denote this hyperfinite subcover by $\mathcal{F} = \{U(x_i, \frac{\delta_{x_i}}{2}) : i \le N\}$ where $\{\frac{\delta_{x_i}}{2} : i \le N\}$ is a hyperfinite set. Let $\delta_A = \min\{\frac{\delta_{x_i}}{2} : i \le N\}$.

Pick any $x, y \in \overline{U}(a_0, 2r_0)$ with $d(x, y) < \delta_A$. We know that $x \in B(x_i, \frac{\delta_{x_i}}{2})$ for some $i \le N$ and ${}^*d(y, x_i) \le {}^*d(y, x) + {}^*d(x, x_i) \le \delta_{x_i}$. Thus both x, y are in some $B(x_i, \delta_{x_i})$. This means that

$$|{}^*g(x, \delta t, \bigcup_{s\in A} B_1(s)) - {}^*g(y, \delta t, \bigcup_{s\in A} B_1(s))| < \epsilon_0. \tag{11.9}$$

Let $\mathcal{M} = \{\delta_A : A \in \mathcal{I}(S)\}$. Note that $\mathcal{M}$ is a hyperfinite set hence there exists a minimum element, denoted by $\delta^{\delta t}$. We can carry out this argument for every $t \in T$. Let δ^t denote the minimum element for time t and consider the hyperfinite set $\{\delta^t : t \in T\}$. This set again has a minimum element δ_0. It is easy to check that this δ_0 satisfies the condition of this lemma. $\square$

Definition 11.11. Let S, S' be two hyperfinite representations of ${}^*X$. The hyperfinite representation S' is a refinement of S if for every $A \in \mathcal{I}(S)$ there exists a $A' \in \mathcal{I}(S')$ such that $\bigcup_{s\in A} B(s) = \bigcup_{s'\in A'} B'(s')$. The set A' is called an enlargement of A.

Let S' be a refinement of S. For any $A \in \mathcal{I}(S)$, note that the enlargement A' is unique. Fix δ_0 in Lemma 11.10 for the remainder of this section. We present the following result.

Lemma 11.12. *There exists a $(\delta_0, 2r_0)$-hyperfinite representation S with $\bigcup_{s\in S} B(s) = \overline{U}(a_0, 2r_0)$ such that S is a refinement of S_1.*

Proof. Fix an arbitrary $(\delta_0, 2r_0)$-hyperfinite representation H such that the collection $\{B_H(h) : h \in H\} = \overline{U}(a_0, 2r_0)$. For every $s \in S_1$, let

$$M(s) = \{B_H(h) : B_H(h) \cap B_1(s) \neq \emptyset\}. \tag{11.10}$$

Note that $M(s)$ is hyperfinite for every $s \in S_1$. Let

$$N(s) = \{B_H(h) \cap B_1(s) : B_H(h) \in M(s)\}. \tag{11.11}$$

Note that $N(s)$ is also hyperfinite for every $s \in S_1$. It is easy to see that $\bigcup_{s\in S_1} N(s) = \bigcup_{s\in S_1} B_1(s) = \overline{U}(a_0, 2r_0)$. Note that $\bigcup_{s\in S_1} N(s)$ is a collection of mutually disjoint * Borel set with diameter no greater than δ_2. Pick one point from each element of $\bigcup_{s\in S_1} N(s)$ and form a hyperfinite set S. This S is a hyperfinite set satisfying all the conditions of this lemma. $\square$

For each $s \in S$, we use $B(s)$ to denote the corresponding *Borel set. By the construction in Lemma 11.12, we can see that every $B(s)$ is a subset of $B_1(s')$ for some $s' \in S_1$ and every $B_1(s')$ is a hyperfinite union of $B(s)$.

By Lemmas 11.10 and 11.12, we have the following result:

Theorem 11.13. *Let S_1, S be the same hyperfinite representations as in Lemma 11.12. Then for any $s \in S$, any $x_1, x_2 \in B(s)$, any $A \in \mathcal{I}(S_1)$ and any $t \in T^+$ we have*

$$|{}^*g(x_1, t, \bigcup_{s\in A} B_1(s)) - {}^*g(x_2, t, \bigcup_{s\in A} B_1(s))| < \epsilon_0. \tag{11.12}$$

An immediate consequence of this theorem is:

Corollary 11.14. *Let S_1, S be the same hyperfinite representations as in Lemma 11.12. For for any $s \in S$, any $y \in B(s)$, any $x \in {}^*X$, any $A \in \mathcal{I}(S_1)$ and any $t \in T^+$ we have $|{}^*g(y, t, \bigcup_{s\in A} B_1(s)) - {}^*f_x^{(t)}(B(s), \bigcup_{s\in A} B_1(s))| < \epsilon_0$.*

We fix S constructed above for the remainder of this section. In summary, S_1 is a $(\delta_1, 2r_0)$-hyperfinite representation of ${}^*X$ for some infinitesimal δ_1 such that $\{B_1(s) : s \in S_1\}$ covers $\overline{U}(a_0, 2r_0)$. S is a refinement of S_1 satisfying the following conditions:

(1) The diameter of $B(s)$ is less than δ_0 for all $s \in S$.
(2) $\bigcup_{s\in S} B(s) = \overline{U}(a_0, 2r_0)$.

We let S be the hyperfinite state space of our hyperfinite Markov process. Note that for any $x \in \mathrm{NS}({}^*X)$ and any $y \in {}^*X \setminus \bigcup_{s\in S} B(s)$, we have ${}^*d(x, y) > r_0$.

We construct $\{X'_t\}_{t\in T}$ on S in a similar way as in Section 8. Let $g'(x, \delta t, A) = {}^*g(x, \delta t, A \cap \bigcup_{s\in S} B(s)) + \delta_x(A){}^*g(x, \delta t, {}^*X \setminus \bigcup_{s\in S} B(s))$ where $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if otherwise. For $i, j \in S$ let $G_{ij}^{(\delta t)} = g'(i, \delta t, B(j))$ be the "one-step" internal transition probability of $\{X'_t\}_{t\in T}$. We use $G_i^{(t)}(\cdot)$ to denote the t-step internal transition measure. By Lemmas 8.12 and 8.13, we know that $G_i^{(t)}(\cdot)$ is an internal probability measure on $(S, \mathcal{I}(S))$ for all $t \in T$.

Similar to Theorem 9.20, we have the following theorem. The two proofs are similar to each other.

Theorem 11.15. *Suppose (DT) and (WF) hold. For any $t \in T$, any $x \in S$ and any near-standard $A \in \mathcal{I}(S_1)$, we have*

$$|{}^*g(x, t, \bigcup_{s\in A_S} B(s)) - G_x^{(t)}(A_S)| \le \epsilon_0 + 5\epsilon_0 \frac{t - \delta t}{\delta t}. \tag{11.13}$$

where A_S is the enlargement of A. In particular, for all $t \in T$, all $x \in S$ and all near-standard $A \in \mathcal{I}(S_1)$ we have

$$|{}^*g(x,t,\bigcup_{s\in A_S} B(s)) - G_x^{(t)}(A_S)| \approx 0 \tag{11.14}$$

Proof. : In the proof of Theorem 9.20, by (SF), we know that for any $s_0 \in S$ and any $t \in T+$

$$(\forall x_1, x_2 \in B(s_0))(\forall A \in \mathcal{I}(S))(|{}^*g(x_1,t,\bigcup_{s\in A} B(s)) - {}^*g(x_2,t,\bigcup_{s\in A} B(s))| < \epsilon_0). \tag{11.15}$$

Under (WF), by Theorem 11.13 and Corollary 11.14 and the fact that S is a refinement of S_1, we know that for any $s_0 \in S$ and any $t \in T^+$

$$(\forall x_1, x_2 \in B(s_0))(\forall A \in \mathcal{I}(S_1))(|{}^*g(x_1,t,\bigcup_{s\in A_S} B(s)) - {}^*g(x_2,t,\bigcup_{s\in A_S} B(s))| < \epsilon_0). \tag{11.16}$$

We use this formula to replace the Eq. (11.15) in the proof of Theorem 9.20. Then the rest of the proof is identical to the proof of Theorem 9.20. $\square$

In Section 8, we have shown that $\{X'_t\}$ is a hyperfinite representation of $\{X_t\}_{t\geq 0}$ in terms of transition probability. We first establish a similar result as Theorem 9.21.

Theorem 11.16. *Suppose (DT) and (WF) hold. For any $x \in \bigcup_{s\in S} B(s)$ let s_x denote the unique element in S such that $x \in B(s_x)$. Then for any $E \in \mathcal{B}[X]$ and any $t \in T$, we have $\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E)) = \overline{G}_{s_x}^{(t)}(\mathsf{st}^{-1}(E)\cap S)$.*

Proof. We first prove the case when $t = 0$. $\overline{{}^*g}(x,0,\mathsf{st}^{-1}(E))$ is 1 if $x \in \mathsf{st}^{-1}(E)$ and is 0 otherwise. Note that $x \in \mathsf{st}^{-1}(E)$ if and only if $s_x \in \mathsf{st}^{-1}(E)\cap S$. Hence $\overline{{}^*g}(x,0,\mathsf{st}^{-1}(E)) = \overline{G}_{s_x}^{(0)}(\mathsf{st}^{-1}(E)\cap S)$.

We now prove the case for $t > 0$. Fix some $x \in \bigcup_{s\in S} B(s)$, some $t > 0$ and some $E \in \mathcal{B}[X]$. By the construction in Theorem 6.11 and Eq. (6.19), we know that for every $t > 0$:

$$\overline{^*g}(x, t, \mathsf{st}^{-1}(E)) = \sup\{\overline{^*g}(x, t, \bigcup_{s \in A} B_1(s)) : A \subset \mathsf{st}^{-1}(E) \cap S_1, A \in \mathcal{I}(S_1)\} \quad (11.17)$$

By Theorem 11.13, we have $|^*g(x, t, \bigcup_{s \in A} B_1(s)) - {}^*g(s_x, t, \bigcup_{s \in A} B_1(s))| < \epsilon_0$. By Theorem 11.15, we know that $|^*g(s_x, t, \bigcup_{s \in A} B_1(s)) - G^{(t)}_{s_x}(A_S)| \approx 0$. Thus we know that $\overline{^*g}(x, t, \bigcup_{s \in A} B_1(s)) = \overline{G}^{(t)}_{s_x}(A_S)$. Hence we have

$$\overline{^*g}(x, t, \mathsf{st}^{-1}(E)) = \sup\{\overline{G}^{(t)}_{s_x}(A_S) : A \subset \mathsf{st}^{-1}(E) \cap S_2, A \in \mathcal{I}(S_1)\}. \quad (11.18)$$

Claim 11.17.

$$\overline{G}^{(t)}_{s_x}(\mathsf{st}^{-1}(E) \cap S) = \sup\{\overline{G}^{(t)}_{s_x}(A_S) : A \subset \mathsf{st}^{-1}(E) \cap S_2, A \in \mathcal{I}(S_1)\}. \quad (11.19)$$

Proof. Let B be an internal subset of S such that $B \subset \mathsf{st}^{-1}(E) \cap S$. For any $b \in B$, there exists a $s_b \in S_1$ such that $b \in B_1(s_b)$. Let $A = \{s_b : b \in B\}$. Then $A \in \mathcal{I}(S_1)$ and it is easy to see that $B \subset A_S \subset \mathsf{st}^{-1}(E) \cap S$. Thus we can conclude that

$$\sup\{\overline{G}^{(t)}_{s_x}(A_S) : A \subset \mathsf{st}^{-1}(E) \cap S_2, A \in \mathcal{I}(S_2)\} = \overline{G}^{(t)}_{s_x}(\mathsf{st}^{-1}(E) \cap S). \quad (11.20)$$

□

Thus we have the desired result. □

The next lemma establishes a weaker form of local continuity of $^*g$.

Lemma 11.18. *Suppose (WF) holds. For any two near-standard $x_1 \approx x_2$ from $^*X$,any $t \in \mathbb{R}^+$ and any $A \in \mathcal{B}[X]$ we have $^*g(x_1, t, {}^*A) \approx {}^*g(x_2, t, {}^*A)$.*

Proof. Fix two near-standard x_1, x_2 from $^*X$. Let $x_0 = \mathsf{st}(x_1) = \mathsf{st}(x_2)$. Also fix $t \in \mathbb{R}^+$ and $A \in \mathcal{B}[X]$. Pick $\epsilon \in \mathbb{R}^+$. By (WF), we can pick a $\delta \in \mathbb{R}^+$ such that

$$(\forall y \in X)(|y - x_0| < \delta \implies (|g(y, t, A) - g(x_0, t, A)| < \epsilon)). \quad (11.21)$$

By the transfer principle and the fact that $x_1 \approx x_2 \approx x_0$ we know that

$$(|{}^*g(x_1, t, {}^*A) - {}^*g(x_2, t, {}^*A)| < \epsilon). \tag{11.22}$$

As ϵ is arbitrary, this completes the proof. □

As Lemma 9.24, the next lemma establishes the link between ${}^*E$ and $\mathsf{st}^{-1}(E)$ for every $E \in \mathcal{B}[X]$.

Lemma 11.19. *Suppose (WF) holds. For any Borel set E, any $x \in \mathrm{NS}({}^*X)$ and any $t \in \mathbb{R}^+$ we have ${}^*g(x, t, {}^*E) \approx \overline{{}^*g}(x, t, \mathsf{st}^{-1}(E))$.*

Proof. The proof uses Lemma 11.18 and is similar to the proof of Lemma 9.24. □

Lemmas 11.18 and 11.19 allow us to obtain the result in Theorem 9.27 under weaker assumptions.

Theorem 11.20. *Suppose (DT) and (WF) hold. For any $s \in \mathrm{NS}(S)$, any non-negative $t \in \mathbb{Q}$ and any $E \in \mathcal{B}[X]$, we have $P^{(t)}_{\mathsf{st}(s)}(E) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(E) \cap S)$.*

Proof. The proof uses Lemmas 11.18 and 11.19 and is similar to the proof of Theorem 9.27. □

In order to extend the result in Theorem 11.20 to all non-negative $t \in \mathbb{R}$, we follow the same path as Section 8. Recall that we needed (OC):

Condition OC. The Markov chain $\{X_t\}$ is said to be *continuous in time* if for any open ball $U \subset X$ and any $x \in X$, we have $g(x, t, U)$ being a continuous function for $t > 0$.

Using the same proof as in Section 8, we obtain the following result.

Theorem 11.21. *Suppose (DT), (OC) and (WF) hold. For any $s \in \mathrm{NS}(S)$, any $t \in \mathrm{NS}(T)$ and any $E \in \mathcal{B}[X]$, we have $P^{(\mathsf{st}(t))}_{\mathsf{st}(s)}(E) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(E) \cap S)$.*

Thus, in conclusion, we have the following theorem.

Theorem 11.22. *Let $\{X_t\}_{t\geq 0}$ be a continuous time Markov process on a metric state space satisfying the Heine-Borel condition. Suppose $\{X_t\}_{t\geq 0}$ satisfies (DT), (OC) and (WF). Then there exists a hyperfinite Markov process $\{X'_t\}_{t\in T}$ with state space $S \subset {}^*X$ such that for all $s \in \mathrm{NS}(S)$ and all $t \in \mathrm{NS}(T)$*

$$(\forall E \in \mathcal{B}[X])(P^{(\mathsf{st}(t))}_{\mathsf{st}(s)}(E) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(E) \cap S)). \tag{11.23}$$

where P and G denote the transition probability of $\{X_t\}_{t\geq 0}$ and $\{X'_t\}_{t\in T}$, respectively.

This theorem shows that, given a standard Markov process, we can almost always use a hyperfinite Markov process to represent it. In [And76], Robert Anderson discussed such hyperfinite representation for Brownian motion. In this paper, we extend his idea to cover a large class of general Markov processes.

11.2. **A Weaker Markov Chain Ergodic Theorem.** In Section 10, we have shown the Markov chain Ergodic theorem under strong Feller condition. In this section, under Feller condition, we give a proof of a weaker form of the Markov Chain Ergodic theorem. In order to do this, we start by showing that $\{X'_t\}_{t\in T}$ inherits some key properties from $\{X_t\}_{t\geq 0}$.

Let π be a stationary distribution of $\{X_t\}_{t\geq 0}$. As in Definition 10.4, we define an internal probability measure π' on $(S, \mathcal{I}(S))$ by letting $\pi'(\{s\}) = \frac{{}^*\pi(B(s))}{{}^*\pi(\bigcup_{s'\in S} B(s'))}$ for every $s \in S$. By Lemma 10.5, for any $A \in \mathcal{B}[X]$ we have $\pi(A) = \overline{\pi'}(\mathsf{st}^{-1}(A) \cap S)$. This π' is a weakly stationary for some internal subsets of S.

Theorem 11.23. *Suppose (DT) and (WF) hold. There exists an infinite $t_0 \in T$ such that for every $A \in \mathcal{I}(S_1)$ and every $t \leq t_0$ we have*

$$\pi'(A_S) \approx \sum_{i\in S} \pi'(i) G^{(t)}_i (A_S). \tag{11.24}$$

where A_S is the enlargement of A.

Proof. The proof is similar to the proof of Theorem 10.6. We use Theorem 11.15 instead of Theorem 9.20. □

Condition CS. There exists a countable basis $\mathscr{B}$ of bounded open sets of X such that any finite intersection of elements from $\mathscr{B}$ is a continuity set with respect to π and $g(x,t,.)$ for all $x \in X$ and $t > 0$.

We shall fix this countable basis $\mathscr{B}$ for the remainder of this section. (CS) allows us to prove the following lemma.

Lemma 11.24. *Suppose (CS) holds. Then we have $\pi(O) = \pi'((^*O \cap S_1)_S)$ where O is a finite intersection of elements from $\mathscr{B}$.*

Proof. Let O be a finite intersection of elements of $\mathscr{B}$ and let $\overline{O}$ denote the closure of O. By the construction of π', we know that $\pi'(\mathsf{st}^{-1}(O) \cap S) = \pi(O) = \pi(\overline{O}) = \pi'(\mathsf{st}^{-1}(\overline{O}) \cap S)$. In order to finish the proof, it is sufficient to prove the following claim.

Claim 11.25. $\mathsf{st}^{-1}(O) \cap S \subset (^*O \cap S_1)_S \subset \mathsf{st}^{-1}(\overline{O}) \cap S$.

Proof. Pick any point $s \in \mathsf{st}^{-1}(O) \cap S$. Then $s \in B_1(s')$ for some $s' \in S_1$. Note also that $s \in \mu(y)$ for some $y \in O$. As O is open, we have $\mu(y) \subset {}^*O$ which implies that $B_1(s') \subset {}^*O$ which again implies that $s \in (^*O \cap S_1)_S$.

Now pick some point $y \in (^*O \cap S_1)_S$. Then $y \in B_1(y')$ for some $y' \in {}^*O \cap S_1$. As y is near-standard, we know that y' is near-standard hence $y' \in \mu(x)$ for some $x \in X$. Suppose $x \notin \overline{O}$. Then there exists an open ball $U(x)$ centered at x such that $U(x) \cap O = \emptyset$. This would imply that $y' \notin {}^*O$ which is a contradiction. Hence $x \in \overline{O}$. This means that $y \in \mu(x) \subset \mathsf{st}^{-1}(\overline{O})$, completing the proof. □

This finishes the proof of this lemma. □

In order to show that the hyperfinite Markov chain $\{X'_t\}_{t \in T}$ converges, we need to establish the strong regularity (at least for finite intersection of open balls) for $\{X'_t\}_{t \in T}$.

We first prove the following lemma which is analogous to Theorem 11.22.

Theorem 11.26. *Suppose (DT), (OC), (WF) and (CS) hold. For any $s \in \mathrm{NS}(S)$ and any $t \in \mathrm{NS}(T)$ we have $g(\mathsf{st}(s), \mathsf{st}(t), O) \approx G_s^{(t)}((^*O \cap S_1)_S)$ where O is a finite intersection of elements from $\mathscr{B}$.*

Proof. By Theorem 11.22, we know that $P_{\mathsf{st}(s)}^{\mathsf{st}(t)}(O) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(O) \cap S)$ and $P_{\mathsf{st}(s)}^{\mathsf{st}(t)}(\overline{O}) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(\overline{O}) \cap S)$ where $\overline{O}$ denote the closure of O. By (CS), we know that $P_{\mathsf{st}(s)}^{\mathsf{st}(t)}(O) = P_{\mathsf{st}(s)}^{\mathsf{st}(t)}(\overline{O})$. Then the result follows from Claim 11.25. $\square$

We now show that $\{X'_t\}$ is strong regular for open balls.

Lemma 11.27. *Suppose (DT), (OC), (WF) and (CS) hold. For every $s_1 \approx s_2 \in \mathrm{NS}(T)$, there exists an infinite $t_1 \in T$ such that $G_{s_1}^{(t)}((^*O \cap S_1)_S) \approx G_{s_2}^{(t)}((^*O \cap S_1)_S)$ for and all $t \leq t_1$ and all O which is a finite intersection of elements from $\mathscr{B}$.*

Proof. Pick $s_1 \approx s_2 \in \mathrm{NS}(S)$ and let O be a finite intersection of elements from $\mathscr{B}$. Let $x = \mathsf{st}(s_1) = \mathsf{st}(s_2)$. By Theorem 11.26, for any $t \in \mathrm{NS}(T)$, we know that $G_{s_1}^{(t)}((^*O \cap S_1)_S) \approx g(x, \mathsf{st}(t), O)$ and $G_{s_2}^{(t)}((^*O \cap S_1)_S) \approx g(x, \mathsf{st}(t), O)$. Hence we have $G_{s_1}^{(t)}((^*O \cap S_1)_S) \approx G_{s_2}^{(t)}((^*O \cap S_1)_S)$ for all $t \in \mathrm{NS}(T)$. Consider the following set

$$T_O = \{t \in T : |G_{s_1}^{(t)}((^*O \cap S_1)_S) - G_{s_2}^{(t)}((^*O \cap S_1)_S)| < \frac{1}{t}\}. \tag{11.25}$$

The set T_O contains all the near-standard $t \in T$ hence it contains an infinite $t_O \in T$ by overspill. As every countable descending infinite reals has an infinite lower bound, there exists an infinite t_1 which is smaller than every element in $\{t_O : O \in \mathscr{B}\}$. $\square$

By using essentially the same argument as in Theorem 7.19, we have the following result for $\{X'_t\}_{t \in T}$. The proof is omitted.

Theorem 11.28. *Suppose (DT), (OC), (WF) and (CS) hold. Suppose $\{X_t\}_{t \geq 0}$ is productively open set irreducible with stationary distribution π. Let π' be the internal probability measure defined in Theorem 11.23. Then for $\overline{\pi'}$-almost every*

$s \in S$ there exists an infinite $t' \in T$ such that

$$G_s^{(t)}((^*O \cap S_1)_S) \approx \pi'((^*O \cap S_1)_S) \tag{11.26}$$

for all infinite $t \leq t'$ and all O which is a finite intersection of elements from $\mathscr{B}$.

This immediately gives rise to the following standard result.

Lemma 11.29. *Suppose (DT), (OC), (WF) and (CS) hold. Suppose $\{X_t\}_{t\geq 0}$ is productively open set irreducible with stationary distribution π. Then for π-almost surely $x \in X$ we have $\lim_{t\to\infty} g(x,t,O) = \pi(O)$ for all O which is a finite intersection of elements from $\mathscr{B}$.*

Proof. Suppose not. Then there exist an set B and some O which is a finite intersection of elements from $\mathscr{B}$ with $\pi(B) > 0$ such that $g(x,t,O)$ does not converge to $\pi(O)$ for $x \in B$. Fix a $x_0 \in B$ and let s_0 be an element in S with $s_0 \approx x_0$. Then there exists an $\epsilon > 0$ and a unbounded sequence of real numbers $\{k_n : n \in \mathbb{N}\}$ with $|g(x_0,k_n,O) - \pi(O)| > \epsilon$ for all $n \in \mathbb{N}$. By Theorem 11.26 and Lemma 11.24, we have $|G_{s_0}^{(k_n)}((^*O \cap S_1)_S) - \pi'((^*O \cap S_1)_S)| > \epsilon$ for all $n \in \mathbb{N}$. Let t' be the same infinite element in T as in Theorem 11.28. By overspill, there is an infinite $t_0 < t'$ such that $|G_{s_0}^{(t_0)}((^*O \cap S_1)_S) - \pi'((^*O \cap S_1)_S)| > \epsilon$. As x_0 and s_0 are arbitrary, we have for every $s \in \mathsf{st}^{-1}(B) \cap S$ there is an infinite $t_s < t'$ such that $|G_{s_0}^{(t_s)}((^*O \cap S_1)_S) - \pi'((^*O \cap S_1)_S)| > \epsilon$. As $\overline{\pi'}(\mathsf{st}^{-1}(B) \cap S) = \pi(B)$, this contradicts with Theorem 11.28 hence completing the proof. $\square$

We now generalize the convergence to all Borel sets. We will need the following definition.

Definition 11.30 ([RS86, Page. 85])**.** Let P_n and P be probability measures on a metric space X with Borel σ-algebra $\mathcal{B}[X]$. A subclass $\mathcal{C}$ of $\mathcal{B}[X]$ is a convergence determining class if weak convergence P_n to P is equivalent to $P_n(A) \to P(A)$ for all P-continuity sets $A \in \mathcal{C}$.

For separable metric spaces, we have the following result.

Lemma 11.31 ([Mol05, Page. 416]). *Let P_n and P be probability measures on a separable metric space X with Borel σ-algebra $\mathcal{B}[X]$. A class $\mathcal{C}$ of Borel sets is a convergence determining class if $\mathcal{C}$ is closed under finite intersections and each open set in X is at most a countable union of elements in $\mathcal{C}$.*

Theorem 11.32. *Suppose $\{X_t\}_{t\geq 0}$ is vanishing in distance and its state space has the Heine-Borel property. Suppose (OC), (WF) and (CS) hold. Suppose $\{X_t\}_{t\geq 0}$ is productively open set irreducible with stationary distribution π. Then for π-almost surely $x \in X$ we have $P_x^{(t)}(\cdot)$ weakly converges to $\pi(\cdot)$.*

Proof. By Theorem 9.6 and Lemma 11.6, we know that $\{X_t\}_{t\geq 0}$ satisfies (DT). Let $\mathscr{B}'$ to be the smallest set containing $\mathscr{B}$ such that $\mathscr{B}'$ is closed under finite intersection. By Lemma 11.29, we know that $\lim_{t\to\infty} P_x^{(t)}(A) = \pi(A)$ for all $A \in \mathscr{B}'$. The theorem then follows from Lemma 11.31. $\square$

By using similar argument as in Theorem 2.16, we obtain the following theorem.

Theorem 11.33. *Suppose $\{X_t\}_{t\geq 0}$ is vanishing in distance and its state space is a separable σ-compact metric space. Suppose (OC), (WF) and (CS) hold. Suppose $\{X_t\}_{t\geq 0}$ is productively open set irreducible with stationary distribution π. Then for π-almost surely $x \in X$ we have $P_x^{(t)}(\cdot)$ weakly converges to $\pi(\cdot)$.*

As one can see, with Feller condition, we can only show that $\{X'_t\}_{t\in T}$ is strong regular for some particular class of sets. In order to prove some result like Theorem 10.16, we need $\{X'_t\}_{t\in T}$ to be strong regular on a larger class of sets.

Open Problem 2. *Suppose (WF) holds. Is it possible to pick a hyperfinite representation S_1 such that $G_x^{(t)}(A_S) \approx G_y^{(t)}(A_S)$ for all $x \approx y$, all $t \in T$ and all $A \in \mathcal{I}(S_1)$?*

12. Push-down Results

In Section 8, we discuss how to construct a corresponding hyperfinite Markov process for every standard general Markov processes satisfying certain conditions.

In this section, we discuss the reverse procedure of constructing stationary distributions and Markov processes from weakly stationary distributions and hyperfinite Markov processes. Generally, we begin with an internal measure on ${}^*X$ and use standard part map to push the corresponding Loeb measure down to X. We start this section by introducing the following classical result.

Theorem 12.1 ([Cut+95, Thm. 13.4.1])**.** *Let X be a Heine-Borel metric space equipped with Borel σ-algebra $\mathcal{B}[X]$. Let M be an internal probability measure defined on $({}^*X, {}^*\mathcal{B}[X])$. Let*

$$\mathcal{C} = \{C \subset X : \mathsf{st}^{-1}(C) \in \overline{{}^*\mathcal{B}[X]}\}. \tag{12.1}$$

Define a measure μ on the sets $\mathcal{C}$ by: $\mu(C) = \overline{M}(\mathsf{st}^{-1}(C))$. Then μ is the completion of a regular Borel measure on X.

Proof. We first show that the collection $\mathcal{C}$ is a σ-algebra. Obviously $\emptyset \in \mathcal{C}$. By Lemma 6.10, we know that $X \in \mathcal{C}$. We now show that it is closed under complement. Suppose $A \in \mathcal{C}$. It is easy to see that $\mathsf{st}^{-1}(A^c) = (\mathrm{NS}({}^*X) \setminus \mathsf{st}^{-1}(A))$. By Theorem 5.1 and the fact that $\overline{{}^*\mathcal{B}[X]}$ is a σ-algebra, $A^c \in \mathcal{C}$. We now show that $\mathcal{C}$ is closed under countable union. Suppose $\{A_i : i \in \mathbb{N}\}$ be a countable collection of pairwise disjoint elements from $\mathcal{C}$. It is easy to see that $\bigcup_{i\in\omega}(\mathsf{st}^{-1}(A_i) = \mathsf{st}^{-1}(\bigcup_{i\in\omega} A_i)$. As $\mathsf{st}^{-1}(A_i) \in \overline{{}^*\mathcal{B}[X]}$ for every $i \in \mathbb{N}$, we have $\mathsf{st}^{-1}(\bigcup_{i\in\omega} A_i) \in \overline{{}^*\mathcal{B}[X]}$. Hence $\bigcup_{i\in\omega} A_i \in \mathcal{C}$.

We now show that μ is a well-defined measure on $(X, \mathcal{C})$. Clearly $\mu(\emptyset) = 0$. Suppose $\{A_i\}_{i\in\omega}$ is a mutually disjoint collection from $\mathcal{C}$. We have

$$\mu(\bigcup_{i\in\omega} A_i) = \overline{M}(\mathsf{st}^{-1}(\bigcup_{i\in\omega} A_i)) = \overline{M}(\bigcup_{i\in\omega}(\mathsf{st}^{-1}(A_i))). \tag{12.2}$$

As A_i's are mutually disjoint, we know that $\mathsf{st}^{-1}(A_i)$'s are mutually disjoint. Thus,

$$\overline{M}(\bigcup_{i\in\omega}(\mathsf{st}^{-1}(A_i))) = \sum_{i\in\omega} \overline{M}(\mathsf{st}^{-1}(A_i)) = \sum_{i\in\omega} \mu(A_i). \tag{12.3}$$

This shows that μ is countably additive.

Finally we need to show that such μ is the completion of a regular Borel measure. By universal Loeb measurability (Theorems 5.1 and 5.9), we know that $\mathsf{st}^{-1}(B) \in \overline{{}^*\mathcal{B}[X]}$ for all $B \in \mathcal{B}[X]$. Consider any $B \in \mathcal{B}[X]$ such that $\mu(B) = 0$ and any $C \subset B$. It is clear that $\mathsf{st}^{-1}(C) \subset \mathsf{st}^{-1}(B)$. As the Loeb measure $\overline{M}$ is a complete measure, we know that $\overline{M}(\mathsf{st}^{-1}(C)) = 0$ since $\overline{M}(\mathsf{st}^{-1}(B)) = 0$. Thus we have $\mu(C) = 0$, completing the proof. □

Note that the measure μ constructed in Theorem 12.1 need not have the same total measure as M. For example, if the internal measure M concentrates on some infinite element then μ would be a null measure. However, if we require $\overline{M}(\mathrm{NS}({}^*X)) = \mathsf{st}(M({}^*X))$ then $\mu(X) = \mathsf{st}(M({}^*X))$. In particular, if M is an internal probability measure with $\overline{M}(\mathrm{NS}({}^*X)) = 1$ then μ is the completion of a regular Borel probability measure on X. Such μ is called a push-down measure of M and is denoted by M_p.

The following corollary is an immediate consequence of Theorem 12.1.

Corollary 12.2. *Let X be a Heine-Borel metric space equipped with Borel σ-algebra $\mathcal{B}[X]$ and let S_X be a hyperfinite representation of X. Let M be an internal probability measure defined on $(S_X, \mathcal{I}[S_X])$. Let*

$$\mathcal{C} = \{C \subset X : \mathsf{st}^{-1}(C) \cap S_X \in \overline{\mathcal{I}[S_X]}\}. \tag{12.4}$$

Then the push-down measure M_p on the sets $\mathcal{C}$ given by $M_p(C) = \overline{M}(\mathsf{st}^{-1}(C) \cap S_X)$ is the completion of a regular Borel measure on X.

The following theorem shows the close connection between an internal probability measure and its push-down measure under integration.

Lemma 12.3. *Let X be a metric space equipped with Borel σ-algebra $\mathcal{B}[X]$, let ν be an internal probability measure on $({}^*X, {}^*\mathcal{B}[X])$ with $\overline{\nu}(\mathrm{NS}({}^*X)) = 1$. let $f : X \to \mathbb{R}$ be a bounded measurable function. Define $g : \mathrm{NS}({}^*X) \to \mathbb{R}$ by $g(s) = f(\mathsf{st}(s))$. Then g is integrable with respect to $\overline{\nu}$ restricted to $\mathrm{NS}({}^*X)$ and we have $\int_X f \mathrm{d}\nu_p = \int_{\mathrm{NS}({}^*X)} g \, \mathrm{d}\overline{\nu}$.*

Proof. As $\overline{\nu}(\mathrm{NS}({}^*X)) = 1$, the push-down measure ν_p is a probability measure on $(X, \mathcal{B}[X])$. For every $n \in \mathbb{N}$ and $k \in \mathbb{Z}$, define $F_{n,k} = f^{-1}([\frac{k}{n}, \frac{k+1}{n}))$ and $G_{n,k} = g^{-1}([\frac{k}{n}, \frac{k+1}{n}))$. As f is bounded, the collection $\mathcal{F}_n = \{F_{n,k} : k \in \mathbb{Z}\} \setminus \{\emptyset\}$ forms a finite partition of X, and similarly for $\mathcal{G}_n = \{G_{n,k} : k \in \mathbb{Z}\} \setminus \{\emptyset\}$ and ${}^*X$. Note that $G_{n,k} = \mathsf{st}^{-1}(F_{n,k})$ for every $n \in \mathbb{N}$ and $k \in \mathbb{Z}$. By Lemma 6.10, $G_{n,k}$ is $\overline{\nu}$-measurable. For every $n \in \mathbb{N}$, define $\hat{f}_n : X \to \mathbb{R}$ and $\hat{g}_n : {}^*X \to \mathbb{R}$ by putting $\hat{f}_n = \frac{k}{n}$ on $F_{n,k}$ and $\hat{g}_n = \frac{k}{n}$ on $G_{n,k}$ for every $k \in \mathbb{Z}$. Thus $\hat{f}_n$ (resp., $\hat{g}_n$) is a simple (resp., *simple) function on the partition $\mathcal{F}_n$ (resp., $\mathcal{G}_n$). By construction $\hat{f}_n \leq f < \hat{f}_n + \frac{1}{n}$ and $\hat{g}_n \leq g < \hat{g}_n + \frac{1}{n}$. It follows that $\int_X f \mathrm{d}\nu_p = \lim_{n\to\infty} \int_X \hat{f}_n \mathrm{d}\nu_p$. By Theorem 12.1, we have $\overline{\nu}(G_{n,k}) = \nu_p(F_{n,k})$ for every $n \in \mathbb{N}$ and $k \in \mathbb{Z}$. Thus, for every $n \in \mathbb{N}$ and $k \in \mathbb{Z}$, we have

$$\int_X \hat{f}_n \mathrm{d}\nu_p = \frac{k}{n}\nu_p(F_{n,k}) = \frac{k}{n}\overline{\nu}(G_{n,k}) = \int_{\mathrm{NS}({}^*X)} \hat{g}_n \mathrm{d}\overline{\nu} \tag{12.5}$$

Hence we have $\lim_{n\to\infty} \int_{\mathrm{NS}({}^*X)} \hat{g}_n \mathrm{d}\overline{\nu}$ exists and $\int_{\mathrm{NS}({}^*X)} g \,\mathrm{d}\overline{\nu} = \int_X f \mathrm{d}\nu_p$, completing the proof. □

12.1. **Construction of Standard Markov Processes.** In Section 8, we discussed how to construct a hyperfinite Markov process from a standard Markov process. In this section, we discuss the reverse direction. Starting with a hyperfinite Markov process, we will construct a standard Markov process from it.

Let X be a metric space satisfying the Heine-Borel condition. Let S be a hyperfinite representation of ${}^*X$. Let $\{Y_t\}_{t\in T}$ be a hyperfinite Markov process on S with transition probability $G_s^{(t)}(\cdot)$ satisfying the following condition:

(1) For all $s_1, s_2 \in \mathrm{NS}(S)$ and all $t_1, t_2 \in \mathrm{NS}(T)$:

$$(s_1 \approx s_2 \wedge t_1 \approx t_2) \implies (\forall A \in \mathcal{I}[S]\overline{G}_{s_1}^{(t_1)}(A) = \overline{G}_{s_2}^{(t_2)}(A)) \tag{12.6}$$

(2)

$$(\forall s \in \mathrm{NS}(S))(\forall t \in \mathrm{NS}(T))(\overline{G}_s^{(t)}(\mathrm{NS}(S)) = 1). \tag{12.7}$$

For every $x \in X$, every $h \in \mathbb{R}^+$ and every $A \in \mathcal{B}[X]$, define

$$g(x,h,A) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(A) \cap S) \tag{12.8}$$

where $s \approx x$ and $t \approx h$. Such $g(x,h,A)$ is well-defined because of Eq. (12.6). By Theorem 12.1 and Eq. (12.7), it is easy to see that $g(x,h,.)$ is a probability measure on $(X, \mathcal{B}[X])$ for $x \in X$ and $h \in \mathbb{R}^+$. In fact, $g(x,h,\cdot)$ is the push-down measure of the internal probability measure $G_s^{(t)}(\cdot)$.

We would like to show that $\{g(x,h,.)\}_{x \in X, h \geq 0}$ is the transition probability measure of a Markov process on $(X, \mathcal{B}[X])$. We first recall Definition 4.17 and Theorem 4.18.

Definition 12.4. Suppose that $(\Omega, \overline{\Gamma}, \overline{P})$ is a Loeb space, that X is a Hausdorff space, and that f is a measurable (possibly external) function from Ω to X. An internal function $F : \Omega \to {}^*X$ is a lifting of f provided that $f = \mathsf{st}(F)$ almost surely with respect to $\overline{P}$.

Theorem 12.5 ([ACH97, Theorem 4.6.4])**.** *Let $(\Omega, \overline{\Gamma}, \overline{P})$ be a Loeb space, and let $f : \Omega \to \mathbb{R}$ be a measurable function. Then f is Loeb integrable if and only if it has a S-integrable lifting.*

We are now at the place to establish the following result.

Lemma 12.6. *Suppose $\{Y_t\}_{t \geq 0}$ satisfies Eqs. (12.6) and (12.7). Then for any $t_1, t_2 \in \mathrm{NS}(T)$, any $s_0 \in S$ and any $E \in \mathcal{B}[X]$, the internal transition probability $\overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E) \cap S)$ is a $\overline{G}_{s_0}^{(t_1)}(\cdot)$-integrable function of s.*

Proof. Fix $t_1, t_2 \in \mathrm{NS}(T)$, $s_0 \in \mathrm{NS}(S)$ and $E \in \mathcal{B}[X]$. By Eqs. (12.6) and (12.7), we know that $g(\mathsf{st}(s), \mathsf{st}(t_2), E) = \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E) \cap S)$ for all $s \in \mathrm{NS}(S)$. The proof will be finished by Theorem 4.18 and the following claim.

Claim 12.7. *The internal function ${}^*g(\cdot, \mathsf{st}(t_2), {}^*E) : S \mapsto {}^*[0,1]$ is a S-integrable lifting of $\overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E) \cap S) : S \mapsto {}^*[0,1]$ with respect to the internal probability measure $G_{s_0}^{(t_1)}(\cdot)$.*

Proof. As $G^{(t_1)}_{s_0}(\cdot)$ is an internal probability measure concentrating on a hyperfinite set, by Corollary 4.14, it is easy to see that ${}^*g(\cdot, \mathsf{st}(t_2), {}^*E)$ is S-integrable. As $g(\mathsf{st}(s), \mathsf{st}(t_2), E) = \overline{G}^{(t_2)}_s(\mathsf{st}^{-1}(E) \cap S)$, it is sufficient to show that ${}^*g(\cdot, \mathsf{st}(t_2), {}^*E)$ is a S-continuous function on $\mathrm{NS}(S)$. Pick some $x_1 \in X$ and $\epsilon \in \mathbb{R}^+$. Let $s_1 \in S$ be any element such that $s_1 \approx x_1$. Let $M = \{s \in S : (\forall A \in \mathcal{I}[S])(|G^{(t_2)}_s(A) - G^{(t_2)}_{s_1}(A)| < \epsilon\}$. By Eq. (12.6), M contains every element in S which is infinitesimally close to s_1. By overspill, there is a $\delta \in \mathbb{R}^+$ such that

$$(\forall s \in S)({}^*d(s, s_1) < \delta \implies (\forall A \in \mathcal{I}[S])(|G^{(t_2)}_s(A) - G^{(t_2)}_{s_1}(A)| < \frac{\epsilon}{2})). \tag{12.9}$$

This clearly implies that

$$(\forall s \in S)({}^*d(s, s_1) < \delta \implies (\forall E \in \mathcal{B}[X])(|\overline{G}^{(t_2)}_s(\mathsf{st}^{-1}(E) \cap S) - \overline{G}^{(t_2)}_{s_1}(\mathsf{st}^{-1}(E) \cap S)| < \epsilon)). \tag{12.10}$$

By the construction of $g(\cdot, \mathsf{st}(t_2), E)$, we have $|g(x, \mathsf{st}(t_2), E) - g(x_1, \mathsf{st}(t_2), E)| < \epsilon$ for all $x \in X$ such that $d(x, x_1) < \frac{\delta}{2}$. Hence $g(\cdot, \mathsf{st}(t_2), E)$ is a continuous function for every $x \in X$ which implies that ${}^*g(\cdot, \mathsf{st}(t_2), E)$ is S-continuous on $\mathrm{NS}(S)$. $\square$

$\square$

We now establish the following result on "Markov property" of $\overline{G}^{(t)}_s(\mathsf{st}^{-1}(E) \cap S)$.

Lemma 12.8. *Suppose $\{Y_t\}_{t \in T}$ satisfies Eqs. (12.6) and (12.7). For any $t_1, t_2 \in \mathrm{NS}(T)$, $s_0 \in \mathrm{NS}(S)$ and $E \in \mathcal{B}[X]$, we have*

$$\overline{G}^{(t_1+t_2)}_{s_0}(\mathsf{st}^{-1}(E) \cap S) \approx \int \overline{G}^{(t_2)}_s(\mathsf{st}^{-1}(E) \cap S)\overline{G}^{(t_1)}_{s_0}(\mathrm{d}s). \tag{12.11}$$

Proof. Pick some $E \in \mathcal{B}[X]$, some $s_0 \in \mathrm{NS}(S)$ and some $t_1, t_2 \in \mathrm{NS}(T)$. For any set $A \in \mathcal{I}[S]$ with $\mathsf{st}^{-1}(E) \cap S \subset A$, we have $\overline{G}^{(t_2)}_s(\mathsf{st}^{-1}(E) \cap S) \leq \overline{G}^{(t_2)}_s(A)$. Hence we have

$$\int \overline{G}^{(t_2)}_s(\mathsf{st}^{-1}(E) \cap S)\overline{G}^{(t_1)}_{s_0}(\mathrm{d}s) \leq \int \overline{G}^{(t_2)}_s(A)\overline{G}^{(t_1)}_{s_0}(\mathrm{d}s). \tag{12.12}$$

By Corollary 4.14, we have

$$\int \overline{G}_s^{(t_2)}(A)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) = \mathsf{st}(\int G_s^{(t_2)}(A)G_{s_0}^{(t_1)}(\mathrm{d}s)) = \mathsf{st}(G_{s_0}^{(t_1+t_2)}(A)) \tag{12.13}$$

Hence, we have

$$\int \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \leq \inf\{\mathsf{st}(G_{s_0}^{(t_1+t_2)}(A)) : \mathsf{st}^{-1}(E)\cap S \subset A \in \mathcal{I}[S]\}. \tag{12.14}$$

Similarly, we have

$$\int \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \geq \sup\{\mathsf{st}(G_{s_0}^{(t_1+t_2)}(B)) : \mathsf{st}^{-1}(E)\cap S \supset B \in \mathcal{I}[S]\}. \tag{12.15}$$

Hence, by the construction of Loeb measure, we have

$$\overline{G}_{s_0}^{(t_1+t_2)}(\mathsf{st}^{-1}(E)\cap S) \approx \int \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s). \tag{12.16}$$

□

We now establish the main result of this section.

Theorem 12.9. *Suppose $\{Y_t\}_{t\in T}$ satisfies Eqs. (12.6) and (12.7). Then for any $h_1, h_2 \in \mathbb{R}^+$, any $x_0 \in X$ and any $E \in \mathcal{B}[X]$ we have*

$$g(x_0, h_1+h_2, E) = \int g(x, h_2, E)g(x_0, h_1, \mathrm{d}x). \tag{12.17}$$

This means that the family of functions $\{g(x,h,\cdot)\}_{x\in X, h\geq 0}$ have the semi-group property.

Proof. Fix $h_1, h_2 \in \mathbb{R}^+$, $x_0 \in X$ and $E \in \mathcal{B}[X]$. Let $s_0 \in S$ be some element such that $s_0 \approx x_0$ and let $t_1, t_2 \in \mathrm{NS}(T)$ such that $t_1 \approx h_1$ and $t_2 \approx h_2$. By the construction of g and Lemma 12.8, we have

$$g(x_0, h_1+h_2, E) = \overline{G}_{s_0}^{(t_1+t_2)}(\mathsf{st}^{-1}(E)\cap S) = \int \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s). \tag{12.18}$$

By Eq. (12.6), we know that $g(x,h_2,E)=\overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)$ provided that $s\approx x$. In Claim 12.7, we know that $g(\cdot,h_2,E)$ is a continuous function hence we have ${}^*g(s,h_2,{}^*E)\approx \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)$ for all $s\in \mathrm{NS}(S)$.

Thus, by Lemma 12.3, we have

$$\int_S \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \tag{12.19}$$

$$= \int_{\mathrm{NS}(S)} \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \tag{12.20}$$

$$= \int_{\mathrm{NS}(S)} \mathsf{st}({}^*g(s,h_2,{}^*E))\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \tag{12.21}$$

$$= \int_{\mathrm{NS}(S)} g(\mathsf{st}(s),h_2,E)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \tag{12.22}$$

$$= \int_X g(x,h_2,E)g(x_0,h_1,\mathrm{d}x). \tag{12.23}$$

Note that the last step follows from Lemma 12.3. Hence we have the desired result. □

As the transition probabilities $\{g(x,h,.)\}_{x\in X,h\geq 0}$ have the semigroup property, we know that $\{g(x,h,.)\}_{x\in X,h\geq 0}$ defines a standard continuous-time Markov process on the state space X with Borel σ-algebra $\mathcal{B}[X]$. In fact, if we define $X:\Omega\times[0,\infty)\to X$ by $X(\omega,h)=\mathsf{st}(Y(\omega,h^+))$ where h^+ is the smallest element in T greater than or equal to h then $\{X_h\}_{h\geq 0}$ is a standard continuous-time Markov process obtained from pushing-down the hyperfinite Markov process $\{Y_t\}_{t\in T}$.

12.2. **Push down of Weakly Stationary Distributions.** Recall from Definition 7.5 that an internal probability measure π on $(S,\mathcal{I}[S])$ is a weakly stationary distribution if there is an infinite t_0 such that

$$(\forall t\leq t_0)(\forall A\in\mathcal{I}(S))(\pi(A)\approx\sum_{i\in S}\pi(\{i\})p^{(t)}(i,A)) \tag{12.24}$$

$p^{(t)}(i,A)$ denote the t-step internal transition probability of a hyperfinite Markov process.

In Section 12.1, we established how to construct a standard Markov process $\{X_t\}_{t\geq 0}$ on the state space X from a hyperfinite Markov process $\{Y_t\}_{t\in T}$ on a state space S satisfying certain properties. Note that S is a hyperfinite representation of X. It is natural to ask: if Π is a weakly stationary distribution of $\{Y_t\}_{t\in T}$, is the push-down Π_p a stationary distribution of $\{X_t\}_{t\geq 0}$? We will show that, if $\{Y_t\}$ satisfies Eqs. (12.6) and (12.7) then Π_p is a stationary distribution on $\{X_t\}_{t\geq 0}$.

For the remainder of this section, let $\{G_s^{(t)}(\cdot)\}_{s\in S,t\in T}$ denote the transition probabilities of $\{Y_t\}_{t\in T}$. Let $\{X_t\}_{t\geq 0}$ be the standard Markov process on the state space X constructed from $\{Y_t\}$ as in Section 12.1. Let $\{g(x,h,\cdot)\}_{x\in X,h\geq 0}$ denote the transition probabilities of $\{X_t\}_{t\geq 0}$. Moreover, let Π be a weakly stationary distribution of $\{Y_t\}_{t\in T}$ such that $\overline{\Pi}(\mathrm{NS}(S))=1$. Let Π_p be the push down measure of Π defined in Theorem 12.1. It is easy to see that Π_p is a probability measure on $(X,\mathcal{B}[X])$.

We first establish the following lemma.

Lemma 12.10. *Suppose $\{Y_t\}_{t\geq 0}$ satisfies Eqs. (12.6) and (12.7). Then for any $t\in \mathrm{NS}(T)$ and any $E\in\mathcal{B}[X]$, the transition probability $\overline{G}_s^{(t)}(\mathsf{st}^{-1}(E)\cap S)$ is a $\overline{\Pi}$-integrable function of s.*

Proof. The proof of this lemma is similar to Lemma 12.6. □

Lemma 12.11. *Suppose $\{Y_t\}_{t\in T}$ satisfies Eqs. (12.6) and (12.7). Then for any $t\in\mathrm{NS}(T)$ and any $E\in\mathcal{B}[X]$, we have*

$$\overline{\Pi}(\mathsf{st}^{-1}(E)\cap S)\approx\int G_s^{(t)}(\mathsf{st}^{-1}(E)\cap S)\overline{\Pi}(\mathrm{d}s). \tag{12.25}$$

Proof. The proof is similar to Lemma 12.8 □

We now show that the push-down measure of the weakly stationary distribution Π is a stationary distribution for $\{X_t\}_{t\geq 0}$.

Theorem 12.12. *Suppose $\{Y_t\}_{t\geq 0}$ satisfies Eqs. (12.6) and (12.7). Let Π be a weakly stationary distribution of $\{Y_t\}_{t\in T}$ with $\overline{\Pi}(\mathrm{NS}(S))=1$. Then the push-down measure Π_p of Π is a stationary distribution of $\{X_t\}_{t\geq 0}$.*

Proof. By Theorem 12.1 and the fact that $\overline{\Pi}(\mathrm{NS}(S)) = 1$, we know that Π_p is a probability measure on $(X, \mathcal{B}[X])$.

Fix $t_0 \in \mathbb{R}^+$ and $A \in \mathcal{B}[X]$. It is sufficient to show that $\Pi_p(A) = \int g(x, t_0, A)\Pi_p(\mathrm{d}x)$. Let t be any element in T such that $t \approx t_0$. By the construction of Π_p and Lemma 12.11, we have

$$\Pi_p(A) = \overline{\Pi}(\mathsf{st}^{-1}(A) \cap S) = \int G_s^{(t)}(\mathsf{st}^{-1}(A) \cap S)\overline{\Pi}(\mathrm{d}s). \tag{12.26}$$

By the construction of g, we know that $g(x, t_0, A) = G_s^{(t)}(\mathsf{st}^{-1}(A) \cap S)$ provided that $s \approx x$. By a similar argument as in Theorem 12.9, we have

$$\int_S \overline{G}_s^{(t)}(\mathsf{st}^{-1}(A) \cap S)\overline{\Pi}(\mathrm{d}s) \tag{12.27}$$

$$= \int_{\mathrm{NS}(S)} \mathsf{st}({}^*g(s, t_0, {}^*A))\overline{\Pi}(\mathrm{d}s) \tag{12.28}$$

$$= \int_X g(x, t_0, A)\Pi_p(\mathrm{d}x). \tag{12.29}$$

Hence completing the proof. □

Suppose we start with a standard Markov process $\{X_t\}_{t\geq 0}$ satisfying (DT), (SF) and (WC). Note that such $\{X_t\}_{t\geq 0}$ may not necessarily have a stationary distribution. An simple example of such $\{X_t\}_{t\geq 0}$ is Brownian motion. The hyperfinite representation $\{X'_t\}_{t\in T}$ of $\{X_t\}_{t\geq 0}$ satisfies Eqs. (12.6) and (12.7). Thus, if there is a weakly stationary distribution Π of $\{X'_t\}_{t\in T}$ with $\overline{\Pi}(\mathrm{NS}(S_X)) = 1$ then there is a stationary distribution of $\{X_t\}_{t\geq 0}$. This provides an alternative approach for establishing the existence of stationary distributions for standard Markov processes. This will be discussed in detail in the next section.

12.3. **Existence of Stationary Distributions.** The existence of stationary distribution for discrete-time Markov processes with finite state space is well-understood (e.g [Ros06, Section 8.4]). The situation is much more complicated for Markov processes with non-finite state spaces. The stationary distribution may not exist at all even for well-behaved Markov processes (e.g Brownian motion). By using the method developed in this paper, we consider the hyperfinite counterpart of the original general-state space Markov process $\{X_t\}_{t\geq 0}$. Assuming the state space is compact, we show that a stationary distribution exists under mild regularity conditions.

We start by quoting the following results for finite-state space discrete-time Markov processes.

Definition 12.13. A $n \times n$ matrix $\mathbb{P}$ is regular if some power of $\mathbb{P}$ has only positive entries.

Theorem 12.14. *Let $\mathbb{P}$ be the transition matrix of some finite-state space discrete-time Markov process $\{Y_t\}_{t\in\mathbb{N}}$. Suppose $\mathbb{P}$ is regular. Then there exists a matrix $\mathbb{W}$ with all rows the same vector $\boldsymbol{w}$ such that $\lim_{n\to\infty} \mathbb{P}^n = \mathbb{W}$. Moreover, $\boldsymbol{w}$ is the unique stationary distribution of $\{Y_t\}_{t\in\mathbb{N}}$.*

Definition 12.15. A $n \times n$ matrix $\mathbb{P}$ is irreducible if for every pair of $i, j \leq n$ there is $n_{ij} \in \mathbb{N}$ such that the (i, j)-th entry of $\mathbb{P}^{n_{ij}}$ is positive.

The following theorem give a sufficient condition for $\mathbb{P}$ being regular.

Theorem 12.16. *Let $\mathbb{P}$ be the transition matrix of some finite-state space discrete-time Markov process $\{Y_t\}_{t\in\mathbb{N}}$. If $\mathbb{P}$ is irreducible and at least one element in the diagonal of $\mathbb{P}$ is positive, then $\mathbb{P}$ is regular.*

For an arbitrary hyperfinite Markov process, we can form its transition matrix as we did for finite Markov process.

Definition 12.17. Let $K \in {}^*\mathbb{N}$. A $K \times K$ (hyperfinite) matrix $\mathbb{P}$ is *regular if some hyperfinite power of $\mathbb{P}$ has only positive entries. A $K \times K$ matrix $\mathbb{P}$ is *irreducible if for any $i, j \leq K$ there is $n_{ij} \in {}^*\mathbb{N}$ such that the (i, j)-th entry of $\mathbb{P}^{n_{ij}}$ is positive.

Similarly, we have the following result for hyperfinite Markov processes.

Theorem 12.18. *Let $\mathbb{P}$ be the hyperfinite transition matrix for some hyperfinite Markov process $\{Y_t\}_{t\in T}$ with state space S. Suppose $\mathbb{P}$ is *regular. Then there* *exists a unique* **stationary distribution* Π *for* $\{Y_t\}_{t\in T}$*, i.e for every* $s \in S$*, we have* $\Pi(\{s\}) = \sum_{k\in S} \Pi(\{k\}) P^{(\delta t)}_{ks}$.

Proof. The proof follows from the transfer of Theorem 12.14. □

Note that if Π is *stationary then Π is weakly stationary as in Definition 7.5. The following theorem gives a sufficient condition for regularity of $\mathbb{P}$.

Theorem 12.19. *Let* $\mathbb{P}$ *be the transition matrix of some hyperfinite Markov process* $\{Y_t\}_{t\in T}$ *with state space* S*. If* $\mathbb{P}$ *is* **irreducible and at least one element in the diagonal of* $\mathbb{P}$ *is positive, then* $\mathbb{P}$ *is* **regular.*

By *irreducible, we simply mean that for any $i, j \in S$ there exists $n \in {}^*\mathbb{N}$ such that $P^{(n)}_{ij} > 0$. The proof of this theorem follows from transfer of Theorem 12.16.

We now turn our attention to standard continuous-time Markov process $\{X_t\}_{t\geq 0}$ and its corresponding hyperfinite Markov process $\{X'_t\}_{t\in T}$. We have the following result:

Theorem 12.20. *Let* $\{X_t\}_{t\geq 0}$ *be a Markov process on a compact metric space* X *and let* $\{X'_t\}_{t\in T}$ *be a hyperfinite Markov process on* S_X *satisfying Eq.* (11.23)*. Let* $\mathbb{P}$ *be the hyperfinite transition matrix of* $\{X'_t\}_{t\in T}$*. If* $\mathbb{P}$ *is* **regular, then there exists a stationary distribution for* $\{X_t\}_{t\geq 0}$*.*

Proof. By Theorem 12.18, there exists a unique *stationary distribution Π for $\{X'_t\}_{t\in T}$. Let Π_p denote the push-down measure of Π. As X is compact, by Theorem 12.12, Π_p is a stationary distribution of $\{X_t\}_{t\geq 0}$. □

Given a standard Markov process $\{X_t\}_{t\geq 0}$. It is not difficult to find the hyperfinite transition matrix of $\{X'_t\}_{t\in T}$. Thus Theorem 12.20 provides a way to look for stationary distributions.

Example 12.21 (Brownian motion)**.** Let $\{X_t\}_{t\geq 0}$ be the standard Brownian motion. Clearly $\{X_t\}_{t\geq 0}$ satisfies all the conditions in Theorem 9.35. Let $\{X'_t\}_{t\in T}$ be the corresponding hyperfinite Markov process. The transition matrix of $\{X'_t\}_{t\in T}$ is regular (in fact $G_{s_1}^{(\delta t)}(\{s_2\}) > 0$ for all $s_1, s_2 \in S$). By Theorem 12.18, there exists a *stationary distribution Π of $\{X'_t\}_{t\in T}$.

Standard Brownian motion does not have a stationary distribution. It does have a stationary measure which is the Lebesgue measure on $\mathbb{R}$. From a nonstandard prospective, as we can see from this example, there exists a *stationary distribution of $\{X'_t\}_{t\in T}$. However, this *statioanry distribution will concentrate on the infinite portion of $^*\mathbb{R}$ since otherwise its push-down will be a stationary distribution for the standard Brownian motion.

13. Merging of Markov Processes

In Section 10, we discussed the total variance convergence of the transition probabilities to stationary distributions for Markov processes satisfying certain properties. In particular, we required our Markov chain to be productively open set irreducible and to satisfy (DT), (SF), (OC) and (CS). However, such Markov processes do not necessarily have a stationary distribution. A simple example is standard Brownian motion. However, the transition probabilities of the standard Brownian motion "merge" in the following sense.

Definition 13.1. A Markov process $\{X_t\}_{t\geq 0}$ has the merging property if for every two points $x, y \in X$, we have

$$\lim_{t\to\infty} \parallel P_x^{(t)}(\cdot) - P_y^{(t)}(\cdot) \parallel = 0 \tag{13.1}$$

where $P_x^{(t)}(\cdot)$ denotes the transition measure and $\parallel P_x^{(t)}(\cdot) - P_y^{(t)}(\cdot) \parallel$ denotes the total variation distance between $P_x^{(t)}(\cdot)$ and $P_y^{(t)}(\cdot)$.

Saloff-Coste and Zúñiga [SCZ11] discuss the merging property for time-inhomogeneous finite Markov processes. In this section, we focus on time-homogeneous general Markov processes. For merging result of general probability measures, see [DDF88].

In this section, we give sufficient conditions to ensure that Markov processes have the merging property. The following definition is analogous to Definition 7.11.

Definition 13.2. Given a Markov process $\{X_t\}_{t\geq 0}$ on some state space X and fix some $x_1, x_2 \in X$. An element $(y_1, y_2) \in X \times X$ is an absorbing point of (x_1, x_2) if for all $n \in \mathbb{N}$

$$Q_{(x_1,x_2)}(\exists t\ Z_t \in U(y_1, \frac{1}{n}) \times U(y_2, \frac{1}{n})) = 1. \tag{13.2}$$

where Q denote the probability measure of the product Markov chain $\{Z_t\}_{t\geq 0}$ of $\{X_t\}_{t\geq 0}$ and a i.i.d copy of $\{X_t\}_{t\geq 0}$, and $U(y, \frac{1}{n})$ is the open ball centered at y with radius $\frac{1}{n}$.

Fix an infinitesimal ϵ_0 such that $\epsilon_0 \cdot (\frac{t}{\delta t}) \approx 0$ for all $t \in T$. As in Section 9, we construct a hyperfinite Markov process $\{X'_t\}_{t\in T}$ on some (δ_0, r_0)-hyperfinite representation of ${}^*X$ where δ_0 and r_0 are chosen with respect to this ϵ_0. Moreover, by Proposition 3.12 and Theorem 6.6, we can assume our hyperfinite state space S contains every $x \in X$. The hyperfinite transition probabilities for $\{X'_t\}_{t\in T}$ are defined in the same way as in the paragraph before Lemma 8.12 and are denoted by $\{G_i^{(t)}(\cdot)\}_{i\in S, t\in T}$.

Lemma 13.3. *Suppose $\{X_t\}_{t\geq 0}$ satisfies (DT), (SF) and (OC). Suppose $(y_1, y_2) \in X \times X$ is an absorbing point of some $x_1, x_2 \in X$. Then (y_1, y_2) is a near-standard absorbing point of x_1, x_2 for the hyperfinite Markov chain $\{X'_t\}_{t\in T}$.*

Proof. As $\{X_t\}_{t\geq 0}$ satisfies (DT), (SF) and (OC), by Theorem 9.35, we have

$$P_{\mathsf{st}(s)}^{(\mathsf{st}(t))}(E) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S) \tag{13.3}$$

hence implies that $\{X'_t\}_{t\in T}$ satisfies Eqs. (12.6) and (12.7). Let $\{X_t^p\}_{t\geq 0}$ denote the standard Markov process obtained from pushing down $\{X'_t\}_{t\in T}$ as in Section 12.1. By the construction of $\{X_t^p\}_{t\geq 0}$, we know that $p_x^{(t)}(E) = P_x^{(t)}(E)$ for all $x \in X$, $t \geq 0$ and $E \in \mathcal{B}[X]$ where p and P denote the probability measure for $\{X_t^p\}_{t\geq 0}$ and $\{X_t\}_{t\geq 0}$, respectively.

Now fix some $x_1, x_2 \in X$. There exists $(y_1, y_2) \in X \times X$ which is an absorbing point for x_1, x_2. Fix an open ball $U_1 \times U_2$ centered at (y_1, y_2). By Definition 13.2, we know that $Q_{(x_1,x_2)}(\exists t > 0\ Z_t \in U_1 \times U_2) = 1$. This implies that

$$q_{(x_1,x_2)}(\exists t > 0\ Z_t^p \in U_1 \times U_2) = 1 \tag{13.4}$$

where q denote the probability measure of the product Markov chain $\{Z_t^p\}_{t\geq 0}$ obtained from $\{X_t^p\}_{t\geq 0}$ and its i.i.d copy. By the construction of $\{X_t^p\}_{t\geq 0}$, we know that

$$\overline{F}_{(x_1,x_2)}(\exists t \in \mathrm{NS}(T)\ Z'_t \in (\mathsf{st}^{-1}(U_1) \times \mathsf{st}^{-1}(U_2)) \cap (S \times S)) = 1 \tag{13.5}$$

where F denote the probability measure of the product hyperfinite Markov chain $\{Z'_t\}_{t\in T}$ obtained from $\{X'_t\}_{t\in T}$ and its i.i.d copy. As $\mathsf{st}^{-1}(U) \subset {}^*U$ for any open set U, we know that $\overline{F}_{(x_1,x_2)}(\exists t \in \mathrm{NS}(T)\ Z'_t \in ({}^*U_1 \times {}^*U_2) \cap (S \times S)) = 1$. As our choice of $U_1 \times U_2$ is arbitrary, this shows that (y_1, y_2) is a near-standard absorbing point of x_1, x_2. $\square$

The proof of the following theorem is similar to the proof of Theorem 7.19.

Theorem 13.4. *Suppose $\{X_t\}_{t\geq 0}$ satisfies (DT), (SF) and (OC) and for every $x_1, x_2 \in X$ there exists a absorbing point $(y, y) \in X \times X$. Then for every $x_1, x_2 \in X$, every infinite $t \in T$ and every $A \in {}^*\mathcal{B}[X]$ we have $G_{x_1}^{(t)}(A) \approx G_{x_2}^{(t)}(A)$.*

Proof. Let $\{X'_t\}_{t\in T}$ be a corresponding hyperfinite Markov chain of $\{X_t\}_{t\geq 0}$. Let $\{Y_t\}_{t\in T}$ be a i.i.d copy of $\{X'_t\}_{t\in T}$ and let $\{Z_t\}_{t\in T}$ denote the product hyperfinite Markov chain of $\{X'_t\}_{t\in T}$ and $\{Y_t\}_{t\in T}$. We use G' and $\overline{G'}$ for the internal probability and Loeb probability of $\{Z_t\}_{t\in T}$.

Fix $x_1, x_2 \in X$. By assumption, there exists a standard absorbing point y. Pick an infinite $t_0 \in T$ and fix some internal set $A \subset S$. Define

$$M = \{\omega : \exists t < t_0 - 1, X'_t(\omega) \approx Y_t(\omega) \approx y\}. \tag{13.6}$$

By Lemma 13.3, for all $n \in \mathbb{N}$, we have

$$\overline{F}_{(x_1,x_2)}(\exists t \in \mathrm{NS}(T)\ Z'_t \in ({}^*U(y, \frac{1}{n}) \times {}^*U(y, \frac{1}{n})) \cap (S \times S)) = 1. \tag{13.7}$$

where F denote the internal transition probability for the product hyperfinite Markov chain $\{Z'_t\}_{t\in T}$ obtained from $\{X'_t\}_{t\in T}$ and its i.i.d copy. By Lemma 7.8, we know that $\overline{F}_{(x_1,x_2)}(M)=1$. By Theorem 9.25, we know that $\{X'_t\}_{t\in T}$ is strong regular. Thus we have:

$$|\overline{G}^{(t_0)}_{x_1}(A)-\overline{G}^{(t)}_{j}(A)| \tag{13.8}$$

$$=|\overline{F}_{(x_1,x_2)}(X'_{t_0}\in A)-\overline{F}_{(x_1,x_2)}(Y_{t_0}\in A)| \tag{13.9}$$

$$=|\overline{F}_{(x_1,x_2)}((X'_{t_0}\in A)\cap M)-\overline{F}_{(x_1,x_2)}((Y_{t_0}\in A)\cap M)| \tag{13.10}$$

$$=0. \tag{13.11}$$

□

We now establish the following merging result for the standard Markov process $\{X_t\}_{t\geq 0}$.

Theorem 13.5. *Suppose $\{X_t\}_{t\geq 0}$ satisfies (DT), (SF) and (OC) and for every $x_1,x_2\in X$ there exists a standard absorbing point y. Then $\{X_t\}_{t\geq 0}$ has the merging property.*

Proof. Pick a real $\epsilon>0$ and fix two standard $x_1,x_2\in X$. By Theorem 13.4, we know that $|G^{(t)}_{x_1}(A)-G^{(t)}_{x_2}(A)|<\epsilon$ for all infinite $t\in T$ and all $A\in {}^*\mathcal{B}[X]$. Let $M=\{t\in T:(\forall A\in {}^*\mathcal{B}[X])(|G^{(t)}_{x_1}(A)-G^{(t)}_{x_2}(A)|<\epsilon)\}$. By the underspill principle, there exists a $t_0\in \mathrm{NS}(T)$ such that $|G^{(t_0)}_{x_1}(A)-G^{(t_0)}_{x_2}(A)|<\epsilon$ for all $A\in {}^*\mathcal{B}[X]$. Pick a standard $t_1>t_0$ and let $t_2\in T$ be the first element greater than t_1.

Claim 13.6. $|G^{(t_2)}_{x_1}(A)-G^{(t_2)}_{x_2}(A)|<\epsilon$ *for all* $A\in {}^*\mathcal{B}[X]$.

Proof. Pick $t_3\in T$ such that $t_0+t_3=t_2$ and any $A\in {}^*\mathcal{B}[X]$. Then we have

$$|G^{(t_2)}_{x_1}(A)-G^{(t_2)}_{x_2}(A)| \tag{13.12}$$

$$\approx|\sum_{y\in S}G^{(t_1)}_{x_1}(\{y\})G^{(t_2)}_{y}(A)-\sum_{y\in S}G^{(t_1)}_{x_2}(\{y\})G^{(t_2)}_{y}(A)| \tag{13.13}$$

Let $f(y) = G_y^{(t_2)}(A)$. By the internal definition principle, we know that $G_y^{(t_2)}(A)$ is an internal function with value between ${}^*[0,1]$. By Lemma 7.24, we know that

$$|G_{x_1}^{(t_2)}(A) - G_{x_2}^{(t_2)}(A)| \lessapprox \| G_{x_1}^{(t_1)}(\cdot) - G_{x_2}^{(t_1)}(\cdot) \| . \tag{13.14}$$

Since this is true for all internal A, we have established the claim. $\square$

By the construction of Loeb measure, we know that

$$(\forall B \in \mathcal{B}[X])(|\overline{G}_{x_1}^{(t_2)}(\mathsf{st}^{-1}(B) \cap S) - \overline{G}_{x_2}^{(t_2)}(\mathsf{st}^{-1}(B) \cap S)| < \epsilon). \tag{13.15}$$

By Theorem 9.35 and the fact that $t_2 \approx t_1$, we know that $|P_{x_1}^{(t_1)}(B) - P_{x_2}^{(t_1)}(B)| < \epsilon$ for all $B \in \mathcal{B}[X]$. This shows that $\{X_t\}_{t \geq 0}$ has the merging property. $\square$

14. Miscellaneous Remarks

(1) There has been a rich literature on hyperfinite representations. In this paper, we cut ${}^*X$ into hyperfinitely "small" pieces (denoted by $\{B(s) : s \in S_X\}$) such that ${}^*g(x, 1, A) \approx g(y, 1, A)$ for all $A \in {}^*\mathcal{B}[X]$ for if x and y are in the same "small" piece $B(s)$. This also depends on (DSF) which states that the transition probability is a continuous function of starting points with respect to total variation norm. In [Loe74], Loeb showed that, for any Hausdorff topological space X, there is a hyperfinite partition $\mathcal{B}_F$ of ${}^*X$ consisting of *Borel sets which is finer than any finite Borel-measurable partition of X. That is, there exists $N \in {}^*\mathbb{N}$ and $\{A_i : i \leq N\} \in \mathscr{P}({}^*\mathcal{B}[X])$ such that

- For any $i, j \leq N$, we have $A_i \neq \emptyset$ and $A_i \cap A_j = \emptyset$.
- ${}^*X = \bigcup_{i \leq N} A_i$.
- For every bounded measurable function f, we have

$$\sup_{x \in A_i} {}^*f(x) - \inf_{x \in A_i} {}^*f(x) \approx 0 \tag{14.1}$$

for every $i \leq N$.

Now consider a discrete-time Markov process with state space X. There is a hyperfinite set $S \subset {}^*X$ and a hyperfinite partition $\{B(s) : s \in S\}$ of ${}^*X$ consisting of *Borel such that for all $s \in S$, any $x, y \in B(S)$ and any $A \in \mathcal{B}[X]$ we have $|{}^*g(x,1,{}^*A) - {}^*g(y,1,{}^*A)| \approx 0$. However, it is not clear whether $|{}^*g(x,1,B) - {}^*g(y,1,B)| \approx 0$ for all $B \in {}^*\mathcal{B}[X]$. A affirmative answer to this question may imply that (DSF) can be eliminated in establishing the Markov chain ergodic theorem for discrete-time Markov processes.

(2) The following nonstandard measure theoretical question is related to the previous point. Let X be a topological space and let $(X, \mathcal{B}[X])$ be a Borel-measurable space. The question is: is an internal probability measure on $({}^*X, {}^*\mathcal{B}[X])$ determined by its value on $\{{}^*A : A \in \mathcal{B}[X]\}$? For nonstandard extensions of standard probability measures on $(X, \mathcal{B}[X])$, the answer is affirmative by the transfer principle. For general internal probability measures on $({}^*X, {}^*\mathcal{B}[X])$, the answer is false. We can have two internal probability measures concentrating on two different infinitesimals. They are very different internal measures but they agree on the nonstandard extensions of all standard Borel sets. We are interested in the case in between.

Open Problem 3. *Let X be a topological space and let $(X, \mathcal{B}[X])$ be a Borel-measurable space. Let P be a probability measure on $(X, \mathcal{B}[X])$ and let P_1 be an internal probability measure on $({}^*X, {}^*\mathcal{B}[X])$. Suppose $P_1({}^*A) \approx {}^*P({}^*A)$ for all $A \in \mathcal{B}[X]$, is it true that $\overline{P_1} = \overline{{}^*P}$?*

We do have the following partial result.

Lemma 14.1. *Let us consider $([0,1], \mathcal{B}[[0,1]])$ and let P be a probability measure on it. Let P_1 be an internal probability measure on $({}^*[0,1], {}^*\mathcal{B}[[0,1]])$ such that $P_1({}^*A) \approx {}^*P({}^*A)$ for all $A \in \mathcal{B}[[0,1]]$. Then $\overline{P_1}(I) = \overline{{}^*P}(I)$ where I is an interval contained in ${}^*[0,1]$.*

Proof. It is easy to see that $\overline{P_1} = \overline{{}^*P}$ if P has countable support. Suppose P has uncountable support. Then there is an interval $[a,b] \subset [0,1]$ such that $P([a,b]) > 0$ and $P(\{x\}) = 0$ for all $x \in [a,b]$. Thus, without loss of generality, we can assume P is non-atomic on $[0,1]$. Let $(x,y) \subset {}^*[0,1]$ be a *interval with infinitesimal length. There is a $a \in [0,1]$ such that $(x,y) \subset {}^*(a, a+\frac{1}{n})$ for all $n \in \mathbb{N}$. As $\lim_{n\to\infty} P((a, a+\frac{1}{n})) = 0$, we know that $P_1((x,y)) \approx 0$. Pick $x_1, x_2 \in {}^*[0,1]$. Without loss of generality, we can assume $x_1 < x_2$. We then have $P_1((x_1,x_2)) \approx P_1((\mathsf{st}(x_1), \mathsf{st}(x_2)) \approx {}^*P((\mathsf{st}(x_1), \mathsf{st}(x_2)) \approx {}^*P((x_1,x_2))$. □

It should not be too hard to extend this lemma to more general metric spaces. Note that the collection of *intervals forms a basis of ${}^*[0,1]$. An affirmative answer to Open Problem 3 may follow from a variation of Theorem 9.37.

(3) It is possible to weaken the conditions mentioned in the Markov chain ergodic theorem (Theorem 10.16). In particular, it would be interesting to reduce (SF) to (WF). In Section 11, we constructed a hyperfinite representation $\{X'_t\}_{t\in T}$ of $\{X_t\}_{t\geq 0}$ under the Feller condition. The problem with the Markov chain ergodic theorem is: we do not know whether $\{X'_t\}_{t\in T}$ is strong regular. Recall that $\{X'_t\}_{t\in T}$ is strong regular if for any $A \in \mathcal{I}[S]$, any $i,j \in \mathrm{NS}(S)$ and any $t \in T$ we have:

$$(i \approx j) \implies (G_x^{(t)}(A) \approx G_y^{(t)}(A)). \tag{14.2}$$

where S denotes the state space of $\{X'_t\}_{t\in T}$. This is related to the following question: Suppose $\{X_t\}_{t\geq 0}$ satisfies (WF). For any $B \in {}^*\mathcal{B}[X]$, any $x,y \in \mathrm{NS}({}^*X)$ and any $t \in T$, is it true that ${}^*g(x,t,B) \approx {}^*g(y,t,B)$? An affirmative answer of this question will imply that $\{X'_t\}_{t\in T}$ is strong regular. By the transfer of (WF), it is not hard to see that ${}^*g(x,t,{}^*A) \approx {}^*g(y,t,{}^*A)$ for all $x \approx y \in \mathrm{NS}({}^*X)$, all $t \in \mathbb{R}^+$ and all $A \in \mathcal{B}[X]$. Thus, an affirmative answer to Open Problem 3 should allow us to reduce (SF) to (WF) in the Markov chain ergodic theorem (Theorem 10.16).

(4) In Section 11.2, we showed that the transition probability converges to the stationary distribution weakly. We achieve this by showing that the transition probability converges to the stationary distribution for every open ball which is also a continuity set. It is reasonable to expect such convergence holds for all open balls, even all open sets. Such a result will "almost" imply the Markov chain ergodic theorem by the following result.

Lemma 14.2. *Let $(X, \mathcal{T})$ be a topological space and let $(X, \mathcal{B}[X])$ be a Borel-measurable space. Let $\{P_n : n \in \mathbb{N}\}$ and P be Radon probability measures on $(X, \mathcal{B}[X])$. Suppose*

$$\lim_{n\to\infty} \sup_{U\in\mathcal{T}} |P_n(U) - P(U)| = 0. \tag{14.3}$$

Then $(P_n : n \in \mathbb{N})$ converges to P in total variation distance.

Proof. Pick $\epsilon > 0$. There is a $n_0 \in \mathbb{N}$ such that $\sup_{U\in\mathcal{T}} |P_n(U) - P(U)| < \frac{\epsilon}{4}$ for all $n > n_0$. Let $\mathcal{K}(X)$ denote the collection of compact subsets of X. Then we have $\sup_{K\in\mathcal{K}(X)} |P_n(K) - P(K)| < \frac{\epsilon}{4}$ for all $n > n_0$. Fix $B \in \mathcal{B}[X]$ and $n_1 > n_0$. Without loss of generality, we can assume that $P_{n_1}(B) \geq P(B)$. As P_{n_1} is Radon, we can choose K compact, U open with $K \subset B \subset U$ such that $P_{n_1}(U) - P_{n_1}(K) < \frac{\epsilon}{4}$. We then have

$$|P_{n_1}(B) - P(B)| \tag{14.4}$$

$$\leq |P_{n_1}(U) - P(K)| \tag{14.5}$$

$$\leq |P_{n_1}(U) - P_{n_1}(K)| + |P_{n_1}(K) - P(K)| \tag{14.6}$$

$$\leq \frac{\epsilon}{2}. \tag{14.7}$$

This implies that $\sup_{B\in\mathcal{B}[X]} |P_{n_1}(B) - P(B)| < \epsilon$. Thus we have $(P_n : n \in \mathbb{N})$ converges to P in total variation distance. ∎

Note that the lemma remains true if we replace convergence in total variation by $\lim_{n\to\infty} P_n(A) = P(A)$ both in condition and conclusion.

(5) Discrete-time Markov processes with finite state space can be characterized by its transition matrix. The same is true for hyperfinite Markov processes. The Markov chain ergodic theorem as well as the existence of stationary distribution are well understood for discrete-time Markov processes with finite state space. In Theorem 12.20, we establish a existence of stationary distribution result for general Markov processes via studying its hyperfinite counterpart. Let $\{X_t\}_{t\geq 0}$ be a standard Markov process and let $\{X'_t\}_{t\in T}$ be its hyperfinite representation. Under moderate conditions, we showed that there is a *stationary distribution Π for $\{X'_t\}_{t\in T}$. Note that every *stationary distribution is a weakly stationary distribution. By Theorem 7.26, under those conditions in Theorem 10.16, we know that the internal transition probability of $\{X'_t\}_{t\in T}$ converges to the *stationary distribution Π. This shows that the Loeb extension of Π is the same as the Loeb extension of any other weakly stationary distributions. However, it seems that a weakly stationary distribution would differ from a *stationary distribution in general. We raise the following two questions.

Open Problem 4. *Is there an example of a hyperfinite Markov process where its *stationary distribution differs from some of its weakly stationary distribution?*

Open Problem 5. *Is there an example of a hyperfinite Markov process where the internal transition probability does not converge to the *stationary distribution in the sense of Theorem 7.26?*

(6) For general state space continuous-time Markov processes, the Markov chain ergodic theorem applies to Harris recurrent chains. A Harris chain is a Markov chain where the chain returns to a particular part of the state space infinitely many times.

Definition 14.3. Let $\{X_t\}_{t\geq 0}$ be a Markov process on a general state space X. The Markov chain $\{X_t\}$ is Harris recurrent if there exists $A \subset X$, $t_0 > 0$, $0 < \epsilon < 1$, and a probability measure μ on X such that

- $P(\tau_A < \infty | X_0 = x) = 1$ for all $x \in X$ where τ_A denotes the stopping time to set A.
- $P_x^{(t_0)}(B) > \epsilon\mu(B)$ for all measurable $B \subset X$ and all $x \in A$.

The set A is called a small set.

The first equation ensures that $\{X_t\}$ will always get into A, no matter where it starts. The second equation implies that, once we are in A, X_{n+t_0} is chosen according to μ with probability ϵ. For two i.i.d Markov processes $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ starting at two different points in A, then the two chains will couple in t_0 steps with probability ϵ.

Let $\{X_t\}_{t\geq 0}$ be a continuous-time Markov process on a general state space X and let $\delta > 0$. The δ-skeleton chain of $\{X_t\}_{t\geq 0}$ is the discrete-time process $\{X_\delta, X_{2\delta}, \dots\}$. As the total variation distance is non-increasing, the convergence in total variation distance on the δ-skeleton chain will imply the Markov chain ergodic theorem on $\{X_t\}_{t\geq 0}$. The following version of the Markov chain ergodic theorem is taken from Meyn and Tweedie [MT93a]. Note that the skeleton condition is usually hard to check.

Theorem 14.4 ([MT93a, Thm. 6.1])**.** *Suppose that $\{X_t\}_{t\geq 0}$ is a Harris recurrent Markov process with stationary distribution π. Then $\{X_t\}$ is ergodic if at least one of its skeleton chains is irreducible.*

Recall that the Markov chain ergodic theorem states that, under moderate conditions, the transition probabilities will converge to its stationary distribution for almost all $x \in X$. The property of Harris recurrent allows us to replace "almost all" by all. For a non-Harris chain, it needs not converge on a null set.

Example 14.5 ([RR06, Example. 3])**.** Let $X = \{1, 2, \ldots\}$. Let $P_1(\{1\}) = 1$, and for $x \geq 2$, $P_x(\{1\}) = \frac{1}{x^2}$ and $P_x(\{x+1\}) = 1 - \frac{1}{x^2}$. The chain has a stationary distribution π which is the degenerate measure on $\{1\}$. Moreover, the chain is aperiodic and π-irreducible. On the other hand, for $x \geq 2$, we have

$$P[(\forall n)(X_n = x + n)|X_0 = x] = \prod_{i=x}^{\infty}(1 - \frac{1}{i^2}) = \frac{x-1}{x} > 0 \tag{14.8}$$

Hence the convergence only holds if we start at $\{1\}$.

The Markov chain ergodic theorem developed in this paper (Theorem 10.16) do not have such restrictions. It does not require the skeleton condition on the underlying Markov process nor does it require the Markov chain to be Harris recurrent.

Acknowledgement

We thank Robert Anderson, H. Jerome Keisler, Edwin Perkins and Daniel Roy for helpful discussions and insights, and thank Patrick Fitzsimmons and Krzysztof Latuszynski for their suggestions on related works in the literature. We would also like to thank the anonymous referee for insightful suggestions.

References

[ACH97] L. O. Arkeryd, N. J. Cutland, and C. W. Henson, eds. *Nonstandard analysis.* Vol. 493. NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences. Theory and applications. Kluwer Academic Publishers Group, Dordrecht, 1997, pp. xiv+366.

[And76] R. M. Anderson. "A nonstandard representation for Brownian motion and Itô integration". *Bull. Amer. Math. Soc.* 82.1 (1976), pp. 99–101. ISSN: 0002-9904.

[And82] R. M. Anderson. "Star-finite representations of measure spaces". *Trans. Amer. Math. Soc.* 271.2 (1982), pp. 667–687.

[Ber+02] J. Berger, H. Osswald, Y. Sun, and J.-L. Wu. “On nonstandard product measure spaces”. *Illinois J. Math.* 46.1 (2002), pp. 319–330. ISSN: 0019-2082.

[Bil95] P. Billingsley. *Probability and measure*. Third. Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1995, pp. xiv+593. ISBN: 0-471-00710-2.

[BW69] A. R. Bernstein and F. Wattenberg. “Nonstandard measure theory”. In: *Applications of Model Theory to Algebra, Analysis, and Probability (Internat. Sympos., Pasadena, Calif., 1967)*. Holt, Rinehart and Winston, New York, 1969, pp. 171–185.

[Cut+95] N. J. Cutland, V. Neves, F. Oliveira, and J. Sousa-Pinto, eds. *Developments in nonstandard mathematics*. Vol. 336. Pitman Research Notes in Mathematics Series. Papers from the International Colloquium (CIMNS94) held in memory of Abraham Robinson at the University of Aveiro, Aveiro, July 18–22, 1994. Longman, Harlow, 1995, pp. x+260.

[Dav71] R. O. Davies. “Measures not approximable or not specifiable by means of balls”. *Mathematika* 18 (1971), pp. 157–160. ISSN: 0025-5793.

[DD16] H. Duanmu and R. Daniel. “Nonstandard complete class theorems”. In preparation. 2016.

[DDF88] A. D’Aristotile, P. Diaconis, and D. Freedman. “On merging of probabilities”. *Sankhyā Ser. A* 50.3 (1988), pp. 363–380. ISSN: 0581-572X.

[GS01] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. Third. Oxford University Press, New York, 2001, pp. xii+596. ISBN: 0-19-857223-9.

[Hen72] C. W. Henson. “On the nonstandard representation of measures”. *Trans. Amer. Math. Soc.* 172 (1972), pp. 437–446. ISSN: 0002-9947.

[Hen79] C. W. Henson. “Analytic sets, Baire sets and the standard part map”. *Canad. J. Math.* 31.3 (1979), pp. 663–672. ISSN: 0008-414X.

[Kal02] O. Kallenberg. *Foundations of modern probability.* 2nd. New York: Springer, 2002, pp. xx+638.

[Kei84] H. J. Keisler. “An infinitesimal approach to stochastic analysis”. *Mem. Amer. Math. Soc.* 48.297 (1984), pp. x+184.

[KS04] H. J. Keisler and Y. Sun. “A metric on probabilities, and products of Loeb spaces”. *J. London Math. Soc. (2)* 69.1 (2004), pp. 258–272. ISSN: 0024-6107.

[Loe74] P. A. Loeb. “A nonstandard representation of Borel measures and σ-finite measures”. In: *Victoria Symposium on Nonstandard Analysis (Univ. Victoria, Victoria, B.C., 1972).* Springer, Berlin, 1974, 144–152. Lecture Notes in Math., Vol. 369.

[Loe75] P. A. Loeb. “Conversion from nonstandard to standard measure spaces and applications in probability theory”. *Trans. Amer. Math. Soc.* 211 (1975), pp. 113–122.

[LR87] D. Landers and L. Rogge. “Universal Loeb-measurability of sets and of the standard part map with applications”. *Trans. Amer. Math. Soc.* 304.1 (1987), pp. 229–243.

[Lux69] W. A. J. Luxemburg. “A general theory of monads”. In: *Applications of Model Theory to Algebra, Analysis, and Probability (Inte rnat. Sympos., Pasadena, Calif., 1967).* Holt, Rinehart and Winston, New York, 1969, pp. 18–86.

[LW15] P. A. Loeb and M. P. H. Wolff, eds. *Nonstandard analysis for the working mathematician.* Second. Springer, Dordrecht, 2015, pp. xv+481. ISBN: 978-94-017-7326-3; 978-94-017-7327-0.

[Mol05] I. Molchanov. *Theory of random sets.* Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2005, pp. xvi+488. ISBN: 978-185223-892-3; 1-85233-892-X.

[MS10] N. Madras and D. Sezer. "Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances". *Bernoulli* 16.3 (2010), pp. 882–908. ISSN: 1350-7265.

[MT09] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Second. With a prologue by Peter W. Glynn. Cambridge University Press, Cambridge, 2009, pp. xxviii+594. ISBN: 978-0-521-73182-9.

[MT93a] S. P. Meyn and R. L. Tweedie. "Stability of Markovian processes. II. Continuous-time processes and sampled chains". *Adv. in Appl. Probab.* 25.3 (1993), pp. 487–517. ISSN: 0001-8678.

[MT93b] S. P. Meyn and R. L. Tweedie. "Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes". *Adv. in Appl. Probab.* 25.3 (1993), pp. 518–548. ISSN: 0001-8678.

[PT91] D. Preiss and J. Tišer. "Measures in Banach spaces are determined by their values on balls". *Mathematika* 38.2 (1991), 391–397 (1992). ISSN: 0025-5793.

[Rob66] A. Robinson. *Non-standard analysis.* North-Holland Publishing Co., Amsterdam, 1966, pp. xi+293.

[Ros06] J. S. Rosenthal. *A first look at rigorous probability theory.* Second. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2006, pp. xvi+219. ISBN: 978-981-270-371-2; 981-270-371-3.

[RR04] G. O. Roberts and J. S. Rosenthal. "General state space Markov chains and MCMC algorithms". *Probab. Surv.* 1 (2004), pp. 20–71. ISSN: 1549-5787.

[RR06] G. O. Roberts and J. S. Rosenthal. "Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains". *Ann. Appl. Probab.* 16.4 (2006), pp. 2123–2139. ISSN: 1050-5164.

[RS86] J. P. Romano and A. F. Siegel. *Counterexamples in probability and statistics.* The Wadsworth & Brooks/Cole Statistics/Probability Series.

Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, 1986, pp. xxiv+303. ISBN: 0-534-05568-0.

[SCZ11] L. Saloff-Coste and J. Zúñiga. "Merging and stability for time inhomogeneous finite Markov chains". In: *Surveys in stochastic processes*. EMS Ser. Congr. Rep. Eur. Math. Soc., Zürich, 2011, pp. 127–151.

[Ste94] Ł. Stettner. "Remarks on ergodic conditions for Markov processes on Polish spaces". *Bull. Polish Acad. Sci. Math.* 42.2 (1994), pp. 103–114. ISSN: 0239-7269.

[Sun98] Y. Sun. "A theory of hyperfinite processes: the complete removal of individual uncertainty via exact LLN". *J. Math. Econom.* 29.4 (1998), pp. 419–503. ISSN: 0304-4068.

University of Toronto,Department of Statistics

University of Toronto,Department of Statistics

University of Toronto,Department of Mathematics